www.mruni.eu

Sergei Kornilov

DOCTORAL DISSERTATION

ASSESSING ORGANIZATIONAL EFFICIENCY UNDER MACROECONOMIC UNCERTAINTY IN DECISION SUPPORT SYSTEMS: ENSEMBLE METHODS IN MACHINE LEARNING WITH TWO-STAGE NONPARAMETRIC EFFICIENCY MODELS

SOCIAL SCIENCES,
ECONOMICS (S 004)
VILNIUS, 2020

MYKOLAS ROMERIS
UNIVERSITY

MYKOLAS ROMERIS UNIVERSITY

Sergei Kornilov

ASSESSING ORGANIZATIONAL EFFICIENCY UNDER MACROECONOMIC UNCERTAINTY IN DECISION SUPPORT SYSTEMS: ENSEMBLE METHODS IN MACHINE LEARNING WITH TWO-STAGE NONPARAMETRIC EFFICIENCY MODELS

Doctoral Dissertation
Social Sciences, Economics (S 004)

Vilnius, 2020

This dissertation was prepared during the period 2011-2017 at Tallinn University of Technology (Estonia) and during the period 2019-2020 at Mykolas Romeris University under the doctoral program right conferred to Vytautas Magnus University, ISM University of Management and Economics, Mykolas Romeris University and Šiauliai University on 22 February 2019 by the Order No. V-160 of the Minister of Education, Science and Sport of the Republic of Lithuania.

Dissertation is defended on a non-resident basis.

Scientific consultant:

Prof. Dr. Asta Vasiliauskaitė (Mykolas Romeris University, Social Sciences, Economics S 004).

Scientific supervisor at Tallinn University of Technology in 2011-2017:

Prof. Dr. Tatjana Põlajeva (Tallinn University of Technology, Estonia, Social Sciences, Economics S 004).

The doctoral dissertation is defended at the Council of Economics of Vytautas Magnus University, ISM University of Management and Economics, Mykolas Romeris University and Šiauliai University:

Chairperson:

Prof. Dr. Violeta Pukelienė (Vytautas Magnus University, Social Sciences, Economics S 004)

Members:

Prof. Dr. Natalja Lace (Riga Technical University, Latvia, Social Sciences, Economics S 004);

Prof. Habil. Dr. Žaneta Simanavičienė (Mykolas Romeris University, Social Sciences, Economics S 004);

Assoc. Prof. Dr. Sigita Urbonienė (Vytautas Magnus University, Natural Sciences, Mathematics N 001);

Assoc. Prof. Dr. Inga Žilinskienė (Mykolas Romeris University, Technological Sciences, Informatics Engineering T 007).

The doctoral dissertation will be defended at a public meeting of the Council of Economics on 30 June 2020 at 12:00 at Mykolas Romeris University, auditorium I-414.

Address: Ateities str. 20, 08303 Vilnius, Lithuania.

MYKOLO ROMERIO UNIVERSITETAS

Sergei Kornilov

SISTEMŲ, PADEDANČIŲ PRIIMTI SPRENDIMUS, ORGANIZACINIO EFEKTYVUMO VERTINIMAS ESANT MAKROEKONOMINIAM NEAPIBRĖŽTUMUI: APIBENDRINTI MOKYMOSI METODŲ ALGORITMAI TAIKANT DVIEJŲ PAKOPŲ NEPARAMETRINIUS EFEKTYVUMO MODELIUS

Mokslo daktaro disertacija
Socialiniai mokslai, ekonomika (S 004)

Vilnius, 2020

Mokslo daktaro disertacija rengta 2011-2017 metais Talino technologijos universitete (Estijos Respublika) ir 2019-2020 metais Mykolo Romerio universitete pagal Vytauto Didžiojo universitetui su ISM Vadybos ir ekonomikos universitetu, Mykolo Romerio universitetu, Šiaulių universitetu Lietuvos Respublikos švietimo, mokslo ir sporto ministro 2019 m. vasario 22 d. įsakymu Nr. V-160 suteiktą doktorantūros teisę.

Disertacija ginama eksternu.

Mokslinė konsultantė:

Prof. dr. Asta Vasiliauskaitė (Mykolo Romerio universitetas, socialiniai mokslai, ekonomika S 004).

Mokslinė vadovė Talino technologijos universitete 2011-2017 metais:

Prof. dr. Tatjana Põlajeva (Talino technologijos universitetas, Estijos Respublika, socialiniai mokslai, ekonomika S 004).

Mokslo daktaro disertacija ginama Vytauto Didžiojo universiteto, ISM Vadybos ir ekonomikos universiteto, Mykolo Romerio universiteto, Šiaulių universiteto Ekonomikos mokslo krypties taryboje:

Pirmininkė:

prof. dr. Violeta Pukelienė (Vytauto Didžiojo universitetas, socialiniai mokslai, ekonomika S 004).

Nariai:

prof. dr. Natalja Lace (Rygos technikos universitetas, Latvijos Respublika, socialiniai mokslai, ekonomika S 004);

prof. habil. dr. Žaneta Simanavičienė (Mykolo Romerio universitetas, socialiniai mokslai, ekonomika S 004);

doc. dr. Sigita Urbonienė (Vytauto Didžiojo universitetas, gamtos mokslai, matematika N 001);

doc. dr. Inga Žilinskienė (Mykolo Romerio universitetas, technologijos mokslai, informatikos inžinerija T 007).

Mokslo daktaro disertacija bus ginama viešame Ekonomikos mokslo krypties tarybos posėdyje 2020 m. birželio 30 d. 12 val. Mykolo Romerio universiteto I-414 aud.

Adresas: Ateities g. 20, 08303 Vilnius.

To my son Artemi.

TABLE OF CONTENTS

LIST OF TABLES 9

LIST OF FIGURES 10

KEY TERMS 11

ABBREVIATIONS 13

INTRODUCTION 15

I. MAIN CHARACTERISTICS OF UNCERTAINTY, EFFICIENCY ASSESSMENT IN DECISION SUPPORT SYSTEMS AND METHODOLOGICAL APPROACHES 27

1.1. The decision-making context in heterogeneous economic processes 27

1.1.1. Bounded rationality under macroeconomic complexity 27

1.1.2. Recursive decision-making model 30

1.1.3. Information accessibility phenomena for decision support systems 32

1.1.4. The problematic of ergodicity and stability of stochastic processes 35

1.2. Methodological context of assessment organizational efficiency 38

1.2.1. Factors of economic complexity 38

1.2.2. Context of heterogeneous economic agents 40

1.3. Foundations of uncertainty in economics theories 42

1.3.1. Uncertainty phenomena in macroeconomic studies 42

1.3.2. Uncertainty in error minimization models 47

1.3.3. Structural risk minimization 51

1.3.4. Uncertainty in expert meta-analysis 53

1.4. Classification of performance, effectiveness and efficiency 56

1.5. Ensemble machine learning approach in decision-making process 67

1.6. Integration of methods into decision support systems 72

1.7. Results and generalization of the literature analysis 76

II. PROPOSED MODEL FOR THE ASSESSING EFFICIENCY IN DECISION SUPPORT SYSTEMS UNDER UNCERTAINTY 81

2.1. Model definition for the assessing efficiency under uncertainty factors 81

2.2. Taxonomy of datasets selection for decision support system 84

2.3. Data mining techniques 89

2.4. Two-stage DEA nonparametric efficiency analysis 92

2.5. Ensemble methods in machine learning 101

2.5.1. Support vector machines 101

2.5.2. Artificial neural networks 105

2.5.3. Random forest 106

2.6. Research limitations 108

III. TESTING THE PROPOSED MODEL FOR THE ASSESSEMENT EFFICIENCY IN DECISION SUPPORT SYSTEMS UNDER UNCERTAINTY 111
3.1. Practical implementation steps of decision support systems 111
3.2. Datasets acquisition and analysis for decision support systems 114
3.2.1. Data model for decision support systems 114
3.2.2. Datasets structure analysis 120
3.2.3. Preliminary results of data model analysis 122
3.3. Efficiency assessment by nonparametric model 125
3.3.1. Nonparametric model selection 125
3.3.2. Decomposition of factors influencing efficiency 131
3.4. Feature set selection and efficiency influencing factors 135
3.4.1. Sensitivity analysis of the factors influencing efficiency 135
3.4.2. Results of analysis of features selection 137
3.5. Nonparametric efficiency models with ensemble machine learning 141
3.5.1. Results of ensemble machine learning techniques 141
3.5.2. Discussion of models for the decision support systems 146
3.6. Empirical Research Results and Discussion 149

CONCLUSIONS AND RECOMMENDATIONS 151
REFERENCES 155
APPENDICES 177
SANTRAUKA 219
MOKSLINIŲ PUBLIKACIJŲ SĄRAŠAS 242
SUMMARY 245

LIST OF TABLES

Table 1. Various measurement approaches of efficiency, performance and effectiveness by source of information and area of implication 57
Table 2. Significance analysis of DEA researchers according to their g-index, h-index 60
Table 3. Scientific articles using DEA method published in the Baltic Sea Region 61
Table 4. Methodological review of machine learning techniques employed for efficiency assessment in economic studies 63
Table 5. Components overview of machine learning methods and techniques for elaboration decision support systems model 70
Table 6. Organization's resources orientated efficiency parameters 94
Table 7. Market performance orientated efficiency parameters 95
Table 8. SVM Kernel functions 104
Table 9. Sequences of practical assessment using proposed methodological approaches 112
Table 10. Data model's logical principle structured in data tables 116
Table 11. Country specific variables composition with data sources and abbreviation 118
Table 12. Country specific data sources 119
Table 13. Datasets for World Uncertainty Index and Economic Policy Uncertainty 120
Table 14. Results of estimating the datasets on random effects, cross-sectional dependences, serial correlation, stationary and heteroscedasticity 121
Table 15. Baltic stock market uncertainty and volatility (log) 124
Table 16. Selection and validation of nonparametric models (A1, A2, A3, A4, A5, A6) 128
Table 17. Validation of efficiency models (A1, A2, A3, A4, A5, A6) 129
Table 18. Mean and Standard deviation in inputs 130
Table 19. Feature selection of efficiency and uncertainty variables using ordinary least squared, fixed effects, random effects and pooling models 136
Table 20. Feature set selection by factors for decision making process 137
Table 21. Influential power for externalities selection in ensemble machine learning 138
Table 22. Firm-level feature set definition for ensemble machine learning 139
Table 23. Factors of market performance output orientated decision making process 140
Table 24. Categorization of performance and efficiency 141
Table 25. Fitting machine learning algorithms 144
Table 26. Analysis of the confusion matrix 146
Table 27. Results of SVM kernels 147

LIST OF FIGURES

Figure 1. Logical Structure of the research 22
Figure 2. Multi-disciplinary aspect of fundamental economic complexity 29
Figure 3. Recursive perception model 31
Figure 4. The stages of the industrial revolution 34
Figure 5. Uncertainty by channels and variables 43
Figure 6. Assessing uncertainty in models 47
Figure 7. Taxonomy of benchmarking techniques 58
Figure 8. Performance assessment approaches 62
Figure 9. Machine learning and its types 68
Figure 10. Dynamically adjustable decision-making model 72
Figure 11. Trends in decision support techniques 74
Figure 12. Generalization of processes of efficiency assessment under uncertainty 79
Figure 13. Algorithm for proposed decision-making systems 83
Figure 14. Data model 85
Figure 15. Handling missing data 91
Figure 16. Two-stage nonparametric efficiency estimation approach 93
Figure 17. DEA models classification 96
Figure 18. Non-discretionary exogenous factors 100
Figure 19. Artificial Neural Network 105
Figure 20. Acquiring data for analysis 114
Figure 21. Representative organizations datasets count by market place and lists 122
Figure 22. Organizations count by industry and sectors 122
Figure 23. Conditional two-stage efficiencies and market volatility 123
Figure 24. Correlation map of models efficiency and environmental variables representing non-linear economic effects on economic agents through various channels 125
Figure 25. Stepwise two-stage efficiency assessment 127
Figure 26. Adjustments of input factor in nonparametric models (Log transformed) 130
Figure 27. Scale and Efficiency changes (Log transformed) 131
Figure 28. Productivity index by stock market and industries (Log transformed, yearly) 132
Figure 29. Level variables correlation 134
Figure 30. The modeling processes of ensemble machine learning using various algorithms 142
Figure 31. Ensemble machine learning V-fold CV risk estimates with 95% CIs 145

KEY TERMS

Aleatoric Uncertainty Also known as Knightian, risk refers to the inherent uncertainty due to the probabilistic variability. This type of uncertainty is Irreducible, in that there will always be variability in the underlying variables. These uncertainties are characterized by a probability distribution (Oberkampf et al. (2001)).

Bounded Rationality Rationality of individuals is limited by the information they have, the cognitive limitations of their minds, and the finite amount of time they have to make decisions. Bounded rationality expresses the idea of the practical impossibility of exercise of perfect rationality (Simon (1997)).

Cross-Validation A statistical method of evaluating generalization performance that is more stable and thorough than using a split into a training and a test set. In cross validation, the data is instead split repeatedly and multiple models are trained. The most commonly used version of cross-validation is k-fold cross-validation, where k is a user-specified number (Müller (2016)).

Dummy Variable A variable that is created by the analyst to represent group membership on a variable (Jaccard et al. (2001)).

Economic Agent An economic decision maker who can recognize that different factors influence and motivate different economic groups (James E Hartley (2002)).

Economic Complexity The composition of a country's productive output and represents the structures that emerge to hold and combine knowledge. (Erkan and Yildirimci (2015)).

Economies of Scale An increase in all inputs leads to a more-than-proportional increase in the level of output. The cost advantages that enterprises obtain due to their scale of operation, with cost per unit of output decreasing with increasing scale. (Samuelson (2010)).

Econophysics An interdisciplinary research field, applying theories and methods originally developed by physicists in order to solve problems in economics, usually those including uncertainty or stochastic processes and nonlinear dynamics.(Black et al. (2018)).

Efficiency It refers to doing that with the least possible cost and resources and doing it in the shortest amount of time. Technical Efficiency refers to the ability of an entity to get the maximum output for a given set of inputs, with reference to a production function. Allocative Efficiency concerns the ability of an entity to use the inputs and produce outputs in optimal proportions given their prices (Cooper et al. (2011)).

Epistemic Uncertainty Also known as Knightian uncertainty, expressing incomplete or potentially biased knowledge of the forecasters. Known and resolvable lack of knowledge, which cannot be addressed owing to the lack of empirical data in the absence of previous occurrences (Black (2018)).

Feature Selection The process of selecting dimension, metrics to be used in machine learning models (Kreienkamp (2014)).

Heterogeneous Agent A theoretical construction defined by bounded rationality follows learning processes that influence the aggregate dynamics of the system (Galati (2013)).

Isolation Effect The phenomenon whereby people value a thing differently depending on whether it is placed in isolation and whether it is placed next to an alternative (Kahneman et al. (1979))

Kernel function	The functions that, given the original feature vectors, return the same value as the dot product of its corresponding mapped feature vectors. Kernel functions do not explicitly map the feature vectors to a higher-dimensional space, or calculate the dot product of the mapped vectors. Kernels produce the same value through a different series of operations that can often be computed more efficiently. (Dangeti (2017)).
Knightian Uncertainty	The phenomenon arises when economic agents cannot reasonably assess the likelihood of all possible future states of nature or characterise the probability distribution of their possible impacts (Black (2018)).
Macroeconomic Uncertainty	The conditional time-varying standard deviation of a factor that is common to the forecast errors for various macroeconomic indicators such as unemployment, industrial production, consumption expenditure, among others (Jo (2017)).
Ontological Uncertainty	Related to pure randomness (unpredictability) of future events. A state of complete ignorance: agents don't know what they don't know (Black (2018)).
Overfitting	The production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably (Dangeti (2017)).
Performance	It refers to the quantitative measure of attainment to a set standard.
Probability Model	A mathematical representation of a random phenomenon. It is defined by its sample space, events within the sample space, and probabilities associated with each event (Kosmidou (2010)).
Proxy Variable	A variable that is used to measure an unobservable quantity of interest. Although a proxy variable is not a direct measure of the desired quantity, a proxy variable is strongly related to the unobserved variable of interest (Denzin (2017)).
Representative Agent	An assumption used by economists to model the macroeconomy. The general idea is to solve a well-specified microeconomic problem, and then use the relationships between the variables in that model as a description of the macroeconomy (Hartly (2002)
Risk	A probability or threat of damage, injury, liability, loss, or any other negative occurrence that is caused by external or internal vulnerabilities, and that may be avoided through preemptive action (Pal (2017)).
Slack Variable	A real variable which is introduced to take up the slack in an inequality constraint, i.e., to convert an inequality constraint to an equality constraint (Pierre (2013)).
Sociophysics	A field of science which uses mathematical tools inspired by physics to understand the behavior of human crowds. In a modern commercial use, it can also refer to the analysis of social phenomena with big data.
Supervised Learning	Model creation where the model describes a relationship between a set of selected variables (attributes or features) and a predefined variable called the target variable. The model estimates the value of the target variable as a function (possibly a probabilistic function) of the features (Provorst (2013)).
Unsupervised Learning	Auto-associative networks can also serve to identify the central tendencies or prototypes found in a range in input data, even where the prototype itself was never encountered in training (Müller (2016)).

ABBREVIATIONS

ADF	Augmented Dickey-Fuller
AE	Allocative Efficiency
BCC	Banker, Charnes and Cooper model
BSR	Baltic Sea Region
CCR	Charnes, Cooper & Rhodes model
CRS	Constant Return to Scale
CV	Cross-Validation
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
EPU	Economic Policy Uncertainty index
ERM	Empirical Risk Minimization
GDPR	The General Data Protection Regulation (EU) 2016/679
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
MCMC	Markov Chain Monte Carlo
MSE	Mean Squared Error
ODM	Oracle Data Mining
PDF	Probability Density Function
RBF	Radial basis function kernel
SCC	Squared Correlation Coefficient
SPF	Survey of Professional Forecasters
SRM	Structural Risk Minimization
SVM	Support Vector Machines
TE	Technical Efficiency
TFP	Total Factor Productivity
UDEA	Uncertain Decision Making Unit
VRS	Variable Return to Scale
XGB	Extreme Gradient Boosting

INTRODUCTION

Relevance of the topic

The modern environment, where we all live in, is the subject of constant changes over time. There are evidences both from mass-media and science, that the modern economic setting is characterized by increasing information flow gathered for decision making process, growing global competition on the macroeconomic level and limited physical resources. It is hard to deny the increasing role of information and knowledge in the XXI century. One particular concern could emerge from decision-making process, which is getting more and more complicated with time passing by. The decision-making process has the result to make the most efficient decision, which implies the minimal allocation of resources and maximum output *ceteris paribus*. Thus, the assessing of effectivity plays enormous role in the decision-making process. The cost of wrong decision is enormous due to increased competitors on the global scale. However, a clear effect of convergence in economics generally is observed on global scale, but the technological advance is the factor, which determines competitive advantage over decades. With developing of technologies, the opportunity cost is getting higher at the explosive scale. For example, it is hard to make any credible assumption in rise of a hypothetical emerging economy, which is able to achieve the technological advance on global scale in a hi-tech area without having prior fundamental knowledge and expertise. In order to avoid possible opportunity costs, the decision-makers are demanding more sophisticated yet reliable prediction techniques. In this context, the issue of handling decisions of generally heterogeneous economic agents under factors of uncertainties emerges as the front-line problematic of scientific research.

The Author asserts throughout the thesis, that none of the above mentioned issues should be regarded isolated manner. This assumption finds its justification in the literature body of economic science virtually since the very beginning. But only in the past decades the scientific methodology reinforced with technologies of machine learning could bring a feasible analytical framework for assessment uncertainty on different levels and capture nonlinearities in processes not only within generalized scientific techniques prevailing with ensemble expectation assumptions underneath. Through innovations caused by financial technologies it becomes possible to shed light on the idea of economic growth and its connection with investments from different perspective in terms of their ability to create or absorb technological innovations within on-going infinite technological progress.

Theoretically seen modern economic settings consist of a large number of smaller complex subsets. In the preceding paper, Kornilov and Polajeva (2016) have already investigated a complex nature of economic processes. The study shows, that increased levels of complexity affected by uncertainty in many ways and thus increase risk factors. Each economic subset is modular in terms of being made up of a large number of functionally specific parts. It is open in the sense that these parts deal with a certain degrees of freedom. Any scientific approach should be able to recognize that agents are naturally heterogeneous, what rose complexity of economic process demanded more sophisticated methods, which

should explain individual set of knowledge collectively, creation of an aggregated outcome and their reaction to this outcome. This differs from other approaches tend itself to expression in equation form, whereas by definition a general pattern that does not change. Modern scientific literature draws attention to important issues in economic studies, including spatial integration and economic complexity. The economic subsets have become increasingly associated with the widely known concept as knowledge economy.

The modern business settings are defined by rapid and radical changes caused by information accessibility. The modern economic science is being in turbulence nowadays (Chuen and Linda (2018), Dunis *et al.* (2016)). The recent trends of financial automation explain increasing progress to have computational applications for forecasting, modelling and trading financial markets and information. New trends of cryptocurrency and digital finance has to be analyzed in the future science but today's evidences have already exhibited that its phenomena. It has been already forged as the result we might see as the convergence of profit motives with social objectives creating a class of large companies in financial technologies. Technological exchange among sectors is intense nowadays, so the underlying innovations may be applied to a wide range of industries simultaneously. The *technological convergence* is another important factor, which is a relative new to the economic science and it is definitely the subject of future in-depth investigation.

The relevance of the topic is supported by the integration processes among EU member states focused of increasing economy efficiency and eliminating economic disparities among nations. The integration development consists of the underlying dynamics of globalization in terms of markets and capital as well as the move towards closer international co-operation through the further development of trade unions and policy co-ordination. Such arrangements represent different modes of economic integration processes by eliminating borders of any kind among member states and applying a common policy and structure the economies to trade with other non-members. Therefore, the assessment of efficiency under uncertainty for the policy-makers belongs to the tasks with the highest priority.

Assessing efficiency is highly dependent on reliable evidences, which are the subjects influenced by uncertainty. A number of various methods exist to assess complexity of economic, uncertainty as the factor of economic processes and assessing efficiency. But the attitudes of researchers in the field are detached. Uncertainty have been proven to be unstable factor, with the variations being most vividly seen during the crises. Due to the reasons mentioned here, the assessing efficiency under uncertainty is relevant in both theoretical and empirical aspects. This doctoral dissertation focuses on both aspects.

The recent studies of Onatski and Williams (2003) argues that uncertainty is persistent phenomena in economics and it must be faced continually by policymakers. Black *et al.* (2018), Meinen and Röhe (2017) supports that measuring macroeconomic uncertainty and understanding its impact on economic activity is thus crucial for assessing the current macroeconomic situation. From modern positions a robust and negative effect of uncertainty on economic growth is obvious and these consequences cannot be neglected by the theory (Lensink *et al.* (1999), Levin *et al.* (2005), Ljungqvist and Sargent (2012)). There are a vast number of studies arguing indicators of uncertainty which can be viewed as representative to the evidences of particular policy, involving a wide number of direct and indirect peers

(Ericsson *et al.* (1999), Benhabib *et al.* (2013), Bird *et al.* (2013),Jurado *et al.* (2013), Ernst and Viegelahn (2014), Baker *et al.* (2015), Jurado *et al.* (2015)). The uncertainty factor is so large that the effects of policy decisions on the economy are thought to be ambiguous. In this situation, any plausible expertise on the nature of uncertainty might be very useful. In order to understand how variations in uncertainty might affect the economic process, it is important to find its source.

The large number of studies shows that assessment of efficiency analysis has become an important topic in operational research, public policy, energy-environment management, and regional development. So it is obvious, that two-stage nonparametric methods have been widely used in the recent literature on productive efficiency measurement and in a large literature of studies. Empirical applications choose one group of measurement techniques.

Therefore, the relevance of the topic shows, that first of all the theoretical and practical findings of the thesis underpin the idea of applying machine learning methods for efficiency assessment under uncertainty, which can be utilized for the future policy-makers.

Second, there is a clear need to predict the influence of the heteroscedasticity on the global scale beyond and within the EU. It contributes to the mechanism linking intertwined cross-border components with high degree of freedom into a system, which has economic inputs and outputs as parameters and is able to handle uncertainties on different layer: limited information, bounded rationality and their expectations, and randomness.

Third, but the most important, to argue that assessment of efficiency under uncertainty plays the leading role in a knowledge-driven economic system with continuously increasing complexity. Depending on a number of factors, it is crucial to elaborate a theoretical framework, which can embrace as many factors. Therefore, any research on assessment of efficiency under uncertainty should have a broader scope and should not be limited on country-specific parameters but include configurations in clusters over the borders. The economic development should be captured not solely in economic terms, but should be shaped for knowledge exchange among economic agents. Adding a factor of uncertainty into analysis opens a specific question of incorporating social processes aspects into study.

Research problematic and the level of its investigation

The economic science represents a huge variety of perfect works assessing uncertainty. At the frontier line of experiment-based models derived from recent observations are Elder (2004), Kontonikas (2004), Daal *et al.* (2005), Fountas (2010), Fountas (2010), Henry *et al.* (2007), Neanidis and Savva (2011). The studies are keen to follow deterministic paradigms to cause uncertainties. In this category prevail a wide family of autoregressive conditional heteroscedasticity models both with error variance or in its general form imposing conditionally-autoregressive errors associated with uncertainties. Methodological questions on measure of the uncertainty raised by Giordani and Söderlind (2003), Diebold *et al.* (1997), Clements and Harvey (2011). Classification of Walker *et al.* (2003) gives fundamental notion on it. Berument *et al.* (2009) and Hartmann and Herwartz (2012) extend the standard assumption with stochastic volatility models. Orlik and Veldkamp (2014) and Glass and

Fritsche (2015) argue that uncertainty is an outcome value of acyclical changes in uncertainty while shocks. Zarnowitz and Lambros (1987), Bomberger (1996), Rich and Butler (1998) and D'Amico and Orphanides (2008) argue the epistemic uncertainty by direct estimation of parametric distributions across individuals. Lahiri and Sheng (2010), Siklos (2013), Lahiri *et al.* (2015) extend the model by to numerous improvements and modifications. Walker *et al.* (2003), Dequech (2004) look into epistemic uncertainty caused by experts incomplete knowledge and the variability uncertainty attributed to accidental factors randomly appeared. Lane and Maxfield (2004) extends the variability uncertainty with the ontological uncertainty. Discussion raised by Walker *et al.* (2003) classification goes into inflation uncertainty by Norton (2006), Kowalczyk (2013), Krayer von Krauss *et al.* (2019). Gelman and Hill (2007) introduces multilevel linear and generalized linear model in which the parameters are given a probability model. Jordà *et al.* (2013), Knüppel (2014) considering uncertainty from rational predicting model or their model combinations not necessarily having econometric apparatus underneath and rely on the assessment of risk factors from the distribution of *ex-post* forecast errors.

The same level of scientific investigation exhibit nonparametric efficiency assessment. Originated from Seiford (1997) with 800 publications, the more recent overview by Seiford (2005) mentions some 2800 published articles on DEA. Since fundamental contributions by Farrell (1957), Koopmans (1952), Aigner and Chu (1968), Aigner *et al.* (1977), Broek *et al.* (1980) concept of efficiency methodology in frontier production function estimation has been rapid developed. There are a number of critical reviews emerged by principle weakness of the conventional methods to assess efficiency. Sexton *et al.* (1986) and followed by Smith (1997) identified the impact of misspecification, Stolp (1990) generalized that homogeneity of technology across DMU, uncertainty over the choice of inputs and outputs can affect the performance assessment.

However, Tobback *et al.* (2018) argues that common methods of measuring uncertainty developed by Baker *et al.* (2015) does not have any predictive power for any of its variables but the machine learning approach outperform the traditional ARCH-based models. Brose *et al.* (2014a) and Brose *et al.* (2014b) argue that managing risks and uncertainty depends critically on information. Past decade, a number of research look deep into usage of an optimization algorithms based on a linear programming model to identify controls that need to be tested to address the risks, which can be developed as hybrid approaches for efficiency classification (Pareek (2006), H.-Y. Kao *et al.* (2013)). Various linear optimization techniques has been successfully applied to predict time series and their co-movements (Kara *et al.* (2011), Karaa and Krichene (2012)).

Therefore, the recent studies employ machine learning both for assessment uncertainty and efficiency measurement. Predictive power of various machine learning techniques like neural networks widely confirmed in the literature and found practical implications as by Alejo *et al.* (2013). Attigeri *et al.* (2017) argue empirical approach is used to build models for financial risk assessment with supervised machine learning algorithms. Kruppa *et al.* (2012), Kreienkamp and Kateshov (2014), Addo *et al.* (2018) results indicate that nonlinear techniques work especially well to model expected value. Past a few years many researchers exploit machine learning technique and nonparametric technique to provide

a new method for predicting efficiency by using data envelopment analysis (Xu and Wang (2009), L. Zhou *et al.* (2014), X. Yang and Dimitrov (2017), Zelenkov *et al.* (2017), Alaka *et al.* (2018)). Q. Zhang and Wang (2018) proposed efficiency prediction model which for the first time combines supervised learning for information analysis with nonparametric model, to evaluate the future efficiency of decision making unit.

Research problem

The main focus of the research is to elaborate approaches to carry out a framework for efficiency assessment, which is from one hand is reinforced by economic science and on another hand take advantage of the machine learning algorithms to create plausible estimation result. The in-depth review of theoretical literature body and practical implications, the problematic of the efficiency assessment has unambiguously and clearly tossed a challenge for further investigation of uncertainty conditions as the result of economic complexity and nonlinearities in respected to the decision-making processes. In the past a lot of excellent scientific researches contributed valuable yet profound findings in economic science to shed the light on economic processes and the role of economic agent involved. The most influential scientists awarded Nobel Memorial Prize in economic sciences for the sound contributions to behavioral economics of bounded rationality. However, it has been admitted, there is a gap exists between theoretical findings and practical real-world efficiency assessment of economic agents in decision-making process under uncertainty conditions.

Research objects

The Author will investigate economic agents to which the proposed model to efficiency assessment under uncertainty will be attributed in order to be prove, that uncertainties from various sources can be estimated using machine learning techniques and included into hybrid model of assessment of efficiency under uncertainty.

Research aim

After disclosing of the factors of heterogeneous attitude, economic complexity, uncertainty and the efficiency, the aim to prepare the methodology for assessment of efficiency under uncertainty and test on the dataset from the financial area. The Author believes that the result of the recent technological development and machine learning techniques might emerge in various forms of automated decision-making processes, which will supply policymakers with relevant yet precise information on a particular problem.

The research aim is supported by the following objectives:

1. To justify the necessity of a complex approach without isolating economic agents by attributing them to spatial subsets.
2. To distinguish the source of uncertainty and analyses the proposed methods for assessment efficiency.

3. To investigate the efficiency as a concept and analyses the proposed methods for assessment efficiency using ensemble machine learning techniques.
4. To propose a conceptual model for the assessment of efficiency under uncertainty using linear optimization methods and machine learning algorithms.
5. To conduct an empirical research in assessment efficiency using ensemble machine learning techniques based on the hybrid model.
6. To discuss the results of empirical research of assessment efficiency under uncertainty conditions to suggest recommendations for their application.

Research methods

The research methods used by the Author comprise analysis, synthesis and comparison of scientific literature to characterize uncertainty and efficiency. There is a growing interest in the machine learning techniques. Hence, data-driven approaches are becoming very important in many scientific areas and real-world applications. The main demand exhibited by two important factors. First, the implication of more effective statistical models is needed to explain the complex data dependencies. Second, the scalable learning systems can handle large datasets for creating plausible predictive results.

The economic science has a lot of various scientific methods. The regression analysis belongs to one of the most recognized and widely used techniques in the field of quantitative research to build up both predictive and explanatory models. Various regression techniques are aimed to find and explain relationship between variables and among other variables, explain determined the relationships and provide valuable knowledge for sciences. Another common approach in the quantitative research is the classification, which is used widely in practical prediction assignments in financial areas. The method of the ensemble methods in machine learning techniques based on Random Forest, Support Vector Machines and Artificial Neural Networks approaches has the convenient superiority both in the classification and regression assignments. This affects its widespread application. The ensemble methods in machine learning is intended to find its application particularly to solve the classification problems. However, the latest development trends in machine learning have extended this technique to solve regression problems. In order to find evaluation method, the Author use a hybrid approach to combine two-stage nonparametric model and machine learning techniques in the joint evaluation process taking full advantage of two-stage nonparametric modeling method of absence predetermined weights to input and output parameters. Furthermore, it is possible to assess relative efficiency of organization with the focus on objectivity with acceptable error. The experimental results show that this method has strong objectivity and impartiality, the evaluation method is simple and easy to interpret (Cristianini and Shawe-Taylor (2000), Scholkopf and Smola (2001), L. Zhou *et al.* (2014)).

The proposed model will be trained using cross-validation procedure, which used to evaluate machine learning models on a limited data sample. The model performs training on the 50% of the given dataset, 25% is used for the training purpose and 25% for testing.

The practical software used for analysis is R[1]. The R software package with its extensions is a free software programming language for statistical computing. R provides a wide variety of statistical including but not limited to linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering. The Author uses ensemble methods in machine learning extension package developed by Kuhn (2008), Kuhn and Johnson (2013), which offer tools for classification and regression training, contains a number of tools for developing predictive models using the rich set of models available in R. The package has function on simplifying model training and tuning with a wide variety of modeling practices. It provides methods for *ex-ante* training data, calculating variable significance, and model representations. The computational interaction is used to exhibit the functionality on a real data set and to target the benefits of parallel processing with given types of models.

The practical implementation in R based on the methodology proposed by Kleiber and Zeileis (2008), Leipzig and Li (2011), Adler (2010), Albert (2007), Albert and Rizzo (2012), Kassambara (2013) and Kassambara (2017a), Kaas *et al.* (2008), Beyersmann *et al.* (2011), model diagnoses by Bolker (2008), statistical by Cohen and Cohen (2008). The regression analysis carried out as highlighted by Cowpertwait and Metcalfe (2009), Karian and Dudewicz (2016), Højsgaard *et al.* (2012).

Data and its sources

The research investigates the efficiency of the selected companies listed on the Nasdaq Baltic Index. The datasets for uncertainty are represented by multiple sources:

1. Economic Policy Uncertainty, which is based on mass-media coverage frequency and also defined as the common volatility in the unforecastable component of a large number of economic indicators.
2. Country-specific factors should include market concentration, presence of foreign investments, fiscal indicators. In rapidly changing business settings evolving working environment, the ability to predict future trends and needs in terms of the knowledge and skills required to justify becomes critical for effective decision support system. These trends fluctuate by geography and industry, and so it is important to anticipate the industry and country-specific variables. The sources are:
   - Federal Reserve Economic Data
   - Deutsche Bundesbank Data Repository
   - Organization for Economic Co-operation and Development
   - NASDAQ OMX Global Index Data
   - World Bank World Development Indicators
3. Organizational datasets are provided by NASDAQ OMX

The research is based on the historical data sources from the sources listed above directly and by using the API and download from the respective source[2].

1 www.r-project.org - The R Project for Statistical Computing

2 Annex 20. Data sources

Dissertation volume and structure

Dissertation consists of introduction, three parts, conclusions and recommendations, references and appendices. The volume of the dissertation is 278 pages. It contains 31 figures, 27 tables, 418 references and 19 appendices. Dissertation's logical structure is introduced in the Figure 1.

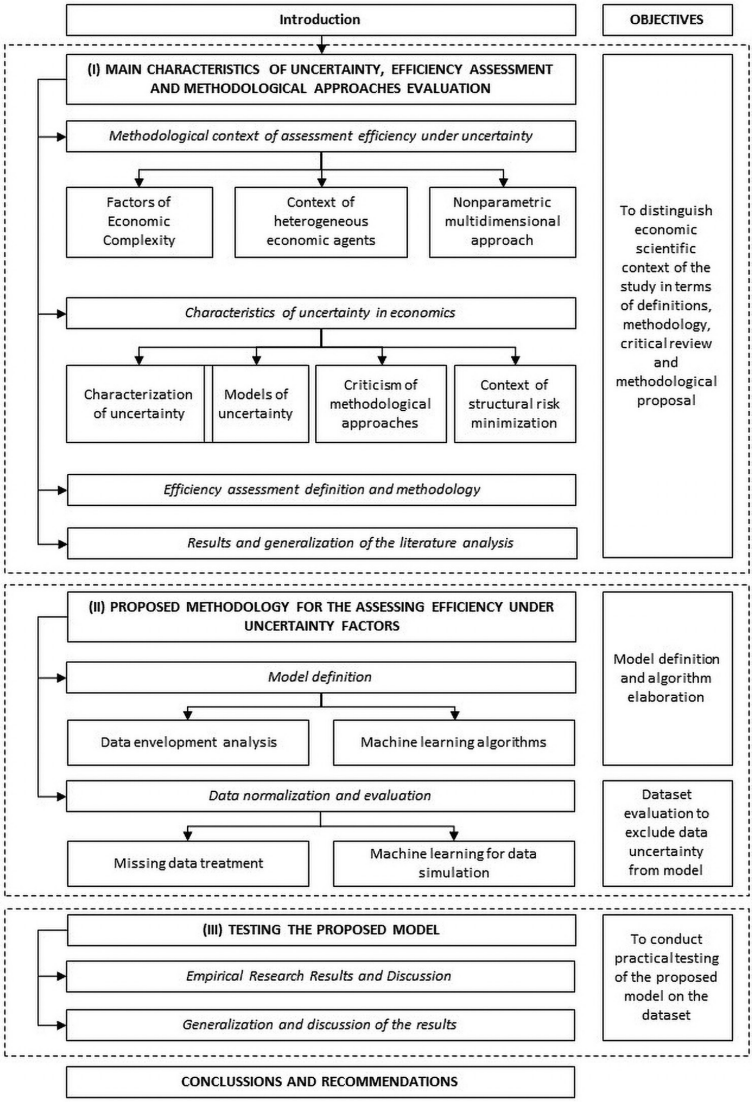


Figure 1. *Logical Structure of the research (Source: The Author's representation)*

In the first part of this dissertation the characteristics of the scientific context. There are the main characteristics, which define the methodological approaches. The nonlinearity assumption clarifies the relations among the components. This is one of the key economic feature means that the whole system cannot be broken down into smaller parts and reassembled to obtain the initial system. It is characterized by unpredictability, random behavior and approximation. They also do not change in a constant proportion with respect to the input. The concept of bounded rationality devised by Herbert Simon defines, that the individual can make decisions that appear irrational from the perspective of conventional economic wisdom. Although rational reasoning seems to be a useful tool in coping with complexity, the concept of rationality as a formal framework for resource bounded agents does not seem to be empirically proven (Müller (2016), Schilirò (2012), Marwala and Hurwitz (2017)). First, second and third research objectives, supporting the research aim, are achieved. In the end of the first part, the main research findings are generalized. The former approach has enticed many researchers to join, forming new streams of research named econophysics and sociophysics (Onozaki (2018)). Each phenomenon cannot be explained in isolated manner but raise new research questions regarding the role of each economic agent in dynamic system transitions. It calls for interdisciplinary view on the problematic. Therefore, the problematic of effectiveness and handing uncertainties is important while making economic real-life decisions are usually made in uncertainty condition. Black *et al.* (2018) argues that uncertainty in its various forms is widely known as the factor that influence economic while difficult to measure. The uncertainty factor is so large that the effects of policy decisions on the economy are thought to be ambiguous. Modeling uncertainty by experts is an *ex-ante* process and can therefore be used for assessing future state of the economy.

Parametric frontier models and nonparametric methods have been widely used in the recent literature on productive efficiency measurement and in a large literature of studies. Nonparametric modelling and their analysis emerged from influential work by Farrell (1957) aimed to develop a comparative measure for production efficiency. However, the tremendous amount of information stored in databases cannot simply be used for further processing (Kelly (1998), Bhavsar and Ganatra (2012), S. Li *et al.* (2012)). Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data set. Analysis of research literature enabled the Author to determine that the efficiency assessment under uncertainty conditions is the result of economic complexity and nonlinearities of the decision-making processes.

The second part describes methods, technological approach and tools for their research, insights and evaluation, such as framework for model, as well as the developed methods. A framework for multidimensional analysis based on machine learning techniques, which enables analysis from different point of views. Machine learning is concerned with the development of techniques and methods which enable the computer to learn and perform tasks and activities. Efficiency assessment are heavily dependent on the data set that is used as an input to the productivity model. As now there are numerous models based on two-stage nonparametric models.

The third part applies the proposed model on the dataset from the previous section. A two-stage nonparametric modelling approach confirms that the use in many industrial

and economic applications justified. However, for large datasets with multiple inputs and outputs nonparametric models might exhibit considerable computing time. Experimental results demonstrate that the proposed method outperforms the earlier methods of regressions.

For the classification tasks hybrid method is apparently a more feasible and effective method in predicting failure comparing to using data with pure machine learning approach. The model employment gives interpretable results, but there is still future work:

1. There is a huge field of optimization based on the constraints of the economic agent. Imposing more restrictions on the variable weights will provide more real-life results.
2. The prediction power of the model depends heavily on the quality of the datasets.
3. Selection of appropriate machine learning technique has effect on analytical results.

The Author is able to assert, that machine learning might outperform regression methods widely used in the scientific researches nowadays. The reason is machine learning methods include multilayer processing with hidden information. Thus, better accuracy performance achieved through subsampling layers.

Research limitations

Several research limitations show that the results of employing the research methodology is heavily dependent on the quality of the data. Nevertheless, it does not diminish the importance of the results in both theoretical and practical levels. On a theoretical level, this research methodology, prepared on the basis of an identified research gap, is one of the first attempts to comprehensively assess not only factors of uncertainty and efficiency separately, but induce state-of-the-art approach to comprehend them as the whole. On an empirical level, this research covers the majority of the issues related to efficiency assessment under uncertainty, taking into account the constraints imposed on such analysis for each cluster of economic agents.

Not all the possible factors are included in the proposed model. Proposed model is only one possible way to achieve plausible results under uncertainty. However, some empirical datasets include 10 horizon of years. The number of macroeconomic factors is limited to 8 major indicators. On the data-mining level there are some assumptions that the data passed initial filters. That means there is no random values in dataset. In real-life where datasets acquired automatically there are probability of missed or wrong values.

Scientific novelty

1. This research is designed to involve the theoretical and empirical aspects of uncertainty, nonlinearities, complexity and bounded rationality as the major assumption of the framework. There are no assumptions in terms of equilibria based theories. Analysis of previous studies shows that often theoretical part is detached from the statistical significant findings. It is obvious, that the economics itself from very early steps accepted equilibria concept and the study of generally balanced growing path. Thus, the statistical findings justified to established theories. Even though, Brian

(2006) pointed conceptually out that traditional studying equilibrium patterns of consistency required further behavioral adjustments.

2. The assessment of efficiency under uncertainty is defined by various sources of uncertainty, which cannot be quantified within other than hybrid model. From formal point of view, various uncertainties from missing data can be generalized with hypothesis of limited information. But there is to admit, that many real-world datasets may contain missing values for various reasons. Taking such data into a model with a lot of missing values can drastically impact the model's quality. The proposed model offers upfront how to deal with missing data using various machine learning techniques. A number of researchers insist on the quality of the data, whereas Lertworasirikul *et al.* (2002) shows that the nonparametric modelling methods require accurate measurement of both the inputs and outputs. In all the situations presented by researchers and practitioners of nonparametric modelling, it is still a relatively subjective approach in filling a gap from missing data. But within the proposed model in this study the treatment of missing data is one of the important tasks.
3. This research is one of the few that employ structural risk minimization principle to estimate uncertainties, whereas instead of minimizing the observed training error proposed machine learning techniques attempts to minimize the generalization error bound so as to achieve generalized performance.
4. This study is one of the first attempts to assess efficiency within both classification and regression model. The Author among other researches investigate ensemble methods in machine learning classifiers in the face of uncertain knowledge sets and show how data uncertainty in knowledge sets can be treated in ensemble methods in machine learning classification by employing robust optimization. Consequently, ensemble methods in machine learning can also be used as a regression method, maintaining all the main features that characterize the algorithm of maximal margin. The Author is agreed with, that the future of the machine learning is in combination of different approaches, because fully supervised algorithms are a useful but perhaps an unnatural assumption due to latent variables in models (D. Chen *et al.* (2013)).
5. The proposed model deals with evaluating efficiencies in the absence of homogeneity gives rise to the issue of how to fairly compare a DMU to other units. A related problem, and one that has been examined extensively in the literature, is the missing data problem addressed directly to appropriate techniques of machine learning (Zhu (2016b)).
6. The Author is the first who explicitly proposed to treat uncertainty not as a dummy variable, but phenomenon dissected within the proposed model on different layers: data-mining uncertainty, analytical framework uncertainty and uncertainty as a factor. Unlike the existing approaches, the combinations of machine learning techniques in this study do not require to think in terms of hypothetical assumption. Mathematically machine learning leads to the identification of implicit restrictions to weights, so there is a fundamental difference in these approaches, emerging from the way in which the data explicitly is gathered. In each process the uncertainty is emerging in different qualities and it should be assessed with respective techniques.

Practical significance of the research results

The study gives clear results on integration of an automated core for any decision support systems. It shows that it is possible to design systems by using modular functions. There are a number of areas of applications, where decision support systems can be applied in:

1. Business Intelligence systems designed the way where data can be promptly observed and filtered by a number of different dimensions in order to obtain immediately insights into recent performance of organizational units.
2. Data mining applications operate with enormous sets of data and facts which have been combined and accumulated through ongoing interaction with counter-parties and environment. These datasets play important role for statistical analysis focused on acquiring meta-information and hidden patterns on utility, preferences, trends, or other associated agents' behavior.
3. Full-scale Enterprise Resource Planning application gives opportunity to conduct organizational workflow based on efficiency a better way with focus on capital investment, inventory, production and logistics.

Only the structured methodology using various methods in approaches both from Data Science and Economic studies might help researchers and policymaker rank the important factors and appreciate the factors underneath a better way. It is not possible to respect the results either from statistical nor theoretical point of view solely, but only as an integrated process with the fully qualified decision support system.

Dissemination of scientific research results

Parts of the research results have been disseminated in papers, published in scientific journals, acknowledged by international scientific conferences.

Publications:

1. S. Kornilov, T. Polajeva (2016) Economic Complexity: the future of the knowledge-based regional development. SGEM 2016. ISBN 978-619-7105-72-8 / ISSN 2367-5659
2. S. Kornilov, S. Ridala, A. Aasma (2016) Dynamics of economic adjustment under uncertainty: MS-VAR model for Baltic Dry Index. SGEM 2016. ISBN 978-619-7105-76-6 / ISSN 2367-5659. Book 2, Vol. 5, P. 189-196, DOI: 10.5593/SGEMSOCIAL2016/B25/S07.025.
3. Sergei Kornilov, Tatjana Põlajeva (2012) Infrastructure development: Economic Growth Effects, VGTU. ISSN 2029-4441

Author's Contribution to the Publications:

– Article (1) – The Author of the thesis was responsible for systemizing literature body, proposing the main idea and data sets for analysis.
– Article (2) - The Author of the thesis was in charge of carrying out the model, data gathering and acted actively in publishing process
– Article (3) - The author of the thesis was in charge of carrying out the model, data gathering and acted actively in publishing process.

I. MAIN CHARACTERISTICS OF UNCERTAINTY, EFFICIENCY ASSESSMENT IN DECISION SUPPORT SYSTEMS AND METHODOLOGICAL APPROACHES

1.1. The decision-making context in heterogeneous economic processes

1.1.1. Bounded rationality under macroeconomic complexity

Economic science is defined as a study of human beings' activities and behavior related to exploiting scarce productive resources to satisfy their fundamentally unlimited needs. The economic science cannot exist without an axiom of scarcity of resources. Its concept of scarcity of resources is deeply rooted in the economic theory and fundamentally grounded in scientific field by the first American recipient of the economics Nobel Prize Samuelson (1983 [1947]), whereas its concept is still actual nowadays (Mankiw (2008) and Samuelson (2010)).

The scarcity of resources concept derives value of resources in order to acquire them. The outcome of an unlimited human desire is considered in form of an opportunity cost. Therefore, one of the key aspects of existing global business settings is allocation of limited resources among economic agents in any form of their representative agent form whether generalized on *Marshallian* conception of a representative agent or a later assumption of heterogeneous agent, which conceptually appeared from criticism these days by Heylighen (2008), Kirman (2010), Kirman and Zimmermann (2012).

Either way, the efficiency assessing of economic agents under macroeconomic uncertainties is a part of economic process on the global scale. The problematic of efficiency forecasting for decision-making under uncertainty yields complex dynamic system questions, which should explain main patterns of how all these components interact and being interconnected. The literature review from the past and recent researches exhibit three ontological factors, which influence outcome under uncertainty:

1. Nonlinearity of process
2. Bounded rationality of economic agents
3. Economic complexity

The nonlinearity assumption clarifies the relations among the components. This is one of the key economic feature means that the whole system cannot be broken down into smaller parts and reassembled to obtain the initial system. It is characterized by unpredictability, random behavior and approximation. Another characteristic is that output does not depend on the input proportion change constantly due to a dynamic and flexible relations between variables.

Recently Puu (2010) investigated the nonlinearity from historical perspective. His study states that a specific problem for economics from formal science point of view is that a true study of nonlinear dynamics requires global analysis. The economics itself from very early

steps accepted equilibria concept and the study of generally balanced growing path. Based on ex-post evidences economists preferred convex structures providing for unique equilibrium states, which could be associated with optimal growth path. Therefore, formal relations are usually described qualitatively, and are seldom stated in explicit form meant to hold over a wide domain of variable values. However, researches of production are exceptional due to the fact that the production functions formulated and estimated from the evidences based on the time series or panel data. Therefore, they exhibit the best explanatory functions which find its application in analysis of global dynamics despite the matter that they were not developed directly for this purpose. However, in financial time series, the main complications arise from the detached of the equilibria concept and presumed nonstationary of the underlying process. The nonlinearity of the regression function is highly dependable on forecasting horizon of the time-scale. Thus much research effort has been devoted to exploring the nonlinearity and to developing specific nonlinear models to improve prediction power. The nonlinearity of the economic model is a key element in generating the possibility of equilibria, which can appear as part of a rational expectations solution in linear models as well (M. Zhang (2008), Puu (2010), Evans and Honkapohja (2012), Nava *et al.* (2018)).

Another key issue is the concept of bounded rationality devised by Herbert Simon. The analysis of the scientific literature body reports that the theory of bounded rationality is a solid analytical approach, which found its application in many diverse areas. It should be noted that the theory of bounded rationality has not replaced the theory of rationality (Marwala and Hurwitz (2017)). In Simon's view the rationality of the individual is bounded, since the quality of information used is poor and the cognitive capacity of the individual is limited. So the individual can make decisions that appear irrational from the perspective of conventional economic wisdom (Schilirò (2012)).

From traditional economic perspective as represented by von Neumann *et al.* (2007 [1947]) individuals generally move in the reality following predetermined patterns of behavior, at the base of which there is the assumption that they always prefer to have a greater wealth than less. Although rational reasoning seems to be a useful tool in coping with complexity, the concept of rationality as a formal framework for resource bounded agents does not seem to be empirically proven (Müller (2016)).

Economic complexity increases the order or regularity between the components interaction and even generates a new order and configuration with the different components behaving autonomously. Thus, radically new processes and component interactions emerge. The recent study of Onozaki (2018) asserts that research on economic complexity has been carried out into two fundamental directions. One is the introduction of stochastic or statistical mechanical frameworks in terms of economic variables, while the other is the development of agent-based models that use computer simulations. The former approach has enticed many researchers to join, forming new streams of research named econophysics and sociophysics. Nevertheless, both methods are similar in its intention to go further beyond the neoclassical paradigm.

Each phenomenon cannot be explained in isolated manner but raise new research questions regarding the role of each economic agent in dynamic system transitions. It calls for interdisciplinary view on the problematic.

In another words, the issue of efficiency assessment and thus growth sustainability based on the human's perception and environmental factors, what makes any research, which implies especially productivity issues, fundamentally complex and multi-disciplinary. It needs to adjust to new knowledge and evolving circumstances. Understanding effectiveness and ways of succeeding involves an understanding of complex adaptive systems and general systems theory but not merely constructing models with given degree of freedom and assumptions. The theory of effectiveness supported by concepts and theories of the economic, social and behavioral sciences and definitely goes far beyond equilibrium concept as presented in the Figure 2.

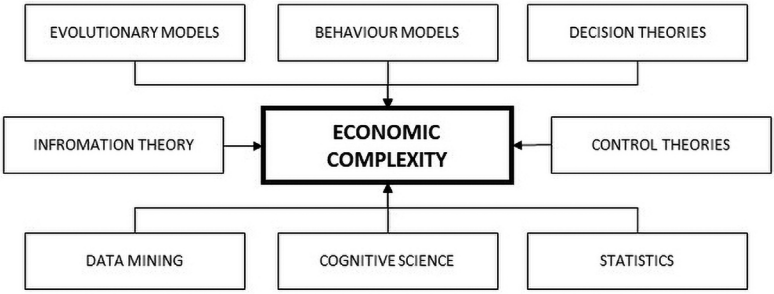


Figure 2. *Multi-disciplinary aspect of fundamental economic complexity (Source: Author's representation)*

The economics have been developing science adjusting methods and theories for the constantly mutable environment settings. The fact of the matter is that the dome is still for discussion, which theoretical and conceptual approach can embrace the processes itself and explain underneath causes and reasons a better way. From the literature body it is obvious to see that equilibrium approaches in different forms have been enormously successful along with the economic science. The main principles of equilibrium have been incorporated into the neoclassical structure widely known nowadays and common accepted to explain macroeconomic processes in every possible detail. Latest dynamic stochastic general equilibrium models are popular nowadays in macroeconomics, which development delivers acceptable forecasts under certain conditions (Del Negro and Schorfheide (2004), Schorfheide (2011)).

However, economic agents are extremely complex in nature. Jofré *et al.* (2007) argues that concepts of *equilibrium* have long been connected with maximization or minimization. In economic and social situations, game theory has provided formulations in which different entities, or agents, with possibly conflicting interests seek to optimize in circumstances where any of actions might have influential consequences for the others. The notion of Nash equilibrium has that form, for instance, as do various models of traffic equilibrium. More complex varieties of equilibrium theory might find its applications for the sake of determination of market prices. Thus, the classical economic equilibrium interpretation of original Walras is meant. The truncation of arguments with

specific estimates, based on the data in the economic model, is intended to transform the unbounded variational inequality that naturally comes up into a bounded one having the same solutions.

There are a number of strategies utilizing a variety of disciplines are needed to acquire a proper understanding. Since Thompson *et al.* (2011 [1967]) generalization of organizational theory, an enormously difficult question of rationality was raised in conjunction with actions taken by economic agents. An attempt of logical formalization by Masuch and Huang (1996) extended the original perception of the theory. Many inheritors of Thompson's theory offer a numerous distinct propositions about the behavior of organizations. Regardless the classification of organizations based on their technologies and environments, there is in common, that any organization should face uncertainty along with risk factors and is enforced to handle uncertainty. Further Anderson (2018 [1988]), Inigo and Albareda (2016), Arthur (2018) foster understanding from modern perspective of how firms get involved in new processes, strategies and behaviors for sustainable development with changing technological innovations. The results of the finding are that organizations nowadays exhibit non-linear, recursive and self-organized features that can should be studied as a complex adaptive system.

1.1.2. Recursive decision-making model

Recursive decision-making process is a backbone of economic processes. Along with environmental changes and globalization processes Inigo *et al.* (2017) shows that the organizational processes are mutating with technological advance, which determines the decision-making of economic agents. Konnov (2007) considers equilibrium concepts and their applications in related fields of economics describing rather complicated systems, where the linear distribution of resources is not possible due to the bounded rationality of economic agents from one hand and the economic complexity phenomena along with macroeconomic uncertainties from another.

The necessity of modeling approach is on the agenda over decades. The latest findings by Khandani *et al.* (2010), Khemakhem and Boujelbene (2017) argue that the large number of decisions made today should rely on models and algorithms rather than human perception. The decision making process should be reinforced by proven information. Since decision-making process of organizations is the subject of contextual experience, Khezrimotlagh and Chen (2018) recognize the necessity in generalized formulation and decomposing of decision-making process by organizations. Allen (2011) and Akkizidis and Stagars (2015) underpin the that adoption of better decisions relayed on technological advance otherwise short-term thinking often undermines the company's success in the long term. It claimed that rule-based event simulation and agent-based modeling should be considered for modeling systems dynamics and simulation of humans' interactions.

The problematic of effectiveness and handing uncertainties in decision-making process is important while making economic real-life decisions are usually made in uncertainty condition. Theoretical framework should be interested not only in first-order effects that determine the risk but also in second-round effects influencing efficiency as an outcome.

Bryant *et al.* (2014) and Firoozye and Ariff (2016) state that many applications of decision-making should include measurement of risk. The understanding and theorization of risk and the weighing of the consequences of risk received much attention by Wen (2014), who regarded the problematic by different measures including probability measure, credibility measure, uncertainty measure and chance measure.

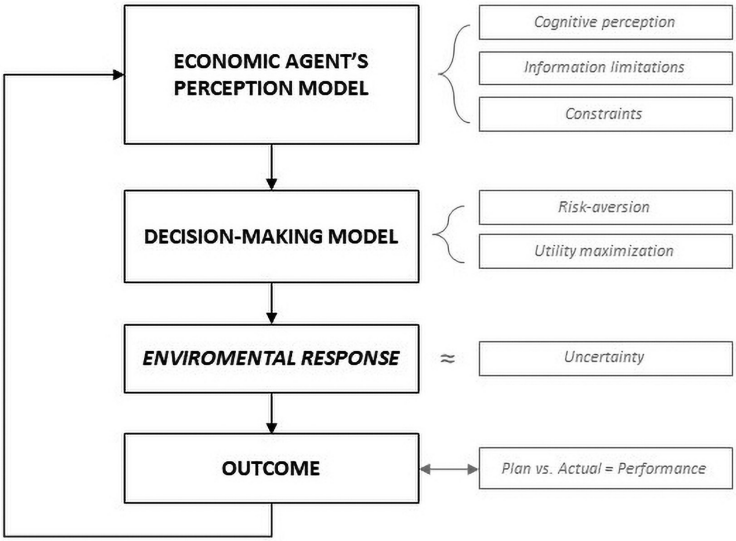


Figure 3. *Recursive perception model*
(Source: Author's representation)

The recursive perception model as in Figure 3 explains that decision-making process include multiple variable, which are influenced by various factors of risk and uncertainty. The economic agent's perception can be modeled with restriction of cognitive and information limitations. Thus, the decision-making model emerges, which outcome results interaction with the environmental response. At the end economic agent reflects the outcome with own desired preferences. In the literature the formulation, that a decision-maker does not know in advance, the consequences of a given action, is represented as a central issue in the decision-making under uncertainty. Decision-making leads to the aspect of effectiveness measurement and forecasting is information from a broader perspective.

The study of Hahn and Huang (1999) stress out that nowadays information is a valuable asset in some ways like other economic commodities. Acquiring of information is a costly process and its possession is valuable for decision-making process. From other perspective Hackeling (2017) underpins that modern technological advance makes possible to achieve the fundamental goal of machine learning is to generalize acquired data array and derive new knowledge without being explicitly programmed. Such questions of machine learning

and forecasting emerged a wave of the interest of not only economists, but found a wide discussion in mass media. Sure it has to do with the fact that technological changes are associated with an example of globalization of information technologies but at the same time, introduces a spatial element in the discussion. The knowledge management is getting a great attention and is considered as an important issue within the research field and policy-making authorities. Alavi and Leidner (2001) argue the related strategies act as a source of competitive advantage for any organization. Mårtensson (2000) supports the idea the knowledge transfer can be regarded as a driver for enhancing productivity and flexibility in organizations. This, the knowledge management is an important competence for any organization. Moving bottom-up the regional development relay to a great extent on the effectiveness of each organization in the economic subset.

Research aim is in a broader sense to find out the drivers for sustain growth and development due to a better resources allocation by improving decision-making process. Narrowing research problem is to determine a broader concept of effectiveness of economic agents in the decision support systems both in private and public sectors. Despite lively discussion around the public expenditure, public investments have not got as much attention from economists. There is an opinion that countries do not grow rich in a sustainable fashion by making more of the same using economies of scale. The countries are seeking continuously for other opportunities to produce by allocating activities that are innovative, effective and profitable.

1.1.3. Information accessibility phenomena for decision support systems

The development of information processing and decision support systems has a long history. Over decades in the late starting with 1960s and 1970s, researches began to focus further attention on finding a combination of operational research, machine learning, and various information systems. In the modern science, there is a common understanding among most researchers that the latest definition of data science known widely also as *big data* with their respective techniques of deep learning, data mining, and sentiment analysis are just innovative key terms for generalization of business analytics. However, the aim remains the same over decades, namely to transform data flow into actionable awareness for more accurate decision-making process in order to support respective policymakers in their particular fields.

The Author shares strong belief of Schwab (2016), K. Zhou *et al.* (2015) that the modern business settings are defined by rapid and radical changes caused by information accessibility. The modern economic science is being in turbulence nowadays. There a number of concept and frameworks describing the same phenomena. One of the prominent definitions has emerged as a concept of the *Industry 4.0* proposed by the German government in November 2011 during the famous Hannover Fair as a strategy for high-tech sector by 2020.

Devezas *et al.* (2017) gives insights into this report, which defines the *Industry 4.0* environment, which includes the strong customization of products under the conditions of

high flexibility of mass production, requiring the introduction of methods of self-organized systems to get the suitable linkage between the material and the virtual worlds.

Since then it has been widely debated and has become a headliner for most global industries and the information industry. No doubt, that it will is an industrial revolution, which will have a great influence on the processes on the global scale.

An empirical analysis conducted by Bartodziej (2016) gives the problem definition of the manufacturing industry, which is currently the subject to huge change. This change is caused by various ongoing global megatrends such as globalization, urbanization, individualization, and demographic change, which will considerably challenge the entire manufacturing environment in the future. On the one hand, an increase of worldwide connected business activities will raise the complexity within manufacturing networks. These challenging requirements will force companies to adapt their entire manufacturing approach including structure, processes, and products.

Originally developed by Shannon (1948) information theory proposed the fundamentals for the future digitalization in many scientific fields telecommunications, genetics, socio-economic and deep space research. Since then, it became clearly that information can be defined in scientific terms and become measurable quantity. Shannon's theory of information provides an analytical framework of information, describing properly what information can be communicated between different elements of a system and in which amount. The Author used widely Google's Ngram service[3] to derive the Figure 4, representing the amount of disposable information using keywords related to technology, production, manufacturing during various stages of industrial revolution. Some of terms linked to complexity, networking and data processing emerged after introduction of *programmable logic controllers*, which are units made to receive information from connected sensors or network units, processes the data, and triggers outputs with predefined parameters.

Buckley *et al.* (2016) gives the latest evolution of financial technology sector from historical perspective, led by start-ups, poses challenges for regulators and market participants, particularly in balancing the potential benefits of innovation with the possible risks of new approaches. The Author shares the opinion of Buckley *et al.* (2016), who recognized the financial technology sector era starts with financial globalization and derives a number of stages. In the late 19th century the convergence of technology in finance combined with others produced the first bricks of financial globalization. The Author supports Geum *et al.* (2012), Gauch and Blind (2014) point of view is that the world is in the era of technological convergence in recent innovation trends, where one notable feature is the merging and overlapping of technologies.

3 Google Ngram Viewer: https://books.google.com/ngrams

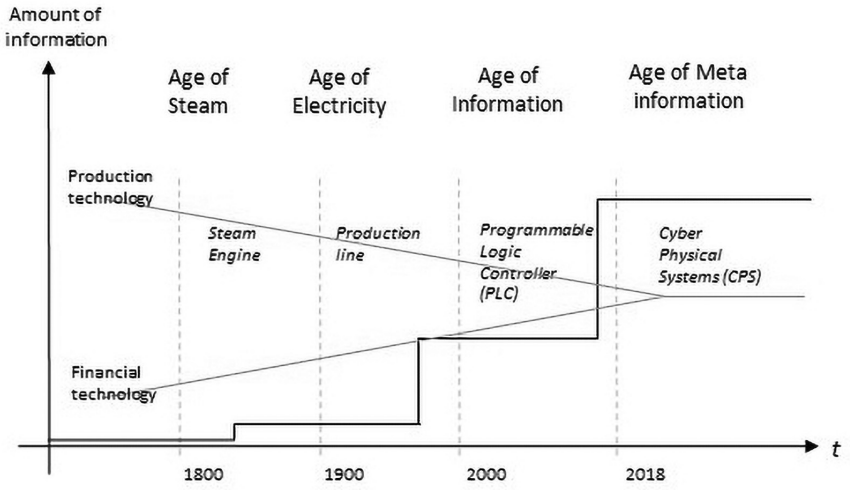


Figure 4. *The stages of the industrial revolution*
*(**Source:** Author's representation, based on Buckley et al. (2016), DFKI (2001)[4])*

Lee and Seshia (2016) gives the latest development of the industrial revolution, which is characterized by cyber-physical systems (CPS), which are an integration of computation with physical processes whose behavior is defined by both cyber and physical parts of the system. It is not sufficient to separately understand the physical processes and the computational components. Their interaction is the subject of analysis. There is less or no use of specific programmable units in the financial sector, but artificial intellect equipped with CPS and machine learning techniques along with pattern recognition is used with virtually no adoption both in manufacturing and financial sectors.

The recent trends in financial automation by Dunis *et al.* (2016), Chuen and Linda (2018), Freund (2018) explain increasing progress to have computational applications for forecasting, modelling and trading financial markets and information. New trends of cryptocurrency and digital finance has to be analyzed in the future science but today's evidences have already exhibited that its phenomena have forged in forms we might see the convergence of profit motives with social objectives creating a class of large companies in financial technologies. Technological exchange among sectors is intense nowadays, so the underlying innovations may be applied to a wide range of industries simultaneously. The financial market agents are looking forward having more sophisticated yet effective financial solutions to emerging challenges. Neural networking is a highly effective, trainable algorithmic approach which emulates certain aspects of human brain functions, and is used extensively in financial forecasting allowing for quick investment decision making. The most innovative technological applications in artificial intellect and data processing are introduced for financial technology and other areas of finance. All of them replace con-

4 German Research Centre for Artificial Intelligence. http://www.dfki.de

ventional time series analysis for forecasting and trading financial instruments. The pattern recognition, timing models, forecasting and trading of financial instruments belong to the technological advance in present days.

Shannon (1948) study remains relevant and plays enormous role still nowadays in economic science and financial sector. The information theory gives the ability to separate meaningful signals from noise using the latest trends and development in the fourth stage of the industrial revolution, to extract precise information from raw data, what is crucial for modern financial markets. Even more generally, a computational neuroscience relies on information theory to provide a benchmark against which the performance of neurons can be objectively measured. Based on their findings far better intelligent approaches can be deployed for an efficient decision making process. Such approaches relay heavily on machine learning techniques. Not surprisingly, a number of researchers on complex networks, network theory, and graph theory, each with a different and often limited focus, have appeared in the past decade.

Notable study of Easley and Kleinberg (2010) goes far beyond boundaries and offers more than generalized study of social networks as such. For example, the interconnection of loans network among financial institutions can be used to decompose the roles of participants in the financial system and reveal the interactions among these roles affect the individual participant's behavior and the system as a whole. First, they prove that the rise of the global communication networks and availability of powerful computing units in low-budget segment have made it possible to collect and analyze network data on a large scale. Therefore, the progress in applied tools along with a variety of new theoretical approached has indorsed to gain new knowledge from many different interconnected sources. Second, by introducing dynamics into network analysis the authors dissect theories of structure and behavior. Graph theory, respectively, is the study of network structure, while game theory provides models of individual behavior in settings where outcomes depend on the behavior of others. Another renowned studies of networks by Newman (2010), Gray (2013), Brown (1983) and recent Brown and Hwang (2012) give insights the study is broadly interdisciplinary by its nature, which can include the measurement and structure of networks in many branches of science, methods for analyzing network data. Practical findings of J. Chen (2005) stress out that certain empirical evidences about information driven instead of a behavioral phenomenon of market players can be explained by an information theory introduced by Shannon's entropy theory of information. For example, drilling down to a single investor's decision case and market patterns are the results of information processing by investors of different sizes with different background knowledge.

1.1.4. The problematic of ergodicity and stability of stochastic processes

Among others, the Author argues in his article Kornilov *et al.* (2016) that in the modern economic processes occasionally exhibit sudden changes in their behavior caused by externalities or dramatic breaks in the government policy. Of particular interest to sciences is the obvious fact that a number of economic variables behave quite differently during economic disturbances and therefore follow different patterns. However, among other ele-

ments of the complex economy the factors of production but not their long-run estimations to grow define the economic dynamics.

In general case, there is a common understanding and widely admitted in the literature body that the scope of problems related with economic growth, financial markets and institutions remain over the entire history generally unsolved within any of existing models. It is caused by the fact that the observed processes are not ergodic nor stationary to full extent.

A substantial body of evidences analyzed by Sarel (1996), Hooker (2002) has found the possibility of nonlinear effects of inflation on economic growth. It proved the function related to inflation growth rates contains a structural break. In case of a structural break, a significant bias in the estimation of inflation effect might occur if nonlinearities will be neglected. Evidences have also reported asymmetric and nonlinear effects on real activity, as well as that structural instabilities exist in those relationships.

The essence of problematic can be derived in details from Banerjee *et al.* (1993) where despite the fact that in economic theory, the concept of equilibrium is well established and well defined, a significant body of methods is developing around the statistical features of equilibrium relationships among time-series processes. Following Cowpertwait and Metcalfe (2009) the concepts of stationarity and particular forms of non-stationarity are crucial to these methods. The fundamentals of stationary defined as a function of a time series model (1) as a function of t:

$$\boldsymbol{\mu(t) = E(x_t)} \qquad (1.1.1)$$

The Equation (1.1.1) shed the light on the nature of a stochastic process. The main purpose for economic science is that any stochastic process should model some economic process with given parameters and interpretable prediction result. From statistical point of view, any stochastic process is meant to generate the infinite array of ensemble samples of all possible observed time series. Therefore, a statistical population can be formulated as an ensemble of stochastic processes in form of time series representing such processes. In this context, it is become clear that uncertainty should be considered in the observed time series in order to produce plausible prediction results. For this case expectation E is considered as expectation of an ensemble average with respect to the distribution of times series.

There is common practice to apply regression techniques to build up prediction models for times-series analysis. The majority of the models require a stationary of an observed stochastic process for any value of t. The requirement for datasets in times-series analysis is represent data in its stationary form where there are no systematic differences in mean and variance values without strictly variations over periods. Hence, If the mean function (1.1.2) is constant, so the time series model is stationary in the mean. The sample estimate of the population mean, μ is the sample mean, $\bar{x}$:

$$\boldsymbol{\overline{x} = \sum_{t=1}^{n} \frac{x_t}{n}} \qquad (1.1.2)$$

Equation does rely on an assumption that a sufficiently long time series characterizes the hypothetical model. Such models are known as ergodic, and a vast number of econometric models are all ergodic mostly. The expectation in this definition is an average taken across the ensemble of all the possible time series that might have been produced by the time series model. It is obvious that a stationarity in the mean of a time series model (1.1.3) is ergodic in the mean taking into account the average for a single time series trend to the mutual mean defined by the length of the increase in time series:

$$\lim_{n\to\infty} \frac{\sum x_t}{n} = \mu \qquad (1.1.3)$$

This implies that the time average is independent of the starting point. Environmental and economic time series are often single realizations of a hypothetical time series model, where the definition of underlying model as ergodic is implied.

Further, Viana and Oliveira (2016) and Coudène and Erné (2016) gives a notion of ergodic theory as the study of the long-term behavior of systems preserving a certain form of energy. From a mathematical point of view, a physical system can be modeled by the data of a space X, a transformation $T: X \to X$, and a measure μ defined on X and invariant under T: for every measurable set $A \subset X$,we have $\mu(T^{-1}(A)) = \mu(A)$. The quadruple consisting of the space X, the measure μ, the σ-algebra consisting of the measurable sets with respect to μ, and the measurable transformation T that preserves μ form what is called a measure-preserving dynamical system.

Borovkov (1998) study of ergodicity and stability of stochastic processes is at the forefront of research and presents results as well as established ideas. The term stability in terms of ergodic and stationary process is used to describe continuity properties of distributions with respect to small perturbations of their local characteristics. It is the key assumption in theorems of ergodicity and stability for a comprehensive number of classes of Markov chains, stochastically recursive sequences and their generalizations. Therefore, considering ergodicity and stability of multi-dimensional Markov chains and Markov processes draw particular attention to large deviation problems and transient phenomenon, which is important for statisticians and applied researchers in the theory of Markov models and their applications.

Pfaff (2008), Shumway and Stoffer (2014) and Pfaff (2016) underpin the fact, that all these assumptions imply that additionally a model error often has to be considered to estimation errors. A non-stationary return process could be exhibited based on a distribution for stationary processes. Therefore, there is a trade-off between using a distribution assumption for stationary processes committing a model error and using a longer sample extent by which the stationarity assumption is more likely to be violated but the estimation error reduces. Another issue pointed out by Greenland *et al.* (2016) is the misuse and wrong interpretation of statistical values in researches. ***P***-values have proven problematic for correct description of complex processes with a single measure.

Therefore, analytical frameworks used in the analysis of stationary models for study economies with sustained growth cannot be employed due non-balanced features, which amended by structural changes. Economic approaches to prediction whatever the case

may be require both ergodicity and stability of stochastic processes. This is the reason why mathematical-based frameworks without any complexity nor behavior assumptions will fail to predict changes in response to rumors, wars, government policy.

1.2. Methodological context of assessment organizational efficiency

1.2.1. Factors of economic complexity

Decision support systems employ efficiency assessment models based on input-output parameters are commonly used to evaluate economic impacts. These models typically evaluate exogenous variables in resource demanding elements with no look at associated effects of recursive simultaneous connections. An analysis from the economic agent perspective is of greater interest to economic that exploit natural resources because their activity is subject to variations or various factors beyond, what formal approach estimates. Herewith proposes a methodology to improve the estimation of the impacts of these variations. Within the methodological context of economic context analysis, a practical methodology is introduced. Hence, the proposed method will improve impact assessments derived from economic agents to environmental events.

The decision support systems should involve the concept of the economic complexity upfront in order to avoid useful yet misleading generalizations. Theoretically seen the modern economic settings consist of a large number of a smaller complex subsets where decision-making process might also consist of various interconnected chains. The theory points out that the complexity of the economic system is defined by its modularity, openness and hierarchic depth. (Kornilov and Polajeva (2016)). Moreover, it is deep in the sense that each module is itself a complex system. Macal and North (2010) and Chan and Steiglitz (2009) give notion of the economic system as modular, open and deep, and because there are many ways for a system to be like this, complexity is inherently emergent. Researchers among others give insights into the meaning and the definition of a complex adaptive system representing the entire economy that development associated with decentralized market economies, such as inductive learning, imperfect competition, formation of trade network, and the open-ended co-evolution of individual behaviors and economic institutions. A renewed interest in empirical works are also on the sources of comparative advantages per Boccaletti *et al.* (2006), Preiser *et al.* (2018), Clayton and Radcliffe (2018)

The analysis of clusters, since its introduction in scientific and policy studies in the early 1990s, had an enormous impact on the decision support systems. The clusters as a whole provide a descriptive way to scrutinize the systemic nature of an economy in terms of various types of industrial activity is related. Starting with the organizations in the industry where the main producers of the primary goods are located, the cluster also embraces suppliers and industries providing different types of specialized inputs and technological processes as well as customers and more indirectly related industries.

Back to eighties of the past century Porter (2008 [1980]) has a considerably facilitated in the development of his theory on competitiveness. The strategy of research was to find some links between the spatial dynamic of some productive systems and his famous dia-

mond of firm rivalry, new entry of competitors, power of upstream suppliers and machinery producers, threat of substitutive products, and factors-demand conditions. Porter has found the five factors of competitiveness, which he originally applied in a macro context to explain the competitive advantage of nations. According to Porter:

> *"Clusters are a prominent feature of the landscape of every advanced economy, cluster formation is an essential ingredient of economic development. Clusters offer a new way to think about economies and economic development".*

Engel and del-Palacio (2009) argues, that Porter's model does not explain why new and apparently unrelated industries have emerged in specialized clusters. The study extends the Porter's definition by Clusters of innovation, which characterized by investments, working force and information, including know-how and intellectual property. More concrete Ellison and Glaeser (1997) describe clusters as non-random geographical agglomeration of firms with similar or closely complementary capabilities supplementing Richardson (1972) annotated by Andersen (1998). However, in the literature body clusters in a wide sense have been introduced under many different terms. The authors have gather some of the synonyms as listed in Table 1 below might give insights into the same essential interpretations as articulated through different authors by using the cluster concept as defined. However, noticed it might have certain differences in peripheral definitions by their implications underlying some minor ideas or assumptions by revealing concepts. Another differences might be in certain concepts often by providing historical or some associations from the historical perspective or applications as the result of current common accepted interpretations that have emerged from the use of a particular term. But the existence of reach vocabulary diversity in terms cannot downgrade the fact that the cluster phenomena as such have raised a great attention during the last decades.

Asheim (2007) and Cooke (2008) prove the fact that the significant number of policy-oriented researches exhibits the fact that cluster building appeals to many policy-makers as the key to national, regional, and even local development policy.

The clusters have become increasingly associated with the widely known concept as knowledge economy. Martin and Sunley (2011) point out that economists operate endogenous production function models to evaluate that the increasing returns to educated labor and Research and Development spending are localized to a great extent, so that regional growth paths may diverge. It has highlighted the importance of spatial contiguity in the human capital accumulation by creating knowledge spillovers. But it hardly sheds light on the reasons and causes such endogenous growth processes become geographically concentrated in particular localities and not others. Notwithstanding the strong knowledge spillovers focus, these theoretical constructions have little to explain on the institutional and social networks through which many such knowledge spillovers might take place. Therefore, wide ranges of approaches broadly described as neo-Schumpeterian are focused on economic localizations in creation knowledge and innovation.

Pyka and Hanusch (2007) define the hallmark of neo-Schumpeterian economics is that they put a strong emphasis on knowledge, innovation and entrepreneur spirit more at the

micro rather than macro level, whereas a central issue in this literature body is that innovation is spatially localized processes. In the neo-Schumpeterian cluster theory focuses on network theories of innovation and the definition of regional innovation, collective learning and local entrepreneurial surrounding. For this reason, much of this literature highlights the importance of districts and clusters heavily rely on high-technologies.

Loasby (1999) argues the main forces of local spillovers often consisted of cost advantages in logistics or a dedicated infrastructure, a pool of experienced and educated labor force, an educational facilities of distinctive relevance, a hub of specialized suppliers. However other aspects should be taken into account analyzing much deeper when including many of not obviously measured factors such as rivalry, information costs, institutional factors and various positive spillovers along with the vertical and horizontal cluster dimensions.

Tesfatsion (2003) gives insights into the meaning and the definition of a complex adaptive system representing the entire economy that phenomena associated with decentralized market economies, such as inductive learning, social network formation, the evolution of individual behavior within specific groups, imperfect competition of economic institutions. A challenging issue motivating research in the area of economic network formation is the manner in which economic interaction networks are determined through deliberative choice of partners as well as by chance.

Hidalgo and Hausmann (2009) point out that economic agents are specialize in different activities. Their development is associated with an increase in the number of processes and with the complexity that emerges from the interactions between them. The view of economic growth and development represents as the complexity of a country's economy by interpreting trade data as a bipartite network in which countries are connected to the products they export. It shows that it is possible to quantify the complexity of a country's economy by characterizing the structure of this network. The measures of complexity are correlated with a country's level of income and it is deviations from this relationship are predictive of future growth. Countries are heading to converge the level of income by the production complexity. Thus, the productive structures indicate that efforts in innovation should focus on creating new conditions that could allow complexity to contribute to sustained growth path and social prosperity as the result.

1.2.2. Context of heterogeneous economic agents

However, the discussion of emerging models in economy has been intensifying and debated in the literature body over the last decades. Brian (2006) pointed conceptually out that traditional studying equilibrium patterns of consistency required further behavioral adjustments. With time passing by economists begins to study the emergence of equilibrium and the general unfolding of patterns in the economy, which motivated to study the economy out of concept of equilibrium. The way of doing economics calls for an algorithmic approach by involving a deeper approach to agents' reactions to change. Guvenen (2011) reviewed macroeconomic models with heterogeneous households.

Algorithmic approach recognizes that agents are naturally heterogeneous, what rose complexity of economic process demanded more sophisticated methods, which should

explain individual behaviors collectively create an aggregate outcome and their reaction to this outcome. Such individual and group behavior creates pattern and pattern in turn influences behavior. This differs from the equilibrium approach tends itself to expression in equation form whereas by definition a pattern that doesn't change. Such simplicity that makes analytical examination possible has a shortcoming. To ensure tractability such models assume in general homogeneous agents or at most two or three classes of agents. The agent behavior assumed that is intelligent but has no incentive to change. Hence it should be assuming that agents and their peers deduce their way into exhausting all information they might find useful, so they have no incentive to change. Out-of-equilibrium systems may converge to or display patterns that are consistent, where standard equilibrium behavior becomes a special case.

Back to the modern macroeconomics formalized by equilibrium equations as illustrated by James E. Hartley (1996) and James E Hartley (2002), who argues the representative agent models abound, where instead of modelling the behavior of millions of different consumers and thou-sands of firms, one usually studies instead the decision problem of the representative economic unit and applies the results to aggregate quantities. Representative agent models allow the researcher to avoid the Lucas critique, they are of help in the construction of Walrasian models, and they may be used to establish micro foundations for macroeconomic analysis. However, Bruun (2004) claims that in the general equilibrium and Keynes theories heavily used the principle of representative agent is rather difficult to formalize heterogeneity by introducing one or few right agents in order to establish a link between the micro and macro world. The view of the economy developed by Kirman (2004) and Kirman (2010) represents heterogeneous interacting agents who collectively organize themselves to generate aggregate phenomena that cannot be regarded as the behavior of some average or representative individual. There is an essential difference between the aggregate and the individual and such phenomena as bubbles and crashes, herd behavior, the transmission of information and the organization of trade are better modelled in the sort of framework suggested here than in more standard economic models.

Another aspect of processes formalization is related to a bounded rationality. The definition of bounded rationality given by Jones (2003) asserts that decision makers are in general rational, so they are goal oriented and adaptive, but because of human beings cognitive and they have emotional constitution which might be occasionally the cause to fail by making important decisions. In politics science this conception has an important implication. In structured situations, at least, we may conceive of any decision as having two components: environmental demands and bounds on adaptability in the given decision-making situation.

Standard statistical techniques give the tools to distinguish systematic from random factors, so in principle it should be possible to distinguish the rational, adaptive portion of a decision from bounds on rationality. Simon *et al.* (1992) looks deeper into phenomena of bounded rationality for economics interpreting the role of individuals' decision making and debating details of information awareness. The aspect of bounded rationality is not reflected in traditional equilibrium-based modelling but it can be seen from the game theory point of new to a certain degree. Matsushima (1997) argues the game theory for economy

is not dealing with bounded rationality directly and the incorporation of the principles based on bounded rationality for agents have been seen by Matsushima as important issue to analyses the agents' behavior.

Another discussion raised by Colander *et al.* (2008) exhibit a strong undercurrent of opposition to modern macroeconomic models that have coalesced around dynamic stochastic general equilibrium models. Colander *et al.* (2008) study also supports the main ideas of Kirman (2004), Brian (2006), Kirman and Zimmermann (2012) of enormous ad hoc assumption in the standard equilibrium models based on introspection, not on any rational empirical evidence or intuitive reasonable criteria. So any meaningful model of the macro economy should analyze not only the characteristics of the individuals but also the structure of their interactions. The advantage of the agent-based modelling and simulation approach for macroeconomics in particular is that it removes the tractability limitations that so limit analytic macroeconomics. Agent-based modelling and simulation modelling allows researchers to choose a form of microeconomics appropriate for the issues at hand, including breadth of agent types, number of agents of each type, and nested hierarchical arrangements of agents. It also allows re-searchers to consider the interactions among agents simultaneously with agent decisions, and to study the dynamic macro interplay among agents.

1.3. Foundations of uncertainty in economics theories

1.3.1. Uncertainty phenomena in macroeconomic studies

In order to implement meaningful decision support systems there is vital important to incorporate the notion of economic uncertainty deeply rooted in the economic theory. Understanding of the nature of economic uncertainties contributes gives clear path how decision support systems should treat uncertainty in economic decision-making process.

In economic theory there is already a long history of studies attributed to risk and uncertainty both macro and microeconomic phenomenon. The importance of these theoretical hypotheses was recognized back to Knight (2012 [1921]) on *Risk, Uncertainty and Profit*, where the pioneering study developed ability to compresence entrepreneurship and the analysis of market mechanisms. The further initiation of an economic science considering uncertainty and behavior towards principles of risk into account is however endorsed to the revolutionary book by von Neumann *et al.* (2007 [1947]), who was without doubt the father of game theory. Since then, the most important advances in micro-economic theory were closely associated to the theories of risk, uncertainty and information. Therefore, nowadays the theory of economic is able to enlighten sophisticated issues of how economic agents, for instance, manufacturers, households and investors would behave under given circumstances. The same way it is possible to inspects how resources would be allocated and market imperfections arise. From the economic point of view there are discussion between two main concepts of uncertainties Knightian and non-Knightian. The Knightian uncertainty predominantly seen in economic studies as a hypothesis related to an aggregate and is not directly measurable. Uncertainty is unobservable and not directly measurable. However, models might rely on proxies in order to evaluate its changes in time.

In the Figure 5, the Author summarizes, that the uncertainty might have various economic effects on economic agents through various channels. The whole economic theory distinguishes three basic aggregated economic agents, who undertake various economic actives including but not limited to production, consumption and exchange. The role of firms is to take care of production decisions including production quantity of goods, means of production and price settings. Firms in general take disposable factors of production in order to sell own goods for consumption to the households or government. Households in general decide on goods consumption and provide own factors of production to firms. The government impose taxes individually or collectively on firms and household by both fiscal and monetary policy. Therefore, the channels are defined by the economic agents' activities.

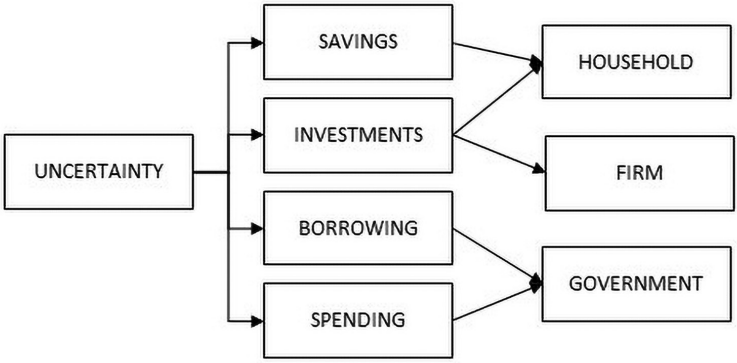


Figure 5. *Uncertainty by channels and variables*
(Source: Author's representation)

The reviewed literature gives detailed insights into this issue. Onatski and Williams (2003) argues that uncertainty is persistent phenomena in economics and it must be faced continually by policymakers. Black *et al.* (2018), Meinen and Röhe (2017) supports that measuring macroeconomic uncertainty and understanding its impact on economic activity is thus crucial for assessing the current macroeconomic situation and forming a view on the outlook. It is a wide view in respect on macroeconomic phenomena such as uncertainty of current and future real GDP growth. It comprises also microeconomic issues of uncertainty about the attitude for firm growth or the projections of household income.

The channels represent macroeconomic systematic uncertainty that cannot be avoided through diversification (Lars Peter Hansen (2013)). The main characteristic of systemic uncertainty that it is not amenable to quantification (Gilboa and Schmeidler (1989), Gilboa *et al.* (2012)). Therefore, L. P. Hansen and Sargent (2001) argues, that statistical models as approximations confront probability of potential for misspecification and seek conservative responses. Beyer *et al.* (2017) argues that the analysis of the channels of transmission can provide valuable insights regarding these interactions and raises the question of how the efficiency and effectiveness of the policies in achieving their objectives may be

affected in influencing the economy. Fischer (1993) asserts that the usual emphasis on economic stability suggests that the main reason macroeconomic factors matter for growth is through uncertainty. The growth might be affected by uncertainty thought two main channels. First, policy-induced macroeconomic uncertainty reduces the efficiency of the price mechanism. This uncertainty, associated with high inflation or instability of the budget or current account, can be expected to reduce the of productivity, and, in contexts where the reallocation of factors is part of the growth process. Second, temporary uncertainty about the macroeconomic tends to reduce the rate of investment, as potential investors wait for the resolution of the uncertainty before committing themselves. This channel proposes that investment tend be get lower while uncertainty is high. Decrease in capital thought its relocation tends to increase domestic instability and gives another channel through which macroeconomic uncertainty shrinks investment in the domestic economy.

Black *et al.* (2018) argues that uncertainty in its various forms is widely known as the factor that influence economic while it is difficult to measure. One common technique is applying proxy variables. There is a number of proxies intended to measure effectively different layers of environmental settings such as financial and political. Often in the empirical literature these proxies represented as a measure of the uncertainty impact on the economic activity in forms of industrial production, GDP, investment or consumption. Hence, that these proxies are definitely vulnerable. The practical measurement of uncertainty should include an encompassing set of datasets. In contrary, Kjellberg and Post (2007) evaluates general macroeconomic uncertainty with the effects of ambiguity on aggregate consumption and residential investment. This can provide some further evidence on the usefulness of available proxies. The cost of the economic crises is enormous. Researchers working with both economic and political issues provided several theories of the crisis. So far, no dominant consensus has been elaborated yet but there are conflicting proposals on how to prevent another crisis.

Romer (2018) tackles the phenomena of uncertainty in his introductory study of the economy on example of models' multiplicity. Starting with the canonical model of optimal growth formulated by Ramsey (1928), Cass (1965) and Koopmans (1965) describes rival firms borrowing capital and employ labor force to generate output and it sell on the market under presumption of having a given number of households infinitely living to provide labor, save capital and consume. For simplicity sake the model does not deal with imperfections, heterogeneous agents and generations. The economy is under perfectly competitive production condition in terms of an aggregated Cobb-Douglas production function (1.3.1) in order to give output $\boldsymbol{Y}$ using capital $\boldsymbol{K}$ and labor $\boldsymbol{L}$:

$$\boldsymbol{Y = F(K, L) = K^{\alpha}(AL)^{1-\alpha}} \qquad (1.3.1)$$

Simultaneously, there is a large number of equal households grow at the rate n. The household's utility function takes the form under condition of renting capital K to firms as the optimization problem of balance between consumption and saving to maximize utility:

$$\boldsymbol{U = \int_{t=0}^{\infty} e^{-pt}\, u\big(C(t)\big)\frac{L(t)}{H}\, dt} \qquad (1.3.2)$$

Solving the given utility function to its functional form it is possible to find a balanced growth path:

$$u\big(C(t)\big) = \frac{C(t)^{1-\theta}}{1-\theta} \qquad (1.3.3)$$
$$\theta > 0$$
$$\rho - n - (1-\theta)g > 0$$

Thus, the concept of *constant relative risk aversion* (CRRA) is a predominant method in macroeconomics, which proposes that consumption preferences are integrally separable over time Equation (1.3.2). From Equation (1.3.3) the CRRA coefficient defined as $-Cu''(C)/u'(C))$ for utility function θ and therefore independent of C. There is no uncertainty in the context of Ramsey model. But there is a clear bias towards avoiding inequality over time periods to keep continuity in the utility function if mathematically seen. In macroeconomics it leads to elasticities effects on balanced growth paths when there is growth in productivity. Again, since there is no explicit uncertainty in this model, but it is obvious that the model should deal with the risks determined by θ from Equation (1.3.3). First of all, households might shift consumption between different periods. Then the investments risks will emerge where returns on borrowed capital in the accumulation phase might decline. The result will be lower than expected accumulated savings. Another chained effect is an annuity risk of low conversion rates.

It became clear from scientific point of view that definition of uncertainty required more sophisticated approach. Thus, in the literature since the early 1960s the issue of optimal accumulation and dynamic efficiency has been recognized as one of the key issues in economics due to the impact of uncertainty on production, saving, investment and economic growth.

Diamond (1965) influenced by Samuelson (1958), Malinvaud (1953) proposed a different perspective of the fundamental theorem characterized by markets moving toward a competitive equilibrium with the Pareto efficient principle under conditions of complete markets, price-taking behavior and non-satiation of preferences. Proposed assumption of a *representative household* and overlapping generations in the equilibrium condition of perfectly competitive market makes Pareto suboptimal even without any market failures and distortions. These key assumptions make major difference between the Ramsey model and the Diamond model. However, persistent entry of new households into the economy setting and infinite planning horizon might provide only approximated framework for investigating the macroeconomic effects and intertemporal allocations on pension schemes and aggregate savings. Fama and French (2002) proved that the main concern remains the same when the uncertainty is represented in terms of various shocks. The Diamond model requires deeper investigation since the constant environmental changes observed on concrete evidences. The evidences are characterized by constant demographic development, which entailed decreasing population growth and increasing life expectancy in the developed countries. Another aspect there is while maintaining capital accumulation, the interest rate will be below the economic growth rate. This will require more investments

to hold market equilibrium requires more investment, what leads to disparities. For the households, the over-accumulation condition will exceed the optimum for maximizing consumption, then at the low interest rates will disable intertemporal transfer of financial resources between generations, what might be a trigger for Ponzi-schemes and Pareto sub-optimum.

The uncertainty represented qualitatively in spreads and developments of interest rates known as the best ultimate measure of the practical feasibility. Therefore, dynamic efficiency requires further fundamental analyses of economic growth. The actual efficiency becomes of great significance for policy implications. The nature of the interest rates is a huge topic in the research field. The interest rates development is the subject of empirical analyses with explicit characteristics of particular financial settings systems. Methods for analysis use the whole range of available techniques from accounting analysis, applying regressions to sophisticated econometric models. L. Hansen and Singleton (1983) argue that the interest rate analysis implies the predictable element of aggregate consumption. The topic of the researches are to determinate consumption growth responds to dissimilarities in the development of the real interest rate. The findings of the researches show significant the intertemporal elasticity of substitution resulting weak responses.

Decades later Abel *et al.* (1989) propose a practical cash-flow-based efficiency measure underpinned by theoretical framework and seem empirically feasible. As the result, the research admits that the dynamic efficiency despite its deepness of research investigation and the growing concern over capital formation, there is no clear answer whether or not actual economies are dynamically efficient. However, the result of the study stated that if the capital market increases the level of consumption, then the economy can be seen as dynamically efficient otherwise economy is inefficient. The very latest evidences from Pozzi (2005) examines in the period 1952-2001 in the US in terms of the significance income uncertainty on aggregate consumption with the result that aggregate income risk might describe only an insignificant fraction of the variance. Thus, the aggregate consumption changes belong the unobserved component.

Expected utility theory by von Neumann *et al.* (2007 [1947]) describes agents tackling risk maximize as expected value of the utility of their wealth. D. M. Kreps and Porteus (1979) and further D. Kreps (2018 [1988]){Kreps, 2018 #771;Kreps, 1979 #773} gives detailed view and application of expected utility as a sequential decision problem solved by a dynamic programming recursion function with given preferences and uncertainty.

Decision support systems integrate ensemble models to comprise all influential factors on the decision-making process. The Figure 6 shows how the uncertainty represented in studies for investigation. The formal economic models deal primarily with four types of them used in economic analysis, visual models, mathematical models, empirical models, and simulation models.

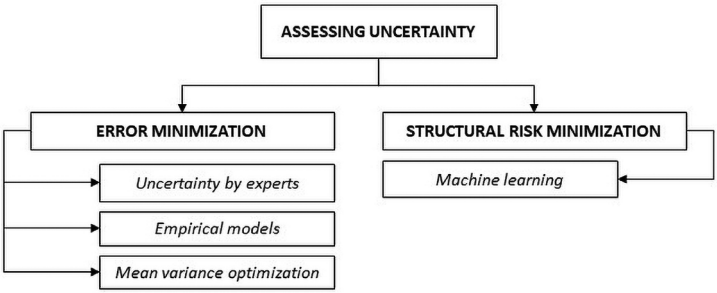


Figure 6. *Assessing uncertainty in models*
(Source: Author's representation)

The Author distinguish that the most formal and abstract of the economic models are the purely mathematical models. The equations systems should be solved simultaneously with an equal or greater number of economic variables. The majority of the models applied in economics belong to the comparative statics models. However, more sophisticated modelling approaches in macroeconomics and business cycle analysis explore dynamic processes. Gollier (2018) in his research underpins the importance and significance of risk and uncertainty in the economic processes especially focused on the decision-making environment. His innovative approach is to shift his attention from simplified utility functions to find solutions in complex issues of decision-making and equilibrium under uncertainty. The sophisticated Neumann and Morgenstern model includes the agent's preferences under uncertainty satisfy condition. This application is widely investigated in the recent literature body by Nishimura and Ozaki (2017), Guiso *et al.* (2018), Gonzalez-Soto *et al.* (2019), Eberlein and Kallsen (2019).

1.3.2. Uncertainty in error minimization models

Formal uncertainty models with error minimization belong to the non-Knightian uncertainties, as the models have to be related to measurable evidences. Taking into account this fact, they cannot be directly attributed for assessing the general state of uncertainty in a macroeconomic. This has to be done with the use of the concept of Knightian macroeconomics uncertainty which related to non-measurable general state of the economy with different approaches.

Historically, while scrutinizing the triggers of decision making, the researches include risk-aversion as the impact factor of general economic uncertainty on investment. Theoretical papers on this issues show that a depressing effect of uncertainty include increased managerial risk-aversion (Bloom (2014), Baker *et al.* (2015), Bloom *et al.* (2007)). Economic agents' expectations might have dramatic effect subsequently on economic developments through different channels by influencing prices, consumption and investment decisions. Surveys among professional forecasters also allow a quantification of aggregate and individual forecast uncertainty (Black *et al.* (2018)). Therefore, channeling expectation through SPF has

very important role in the overall assessment of the risk factors and uncertainties from the point estimates and the perception of risks around point estimates. Additionally, around the point probability distribution has to be estimated in order to bring to light to the uncertainty characterized by individual forecaster. Whereas the average standard deviation of the individual probability distributions given by the forecasters is an aggregated individual measure of uncertainty. The point estimates give insight into the economy evolution, observed shocks or involve in their assumptions (Carroll (2003)). A long-term point inflation expectations can be used to measure effectiveness of monetary policy. Risk perceptions, disclose evidences on the expected distribution of economic shocks with focus on assessing the strength of the longer-term inflation expectations (Kowalczyk *et al.* (2013) and Clements (2014)).

Born *et al.* (2018) argues, that one would like to know the subjective probability distributions over future events from firms and households. However, this is almost impossible to quantify directly, there exists no agreed measure of uncertainty in the literature. But the literate body, the same way macroeconomic uncertainty can be categorized based on its application, measurement methodology and assessment strategy.

Uncertainty by Gabbay and Smets (2013) is a characteristic of the state of knowledge of the agent about which of the possible worlds is the actual world and an added information that expresses the idea that the truth of some propositions is better supported than the truth of others. From formal point of view, it is an extra information that gives weights to the various subsets. Oberkampf *et al.* (2001) argues, that modern theories of uncertainty can represent much weaker statements of knowledge and more diverse types of uncertainty than traditional probability theory. The study distinguishes *aleatory* uncertainty is also referred to in the literature as variability, irreducible uncertainty, inherent uncertainty, and stochastic uncertainty and *epistemic uncertainty* is also referred to in the literature as reducible uncertainty, subjective uncertainty, and model form uncertainty.

Last decade raised a lot of new methodologies and data-driven approaches in the field of uncertainty. Kelleher *et al.* (2015) Modern organizations collect massive amounts of data. Based on the data extraction and analysis, the resulting insights can be applied for a better decision-making process. Predictive analytics is the assignment of creating and using various models and approaches that can forecast based on recognized patterns obtained from historical datasets. Risk factor is in almost every decision made, which can be avoided by utilizing predictive models to predict the risk related to decisions. Niaf *et al.* (2011) propose to deal with these uncertainties by introducing probabilistic labels in the learning stage so as to stick to the real life annotation problem, avoid discarding uncertain data and balance the influence of uncertain data in the classification process. Dutt and Kurian (2013) argues the need for handling uncertainty increases to incomplete information and unpredictability. The study highlights are many techniques to analyze the uncertainty from machine learning perspective: probabilistic analysis, fuzzy analysis, Bayesian Network analysis, soft computing technique and rule based classification technique. They argue, that among these the probability analysis, fuzzy analysis and Bayesian Network analysis are the most technically challenging techniques.

Hirshleifer (1965) introduces the standard model of decision-making under certainty of a choice set $\boldsymbol{C}$ available to decision-maker with an ordering $\preccurlyeq \boldsymbol{C}$ over the choice with

preferences and a behavioral hypothesis $\boldsymbol{C}*$. Therefore, decision-making under uncertainty entails consequences as the result on the choice of possible direct actions A for the choices out of set $\boldsymbol{C}$ in the environment W. The process of making choices under uncertainty relates to the disposable decision-maker's knowledge of the world in which they have to act, which leads to the objective probabilities about awareness of the likelihood of the current situation and it can be formulated as the probability $\boldsymbol{p}(\boldsymbol{c}|\boldsymbol{a})$, where $\boldsymbol{c} \in \boldsymbol{C}$ and $\boldsymbol{a} \in \boldsymbol{A}$, so a probabilistic picture of the uncertainty $\boldsymbol{W}: \boldsymbol{A} \rightarrow \boldsymbol{C}$. Choices made my agents knowing the probabilistic information about the environment and consequences can be described as risk category formulated with probability distribution over outcomes. While decisions under risk agents poses overall knowledge of the likelihood in each state. If decisions under uncertainty implicate choices between actions that have consequences depending on an unknown environmental conditions.

Considering hypothesis of expected value maximization in a risk situation, if $\boldsymbol{a}_* \in \boldsymbol{A}$ is chosen, then $\boldsymbol{EV}(\boldsymbol{a}_*) \geq \boldsymbol{EV}(\boldsymbol{a}), \forall \boldsymbol{a} \in \boldsymbol{A}$, the preferred action is expressed in expected value terms.

If hypothesis of expected utility maximization is taken into account, then in a choice situation, a decision-maker will take an action as $\boldsymbol{a}_* \in \boldsymbol{A} \backsim \boldsymbol{EU}(\boldsymbol{a}_*) \geq \boldsymbol{EU}(\boldsymbol{a}), \forall \boldsymbol{a} \in \boldsymbol{A}$. The expected utility hypothesis is the major descriptive theory of individual choice under conditions of risk or uncertainty (Röthig (2009), Acemoglu (2008), Rachev *et al.* (2008), Müller (2016))

The nature of decision-making is complicated due to involvement of contradicting payoffs which are the general subject to risk and uncertainty (Bryant (2014)). Despite this significant change in context from risk to uncertainty, the central question remains essentially the same of how the economic agents adjust actions and then make a selection considering limitations of probability represented by the concept risen from *Ellsberg* or *Allais* paradoxes (Segal (1987), Matsushima (1997), Gilboa *et al.* (2012), Firoozye and Ariff (2016)).

Wide range of authors seen financial proxies for measurement of macroeconomic uncertainty (Bloom (2009), Haddow *et al.* (2013), Bekaert *et al.* (2013)). All these theories are grounded on the idea of Knightian uncertainty. Microeconomic theory uncertainty can be decomposed in its components such as inflation, labor market, output, consumption and measured separately. Galati and Moessner (2013) argues risks factor in macroprudential policy during systematic crisis. It is underpinned that the financial crisis has stressed out necessity to overcome a merely micro-based attitude to financial regulation. Narrowing to the financial area risk representing difference of actual returns on an asset and their difference from expected return. However, from the economic perspective the non-Knightian uncertainty can be modeled in terms of probability distribution of the *ex-ante* factors.

Jurado *et al.* (2013) , who use large scale dynamic factor model with stochastic volatility to extract joint forecastable component from 279 macroeconomic and financial indicators allowing for idiosyncratic shocks in each of the indices. The study defines macroeconomic uncertainty representatively:

> *"the conditional volatility of a disturbance that is unforecastable from the perspective of economic agents"*

Hwang *et al.* (2016) suggests the cooperative model to avoid lack of precision in the parameters of the linear production problem by modeling fuzzy logic. A far better way is offered to include machine learning techniques. After the process of data mining it is possible to distinguish two sources of randomness and uncertainty. The randomness and uncertainty underlying the process itself, and the uncertainty associated with underlying data collection methods (O'Neil and Schutt (2013), Marwala and Hurwitz (2017).

From modern positions a robust and negative effect of uncertainty on economic growth is obvious and these consequences cannot be neglected by the theory (Lensink *et al.* (1999), Levin *et al.* (2005), Ljungqvist and Sargent (2012)). These results of crises underline the importance of growth sustainability and policy credibility. Negative effects make firms more cautious when investing or disinvesting under uncertainty. In this case the policy effectiveness has multiple first-moment negative effects and uncertainty is also strongly countercyclical at the industry level (Bloom *et al.* (2007) and Bloom *et al.* (2018)). Mosini (2008) and Binder *et al.* (2017) summarize that uncertainty, limited information, bounded behavior and other phenomena cannot be incorporated into general equilibrium theory completely. Thus, the discussion of benchmarking model under uncertainty raised to another level how to learn best practices, organize and coordinate production and motivate performance (Bogetoft and Otto (2010) and Bogetoft (2013)).

Accordingly, the non-Knightian uncertainty concept is coupled with the ex-post effects of business cycles. There are a vast number of studies arguing indicators of uncertainty which can be viewed as representative to the evidences of particular policy, involving a wide number of direct and indirect peers (Ericsson *et al.* (1999), Benhabib *et al.* (2013), Bird *et al.* (2013), Ernst and Viegelahn (2014), Baker *et al.* (2015), Jurado *et al.* (2015)). Rachev *et al.* (2008) gives the definition of risk as a subjective phenomenon involving exposure and uncertainty. The risk factor takes place when uncertainty exists. Arguments of Knight (2012 [1921]) postulate, that the non-Knightian uncertainty from macroeconomic perspective should be presumed as risk factor. The theory of decision-making under risk presumed objective probabilities over states of the environment. The choice under uncertainty where objective probabilities over states are not available decision-makers are therefore typically forced to make subjective judgements about the likelihood of various events and states in order to select an action.

Kahneman and Tversky (1979) exhibit in response to their observation that the choices made by individuals in risky situations have several characteristics that are inconsistent with utility maximization. The certainty effect is that individuals underweight probable outcomes in comparison with outcomes that are certain. The study also observed that this effect can lead to risk-aversion in choices involving certain gains and risk-seeking in choices involving certain losses. Secondly, isolation effect is when individuals facing choices among different prospects disregard components that are common to all prospects under consideration. This effect can cause the framing of a prospect to change choices. Thirdly, individuals display a reflection effect in which choices involving negative prospects and positive prospects are treated equivalently.

From theoretical point of view, Coelli *et al.* (2005), Charnes *et al.* (2013), Paradi *et al.* (2017) disclose methodological approaches that can encompass environmental uncertain-

ties through variables in nonparametric models. The first widely accepted method proposed by Banker and Morey (1986) and Banker *et al.* (2012) involves the environmental proxies categorized by the order of their influence on efficiency. Each organization in a sample can be compared against their peers with the environmental variable, which is respectively less than others. It helps to avoid comparison of organizations which have competitive advantage due to more beneficial environmental settings.

Represented by Drake *et al.* (2006) employed the multi-stage nonparametric modeling to include externalities and environmental influences into analysis. As before the multi-stage nonparametric modeling approach estimates model with conventional inputs and outputs factors. Thus, the estimation of slacks of the resulting model adjust the initial multi-stage nonparametric model and analysis starts over again.

The environmental analysis requires more country-specific and macro-level factors. In order to describe the current business settings in adequate quantitative form, the factors of market sustainability, presence of investment to GDP, fiscal parameters, GDP growth. Fethi and Pasiouras (2010) investigate the level of the relationship between the technical efficiency and governmental regulation imposed on market.

1.3.3. Structural risk minimization

Ensemble methods in machine learning can be seen as a nonparametric approach in terms of parameters defined by the capacity of the model, which is data-driven to match the model capacity to data complexity. This is a basic paradigm of the structural risk minimization (SRM) suggested in seventies of the last century and further developed by Vapnik (1992) and Vapnik (2013). Cao and Tay (2003) research goes deep into the paradigm of SRM copes with uncertainty classification problem by minimizing an upper bound on the expected risk, over each of the hypothesis classes by the generalization error consisting of the sum of the training error and a confidence interval. The SRM principle is originated from computational learning theory. The whole concept is built upon the idea of seeks to minimize an upper bound of the generalization error rather than minimize the training error. A linear structure of learning algorithm is implemented to apprehend non-linear class boundaries through extremely non-linear mapping of the input vectors into the high-dimensional space.

Drucker *et al.* (1997), Cristianini and Shawe-Taylor (2000), Scholkopf and Smola (2001), H.-C. Kim *et al.* (2002), H.-C. Kim *et al.* (2003), Smola and Schölkopf (2004), Welling (2004), Wang (2005), Gu and Han (2013), Ma and Guo (2014) defines the learning problem setting for machine learning as an unknown and nonlinear dependency:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) \qquad (1.4.3.1)$$

between some high-dimensional input vector $\boldsymbol{x}$ and scalar output $\boldsymbol{y}$, without information about the underlying joint probability functions. A distribution-free learning is when information available in a training data set:

$$\boldsymbol{D} = \{(\boldsymbol{x_i}, \boldsymbol{y_i}) \in \boldsymbol{X} \times \boldsymbol{Y}\}, \boldsymbol{i} = \boldsymbol{1}, \boldsymbol{l} \qquad (1.4.3.2)$$

, where $\boldsymbol{l}$ stands for the number of the training data pairs and is therefore equal to the size of the training data set $\boldsymbol{D}$. This machine learning problem is analogous to the classic statistical implication but with a few significant differences in approaches in training. Classic statistical implication has the following fundamental assumptions:

1. Linearity in parametric paradigm in learning from experimental data,
2. A stochastic component of data is the subject of the normal probability distribution,
3. Parameter estimation is the maximum likelihood method reduced to the minimization of the sum-of-errors-squares cost function

Originated from the statistical learning theory developed by Vapnik and Chervonenkis (VC), The concept of *Support Vector Machines* (SVM) represent relatively novel techniques introduced in the framework of SRM, whereas instead of minimizing the absolute value of an error, SVM perform SRM, minimizing VC dimensions. Vapnik proved the correlation between low expected probability and VC dimension of the model. The VC dimension is a property of a set of approximating functions of a learning machine that is used in all important results of statistical learning theory. Since SVM has become widely established as one of the leading approaches to pattern recognition and machine learning. The method utilizes a linear combination of kernel functions for predictions in terms of a pinpointed on a subset of the support vectors in form of training data. By using this induction principle makes the difference with ERM which minimizes only the error on training datasets (Shawe-Taylor *et al.* (1998), Cao and Tay (2003)).

Established on the unique theory of the structural risk minimization principle to estimate a function SVM is shown to be very resistant to the overfitting problem, eventually achieving a high generalization performance. Due to the advantages of SVM algorithm in solving nonlinear problems, it can be used to capture and provide explanatory power of underlying uncertainties (K.-j. Kim (2003), Huang *et al.* (2005), Ahmad *et al.* (2014)). Besides SVM classifiers and regressions there is another approach known as decision trees and related ensemble methods like random forest, which are getting popular tools in the field of machine learning for predictive regression and classification. Breiman (2001) and Torgo (2016) investigate Random Forest as an example of an ensemble model, that is, a model that is formed by a set of simpler models. However, they lack interpretability and can be less relevant in practical applications, where decision-makers and regulators need a transparent linear function that usually corresponds to the link function in logistic regressions (Dumitrescu *et al.* (2018)).

Q. Zhang and Wang (2018) proposed efficiency prediction model which for the first time combines information granulation and machine learning with nonparametric model, to evaluate the future efficiency of decision making unit. The model implements fuzzy information separation in order to separate the input-output time series data. The description of the data characteristics within each time frame is represented by the minimum, average and maximum values and established the IG - SVM model. The model is based on the fuzzy information separation and support vector machine. Following the training process

if time series data, the optimal model of regression can be derived. This model explicitly defines the minimum, average and maximum values of the next future frame predicted. The future efficiency of the DMU can be premeditated by the DEA model.

Abe (2010) argues that the results of the classifier should be clearly interpretable for decision support process. Otherwise a classifier would not find its application if even it has a high degree of generalization ability. However, support vector machines possess high generalization ability weigh against other classifiers algorithms but their interpretability is fairly weak, particularly when using nonlinear kernels. Therefore, advancing in interpretability plays enormous important role for the support vector machines.

1.3.4. Uncertainty in expert meta-analysis

Expert meta-analysis comprises a number of independent researches of the identical issue in order to derive the overall global trends. This approach relates to measurement of economic policy uncertainty rather than to macroeconomic or financial uncertainty and is pioneering by Baker *et al.* (2015) who developed the Economic Policy Uncertainty and alternatively Jurado *et al.* (2015). The Economic Policy Uncertainty index consists of three components, the main of which is frequency of newspapers references to economic policy uncertainty, other two components are based on tax provision and disagreement among professional forecasters. A variation of the Economic Policy Uncertainty index based on news only is also published. News based Economic Policy Uncertainty index considers number of articles that include words *economic* or *economy* and *uncertain* or *uncertainty* and *regulation* or *deficit*, or foreign reserve or congress or legislation or White House' in 10 major US newspapers. The index, therefore, captures the uncertainty related to who, what and when undertakes economic policy actions and what might be an economic effect of this policy. The Economic Policy Uncertainty index provides reliable proxy for economic policy-related uncertainties and is widely used in applied research for identification of policy uncertainty shocks Bernal *et al.* (2016), Istrefi and Piloiu (2014). Among other problems researches brought large step-ups in general understanding of how individual agents decisions interact in a market economy. But beyond that, there a number of economic concepts give wide aggregations of economic complex reality which simply could not be thoroughly yet realistically analyzed in the absence of a risk theoretical framework. It started in the fifties with the portfolio theory of Markowitz *et al.* (2000 [1952]) of the mean-variance approach of investment portfolio diversification led financial option paradigm and the microeconomics of information are merging into a comprehensive theory of contracts and agency problems.

Bomberger (1996) as commented by Rich and Butler (1998) explains disagreement as a measure of uncertainty approach, which widely accepted and therefore offspring probabilistic forecasts by experts by formulation their findings not only about expected outcome of the forecasted variables, but include the probabilities. The reason for popularity is stimulated by increased number of panel-type databases. It led to the result that the forecasting processing became more accessible, created an additional methodological questions on measure of the uncertainty means and uncertainty distribution by Giordani and Söder-

lind (2003), Diebold *et al.* (1997) and Clements and Harvey (2011). Referring to the classification of Walker *et al.* (2003) it is an epistemic uncertainty, generated by incomplete knowledge of the system by the experts, where it encompass a feature of fundamental variability. The uncertainty can be measured by empirical models, where ARCH class models play a central role. In this models conditionally-autoregressive errors are associated with uncertainties Elder (2004), Kontonikas (2004), Daal *et al.* (2005), Fountas (2010), Henry *et al.* (2007), Neanidis and Savva (2011). Berument *et al.* (2009) and Hartmann and Herwartz (2012) extend the standard assumption with stochastic volatility models. However, Orlik and Veldkamp (2014) and Glass and Fritsche (2015) argue that uncertainty is an outcome value of acyclical changes in uncertainty while shocks. Zarnowitz and Lambros (1987), Bomberger (1996), Rich and Butler (1998) and D'Amico and Orphanides (2008) argue the measuring epistemic uncertainty between the forecasters by direct estimation of parametric distributions characterizing the uncertainty across individuals. Lahiri and Sheng (2010), Siklos (2013), Lahiri *et al.* (2015) extend the model by to numerous improvements and modifications. Walker *et al.* (2003), Dequech (2004) look into epistemic uncertainty caused by experts incomplete knowledge and the variability uncertainty attributed to accidental factors randomly appeared. Lane and Maxfield (2004) extends the variability uncertainty with the ontological uncertainty. Discussion raised by Walker *et al.* (2003) classification goes into inflation uncertainty by Norton (2006), Kowalczyk (2013), Krayer von Krauss *et al.* (2019). Gelman and Hill (2007) introduces multilevel linear and generalized linear model in which the parameters are given a probability model. This second-level model has hyperparameters parameters of its own which are also estimated from data.

Another way of gaining uncertainty from models is to use a sensible forecasting modelling based on the distribution assumption of *ex-post* forecast errors. An obvious measure of uncertainty is variance of such distribution Knüppel (2014), Jordà *et al.* (2013). This approach is not well supported so far by economic theory, but is popular among the researchers. The growing and the most resent literature on a various approached in this direction include, Faust and Wright (2007), Monti (2010), Rich and Tracy (2010), Kowalczyk *et al.* (2013), Krüger *et al.* (2017), Jo and Sekkel (2016).

Fildes and Stekler (2002) stress out that macroeconomic forecasts are used extensively in industry and government. Issues discussed include the comparative accuracy of econometric models compared to their time series alternatives, whether the forecasting record has improved over time, the rationality of macroeconomic forecasts and how a forecasting service should be chosen. Typically, these error measures have focused on the point forecast alone. Recently, attention has also been given to the uncertainty around the published forecasts. The question of whether estimates of the uncertainty in a point forecast are well calibrated or, more generally, the estimated probability distribution matches the realized distribution has received relatively little attention as there are little data available.

Abdou and Pointon (2011), Le Bellac and Viricel (2017) are linked to the unobservable character of the behaviors one attempts to model. The main research question does not concern the existing models but make sure their appropriate application and to inform about their limits. There is an interesting methodological suggestion in Harrison and Rutström

(2009) to abandon the search for a unique theory to explain all choices under risk and uncertainty. Their study suggest that we should aim instead for a combination of hypotheses, the weights on which correspond to the percentage of the population exhibiting expected utility behavior, prospect theory behavior. Nevertheless, it also has some disadvantages. It is costly and creates obvious difficulties in completing a competent panel of experts. As the most professional forecasts are recently probabilistic, still relatively unsearched problem of the psychological ability of an individual to express probability statements in an unbiased way has to be addressed. There is some empirical evidence, and also results of psychological experiments, suggesting that this might not be possible Soll and Klayman (2004), Hansson *et al.* (2008), So (2013). Clements (2014) argues that, in the context of survey forecasting, the panel data tend to overestimate the short-term uncertainty and underestimate the long-term one. This result contradicts, to an extent, the psychological literature quoted above making the problem even more complex.

However, the crucial problem here seems to be the joint bias of forecasts formulated by different forecasters. The potential for an existence of such bias is rather obvious, as the panelists either have access to identical sources of information, which influence them in a similar way, or may know each other and their formal or informal discussions about the state of the economy may inadvertently cause correlation of their individual forecasts. Disagreement in survey point forecasts reflects the differences in opinion rather than uncertainty Diether *et al.* (2002), Mankiw *et al.* (2003). The consequence of this can be unexpected, as it results in a relatively small dispersion between the means of the forecasters' distribution and a substantial bias, which reduces the usefulness of the results.

Other methods of assessing modelling uncertainty, also have their advantages and disadvantages. They usually do not require other data than publically available, they represent well past dependencies and are, to an extent, independent from psychologically induced trends and rumors. Current methods give an opportunity to identify the ontological and epistemic elements leading to some approximation of *ex-ante* uncertainties per Charemza *et al.* (2014).

Among the disadvantages the most relevant one seems to be the model dependence. The *aleatoric* uncertainty methods are assuming a perfect model. Clearly this can always be disputed. Also, there are often problems with associating the uncertainty with particular timing. As ontological uncertainty requires collecting data related to a considerable period of time, there is a question of time invariance, in the *ex-post* and, in particular, in the *ex-ante* context, when the uncertainty is used for the assessment of probabilistic forecasts. The most criticism is linked with the limitation in ability to link cause and effect. Application of models have important limitations, which it makes difficult to choose the right one. The reason behind is that the most significant problem to define point predictions rather than generating predictive distributions. However, some of the controversial problems might be resolved by reinforcement machine learning techniques.

Tsang *et al.* (2011) express the idea that instead of abstracting uncertain data by statistical derivatives it is possible to increase the accuracy of prediction by using classification techniques taking into account the probability density function. The classifications of uncertain data become one of the tedious processes in the data-mining domain.

However, the challenging work of Tobback *et al.* (2018) argues that original method of measuring Economic Policy Uncertainty developed by Baker *et al.* (2015) does not have any predictive power for any of its variables using conventional regression methods. The study shows that machine learning approach has a higher predictive power and changes in the level of policy uncertainty during turbulent periods of high uncertainty and risk can forecast variations in the government bond, the credit default swap yield and spread. Brose *et al.* (2014a) and Brose *et al.* (2014b) argue that managing risks and uncertainty depends critically on information. The studies exhibit that, demand on risk models have flourished as evidenced by recent events, the need has never been greater for skills, systems, and methodologies to manage risk information during crises and shocks.

1.4. Classification of performance, effectiveness and efficiency

The difference among efficiency, performance and effectiveness is enormous despite the fact all of these terms supposed to benchmark the input and output factors in terms of resources consumed and output produced. Under the business performance the Author understands the measurement of business performance following Rappaport (1986), who stated that the shareholder value should become the global standard for measuring business performance. The discussion is followed by an enumeration of the shortcomings of the accounting return on investment and accounting return on equity as standards for measuring business performance. Effectiveness can be measured as a combination of efficiency and performance. The explicit measurement of effectiveness is out of the scope of this research. Thus, particular interest of the research is to establish the link between business performance in financial terms, economic growth factors and the decision-making process. Foremost, the Author clearly defines:

1. **Efficiency** are achievements with the least possible cost and resources done within the shortest possible amount of time.
2. **Effectiveness** is the ability to achieve a desired result with an acceptable level of quality and user satisfaction and meeting any relevant standards that must be met.
3. **Performance** is measured by accomplishment to a given set standard.

The Author gathers the key differences in terms and definitions in Table 1. Different interpretation of definitions used for the benchmarking scope distinguish mainly by source in Column (2) of information and implication area in Column (3). Due to scientific nature some approaches proved to work well for scientific research, fail to operate the same way in business applications. The reason is often that the scientific researches propose to high level of aggregation and the results can be hardly interpreted by the business terms and definitions.

Table 1. *Various measurement approaches of efficiency, performance and effectiveness by source of information and area of implication*

Definition	Source of information	Area of implication
Efficiency	Scientific	Science, business
Performance	Business	Finance, business, operations, administration
Effectiveness	Marketing	Operations, business administration

(Source: Author's representation)

Karadgi (2014), Franceschini *et al.* (2007) give details insights from *business* point of view into various performance measurement approaches have been elaborated, particularly from a strategic perspective. These systems highlight the importance of non-financial or operational metrics, and linking the financial and operational metrics, among others.

On the other hand, parametric frontier models and nonparametric methods have been widely used in the recent scientific literature on productive efficiency measurement and in a large literature of studies. Empirical applications have usually dealt with either one or the other group of techniques. Since fundamental contributions by Farrell (1957), Koopmans (1952), Aigner and Chu (1968), Aigner *et al.* (1977), Broek *et al.* (1980) concept of efficiency methodology in frontier production function estimation has been rapid developed. Färe *et al.* (1994) followed by Heathfield (1995) differs two main components: Technical Efficiency and Allocative Efficiency.

Khetrapal and Thakur (2014) include linear programming methods, statistical techniques and process approaches into domain of benchmarking approaches. The benchmarking selection methods used by individually depending on the available data and the aim of the benchmarking process. The benchmarking measurement can have impact on the purpose of efficiency scores as represented in Figure 7. Programming techniques does not require specification of a production or cost function and correlate outputs to inputs without emphasized to econometric estimation. The efficiency frontier is the data-driven approach. Data envelopment analysis (DEA) and Free Disposal Hull (FDH) are two widely used programming technique, which calculates the efficiency in a given set of decision-making units. Index approaches used to determine efficiency (total factor productivity and partial) also calculate efficiency scores, and so are included in programming technique category, although they do not involve in the calculation of efficiency frontier. Econometric techniques, in contrast, require specific assumption about the relationship between the inputs and outputs, and estimate the parameters of a functional form representing this. Econometric techniques should be seen from deterministic or stochastic point of view. The deterministic frontier approach assumes that all the deviation from an estimated frontier is mainly due to technical inefficiency, with no role played by random factors. Unlike the deterministic frontier approach, a stochastic production frontier approach, however, incorporates both noise and inefficiency component into the model specification.

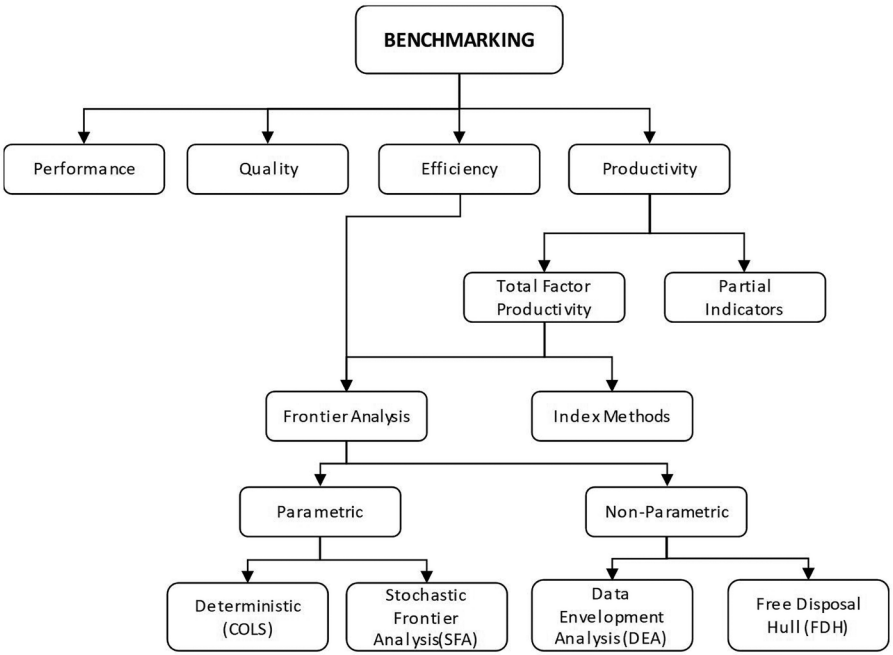


Figure 7. *Taxonomy of benchmarking techniques*
(Source: Author's adoption from Khetrapal and Thakur (2014))

DEA emerged from influential work by Farrell (1957) aimed to develop a comparative measure for production efficiency. This work extended toward DEA, proposed by Charnes *et al.* (1979) who presented a quantitative measure for assessing the relative efficiency of DMUs using a frontier method that aims to determine the maximum volume of outputs, given a set of inputs. It is then possible to assess ex-post the efficiency of a production system using the distance to the production frontier. This is usually a deterministic analysis, which has a close resemblance to nonparametric linear programming. In parallel, a SFA was proposed by Aigner *et al.* (1977) and Broek *et al.* (1980) who presented a parametric and stochastic approach. This approach assumed that the product function includes stochastic components, which describe random shocks, such as climate or geographical factors.

The total economic efficiency implies the analytical frameworks namely, the SFA, the parametric one, and DEA, nonparametric. To a large extent, these are competing methodologies. Coelli *et al.* (2005) argue no formulation has yet been devised that unifies methods in a single analytical framework. Not like Total Factor Productivity (TFP) and indexes-based approaches, both methods require no or very little preference, price or priority information and are able to cope effectively with multiple inputs and outputs, reflect and

respect the characteristics of the industry, deal with noisy data, measurement errors and environment.

In the core of the stochastic frontier approach are taken as a system with a number of inputs and outputs run by DMU. The theory of production possibility sets represents the method of its frontier and production functions. Applying statistical method, it can be possible to estimate the standard error term β for the model, dispersion gives the fraction given by inefficiency. One of the main advantages of the SFA comparing to DEA is the probabilistic nature of the estimated parameters. The methodology assumes the presence of uncertainty in the results and allows estimating it.

The main feature of DEA is normally on the performance of agents' efficiency based of quantitative indicators of input and output. Mathematically, DEA is a linear programming-based methodology for evaluating the relative efficiency of a set of DMUs with multi-inputs and multi-outputs. The DEA assesses the efficiency of given DMU related to anticipated production possibility frontier determined by all DMU set. No *ad-hoc* assumption on the shape of the frontier surface is needed. It makes no expectations concerning the internal operations of a single DMU. This is an advantage of using DEA. Since the original DEA study by Charnes *et al.* (1979) there has been a continuous growth in the field. As a result, a considerable amount of published research and bibliographies have appeared in the DEA literature Seiford (1996), Gattoufi *et al.* (2004), Cook and Seiford (2009)

Since then the method of DEA became well-established methodology for measuring the relative efficiencies of a set of DMU with multiple inputs to make multiple outputs. This nonparametric efficiency approach made possible to compare various units relative to their best peers. It is logical that DEA became over last decades a well-established method for economic agents' comparative studies.

Zhu (2016a) followed by the Author find that the large amount of DEA literature makes it difficult to use any traditional qualitative methodology to sort out the matter. The findings on the literature review were summarized in Table 2 based on Liu *et al.* (2013) survey using a two citation-based methodology of the main path analysis and the *g*-index and *h*-index. It is a clustering method to group the literature through a citation network established from the DEA literature over the period 2000 - 2014. The main path analysis aims to comprehend the DEA development to a more detailed level, while the *g*-index and *h*-index in Column (1,2) is used to compare the effect of DEA authors in Column (3) and journals. Every study is then examined with main path analysis to expose the components in its core in Column (4,5). In the end, it is found that present the prevailing DEA applications and the observed association between DEA methodologies and applications.

Table 2. *Significance analysis of DEA researchers according to their g-index, h-index*

Ranking		Authors	Ranking		Years active		Total number of papers
g-Index	h-Index		g-Index	h-Index	from	to	
1	1	Cooper, WW	82	30	1978	2009	82
2	2	Banker, RD	43	22	1980	2010	43
3	3	Charnes, A	42	25	1978	1997	42
4	4	Seiford, LM	42	22	1982	2009	42
5	5	Grosskopf, S	41	23	1983	2010	69
6	6	Färe, R	40	22	1978	2010	79
7	7	Lovell, CAK	33	17	1978	2007	40
8	8	Thanassoulis, E	40	16	1985	2010	45
9	9	Zhu, J	33	18	1995	2010	70
10	10	Simar, L	30	15	1995	2010	29
11	11	Cook, WD	29	15	1985	2010	63
12	12	Thrall, RM	29	14	1986	2004	27
13	13	Sueyoshi, T	27	18	1986	2010	58
14	14	Golany, B	27	16	1985	2008	26
15	15	Wilson, PW	26	15	1993	2009	26
16	16	Dyson, RG	22	13	1985	2010	22
17	17	Talluri, S	21	13	1997	2007	22
18	18	Athanassopoulos, A	20	13	1995	2004	23
19	19	Pastor, JT	19	12	1995	2010	25
20	22	Forsund, FR	19	9	1979	2010	22

(Source: Liu et al. (2013))

Research activities relating to DEA have grown at a fast rate recently. Exactly what activities have been carrying the research momentum forward is a question of particular interest to the research field.

Originated from Seiford (1997) with 800 publications, the more recent overview by Seiford (2005) views around 2800 published articles on DEA. This large number of studies shows that comparative efficiency analysis has become an important topic in operational research, public policy, energy-environment management, and regional development. The

Author summarized in Table 3 amount of scientific articles using DEA method in the Baltic Sea region by country in Column (1), number of relevant publications as they appear in scientific journals in Column (2). The Colum (3) gives an indicative notion of scientific contribution of each country into the research problematic

Table 3. *Scientific articles using DEA method published in the Baltic Sea Region*

Country	Number of articles	Percentage
Lithuania	141	11,06%
Latvia	53	4,16%
Estonia	21	1,65%
Finland	257	20,16%
Poland	803	62,98%

(Source: Author's representation, based on Google Scholar)

A range of works from Thanassoulis (1993) to Fried *et al.* (2008) investigate various methodological approaches of performance assessment. Nonparametric and linear programming approaches are seen as methods for multiple input-output configuration of DMU. In the single-input case input levels can be regressed on output levels to estimate an explanatory model. If a satisfactory model is found it can be used to predict the input level of each DMU from its output levels. Then, comparing the actual and predicted input levels of a DMU, conclusions can be drawn about its comparative efficiency. In an analogous way regression analysis can be used to assess the performance of DMUs which produce a single output.

The advantage of applying DEA is straightforward, because it is able to accommodate and handle a multiplicity of inputs and outputs. This is a valuable feature because it considers returns to scale in efficiency estimation, admitting the concept of increasing or decreasing efficiency justified on output levels and size (Ali and Lerme (1997), Cook *et al.* (2014)). DEA application does not require to explicitly specify a mathematical form for the production function but still capture uncovering relationships that remain hidden for other methodologies. It leads that the sources of inefficiency can be analyzed and quantified for every evaluated unit.

Literature body in field of machine learning and performance assessment is relatively small compared to other conventional methods described above. The recent development of the performance assessment incorporates machine learning methods such as SVM with kernel functions, Random Forest, K-Nearest Neighbor, and Neural Networks as given in the Figure 8.

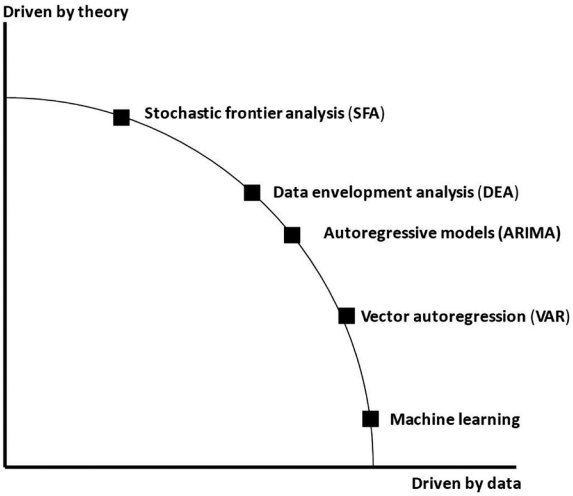


Figure 8. *Performance assessment approaches (Source: Author's representation)*

Recently various machine learning techniques has been successfully applied to predict time series and their co-movements (Kara et al. (2011), Karaa and Krichene (2012)). Very promising studies of Kruppa et al. (2012), Kreienkamp and Kateshov (2014), Addo et al. (2018) results indicate that non-linear techniques work especially well to model expected value. Yeh et al. (2010) extend the prediction of business failure by incorporating the efficiency of a corporation's management. Predictive power of various machine learning techniques like neural networks widely confirmed in the literature and found practical implications as by Alejo et al. (2013).

The Author based on the methodological review summarized in Table 4 the machine learning techniques in Column (1) with their most recognized abbreviation in Column (2) applied for heterogeneous economic agents assessment tasks. The indicative percentage of the research literature in Column (3) gives the research path of various machine learning techniques by incorporating the efficiency. Column (4) indicates whether the proposed in the literature review method is used in this study in the Chapter II. Here is worth to mention, that the overall literature body in field of machine learning and economic research is relatively new compared to other conventional methods described above. However, the literature body used in Data Science has been omitted in this research due to its practical orientation and narrow focus on business tasks. The lack of scientific generalization does not allow to include the Data Science into Economic studies directly without careful review of research aims, problematic, generalizations, and practical application from economic point of view.

The current research covers 80,19% of applicable methods in machine learning for constructing decision support system for efficiency assessment.

Table 4. *Methodological review of machine learning techniques employed for efficiency assessment in economic studies*

Method	Abbreviation	Usage percentage	Current research
Artificial Neural Network	ANN	38,71%	Used
Support Vector Machines	SVM	24,58%	Used
Decision Trees	DT	11,51%	Used
k-Nearest Neighbors	kNN	6,80%	-
Bayesian Networks	BN	6,71%	-
Random Forest	RF	5,40%	Used
Naive-Bayes	NB	3,90%	-
Boosting	ADAXGB	2,40%	-
	Methods coverage by the current research		**80,19%**

(Source: Google Scholar, Author's representation)

Many researchers exploit machine learning technique and nonparametric technique to provide a new method for predicting efficiency by using DEA scores as the only inputs into SVM predict parameter (Xu and Wang (2009), L. Zhou *et al.* (2014), X. Yang and Dimitrov (2017), Zelenkov *et al.* (2017), Alaka *et al.* (2018)). However, studies showed that ANN is slightly better for calculating the correlation estimation between variables, researches indicate that SVM is a machine learning technique with the best accuracy in comparison with other techniques. In order to achieve better results in the accuracy and correlation, can be used ensemble method by combining several techniques in several stages (T. Chen and Guestrin (2016)). N. D. Lewis (2015) the SVM finds the decision hyperplane leaving the largest possible fraction of points of the same class on the same side, while maximizing the distance of either class from the hyperplane. This minimizes the risk of misclassifying not only the examples in the training data set but also the yet-to-be seen examples of the test set.

Tattar (2018) among others states, that ensemble techniques are model output aggregating techniques that have evolved over the past decade and a half in the area of statistical and machine learning. The most applicable modeling problems comprise the problematic of a model choice. The theory offers various methodologies to complete this task. For machine learning models such as neural networks, decision trees, a k-fold CV is useful when the model is built using a part of the data referred to as training data. The accuracy is resulted from the untrained area and later on validation data. In case the model cannot handle complexity proper way, the result could be ineffective. The process of obtaining the best model means that we create a host of other models, which are themselves nearly as efficient as the best model. While applying such sophisticated techniques, the best possible model can deal with the majority of samples and other assisting models can assess the datasets where the

main model is inaccurate. Therefore, ensemble methods are not the final ones but merely extensions to the unsupervised learning problems.

Khan *et al.* (2018) gives evidence for the most classification problems from practical point of view, where the accumulated data for a few object type is overwhelming while the rest of datasets for other classes is incomplete. This leads to the class-imbalance problem. The class-imbalance affects nearly all of the collected classification databases. A multi-class dataset can be defined as imbalanced in case where some of its classes, in the training set, are massively suffering from lack of data comparing to other classes. This skewed classes instances distribution compels the classification algorithms to be biased towards the dominant classes. As the result, the properties of the marginal classes are not sufficiently analyzed.

Bensusan and Kalousis (2001) brings previous experience with classifiers as well as her preferences to the estimation process. As a consequence, the estimation is often vague, in many cases unprincipled and always relying on the human expert. Often, the practitioner can appeal to a well-established technique in the field, CV, to help the estimation or at least to establish which classifiers are likely to work best. Meta-learning is the endeavor to learn something about the expected performance of a classifier from previous applications. Y. Yang (2016) summarizes the basic concept of ensemble learning is to train multiple base learners as ensemble members and combine their predictions into a single output. This approach should have in average a better result than any other ensemble model with uncorrelated error on the validation datasets. Recently, ensemble learning found its extension to clusters for unsupervised learning problem combining different strategies.

Classification, regression and information retrieval are the most important assignments in the data-driven analysis. Each part proposes approaches and algorithms that give assistance for decision making. Hence, the result of applying algorithms can assess performances in a crucial problem. And usually, these algorithms come with several parameters that can modify their behaviors and performances. This makes the importance of appropriately selecting these parameters easily understandable. The number of different machine learning methods has grown over the past years and so the user faced with the question of which method it should use on a given problem. The problem is aggravated by the fact that many machine algorithms require that parameters should be set prior to their application, and besides, given data may be pre-processed in many different ways (Rakotomamonjy (2004), Grąbczewski (2013)).

Mullainathan and Spiess (2017) argues that machine learning not only provides new tools, it solves a different problem. Definitely, machine learning question is moving around the problem of prediction. At the same time, many economic applications try to find estimation parameter a better way. So applying machine learning technique to economics assignments entails definition of the scope related tasks.

A lot of works done in the field of classification algorithms. Liaw and Wiener (2001) methods that generate many classifiers and aggregate their results. Two well-known methods are boosting (see Schapire *et al.* (1998), Mason *et al.* (2000), Cristianini and Shawe-Taylor (2000), Witten *et al.* (2016)) and bagging Breiman (1996) of classification trees using the boosting techniques and the bagging method.

Niu *et al.* (2008) represent SVM as the convenient superiority in the classification. But the insufficiency is required the classed samples ahead of time. Cao and Tay (2003) and Sakouvogui (2019) presume the fundamental assumption of the DEA method where the DMU should all have a functional similarity. Their study proposed technique based on SVM prediction and classification algorithm. The latest studies of combining DEA efficiency techniques with machine learning represented by SVM might have shortcomings including various dependences of the efficiency measures. Pareek (2006), H.-Y. Kao *et al.* (2013) use an optimization algorithm based on a linear programming model to identify controls that need to be tested to address the risks, which can be developed as hybrid DEA and SVM approaches for efficiency classification.

Previous studies have combined the use of DEA and decision trees in analyzing organizational units Samoilenko and Osei-Bryson (2008), Seol *et al.* (2007), Young Sohn and Hee Moon (2004). While Eftekhary *et al.* (2012) suggests data mining techniques, extracting patterns from large databases proposing normalization methods and then normalized selected data sets afterward calculated the accuracy of classification algorithm before and after normalization. In this study DEA is used for ranking normalization methods along with the SVM algorithm was used in classification because this algorithm works based on *n*-dimension space and in case the data sets expect normalized the enhancement of results.

However, the tremendous amount of information stored in databases cannot simply be used for further processing (Kelly (1998), Bhavsar and Ganatra (2012), S. Li *et al.* (2012)). Sophisticated data analysis tools are applied in data mining process to determine previously indefinite, usable patterns and links in large data set. Such approaches should include statistical models, mathematical procedures and machine learning techniques. Consequently, data mining consists of more than collection and managing data, it also includes analysis and validation.

There are a number of critical reviews emerged by principle weakness of the DEA method. A desire to elaborate a better DEA approach by reducing its disadvantages and fortifying its advantages is the major cause for many discoveries in the recent literature.

The currently most often DEA-based method to obtain unique efficiency rankings is originated by Sexton *et al.* (1986). Sexton *et al.* (1986) and followed by Smith (1997) identified the impact of misspecification on model results, which in contrary to econometric methods is intractable. DEA do not offer diagnostics with which to judge the suitability of the chosen model.

Stolp (1990) generalized that homogeneity of technology across DMUs, uncertainty over the choice of inputs and outputs can affect the performance assessment. Banker *et al.* (1984), Banker *et al.* (1996) suggest to overcome inaccurate and imprecise inputs and outputs in DEA models by using simulation techniques. Cooper and Tone (1997) suggested to investigate deterministic DEA approaches combined with stochastic regressions to open additional possibilities for development. Such combinations can be effected in a variety of ways, but the studies examined involved a two-stage approach. In stage one, DEA is applied to the data in order to distinguish which observations are associated with efficiently and which are associated with inefficiently performing DMUs. In stage two, the results of stage one are incorporated as 'dummy variables' in the regressions to be estimated.

There is a question for evaluating the relative efficiencies of a set of homogeneous DMU so each of them has the same input and output measures. But in number of applications, the assumption of homogeneity among DMUs may not apply. Lack of homogeneity by evaluating efficiencies rise to the question on the fair approach be comparing a single DMU to other units. A related problem, and one that has been examined extensively in the literature, is the missing data problem addressed directly to appropriate techniques of machine learning (Zhu (2016b)).

Lertworasirikul *et al.* (2002) shows that the proposed DEA methods still require accurate measurement of both the inputs and outputs. Nevertheless, the given input and output data in real-world is sometimes inaccurate or vague. Inaccurate evaluations can be derived from uncountable, incomplete and hidden information. A considerable number of studies suggested different unclear methods for coping with this impreciseness and uncertainty in DEA models.

Cook and Zhu (2006) stress out that despite its recognized popularity, the technique has also various boundaries and limitations, in general, the construction of conventional projection on the efficiency frontier, the lack of description of heterogeneous behavior in performance among many efficient agents. A shortcoming in a standard DEA model is that all efficient DMUs have the same estimation with no way to separate them. This has led to focused research to further discriminate between efficient DMUs, in order to arrive at a ranking, or even a numerical rating of these efficient DMUs, without affecting the results for the non-efficiency.

Emrouznejad and Anouze (2010) proposed an alternative approach to retain fuzziness of the model by maximizing the membership functions of inputs and outputs. Emrouznejad and Tavana (2013) provide further the necessary background to work with existing fuzzy DEA models, which problematic is in the presence of noise (Fried *et al.* (2008)).

Hatami-Marbini *et al.* (2011) points out that crisp input and output datasets are fundamentally crucial in conventional DEA. However, the observed input and output data in real-world is sometimes unclear or ambiguous. There is a number of studies suggested to deal with various fuzzy methods to cope with the unclear and ambiguous datasets in DEA method.

Kolaczyk and Csárdi (2014) underpin that the estimation of parameters and data analysis are essential elements of network research. Hence, there is a clear demand for network analysis, both conventional and sophisticated, varying from applications to methodology. As with other areas of statistics, there are both descriptive and inferential statistical techniques available modeling and prediction of network-indexed processes, both static and dynamic.

Extension by Joro and Korhonen (2015) introduce how to incorporate preference information the field of DEA, which is closely related with the issued raised almost four decades ago. C. Kao (2016) describes system as composed of many subsystems operating interdependently, while conventional DEA considers the inputs supplied to and the outputs produced from the system in measuring efficiency, ignoring its internal structure. As a result, it is possible that the overall system is efficient, even while all component divisions are not. More significantly, there are cases in which all the component divisions of a DMU

have performances that are worse than those of another DMU, and yet the former still has the better system performance. With an eye on solving these problems, many ideas have been extended from the conventional DEA to build models to measure the efficiency of production systems with different network structures, which are referred to as network DEA. The study presents the underlying theory, model development, and applications of network DEA in a systematic way, to give the readers an idea of what should be done when developing a new model.

Paradi *et al.* (2017) and Aldamak and Zolfaghari (2017) incorporate data science tools for analysts who aim to apply DEA in their data assessment process with discussion of ranking application. Ehrgott *et al.* (2018) argue how to cope with data assumed to be not known precisely. The study considers situation in which data is uncertain, so the efficiency scores increase monotonically with uncertainty. This enables inefficient DMU to leverage uncertainty to counter their assessment of being inefficient.

A wide range of authors Färe and Grosskopf (2012), Ouellette and Yan (2008) and Z. Li *et al.* (2017) describe that dynamic models have inherent advantages over static models in the context of event prediction because conditions and behaviors change over time, so predictions need to be adjusted by incorporating as much information as possible. Study extend DEA with dynamic scores which provide insights into the efficiency of a company relative to others over time with focus on the Malmquist productivity index.

1.5. Ensemble machine learning approach in decision-making process

The Author argues for ensemble methods in machine learning approach, where various techniques are combined in order to deliver the best possible estimation. Fundamentally a vast number of economic applications deal with a parameter estimation in order to deliver good estimates of parameters $\boldsymbol{\beta}$ that explain the relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$. It is key issue to underpin that machine learning algorithms regression coefficients and their estimates are rarely consistent. Machine learning tackles the problem of prediction to produce predictions of $\boldsymbol{y}$ from $\boldsymbol{x}$. The advantage of machine learning is that it can expose generalizable patterns by ability to uncover complex structures that was not stipulated in advance. Machine learning can fit flexible yet complex data settings without overfitting. From this perspective applying of machine learning to economics demands finding relevant $\boldsymbol{y}$ tasks. One set of such problems are in new set of data for measuring economic activity using classifying approaches. In another field of investigation, the major interest is actually a parameter $\boldsymbol{\beta}$ with extrapolation procedures contain a prediction task.

Machine learning methods is dealing with the theoretical set of techniques and approaches which translate to computer how to absorb knowledge, find dependencies and perform certain assignments. Machine learning has a lot of similarities with conventional statistics. Summarized as in the Figure 9 it is possible to define a typical process of learning machine is finding a mathematical formula, which, when applied to a collection of inputs, produces the desired outputs (Burkov (2019)). This mathematical formula also generates the correct outputs for most other inputs on the condition that those inputs come from the same or a similar statistical distribution as the one the training data was drawn from.

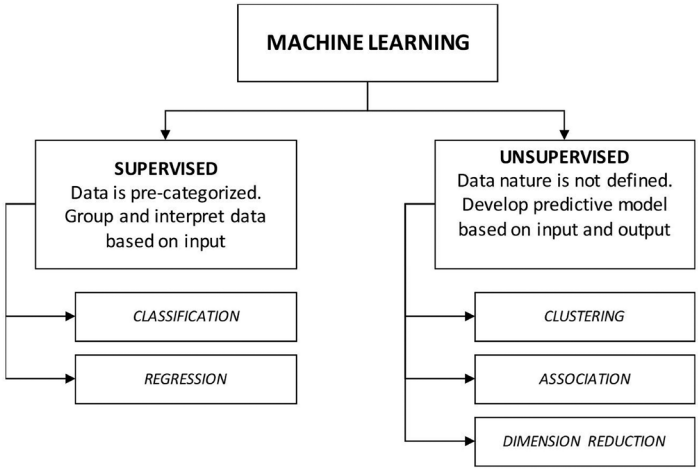


Figure 9. *Machine learning and its types*
(Source: Author's representation adopted from Praveena and Jaiganesh (2017))

Supervised learning is the core approach of the machine learning techniques. The process is defined by a certain known variable with the aim to recognize given variable by the influence of other variables. Hence, machine learning revolves with respect to the output variable, whereas particularly the supervised learning denoted often as the learning process with the teacher. All target variables are not alike, and they often fall under one of the following four types (Ayyadevara (2018)):

1. **Classification problem** is to classify observations into one of k types of class. Such variable is referred to as a categorical variable in statistics. Therefore, the target variable might be a continuous variable in numeric representation.
2. **Regression problem** is the purpose of the machine to learn the variables in terms of other associated variables, and then predict it for unknown cases in which only the values of associated variables are available.

In contrary, unsupervised machine learning purposes to uncover previously unknown patterns in data, but most of the time these patterns are poor approximations of what supervised machine learning can achieve. The ability of imbalanced data to significantly compromise performance of most standard learning algorithms is the fundamental issue of imbalanced learning problem. Most standard algorithms assume or expect balanced class distributions or equal misclassification costs. Therefore, when presented with complex imbalanced data sets, these algorithms fail to properly represent the distributive characteristics of the data and resultantly provide unfavorable accuracies across the classes of the data. Important classification algorithms that are designed to deal with uncertainty principally exploiting the following procedures to handle imbalanced data (He and Garcia (2008), Aggarwal and Yu (2009), García *et al.* (2012). Some applications of unsupervised machine learning techniques include:

1. **Clustering** allows automatically split the dataset into classes according to similarity. Frequently, however, cluster analysis overvalues the correlation between classes, where data points are not treated separately. Therefore, cluster analysis has limitations in application of areas such as customer segmentation and targeting.
2. **Association analysis** identifies sets of items that frequently occur together in dataset.
3. **Dimension reduction** is commonly used for data preprocessing, such as reducing the number of features in a dataset or decomposing the dataset into multiple components.

The Author intends to build up own model based on the analysis in Table 5. The proposed approach should take the advantages of both statistical research methods and machine learning to create models from data from different perspectives. Regression analysis will provide a form of data reduction to operate with the mean and standard deviation for descriptive and inferential statistics. By the mean of the descriptive statistics a clear way of understanding of complex data can be done. In order to make statements about data, inferential statistical methods will be applied. Ensemble machine learning focus on prediction by means of learning algorithms to discover patterns in bulky data. Ensemble machine learning methods are predominantly effective even if the datasets are obtained without a carefully controlled experimental design and in the presence of dense nonlinear interactions. There are summarized the most recent methods in Column (1) for the approaching the efficiency assessment under uncertainty conditions in decision support systems. The methods might be generalized by its purpose in Column (2). The purpose can be seen in terms of data mining techniques, where no generalization needed but rather accuracy and efficiency of the result. Machine learning techniques require more generalization of its principles. However, from economic science point of view, machine learning still suffers from lack of generalization due to its nature arise from data processing and pattern recognitions. In general-class approaches such as descriptive analytics, the most traditional methods are reinforced with recent developments in machine learning and it got extended its capabilities for enhanced knowledge discovery and improved decision-making. The growing class is the predictive analytics focused on the building and assessment of models that seek to make empirical predictions with weaker theoretical framework. Formal models prove their hypotheses thought validation of findings by checking model fit using goodness-of-fit tests and residual analysis. The problem, however, real data application show that when the correlation between the dependent and independent variables in a regression analysis has nonlinear feature, tests on goodness-of-fit do not reject linearity unless the nonlinearity was extreme, what influence predictive power of the model. Thus formal models without validating their findings with predictive analytics mostly the data-driven techniques may result in misrepresentative and biased conclusions if even those findings pass goodness-of-fit tests and residual checks. In Column (5) there is a referential source for the applicable method.

Table 5. *Components overview of machine learning methods and techniques for elaboration decision support systems model*

Method	Implication	Purpose	Source
Association analysis	Data mining	This method attempts to find the relations between entities based on transactions or events that involve them. Discovers patterns of frequent subsequences in a database of events or transactions.	Provost and Fawcett (2013)
Clustering	Data mining	Well-known method of group a set of objects such that those in the same group are more similar based on a certain criterion to each other than to the objects in other groups.	Doumpos *et al.* (2018)
Decision tree	Machine learning	Decision trees can be used for regression and classification purposes. Classification accuracy and size of a decision tree are used to determine its quality. Decision trees recursively separate observations into branches to construct a tree for improving prediction accuracy.	Dangeti (2017), Hackeling (2017)
Artificial neural networks	Machine learning	Modeling very complex non-linear functions and predicting new observations from other observations after executing a so-called process of learning from existing data.	da Silva *et al.* (2016), Miller and Forte (2017)
Support vector machine	Machine learning	Algorithm creates an optimal hyperplane that can be used to categorize new observations. While there are numerous linear hyperplanes that can separate the two classes of the response variable.	Vapnik (2013)

(Source: Author's representation)

Härdle *et al.* (2006) finds the preferable application of SVMs in the field of bankruptcy and solvency prediction. Among others researchers Fan and Palaniswami (2000) compared SVM approach for decision support systems with widely used Neural Network and Multivariate Discriminant Analysis. The analysis tells that SVM in general achieved a better prediction quality (70.35-70.90%), followed by Neural Network (66.11–68.33%), followed by Multivariate Discriminant Analysis (59.79–63.68%). Shin *et al.* (2005), Min and Lee (2005) The prediction accuracy of SVM is evaluated by in the comparative study with also Multivariate Discriminant Analysis, logistic regression analysis, and back-propagation neural networks. The results of the research argue that SVM leave behind the other methods in terms of prediction accuracy. Van Gestel *et al.* (2003) relied on the least squares

modification of SVMs, which exposed significantly better outcome in solvency prediction compared to the traditional analytical frameworks.

The Author among other researches investigate SVM classifiers in the face of uncertain knowledge sets and show how data uncertainty in knowledge sets can be treated in SVM classification by employing robust optimization (Jeyakumar *et al.* (2014)).

The last decade an increasing number of researchers use SVM to solve a variety of practical problems in classification, clustering and regression either linear or nonlinear (Bishop and Tipping (2000), Ma and Guo (2014), Murty and Raghava (2016)). The main theory is that the linear models typically are learnt based on a linear discriminant function that separates the feature space into two half-spaces, where one half-space corresponds to one of the two classes and the other half-space corresponds to the remaining class. So, these half-space models are ideally suited to solve binary classification or two-class classification problems. There are a variety of schemes to build multiclass classifiers based on combinations of several binary classifiers. Consequently, Drucker *et al.* (1997), Smola and Schölkopf (2004) argue that SVM can also be used as a regression method, maintaining all the main features that characterize the algorithm of maximal margin. Thus, SVM can be defined as methods which use hypothesis space of a linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory. A learning bias is derived from statistical learning theory. SRM minimizes an upper bound on the expected risk, whereas ERM minimizes the error on the training data. It gives SVM advantage and a greater ability to generalize, which is the goal in statistical learning. SVM were developed to solve the classification problem, but recently they have been extended to solve regression problems.

SVM can also be applied to regression problems by the introduction of an alternative loss function. A distance measure should be included into the modified loss function. The regression can be linear and nonlinear. A nonlinear model is required to adequately model data. In the same manner as the non-linear SVC approach, a nonlinear mapping can be used to map the data into a high dimensional feature space where linear regression is performed. The kernel method is used to approach the problematic of dimensionality. Similar manner, the machine learning regression considers the problem of the noise distribution based on prior knowledge.

The advantages of SVM are the effective in cases where number of dimensions is greater than the number of samples. Various kernel functions can be employed for the decision-making function. However, there is a significant disadvantage of SVM, because it does not directly provide probability estimates but using a CV is the solution.

The future of the machine learning is in combination of different approaches, because fully supervised algorithms are a useful but perhaps an unnatural assumption due to latent variables in models (D. Chen *et al.* (2013)). In reality, there is not always possible to have complete supervision because there are always some variables relevant to the problem that not annotated in datasets.

1.6. Integration of methods into decision support systems

Hence, the research aim becomes clear. It is needed to shed light on the idea of economic decision process and its connection with investments from different perspective in terms of their ability to create or absorb technological innovations within on-going infinite technological progress. The Author believes that the result of the recent technological development and machine learning techniques might emerge in various forms of automated decision-making processes, which will supply policymakers with relevant yet precise information on a particular problem.

The Dynamically adjustable decision-making model is presented in Figure 10. The model corresponds well with other recursive models, where agents make their decision in respect to the environmental response. Each organization has certain disposable inputs, desired outputs and assumed targets, which should be regarded separately due to its nature. Hence, a two-stage efficiency nonparametric analysis is strongly advocated. The results of the analysis are integrated part of decision process, which can be reinforced with ensemble methods in machine learning.

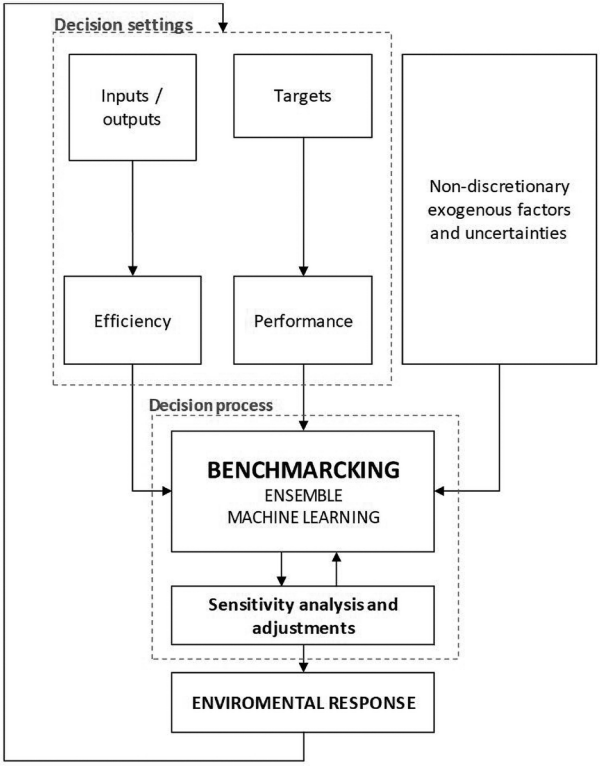


Figure 10. *Dynamically adjustable decision-making model (Source: The Author's representation)*

Therefore, the definition of *decision support systems* emerges for organizations applying different business intelligence techniques of statistical and scoring modeling, neural networks, various expert systems, agent-based systems, neuro-fuzzy systems, various case-based systems, or simply guidelines that have been developed through data-driven experience.

The two-stage nonparametric models hence evaluate the relative efficiency of decision making units through multiple inputs and outputs aimed to give non-biased yet independent measures without having implied any particular postulations about datasets. Here is the major advantage of the usage of the nonparametric models allowing to clearly recognize and categorize the influential factors leading to the successful decision.

The two-stage decision support systems practically seen are about to find the best trade-off between risk and return in a given set of observation is a complex task because of the magnitude of the existing options. According to Cornett and Saunders (2003), Saunders *et al.* (2006) the financial institutions aim to increase returns for its shareholders considering the cost of increased risk. The financial institutions are dealing with various risks, which are the subject to solve by effective management in order to gain the best performance. Jorion (2000) defines market risk as the risk of losses due to volatility fluctuations or variations in the level of market prices. Not only pure operational and credit risks but regulatory attention and laws imposed by governmental bodies require financial institutions to have capital at disposal to cover risks arise from all these channels.

Since the global financial turbulence, decision support systems in financial institutions has gained more attention in focus on how to detect, measure, report and manage various categories of risks. Van Liebergen (2017), Helbekkmo *et al.* (2013) argue that policymakers are increasingly looking forward having employed artificial intellect and machine learning techniques to manage reporting data and unstructured information. There are a number of estimation, but in less than decade, the risk estimation and perception will be fundamentally different from what exists on the market today because of automated decision support systems. The main reasons are enlargement and deepening of laws and regulations, customer set up higher expectations. All these factors give a new notion of risk types expected to perform the corresponding changes within risk management. Artificial intellect and machine learning methods are the cutting-edge technologies with a promising future implications in area of risk management. It might help to develop models defined by accuracy and complexities coping a better way with nonlinear patters within large datasets. It is anticipated that innovative techniques will be applied across multiple areas of policy-making.

Awad and Khanna (2015) explains various machine learning techniques coming together from computer science, engineering and statistics to be integrated into decision support systems of next generation. It has been presented as an important means of combining and interpreting the hidden relationship among data patterns and optimization tasks. Machine learning takes advantage the power of problem generalization, which is an essential part of concept formulation through human knowledge. The learning curve consists of knowledge base that is tithed together by feedback to improve performance.

Benchmarking in decision support systems plays extremely important role. Research conducted by Rigby and Bilodeau (2011) on the decision-making instruments emphasize the im-

portance of benchmarking among other important techniques listed in the Figure 11 among most popular selected approaches by years. There are numerous surveys between 2000 and 2010 considered benchmarking as the most important tool among *strategic planning, customer relationship and management, mission and vision statements* and *balanced score card*. Here is the clear trend in resources optimization by benchmarking and outsourcing during observed period. It might be a response on the economic recessions. That fact can explain, why the benchmarking has consistently been in top five most widely-used tools in the past decade in the range from 67% to 82% of the observed companies (Figure 11).

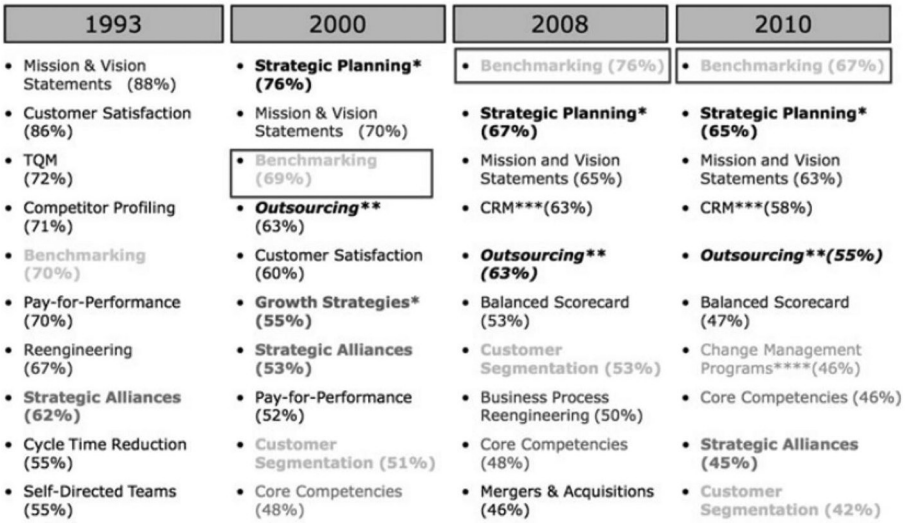


Figure 11. *Trends in decision support techniques (Source: Adopted from Rigby and Bilodeau (2011))*

The decision support systems might be described as an integrated flow of data processing, analysis and enhancing decision reinforced by results. The decision support systems might encompass the learning process. Brose *et al.* (2014a) endeavors detail of the relational information technologies, which are widely used for more than forty years over the entire industry and science. Based on the data processing technologies there is a wide range of tools developed for pre-processing and post-processing data. This process is focused on data transformation from processing transactions into a relational form. Further data in its relational form might appear in form of complex statistical processing or become a part of business intelligence.

The recent development in databases and data management introduces *On-Line Transaction Processing* (OLTP), which are designed to work with a large number of transactions in real time. OLTP are heavily concentrated on the needs of analytical processing of data aggregation, manipulation, filtering and data reporting. In order to handle time-

consuming shortcomings of using OLTP for analytic purposes, data science offers various *dimension data modeling* techniques. This is a fundamental requirement for building effective decision support systems addressed the needs of a particular business function. As a prominent example, credit scoring systems can be built to shed a light at a wide range of issues around a counterparty credit reliability.

Decision support systems can be validated against known results as well as against expert knowledge. However, it is not possible to neglect the significance of data consistency and reliability. But the same time, the types of measures and policies should be adjusted to the environmental challenges and uncertainties. One of the solution is to employ the analysis of the extremely large volumes of a various datasets known also as *big data* extracted from new digital data sources including but not limited to unstructured text, keywords analysis, advertising campaigns and events, financial reports and transactions, stock exchange tickers, users web interactions, logs of agents' behavior. One of the most promising analysis is a metadata analysis, which can be described as data analysis of data. They represent evidences about indirect parameters such as definitions, occurrences, sources and any other indices that contribute to interpretation of the underlying hidden patterns faster, make it more effective and therefore more reliable. Data mining analysis is based on machine learning and statistical analysis. Data mining techniques are applied in a wide range of domains where large amounts of data are available for the identification of unknown or hidden information. (Han *et al.* (2011), Siguenza-Guzman *et al.* (2015), Mittal *et al.* (2016)).

In order to elaborate criteria for efficient and effective decision support systems, data systematization practice should be able to incorporate and be relied on analytical infrastructure that specifically designed to provide reliable and appropriate access to accurate and detailed datasets. The datasets might with no doubt be sourced and obtained from a wide variety of input sources and databases. The intermediate results of analytical scenarios are the backbone of decision support systems. Decision support systems are typically drilled down to a single decision-maker within organization for example focused on fraud detection systems, transaction credit approval systems as well as for risk management and strategic planning. Practical aspect of implementation of decision support systems should include other applications as the following:

1. **Business Intelligence systems** designed the way where data can be promptly observed and filtered by a number of different dimensions in order to obtain immediately insights into recent performance of organizational units.
2. **Data mining applications** typically operate with enormous sets of data and facts which have been combined and accumulated through ongoing interaction with counter-parties and environment. These datasets play important role for statistical analysis focused on acquiring meta-information and hidden patterns on utility, preferences, trends, or other associated agents' behavior.
3. Full-scale **Enterprise Resource Planning** application gives opportunity to conduct organizational workflow a better way focused on including but not limited to capital investment, inventory, production and logistics.

Therefore, an automated core of any decision support systems should be designed by using modular functions to endorse an intelligent response control system. Machine learning techniques are indispensable in modeling the knowledge function to create rules, data design, constraints, and patterns in a structured manner. As the result of the learning process, novel knowledge is created based on using existing structures and future new learnings. The obtained knowledge should be validated by a feedback control recursive function, which estimates parameters reasonably. The supporting functions that enable an intelligent feedback control recursive function comprise:

1. A **sensor input function** to take parameters of the internal or external environment in terms of aberrant behavior.
2. An **adjusting function** to balance the effects of environmental instabilities by changing the system elements, thereby maintaining optimal and sustainable operations.
3. An **analytical function** to analyze the obtained data to control if any of the crucial variables are within reasonable bounds, or limits.
4. A **forecasting function** to gain information on the changes that need be done to the current settings to find optimal balanced state of the new environment.
5. A **knowledge function** that encompasses the vectors of possible behaviors and actions that can be initiated as the response to the new environment. The forecasting algorithm benefits from this knowledge to find out the appropriate action to cope with the disturbance. The knowledge function is created by the generalization meta-analysis of ongoing tasks under assumption of a richer hypothesis space.

The decision support systems should thus embrace necessity in effective management in any organization with sufficient integrated subsystems and modules, which constitute an organizational structure for the decision-making process. The decision support systems interact with the operational environment and settings, which are influenced by the managerial decision-making process. Effective promotion of such decision-making policy will extend level of collaboration and coordination within organization. Learning intelligence for earlier forecasting of changes in the environment and uncertainties helps in capturing a complete assessment of the operational environment. It will bring benefits in formulating alternate strategies, which are necessary for foresee to transforming environmental settings to keep the effective and sustainable development path. Such policy guides the organization toward a strategic aim by proposing a better decision-making policy functions.

1.7. Results and generalization of the literature analysis

Analysis of research literature enabled the Author to determine that the efficiency assessment under uncertainty conditions is the result of economic complexity and nonlinearities of the decision-making processes. The reason is that the expected outcome of analytics within the context of complementing traditional statistical analysis is figuring out of new correlations that emerge from large and uncertain data. This approach then be used to develop new theories for further statistical analysis and testing.

The uncertainty factor is so large that the effects of policy decisions on the economy

are thought to be ambiguous. In this situation, any plausible expertise on the nature of uncertainty might be very useful. In order to understand how variations in uncertainty might affect the economic process, it is important to find its source. Various categories of uncertainty might have diverse affect. The individual sectors of households, firms and government have different scales of persistence. The sources of uncertainty are in poor quality of data, unpredictable shocks hitting the economy, econometric errors in estimation, and a lack of understanding of the fundamental economic mechanisms.

Decision support systems are nowadays an attractive research issue in the practical field and from scientific point of view. A better decision-making process contributes to overall efficiency and performance by articulating strategic information about the current operations and environmental settings. Awareness of a bigger picture it may affect a management decision-making process. Heterogeneous environmental settings and nonlinearities of processes might also in turn affect the stock market, consumers' preferences, and even competitors' policy. All of these considerations and presumptions lead to concentrate specific research efforts in both business and science.

From the literature review and methodology there is a clear shift to more intelligent decision support systems. A considerable number of innovative approaches have been used in corporate decision-making processes, most of which employ a wide of information sources from financial ratios, financial statements to mathematical modeling and evaluations. Among all these methods, the pioneering multiple discriminant analysis over decades found its application developing a model that utilizes ratios in a linear system to obtain a score. The score analysis is intended to classify organizations into categories at various risk criteria of failure, healthy, and the middle status. However, most ratio analysis methods rely on financial statements and derived ratios assumed crucial factor and weighted more relative to other factors. The problem is that peer-to-peer analysis of manufacturing companies needs to be scaled. Various combination of analysis based on discriminant analysis or multiple discriminant analysis consider the linear combination of two or more independent variables that will differentiate best between pre-defined groups. In the advanced two-group case, separation analysis can also be brought by multiple regression.

Modeling uncertainty by experts is an *ex-ante* process and can therefore be used for assessing future state of the economy. It is based on the transparent and intuitive assumptions and it is easy to interpret. It can also be associated with particular periods of time in a natural way, as the surveys are usually well grounded in time, both in terms of the periods after the projection estimated and the period for which such projection is intended to be applied. There has been recently a substantial methodological and technical progress in this type of research, so that the quality of survey based methods is improving.

There is a number of prerequisite arisen from its nature of a complex economic system development. The analytical framework should provide the conventional methods describing the random behavior of the heterogeneous economic agents, the changing structure of entire markets and the institutions, considering the influence the heteroscedasticity of the global processes beyond and within the European Union or at the global scale and provide the mechanism to link intertwined components into a framework. From the global economic perspective markets denotes the complex systems represented as a network, which

diverges from an initial state even by local events can spawn large-scale pat-terns and the global shocks. The whole system can be spitted into subsystems organized hierarchically. The fundamental problem of the economic analysis is that the complex system evolves through time emerged from responses on external factors, interactions among agents characterized often by bounded rationality, but not equilibrium enacted by policies exogenously. The economy itself does not subsist separately from the environment.

From practical point of view, the integration process has a long history, where since the second half of the seventies of the past century, a period of economic recession, a growing vast interest for technological change in economic science can be observed in all industrial countries. The reason for this phenomenon is the faith that the economic prosperity of these countries will depend to a large scale on their ability to create revolutionary products and processes and to bring them viable. Therefore, the stimulation through providing opportunities and creating of various policy of all kinds of Research and Development activities is seen as priority governments of these countries.

Widely accepted point of view is that countries do not develop in a sustainable path by making more of the same using economies of scale. Instead of it the countries are seeking for changes what might create innovative activities that are more productive and profitable. The heterogeneousness in processes leads to increased sophistication over time. Therefore, the role of infrastructure plays an important role, whereas the question of how effective large spending on infrastructure raises a vast wave in mass-media accompanied by scientific researches in this area. The strong intensity in economic policy on technological change or innovation was guided and supported by scientific research. Albeit policy measures strongly preferred technological research, many investigations used to take care of the relation between technological challenge and economic development. The nature of such researches comprises a large scientific area such as the proposals with the Kondratiev long waves, Schumpeter's hypothesis about entrepreneurship and the concept of product life cycle.

From the literature review Author found that the major differences between the traditional scientific and the emerging analytical research:

1. In nonparametric research and machine learning data-driven approach may precede theory or a model.
2. The development of theoretical frameworks focused on the nonlinear complex correlations and patterns present in the data rather than on hypothesizing.
3. Metrics used and general approaches are different from formal methods

The process of the efficiency assessment in decision support system has a number of processes, which imply actions on different layers. Its graphic representation can be seen in the Figure 12. The Author following the study of Aggarwal and Yu (2009), Aggarwal (2014) proposes to treat uncertainty as phenomenon dissected on different layers: data-mining uncertainty, analytical framework uncertainty and uncertainty as a factor. Most approaches which used to incorporate uncertainty are based on restricting the weights in the multiplier model. Unlike the existing approaches, the combinations of machine learning techniques in this study do not require to think in terms of hypothetical assumption. Mathematically machine learning leads to the identification of implicit restrictions to weights, so there is

a fundamental difference in these approaches, emerging from the way in which the data explicitly is gathered. In each process the uncertainty is emerging in different qualities and it should be assessed with respective techniques. The unique research of Wen (2014) analyzes in details efficiency assessment and related decisions-making process, which usually in real-life made in uncertainty. The central role of the proposed research is the non-parametric models under various assumptions of the probability theory, credibility theory, chance theory and uncertainty theory. Simon *et al.* (1992), Simon (1997) and Bloom (2014) investigate the framework models of bounded behavior under uncertainty conditions.

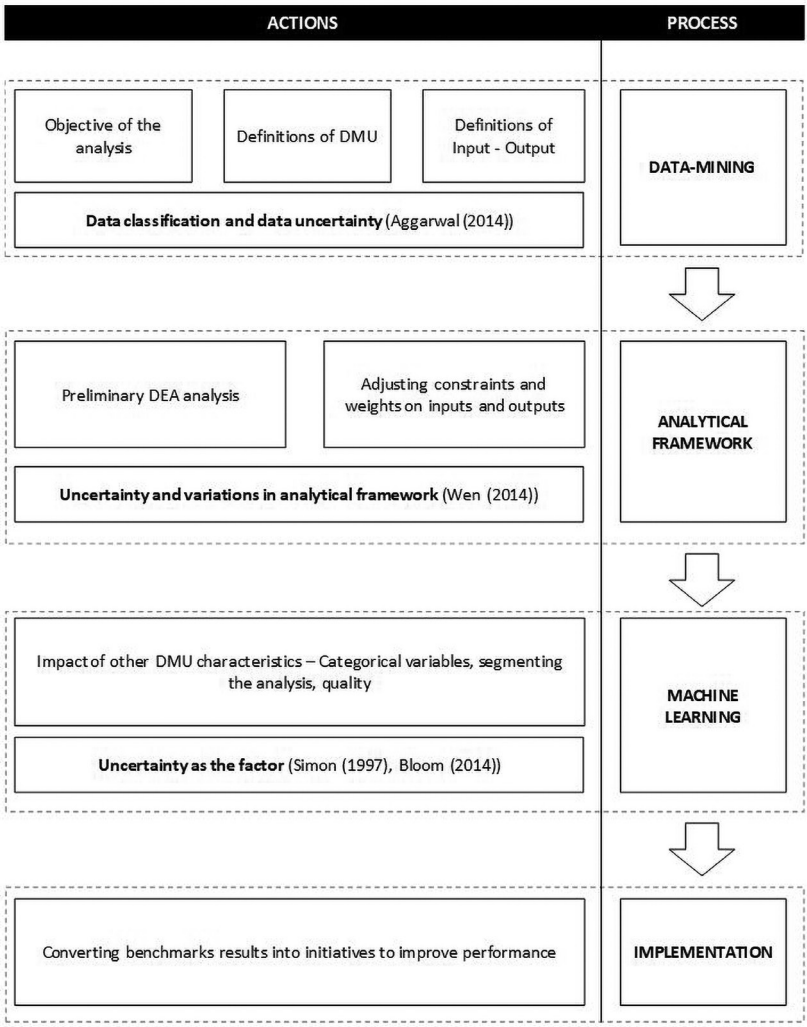


Figure 12. *Generalization of processes of efficiency assessment under uncertainty (Source: Author's representation)*

Data mining is related with knowledge management and it can be defined as the process of analyzing large information databases and of discovering implicit, but potentially useful information. Data mining is able to find hidden layers and associations of unknown patterns and trends by working with large amounts of data. The huge amount of data in recent business settings and science demands more sophisticated tools. Although there are achievements in data mining technology, which simplified large data collection, it is still a strong need for better techniques that can enable to transform the datasets into applicable information and knowledge.

The importance of sample-specificity and prediction error for the assessment of efficiency functions becomes straightway obvious when estimating a production function using data-mining. Emerged machine learning concepts allow to investigate the extent to which an estimated production function characterizes both the set of observed formations and the set of hidden variables for a particular industry. The aim of efficiency is characterized by maximizing output given a certain amount of input or by minimizing input given a certain amount of output. In empirical efficiency analysis, we most often apply a rather relative than absolute concept of efficiency. The concept of DEA reveals the units which are supposed to be able to improve their performance and the units which cannot be recognized as poor performers.

The agents are continuously adjusting their behavior in the dynamically changing environments through generation of new patterns of behavior and raised complexity of the interactions. All these criteria impose limitations on the theoretical framework, which might be applied. From this point of view, the agent-based modelling and machine learning represents a simulation modelling technique that might help develop the estimation framework considering the natural patterns agents and sufficient level of flexibility. One of the advantages of considering the data-driven modelling allows incorporating the asynchronous approach implying events influence decisions are happening at different time frames and different order, what might be an appropriate assumption for the framework.

II. PROPOSED MODEL FOR THE ASSESSING EFFICIENCY IN DECISION SUPPORT SYSTEMS UNDER UNCERTAINTY

2.1. Model definition for the assessing efficiency under uncertainty factors

This section describes methods, technological approach and tools for research, insights and evaluation, such as framework for the model. A framework for multidimensional analysis based on machine learning techniques, which enables analysis from different point of views.

Although nonparametric efficiency models were originally intended for use in microeconomic environments to measure the performance on the microeconomic level it is ideally suited to performance analysis for aggregated datasets. Traditionally macroeconomic performance is measured as the extent to which policy makers reach their macroeconomic objectives. Policy objectives represented then usually as a sum of the GDP growth rate, the inflation rate, the unemployment rate and the surplus or deficit on the current account of the balance of payments. Lovell *et al.* (1995) claim that other dimensions should be incorporated in economic performance analysis.

The initial model introduced by Scholkopf and Smola (2001), Emrouznejad (2006) and Emrouznejad and Shale (2009) of combining linear programming approaches. The proposed model based on Vilela *et al.* (2018) a two-stage model for forecasting time series. At the initial phase the classification methods are applied to order the time series into its various classes and contexts. The second stage makes use of ensemble approach of SVM, NN and Decision Trees, one for each context, to forecast future values of the series.

Kuhn (2008), Kuhn and Johnson (2013) rely on the *caret* package, which offers tools for classification and regression training, contains a number of tools for predictive models using the huge set of models available in R. The package has function on simplifying model training and tuning with a wide variety of modeling practices. It provides methods for *ex-ante* training data, calculating variable significance, and model representations. The computational interaction is used to exhibit the functionality on a real data set and to target the benefits of parallel processing with given types of models.

The practical implementation based on the methodology of Kleiber and Zeileis (2008), Leipzig and Li (2011), Adler (2010), Albert (2007), Albert and Rizzo (2012), Kassambara (2013) and Kassambara (2017a), Kaas *et al.* (2008), Beyersmann *et al.* (2011), model diagnoses by Bivand *et al.* (2013), Bolker (2008), statistical by Cohen and Cohen (2008), regression analysis by Cowpertwait and Metcalfe (2009), Karian and Dudewicz (2016), Højsgaard *et al.* (2012).

Model definition for the assessing efficiency under uncertainty factors should include influence of the environmental subsets evaluation, which the entire sample is made of. For decision support system each economic subset exhibits modularity in terms of being created from a large number of complex yet functionally specific parts. The openness within

sample should mean in the sense that these parts deal with degrees of freedom. Therefore, the projection of observed subset points divided into their eventual frontiers and solved by single estimated organization differentiated by the mean efficiency of other subsets. However, the scope of analysis is limited by an aggregated environmental variable. Further development of considering environmental variable into nonparametric model is a non-discretionary inputs in case of positive impact on efficiency or outputs if there is a negative impact on efficiency. The non-discretionary factors require a priori knowledge of the direction of the impact. This is a main disadvantage, which makes difficult to setup a full-scale decision support system based on prior datasets because the nonparametric model cannot handle structural breaks within the environmental variable. The leap forward in approaches of including environmental uncertainties into nonparametric model in decision support systems is the two-stages approach. In this setting the nonparametric model is involved with conventional and well-established inputs and outputs in decision support systems in the first stage. The efficiency assessment is conducted in the second stage by regression on environmental variables. Thus results of the nonparametric model and their outcome regressed on explanatory variables using common regression techniques. The two-stage approach has found its application in the literature body, but it does not succeed to assess efficiency in a clear and actionable way using conventional regression techniques.

The comprehensive evaluation for the economic agents is an important tool to achieve the objective effective resource allocation. It can help to improve decision-making process in order to strengthen the management and provide basis for decision-making. The literature review exhibits many evaluation methods. But most of the methods need to decide the weight of the indices first, scores weighted indices and biased estimations. These methods obviously have a certain subjective fairness of evaluation results are not obvious.

The machine learning system based on ensemble techniques which application is growing in past decades can solve the given problematic appropriately. Random Forest, SVM and ANNs have the convenient superiority in the classification and regression. This affects its widespread application. In order to find evaluation method, the Author use a hybrid approach to combine DEA and machine learning in the joint evaluation process taking full advantage of DEA method of absence predetermined weights to input and output parameters. Furthermore, it is possible to assess relative efficiency of DMU with the focus on objectivity with acceptable error. The experimental results show that this method has strong objectivity and impartiality, the evaluation method is simple and easy to interpret.

Analysis of scientific literature and research gaps identified as the result of the bibliographic study enabled proposing a model to assess the factors influencing efficiency which is presented in the Figure 13. There is the algorithm for carried out the decision support systems, which are theoretically grounded in the Chapter I, Subsection *1.6 Integration of methods into decision support systems* on page 70.

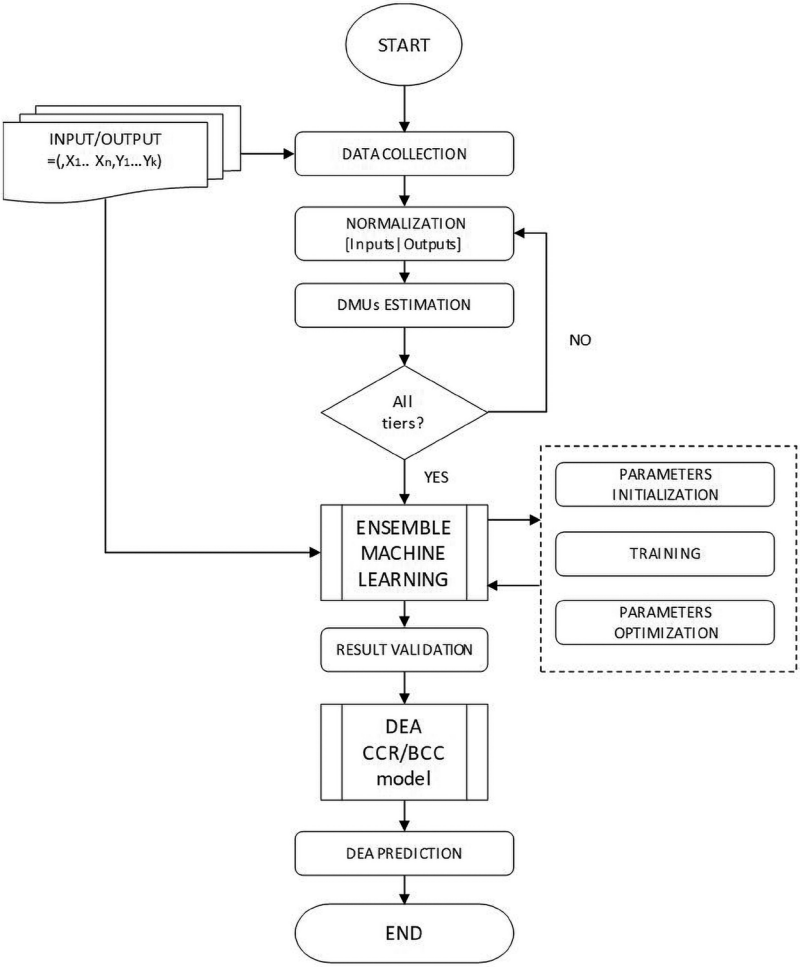


Figure 13. *Algorithm for proposed decision-making systems*
(Source: Author's representation)

The Farrell's approach is supported by DEA and SFA models. There is a number of DEA models implementing the Farrell Efficiency such Charnes *et al.* (1979) methods CCR and Banker *et al.* (1984) BCC models amended by Banker *et al.* (1996). The resulting efficiency is always at least equal to the one given by the CCR model, and those DMUs with the lowest input or highest output levels are rated efficient. The proposed BCC model allows for VRS contrasting to the CCR model, That is the main idea of the Farrell (1957) efficiency measures meaning the proportional changes in all inputs and outputs. Therefore, the Farrell (1957) input efficiency tells to what extent input can be proportionally reduced at the same production output.

2.2. Taxonomy of datasets selection for decision support system

Evans and Honkapohja (2012) provides a systematic treatment of the learning approach to modeling expectations formation in macroeconomics. This approach goes beyond rational expectations, the current standard hypothesis about expectations in macroeconomic theory. They focus on adaptive learning in which, at each moment of time, agents make forecasts using forecast functions formulated on the basis of available data. The common practice that these forecast functions can be revised accordingly and updated with time passing by as soon as there is a new data available. The body of the study is devoted to the statistical or econometric approach to learning which further postulates that econometric techniques are used to estimate the parameters of the forecast functions.

There are a number of evidences, that the majority of data sets collected by researchers in all disciplines are multivariate, meaning that several measurements, observations, or recordings are taken on each of the units in the data set (Everitt and Hothorn (2011)). Latest Zolbanin and Delen (2018) argue that the availability of data in massive collections in recent past not only has enabled data-driven decision-making, but also has created new questions that cannot be addressed effectively with the traditional statistical analysis methods. The traditional scientific research not only has prevented business scholars from working on emerging problems with big and rich data-sets, but also has resulted in irrelevant theory and questionable conclusions; mostly because the traditional method has mainly focused on modeling and explanation than on the practical problem and the data. Provost and Fawcett (2013) argues the fundamental principles of data science: data, and the capability to extract useful knowledge from data, should be regarded as key strategic assets. However, there is a common perception from business perspective is that manipulating data is often processed without solid comprehension of underlying methods.

Hence, introduction and deployment of machine learning models involves a series of steps that are almost similar to the statistical modeling process, in order to collect, validate and train model with hyper-parameters (Dangeti (2017)). An often methodological mistake occurs by taking into account the parameters of a prediction function and testing it on the same dataset. Hence, a model that would just repeat the labels of the samples that it has just seen would have a perfect score but would fail to predict anything useful on unseen data. This situation is called overfitting. In order to avoid overfitting, algorithm requires a larger number of training patterns. The kernel method that permits to deal with the low-dimensional input space instead of the high-dimensional feature space. The model performs training on the 50% of the given dataset, 25% is used for the training purpose and 25% for testing. The major disadvantage of this method is that we perform training on the 50% of the dataset, it may possible that the remaining data contains some hidden layers, which are left while model training.

The research investigates the efficiency of the selected companies listed on the Nasdaq Baltic Index. Answering the question of efficiency might assist greatly to the decision-makers during the crisis. The research goes far beyond estimation of the companies' performances from the financial point of view. Efficiency assessment are heavily dependent on the dataset quality that is used as an input to the productivity model. As now there are nu-

merous models based on nonparametric model. However, there are certain characteristics of data that may not be acceptable for the execution of nonparametric models.

The general structure of the datasets layers is summarized in the Figure 14 where different layers of datasets are decomposed by its assembled parts. Current methodology of data representation is often given in single data table. But recently most of these methodologies are protracted to relational cases. Relational data mining combines various data mining techniques with multiple dataset for extracting the knowledge from it.

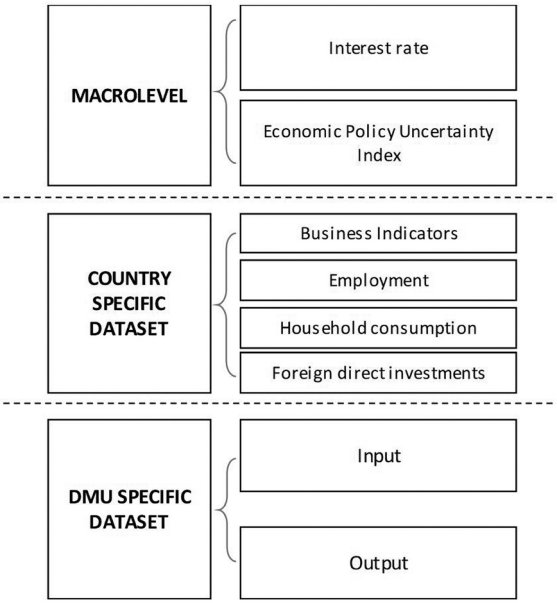


Figure 14. *Data model*
(Source: Author's representation)

The datasets for uncertainty are represented by multiple sources. Baker *et al.* (2015) and alternatively Jurado *et al.* (2015) developed an index of Economic Policy Uncertainty, which is based on mass-media coverage frequency and also defined as the common volatility in the unforecastable component of a large number of economic indicators. Several types of evidence indicate that the index proxies for movements in policy-related economic uncertainty. Using firm-level data, the study found that policy uncertainty is associated with greater stock price volatility and reduced investment and employment in policy-sensitive sectors like defense, health care, finance, and infrastructure construction. At the macro level, innovations in policy uncertainty foreshadow declines in investment, output, and employment.

The Harmonized Index of Consumer Prices (EU-HICP) is calculated in each Member State of the European Union to allow the comparison of consumer price trends in the different Member States. It measures the change over time in the prices of consumer goods

and services acquired, used or paid for by euro area households. The term harmonized denotes the fact that all the countries in the European Union follow the same methodology. This ensures that the data for one country can be compared with the data for another. The EU-HICP is compiled by Eurostat and the national statistical institutes in accordance with harmonized statistical methods. As Eurostat defines, the Industrial Production Index (IPI) shows the output and activity of the industry sector. The main indicator is to evaluate monthly changes in the output amount. The entire datasets are represented according to the Statistical classification of economic activities in the European Community. Industrial production is represented under the position *Fixed base year, per year, last year, type volume-index*. The index is seasonally adjusted. Growth rates related to the previous month estimated panel datasets with seasonally adjusted figures. Further, the growth rates related to the same month of the previous year are estimated from panel datasets with adjusted figures.

According to the OECD definition *Unemployment rate* is the number of unemployed people as a percentage of the labor force, where the latter consists of the unemployed plus those in paid or self-employment. Unemployed are defined those who claimed without any work under the condition that they are available to be employed and that they are active seekers in the last four weeks. Once unemployment rate is relatively high, the labor force shrinks if people stop seeking job actively anymore. This implies that the unemployment rate may fall, or stop rising, even though there has been no underlying improvement in the labor market.

The OECD defines household spending is the amount of final consumption expenditure made by resident households to meet their everyday needs: food, clothing, private housing or rent, energy, transport, durable goods, health costs, leisure, and miscellaneous services. The estimated share is normally around 60% of GDP. It is therefore an important variable for economic analysis of demand side. Household spending, including government transfers, is the actual individual consumption that it is equal to households' consumption expenditure plus those (individual) expenditures of general government and non-profit institutions serving households to support in form of medical care and education. *Housing, water, electricity, gas, and other fuels* is one out of the twelve categories included as part of individual consumption expenditures. Housing and energy expenditures consist of actual rentals for housing, imputed rentals for owner-occupied housing, housing maintenance and repairs, as well as costs for water, electricity, gas and other fuels. All OECD countries should provide their datasets in line with the 2008 System of National Accounts (SNA).

Per OECD definition, Foreign Direct Investment flows record the value of cross-border transactions related to direct investment during a given period of time, usually a quarter or a year. Typically, financial flows are made of reinvestment of earnings, equity transactions and transactions related with intercompany debt. In the reporting economy, such flows are in form of transactions increasing the investment indicator. The foreign investors participate in resident enterprises through purchases of equity or reinvestment of earnings, minus any transactions that reduce the investment that foreign investors already have in own enterprises in forms of sales of equity or borrowing by the resident investor. Inbound flows are in form of transactions that increase the investment that foreign investors have in

resident enterprises less transactions that decrease the investment of foreign investors in resident enterprises.

In the scientific literature as well as in industry there is a continuous discussion regarding the proper definition and selection of inputs and outputs. Cook *et al.* (2014) underpins that DEA is not a form of regression model, but it is a frontier-based linear programming-based optimization technique. Therefore, it does not make any sense to set a sample size requirement to DEA model but the quality and interpretability of the parameters. The model should be seen as a specific individual performance benchmarking tool. Therefore, a mixture of ratios or percentiles and raw data is permissible in DEA applications.

Wagner and Shimshak (2007) argues, that a number of variables to be considered for efficiency analysis is normally very large. Any possible resource both tangible or intangible employed by organization can be treated as an input variable. Therefore, the output variables defined as a performance or an outcome of operational activity, which transforms resources into organizational performance. Uncertainty in terms of environmental variables affect the operational process are also should be aggregated and considered, because they influence the availability of resources. As mentioned above, there us a number of approaches proposed in the literature suggesting strongly to limit the number of variables relative to the number of organizations in dataset. The larger the number various variables, the higher linear programming solution space dimensions will become with the less discriminating power. The effect in terms of DEA analysis will shift a larger number of compared peers to the efficient frontier with high efficiency scores.

The financial point of view at efficiency require estimation under the input-oriented approach due to underlying assumption that financial institution poses higher control over inputs as a general rule rather than outputs. In the time of technological convergence, the competitive advantage is characterized by a better input resource management rather than scale effects. However, there are also some evidences of adoption the output-oriented approach. More recently, some studies provide further modifications of the interdisciplinary approach, which is profit-oriented and defines revenue components as outputs and cost components as inputs. Drake *et al.* (2006) gives precise definitions:

> *"...from the perspective of an input-oriented DEA relative efficiency analysis, the more efficient units will be better at minimizing the various costs incurred in generating the various revenue streams and, consequently, better at maximizing profits…"*

Therefore, there is a general understanding of the major categories of inputs and outputs. Despite the facts, it does not necessarily involve the consistency and regularity with respect to the specific inputs outputs variables used by different researches.

Following Viebig *et al.* (2008), Rawley and Benton (2009), Massari *et al.* (2016) there is a traditional approach in valuation methodologies include the asset-based approach, the discounted cash flow method, and the comparable peers approach can be elaborated. But a number of shortcomings have been admitted in terms of companies' comparability. Therefore, the application of DEA method to obtain companies comparability can be considered as an extension of the market-based approaches. All these factors analyzed are the market

value for assets, intellectual property, patents, trademarks, copyrights and inventory. The addition of these values will give an approximated evaluation of the market value.

Another important part of ratios is related to liquidity representing organizations' short term financial situation or solvency. Liquidity ratios are related to the ability of organizational resources to fulfill short term cash requirements. A lack of liquidity imposes failure of an organization to take advantage of beneficial opportunities. Organizations facing short term liquidity risk are distressed by the lack of operational funds inflows and outflows along with its perspective affecting future performance negative way. The definition of *short term* is traditionally seen in terms of a period up to one year identified with the operational business cycle. All counter-parties are influenced by short term liquidity difficulties identified by inability to execute obligations against counter-parties. In case of unlimited liability of owner, a limited liquidity jeopardizes their personal assets. For financial institutions finance such organization, a lack of liquidity will delay interests' and principal payments' payoffs and it will cause risks of other losses. When any organization fails to serve its current obligations, the further future development is doubtful and the risks factors getting increased.

Set of parameters widely used in financial analysis is related with working capital to assess short term liquidity. The definition of the working capital implies the surplus of current assets over current liabilities. When current liabilities are above current assets, then organization has a lack of working capital. Working capital is a significant measure for liquid assets and liquid reserve of an organization facing uncertainties including balance of cash flows. In case of negative working capital organization might have difficulties in conducting payments.

Various factors influence liquidity are operating activities by decomposing working capital in account receivable and inventory. Nowadays financial credit instruments play enormous role in operation activities. Therefore, credit management become a significant part of working capital. In order to analyze organizational efficiency, the measurement of the quality of receivables accounts is also important. Furthermore, an increase in assets decrease sales performance, so it may create a liquidity problem since loans and advances settlement as a rule arise from transformation of current assets into cash.

Many studies revolving about the characteristics of the current ratios admit, that there is a limited ability to identify the cash flows for operations. It initiated a search for a dynamic measure indicators of liquidity for a better insight of liquidity risk. Since liabilities require cash, a ratio comparing operating cash flow to current liabilities overwhelms the static disposition of the current ratio.

Leverage ratios are linked with long term efficiency. The advantages of return to financial leverage on a long term debt positions lead to benefits to equity holders. But the same time, the underlying risk probability with increased leverage is the risk of certain operational payments can be postponed in crisis time, while the charges related to debt have adverse effect in this case. Therefore, an over-leverage effects might jeopardize financing flexibility, which affects the ability to appeal additional funds in periods of market volatility.

Worth to mention that there are several variations in debt ratios based on various assumptions. Here capital structure indicators accumulate the overall performance of respec-

tive organization. The ration of liabilities and equity capital helps to assessing solvency perspective. In case of the larger proportion of debt in it and the greater the fixed payments of interests all of these factors increase the likelihood of insolvency in turbulences.

Profitability ratios conduct performance analysis in several ways. Sales, revenue indicators, gross profit and net income are generally accepted performance indicators according to regulations. It should be articulated clearly that none of these indicators taken separately into analysis can be plausible and convincible proxy for organization's performance due to interdependency of operation and heterogonous environmental process. Profitability indicators make use of the margin analysis to scrutinize the return on sales activities considered capital retained. Profit margins reveal the organization's ability to come with a product on the market at a low cost or a high price. Obviously, profit margins itself are not straightforward measures of profitability to any extent since they are biased on operating revenue, but not on the investment activities or the investors' equity. The indicators based on the organizational earnings can improve profitability analysis which is isolated from operational activities and concentrated on long-run perspective.

Among direct profitability indicators, return on Assets (ROA) and Return On Investments (ROI) is no doubtfully the most widely established parameter of organizational performance, which is a rational indicator of an organization's long-term financial perspective. It uses aggregated factors from both financial income statement and from the balance to evaluate profitability. It enables assess a better way the returns and risks related with an organization from operating decision's perspective and environmental effects.

In the current study, the comparative evaluation among the organizations is an important consideration. In addition, the inputs are an outcome of managerial decision, which are the subject of bounded rationality. Thus, input-based formulation is recommended for this particular study but compared against output-based approach. The objective of the analysis is to elaborate a benchmark for DMUs across industries but stock listed, because stock market is an important part of the economy. The stock market plays a play a major role in the growth of the industry and commerce of the region that eventually affects the economy of the country to a great extent. That is reason that the government, industry and even the central banks of the country keep a close watch on the stock market. The stock market is important from both the policy-maker's point of view as well as the investor's point of view.

The Author proposes to incorporate the most significant variables into the model to let ensemble machine learning algorithms find the best possible pattern for a better prediction and classification outcome that existing regression-based approaches.

2.3. Data mining techniques

The approach developed by Xie *et al.* (2016) shows that dimensionality reduction have a central role to many data-driven application domains and has been studied extensively in terms of distance functions and grouping algorithms. Celebi and Aydin (2016) suggest the model-based approach to dimensionality reduction, where each group is represented by a parametric distribution, and then a finite mixture model is used to model the observed

data. Parameters are estimated by optimizing the fit, expressed by the likelihood, between the data and the model.

Doumpos *et al.* (2018) shows that data-driven approaches based of models constructed on historical data analysis techniques are useful tools for identifying meaningful and homogeneous groups of customers. Kassambara (2017b) says that the dimensionality reduction is one of the important data mining methods for discovering knowledge in multidimensional data. The goal of such approach is to identify pattern or groups of similar objects within a data set of interest. In the literature, it is referred as pattern recognition or unsupervised machine learning. However, Marsland (2011) underpins that the basic idea is that by having lots of learners that each get slightly different results on a dataset some learning certain things well and some learning others and putting them together, the results that are generated will be significantly better than any one of them on its own.

H. F. Lewis and Sexton (2004) and N. D. Lewis (2015) show that, in the dataset, where the actual baselines were known for each input and output factor. The values within each column were subtracted from their respective baseline. These findings give adequate characteristics for the inputs. However, the outputs parameters are still needed to include larger datasets. These datasets were transposed against their actual values. Thus, the integrity of the relationships of the data was maintained. There is also a caution when taking the inverse of data as a translation. This translation may also cause a variation in the efficiency scores. Thus, decision-making bias takes a great influence again, depending on the count of decision-makers involved. Dealing with translation error is to explicitly include before and after performance. That is, instead of subtracting the data from period to period, or adjusting with inverses, the purest method may be to use the previous period's performance as an input and this period's performance as an output for those measures where larger values are better, and the opposite for those measures where smaller values are better. This will require the additional input and output factors to be included and may hurt the discriminatory power of some productivity models if not enough DMUs exist.

Ali and Seiford (1990) shows that a displacement does not alter the efficient frontier for certain DEA formulations and thus these approaches are translation invariant. Thus, in the additive model unbounded value coefficients are added to any input and output parameters. It helps to solve the issue of negative or zero-valued problems in terms of any output in the BCC model. Bowlin (1998) addresses the substitution of a very small positive value for the negative value if the variable is an output. He suggests this method due to the nature of DEA models, which is define each DMU as good as possible. It leads to overestimate the outputs of the best performing DMUs. Thus, an output variable with a very small value would not be expected to contribute to a high efficiency score which would also be true of a negative value. Thus, this method of translation proved in general incorrect affect the efficiency score. Obviously, the given value cannot be larger than any other output value in the given data set.

When evaluating organizations and attempting to get the necessary parameters it is vital important to handle the situations where data is missing.

One approach is to get a best estimate as it would be the value for the missing data point. This is the simplest way to accomplish missing data, but is very subjective. An expected

value may be determined in this situation. One method to get an expected value is to apply the subjective values into an expected value calculation based on this probability distribution. So the β distribution expected value calculation is defined as:

$$V_e = \frac{V_o + 4V_m + V_p}{6} \qquad (2.2.2.1)$$

where V_p is the pessimistic value, V_o is the optimistic value, V_m is the most like value in respect to the estimated V_e.

Since Little (1988) suggested and Donders *et al.* (2006) further developed the opportunities to handle missing data are still quite limited. In all the situations presented by researchers and practitioners of DEA, it is still a relatively subjective approach in filling a gap from missing data. Before any data attribution, the source of missing data should be found and treated:

1. Missing at Random (MAR)
2. Missing Completely at Random (MCAR)
3. Missing not at Random (MNAR)

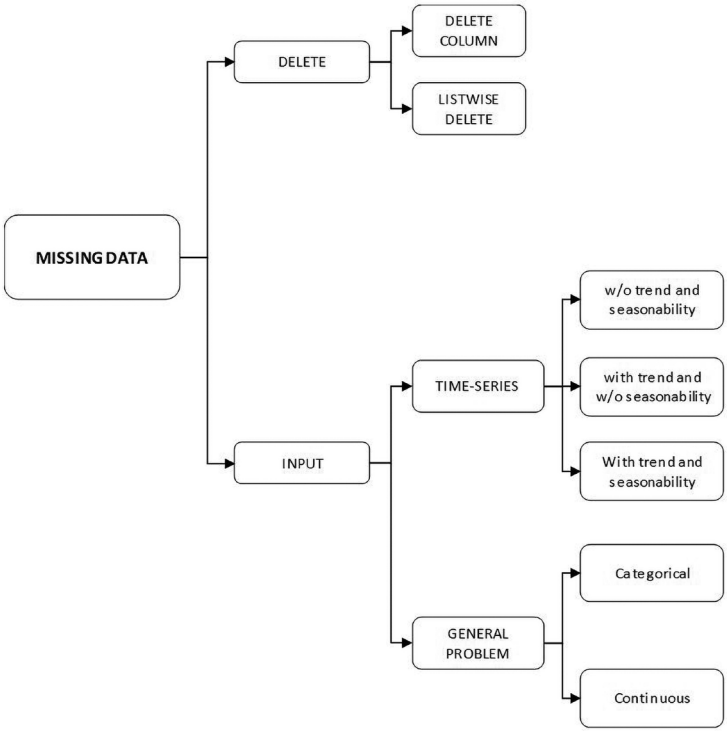


Figure 15. *Handling missing data*
(Source: Author's representation)

Listwise deletion removes all data for an observation that has one or more missing values. Particularly if the missing data is limited to a small number of observations. However, in most cases, it is often disadvantageous to use listwise deletion. This is because the assumptions of MCAR are typically rare to support. As a result, listwise deletion methods produce biased parameters and estimates. Pairwise deletion examines all cases in which the variables of interest are present and thus maximizes all data available by an analysis basis. A strength to this technique is that it increases power in your analysis but it has many disadvantages. It assumes that the missing data are MCAR. Computing the overall mean, median or mode is a very basic imputation method, it is the only tested function that takes no advantage of the time series characteristics or relationship between the variables. It is very fast, but has clear disadvantages. One disadvantage is that mean imputation reduces variance in the dataset.

To begin, several predictors of the variable with missing values are identified using a correlation matrix. The best predictors are selected and used as independent variables in a regression equation. The variable with missing data is used as the dependent variable. Cases with complete data for the predictor variables are used to generate the regression equation; the equation is then used to predict missing values for incomplete cases. In an iterative process, values for the missing variable are inserted and then all cases are used to predict the dependent variable. These steps are repeated until there is little difference between the predicted values from one step to the next, that is they converge.

It provides good estimates for missing values. However, there are several disadvantages of this model which tend to outweigh the advantages. For example, because the replaced values were predicted from other variables they tend to bias and so standard error is deflated. One must also assume that there is a linear relationship between the variables used in the regression equation when there may not be one.

2.4. Two-stage DEA nonparametric efficiency analysis

The DEA modelling is a mathematical programming method for the analyzing of production frontiers. The measurement of efficiency is therefore relative to these frontiers, where each organization in subset is assigned an efficiency score. One of the eminent advantages of DEA is that it performs even with small subsets and for implementation of this method any prior assumptions are not needed about the distribution of inefficiency. Nonparametric modeling does not demand any functional form on the data in determining the most efficient peer. However, nonparametric modeling has also own limitations such as that DEA assumes data to be validated against measurement error and the method is sensitive to outliers.

The two-stage models in Figure 16 use data outputs and inputs in the first stage, and use data on observable exogenous factors in the second stage, the objective being to determine the impact of the observable exogenous factors on initial evaluations. In case of applying machine learning approach the DEA-based model is capable of attributing some portion of the variation to the effect of statistical noise. Especially in terms of analyzing financial statements, the relationship between financial information and organizational value is established through a two-stage fundamental process:

1. A **predictive organizational efficiency measure** to link current performance data to resources allocation in forms of economic value for firms.
2. Valuation connection that projects organizational efficiency to **market performance**.

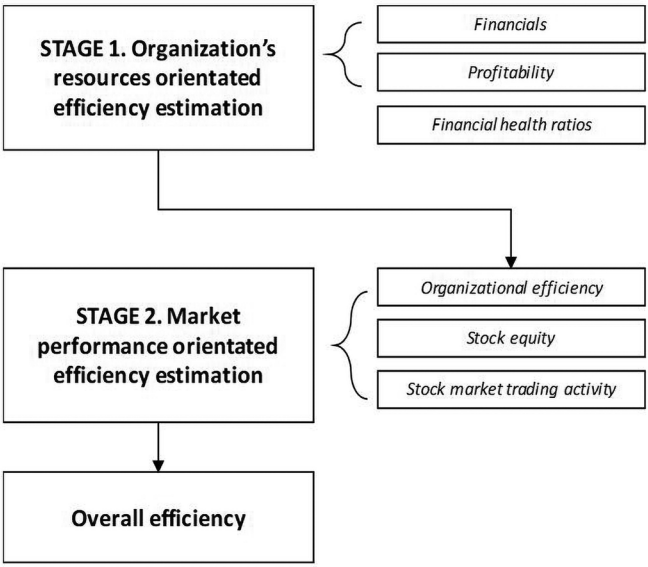


Figure 16. *Two-stage nonparametric efficiency estimation approach (Source: The Author's representation)*

Based on H.-H. Chen (2008) confirms that the parameters consideration relies on *size* effect, where smaller in operations size organizations used to yield higher returns, and that returns are higher for stocks with low equity ratios. Hence, by the application of DEA method, companies can be classified into either an *inefficient* or *efficient* group. For each it is assumed that a relevant upper or lower threshold of its market value exists. Likewise, the market value of an *inefficient* company represents range of values, which are estimated by using reference set defined by DEA method. The results bellow exhibit that using DEA method in valuing financial entities is a promising method, which is important in company valuation. Based on this assumption above the Author proposes the following datasets for the DMU resources orientated efficiency estimation at the Stage 1 in Table 6 and grouped by implication area in Column (1) and analytical parameter in Column (2). Each group represents a set of variables, which indicate the most prominent and comparable features. The variables should correspond to common financial standards. Therefore, the variables are comparable across the industries and markets. The comparability of variables in their standards and levels is one of the most important requirement for efficiency assessment using nonparametric models. The Stage 1 Efficiency comprise the firm level efficiency with the assumption that the organization have more control over inputs, rather than outputs.

That means that in order to achieve a better performance, organization as the rule practice the cost optimization strategy rather than profit maximization approach.

Table 6. *Organization's resources orientated efficiency parameters*

Group	Parameter	Variables
Profitability	OUT	Entity profitability according to standards
Financials indicators	IN	Gross Margin
		Operating Margin
		Operating Cash Flow
		EBIT
		Capital Expenditure
		Free cash flow
Leverage and solvency indicators	IN	Return on Assets
		Return on Equity
		Net Margin
		Financial Leverage
		Long term debt to assets
		Long term debt to tangible assets
		Interest coverage ratio
Profitability indicators	IN	Working Capital
		Total Equity

(Source: Author's representation)

The DMU specific dataset based on the concept of Rappaport (1986) stated that the shareholder profitability value is the merit for business performance. The discussion is followed by an enumeration of the shortcomings of the accounting return on investment and accounting return on equity as standards for measuring business performance. According to Anadol *et al.* (2014) asserts that the value is the uppermost price that an informed buyer wants to pay off in a free market. Depending on the need and requirements the valuation may come from a variety of settings will require different valuations. Therefore, company valuation methods traditionally are based on the common ratios and values.

In Table 7 there are the market performance orientated efficiency parameters in order to involve uncertainties and externalities into efficiency estimation. However, it is reasonable to regard these efficiencies separately due to the nature of the efficiency. In the first case (Stage 1), the input allocation and firm level efficiency is the key for the firm level efficiency

assessment, where externalities do not have direct impact on the technology applied in production. The Stage 2 efficiency has output oriented approach. The Stage 2 efficiency is related with the market capitalization and market stock equity. Therefore, the input parameters are greatly influenced by uncertainties and market volatilities. stockholders are interested in profit maximization. Hence, they do not influence firm level efficiency directly.

Table 7. *Market performance orientated efficiency parameters*

Name	Parameter	Variables
Capitalization	OUT	Value of all a company's shares of stock
Efficiency 1	IN	Resource input orientated efficiency
Market coefficient	IN	Coefficient of investors activity
Equity ratio	IN	The assets remaining once all liabilities have been settled
Stock volatility	IN	Market uncertainties

(Source: Author's representation)

Earlier Charnes *et al.* (1979), Broek *et al.* (1980), Banker *et al.* (1984) and later Kaoru Tone (2001), K. Tone and Tsutsui (2009), K. Tone and Tsutsui (2010) point out among other researchers that despite the fact DEA-based models represent well-known approaches for accessing the relative efficiency of a set of identical DMUs, there are shortcomings to overcome. There is a number of models proposed to include various factors into a DEA-based evaluation. These models can be grouped roughly into one-stage models and two-stage models. One-stage models use data on outputs, inputs and observable factors all at once, the objective being to control for observable exogenous factors in the evaluation. Nevertheless, these models are deterministic and they are not able to attribute the effect of statistical noise.

Coelli *et al.* (2005) underpins that having a limited number of observations along with many efficiency factors will result in many organization appearing on the frontier. Another issue is that considering inputs-outputs as homogenous factors in their heterogeneous state may bias the results. Nonparametric method does not take into account for differences in the environmental settings, what clearly leads to misleading results. Nonparametric approach does not distinguish multi-period optimization or risk factors in decision making process.

The fundamental purpose of the analysis is to categorize hidden causal factors taken from financial accounting statements that can be used to explain the performance market stock value. In terms of decision support systems, DEA should be assessed under pre-assumption either *constant returns to scale* (CRS) or *variable returns to scale* (VRS). In the influential research study, Charnes *et al.* (1979) brought a nonparametric input-orientated model under CRS assumption. This model returns a score that give a notion of the *overall technical efficiency* (OTE) of peer organization. Banker and Morey (1986) developed further approach VRS by differentiation into two components such as pure *technical efficiency*

and *scale efficiency*. The first definition referred to the ability of organization to utilize own given resources, while the second means the exploiting scale economies by performing operation with the production CRS frontier. In most of the researches the DEA approach is carried out under the VRS condition, arguing that CRS assumption is adequate under the circumstance where all organization performing at the optimal production function point. Worth to mention that there is a number of excellent studies which go specifically for CRS assumption. Thus, in recent studies the results are obtained under both CRS and VRS condition in order to find the upmost discrimination power.

Charnes *et al.* (1979) provides positivity requirement of DEA models, meaning that DEA models are not capable of completing an analysis with negative numbers and all numbers must be non-negative and preferably strictly positive. The methods for eliminating the problems of non-positive values is done by adding large positive constant to the values of the input or output that has the non-positive number. Bowlin (1998) advise to make the negative numbers or zero values a smaller number in magnitude than the other numbers in the data set to overcome some of the complications of this limitation. However, results due to translation variance may change depending on the scale used by the models. Ali and Lerme (1997) argues that ratio-based DEA models are translation invariant as the BCC model with some limitations. A number of phases should be executed in order to minimize translation errors and to ensure that data scaling do not influence the final results of the performed analysis. Foremost, every interaction should be made to obtain and apply actual values for the initial starting point instead of using arbitrary and subjective values.

Cook *et al.* (2014), Zhu (2016b) argue that the choice of a DEA model between BCC-CCR is the crucial point for any researcher in this field. In order to estimate a DMU's efficiency it should be upfront well-defined DMUs activity. It gives the answer which efficiency the DEA model should capture.

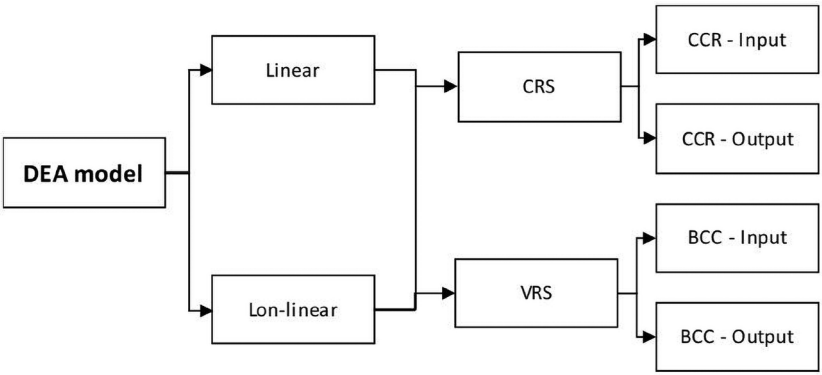


Figure 17. *DEA models classification*
Source (Adopted by the Author, Ali et al. (1995))

The Author represents in the Figure 17 the main approaches by models selection based Charnes *et al.* (1979), Charnes *et al.* (2013) assumptions of:

1. The CCR model exhibits a linear constant returns-to-scale effect of the envelopment function scale.
2. The BCC are known for variable returns-to-scale envelopment function scale.

Let's assume $\boldsymbol{m}$ inputs, in order to produce $\boldsymbol{s}$ outputs under the condition that $\boldsymbol{DMU_j}$ $\boldsymbol{(j=1,\ldots,n)}$ is using $\boldsymbol{x_{ij}}$ of input variable $\boldsymbol{i}$ and produce output $\boldsymbol{y_{rj}}$ of the variable $\boldsymbol{r}$. Assume, that $\boldsymbol{x_{ij} \geq 0}$ and $\boldsymbol{y_{rj} \geq 0}$ and each DMU has at least one positive input and output variable. Each input and output variable has some weighted confidents $\boldsymbol{v_i}$ and $\boldsymbol{u_r}$:

$$\begin{aligned} \boldsymbol{Input} &= v_i x_{ij} + \cdots + v_m x_{mi} \\ \boldsymbol{Output} &= u_r y_{rj} + \cdots + u_s y_{sr} \end{aligned} \qquad (2.1.1.1)$$

Using the linear programming method, we can define the weight as the following:

$$\frac{\boldsymbol{Output}}{\boldsymbol{Input}} \qquad (2.1.1.2)$$

DEA requires several inputs and outputs to be considered at the same time to measure DMU efficiency which is defined as:

$$\boldsymbol{Efficiency} = \frac{\boldsymbol{weighted\ sum\ of\ outputs}}{\boldsymbol{weighted\ sum\ of\ inputs}} \ \forall\ \boldsymbol{DMU} \qquad (2.1.1.3)$$

The optimal weight coefficient will vary among DMU and input and output coefficient can be represented in a matrix form:

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \cdots & \cdots & \cdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix}$$

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & \cdots & y_{1n} \\ \cdots & \cdots & \cdots \\ y_{s1} & \cdots & y_{sn} \end{pmatrix} \qquad (2.1.1.4)$$

CCR model definition. The CRS assumption is underneath of the CCR model. The main idea of CRS is to apply in the production frontier with multiple input and output data. Considering equations above it is obvious to perform $\boldsymbol{n}$ optimizations in form of weighted coefficients under the maximum condition for input and output in order to estimate the efficiency of $\boldsymbol{n}$ DMU. In general form, let's evaluate $\boldsymbol{DMU_o}$, where $\boldsymbol{o}$ between $\boldsymbol{1,2,\ldots,n}$. Then the following equation needs to be solved for $\boldsymbol{v_i}$ $\boldsymbol{(i = 1,2,\ldots,m)}$ input variables and coefficients and $\boldsymbol{u_r}$ $\boldsymbol{(r = 1,2,\ldots,s)}$ for output:

$$max\ \theta = \frac{u_1 y_{1o} + u_2 y_{2o} + \cdots + u_s y_{so}}{v_1 x_{1o} + v_2 x_{2o} + \cdots + v_m x_{mo}} \qquad (2.1.1.5)$$

Under the constraints that:

$$\frac{u_1 y_{1o} + u_2 y_{2o} + \cdots + u_s y_{so}}{v_1 x_{1o} + v_2 x_{2o} + \cdots + v_m x_{mo}} \leq 1\ (o = 1, \ldots, n) \qquad (2.1.1.6)$$

$$v_1, v_2, \ldots, v_n \geq 0$$

$$u_1, u_2, \ldots, u_s \geq 0$$

Constraints, mean that output cannot exceed input more than 1 for each DMU. The optimal θ is equal to 1. The system can be represented in its linear form:

$$max\ \theta = u_1 y_{1o} + u_2 y_{2o} + \cdots + u_s y_{so} \qquad (2.1.1.7)$$

Under the constraint that:

$$v_1 x_{1o} + v_2 x_{2o} + \cdots + v_m x_{mo} = 1$$

$$u_1 y_{1j} + u_2 y_{2j} + \cdots + u_s y_{sj} \leq v_1 x_{1j} + v_2 x_{2j} + \cdots + v_m x_{mj}$$

$$(j = 1, 2, \ldots, n)$$

$$v_1, v_2, \ldots, v_n \geq 0$$

$$u_1, u_2, \ldots, u_s \geq 0 \qquad (2.1.1.8)$$

The linear programming task is derived theorems formulated by Cooper *et al.* (2007) and Cooper *et al.* (2011) optimal value of θ in linear task does not depend on the input and output variables for each undependable DMU. Assumed that values θ^*, v^*, u^* are found. Then v^*, u^* are the set of preferable weighted coefficients for DMU_o if optimizing:

$$\theta^* = \frac{\sum_{r=1}^{s} u_r^* y_{ro}}{\sum_{i=1}^{m} v_i^* x_{io}} \qquad (2.1.1.9)$$

So now it is possible to estimate the CCR effectivity. In order to attribute to DMU effectivity in the CCR model, the following constraint should be satisfied $\theta^*=1$ and there is at least one solution for $v^* > 0$ and $u^* > 0$, otherwise DMU is inefficient.

There are two possible options in CCR model:

1. **Input** – minimizing input by the given output variables
2. **Output** – maximizing output by the given input variables

Assume, that (x_j, y_j) – input-output vector for DMU_j $(j = 1,2,\ldots,n)$. The activity assumption formulated as $DMU\ (x, y) \in R^{(m+s)}$, $x \in R^m$, $x_j \geq 0$, $x_j \neq 0$ u $y_j \geq 0$, $y_j \neq 0$ $(j = 1,2,\ldots,n)$. $R^{(m+s)}$ – linear vector, where m and s define the input and output variables. The set of possible production possibilities is defined as P under the following constraints:

1. (x_j, y_j) $(j = 1,2,\ldots,n) \in P$.
2. If any $(x, y) \in P$ exists, then the vector $(tx, ty) \in P$ exists, where $t > 0$, representing constant return to scale.
3. If any vector $(x, y) \in P$ exists, then any exists semipositive vector $(x_1, y_1) \in P$, where $x_1 \geq x$ and $y_1 \leq y$.

Thus all possible sets of production for CCR defined:

$$P = \{(x, y) \mid x \geq X\lambda, y \leq Y\lambda, \lambda \geq 0\}, \qquad (2.1.1.10)$$

where λ – semipositive vector R^n.

The BCC model definition. The key feature of CCR model is CRS. Taking this fact into account the BCC model proposed. The production frontier of CCR is a linear function whereas production frontier of BCC has linear and convex characteristic representing the increasing or declining trends. The set of possibilities of BCC model defined in vector form as:

$$P = \{(x, y) \mid x \geq X\lambda, y \leq Y\lambda, e\lambda = 1, \lambda \geq 0\}, \qquad (2.1.1.11)$$

where $X = (x_j) \in R^{m \times n}$ and $Y = (y_j) \in R^{s \times n}$, a $\lambda \in R^n$.
CCR and BCC differs only by constraint $e\lambda = 1$, if $\lambda \geq 0$ adding VRS.

Malmquist productivity index (MPI) was introduced by Caves *et al.* (1982) to estimate the productivity change between two points in terms of ratios of distances function. The Malmquist nonparametric estimation method is designed to estimate changes in their technical advancement by comparing the production frontiers between the points in a time series. DEA can be used to describe the distance function of technical efficiency in the decision making units regardless of the specific production function. So here it becomes possible to establish the distance function and provide the MPI model, which takes the MPI of each firm as the estimate of the TFP of each listed companies by measuring the changes of TFPs in different periods. In order to build the analysis on the efficiency, it is necessary to calculate the efficiency scores for each firm in each year of observation. The applying multiple inputs and outputs assumed into Malmquist DEA model in case when DMUs are observed after certain period of time on a certain interval basis.

The productivity progress can be estimated using this common method mostly applied in assessing the productivity variation in healthcare sector. Two different measurements methods can be presented to estimate the frontier. Heathfield (1995), Färe *et al.* (2013 [1985]) the output oriented MPI change between period (t) and period (t+1) is given by:

$$M_0(y_t, x_t, y_{t+1}, x_{t+1}) = \frac{d_o^{t+1}(y_{t+1}, x_{t+1})}{d_o^t(y_t, x_t)} \times \left[\frac{d_o^t(y_{t+1}, x_{t+1})}{d_o^{t+1}(y_{t+1}, x_{t+1})} \times \frac{d_o^t(y_t, x_t)}{d_o^{t+1}(y_t, x_t)}\right]^{1/2} \qquad (2.1.1.12)$$

Where, $\boldsymbol{M_0(y_t, x_t, y_{t+1}, x_{t+1})}$ represents Total Factor Productivity change, and $\boldsymbol{d_o^t(y_t, x_t), d_o^t(y_{t+1}, x_{t+1}),\ d_o^{t+1}(y_t, x_t), d_o^{t+1}(y_{t+1}, x_{t+1})}$ represent the distance function values.

The MPI decomposed in to two components: efficiency change and technological change. Where, the first ratio outside the brackets indicates to efficiency change and the second ratios indicates to technological change. Thus, the technological change and efficiency change are expressed as followed:

$$\textbf{Efficiency change} = \frac{d_o^{t+1}(y_{t+1}, x_{t+1})}{d_o^t(y_t, x_t)} \qquad (2.1.1.13)$$

$$\textbf{Technological change} = \left[\frac{d_o^t(y_{t+1}, x_{t+1})}{d_o^{t+1}(y_{t+1}, x_{t+1})} \times \frac{d_o^t(y_t, x_t)}{d_o^{t+1}(y_t, x_t)}\right]^{1/2} \qquad (2.1.1.14)$$

The MPI measured by using DEA, and assume CRS output oriented approach. Thus, to compute the distance functions, four linear programing models are needed to solve for each DMU.

Since the beginning of nonparametric techniques for efficiency measurement, the DEA framework has been extended to include non-discretionary inputs that distress organizational efficiency. Golany and Roll (1989) gives the definition of a *non-discretionary factor* (ND) in the DEA. A decision-making organization does not possess influence over such factor even though this factor has importance in evaluating relative efficiencies among other peers.

Ruggiero (1998), Blackburn *et al.* (2014) show clearly that the control of multiple exogenous factors is the huge issue. The treating of the non-discretionary exogenous factors for nonparametric measures requires a relatively large number of observations compared to the parametric models. In this case, two separate frontiers based recognizing possible differences in the production technology should be recognized with discretionary inputs include exogenous socio-economic characteristics or environmental ones. TE is always based on resources and it is affected by DMU perception, some environmental factors along with random errors. The first factor is an endogenous variable, and the two others are exogenous variables. In order to improve the accuracy of efficiency assessment, there is a need to measure the specific impacts on productivity of these all factors.

Exanimating production process represented in Figure 18 as the series of inputs and non-discretionary exogenous factors allow to explicitly model intermediate inputs within DMU.

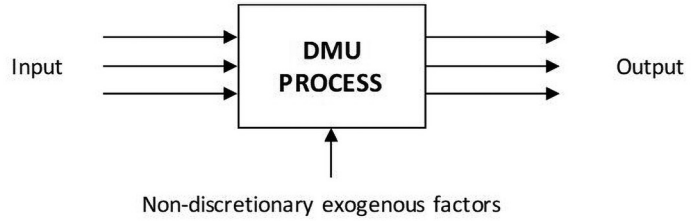


Figure 18. *Non-discretionary exogenous factors (Source: Bogetoft (2013), Author's representation)*

Bogetoft (2013) deals with multiple effects that may interact in complicated ways. Therefor there is a growing number of DEA extensions that avoid assuming that all inputs and outputs are by a single process in a row, the system can be modelled as formulated by distinct sub-processes. The main idea of DEA extension to involve non-discretionary exogenous factors makes possible presence of intermediate processes influenced from exogenously. Färe and Grosskopf (2012) respect non-discretionary exogenous factors which are connected in a network to form the overall frontier or reference technology.

2.5. Ensemble methods in machine learning

2.5.1. Support vector machines

Method of the Support vector machines (SVM) proposed by Vapnik (1992) is a mathematical approach in a supervised learning used for both classification assignments and regressions. They belong to a family of generalized linear classifiers. SVM is a classification and regression prediction algorithm that uses machine learning theory to maximize predictive accuracy while automatically avoiding overfitting.

As a starting point let's assume a linear programming method following Abe (2010), Campbell and Ying (2011), Witzany (2017). There is an assumption linear programing approach and SVM has the same aim to separate sets observations represented by the vectors of explanatory variables, possibly after a transformation, by a hyperplane in an optimal way. The aim of separation of cases in categories:

$$\boldsymbol{s}(\boldsymbol{x_i}) = \sum_{k=1}^{n} \boldsymbol{\beta}_k \boldsymbol{x}_{ik} \qquad (2.5.1.1)$$

Hence, the threshold value $\boldsymbol{C}$ defines categories separation by $\boldsymbol{s}(\boldsymbol{x_i}) < c, \boldsymbol{x_i} \in \boldsymbol{A_B}$ or $\boldsymbol{s}(\boldsymbol{x_i}) > c, \boldsymbol{x_i} \in \boldsymbol{A_B}$ respectively on training and validation samples in order to obtain *ex ante* predictions. In case when the explanatory factors are numerical value then the problem is formulated as a linear separability problem of the two sets of points $\boldsymbol{A}_G$ and $\boldsymbol{A}_b$ in the space $\boldsymbol{R^n}$. The error $\boldsymbol{\varepsilon}$ is defined as $\boldsymbol{s}(\boldsymbol{x_i}) < c, +\boldsymbol{\varepsilon_i}\, \boldsymbol{x_i} \in \boldsymbol{A_B}$ and $\boldsymbol{s}(\boldsymbol{x_i}) > c, +\boldsymbol{\varepsilon_i}\, \boldsymbol{x_i} \in \boldsymbol{A_B}$, where $\boldsymbol{\varepsilon_i} \geq \boldsymbol{0}$.

Then it becomes a typical linear program where the objective is to minimize the total sum of errors $\sum_i \boldsymbol{\varepsilon_i}$ that can be solved by common linear programing methods. The simplification is achieved when a constant error term $\boldsymbol{\varepsilon_i} = \boldsymbol{\varepsilon}$ assumed and minimizing just $\boldsymbol{\varepsilon}$ over potential values of the coefficient vector $\boldsymbol{\beta}$ and the threshold $\boldsymbol{C}$. The advantage of the linear programming approach is that it is possible to introduce new constraints required if the coefficient of one variable is larger than the coefficient of another variable. However, the linear programming approach should not have a trivial solution if all $\boldsymbol{\beta}$=**0** and $\boldsymbol{C}$=**0** .

One intuitive solution is to maximize the gap or margin separating the positive and negative examples in the training data. The optimal hyperplane is then the one that evenly divides the margin between two classes with closest data points to the separating hyperplane. Support vectors are defined as the problem is to find with maximal margin:

$$f(\mathbf{x}) = (\mathbf{w}^T \mathbf{x}_i + b), \qquad (2.5.1.2)$$

$$\mathbf{w}^T \mathbf{x}_i + b = 1$$

$$\mathbf{w}^T \mathbf{x}_i + b > 1$$

By a linearly separable dataset, the learning coefficients ***w*** and ***b*** of SVM as the function of $f(\mathbf{x}) = (\mathbf{w}^T \mathbf{x}_i + b)$ expressed as constrained optimization problem:

$$\text{find } \boldsymbol{w} \text{ and } \boldsymbol{b} \text{ that minimize: } \tfrac{1}{2}\|w^2\| \qquad (2.5.1.3)$$

, subject to: $y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \forall i$ by maximizing *f(x)*. This optimization problem is subject of the Lagrangian function defined as:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} - \sum_{i=1}^{N} \alpha_i [y_i(\mathbf{w}^{\mathrm{T}} \mathbf{x}_i + b) - 1], \text{ such that } \alpha_i \geq 0, \ \forall i \qquad (2.5.1.4)$$

, where $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \ldots \boldsymbol{\alpha_N}$ are Lagrange multipliers and $\boldsymbol{\alpha} = [\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \ldots \boldsymbol{\alpha_N}]^{\mathrm{T}}$. The support vectors are data points of each class closest to separation margin $\mathbf{x}_i$ with $\boldsymbol{\alpha_i} > \mathbf{0}$ solving for optimization conditions:

$$\mathrm{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i \text{ where, } \sum_{i=1}^{N} a_i y_i = 0 \qquad (2.5.1.5)$$

By substituting $\mathrm{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i$ into the Lagrangian function and by using $\sum_{i=1}^{N} a_i y_i = 0$ as a constraint the optimization problem can be expressed in a dual problem using conventional method:

$$\text{Find } \boldsymbol{\alpha} \text{ that maximizes } \quad \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i{}^T \mathbf{x}_j \qquad (2.5.1.6)$$

$$\text{subject to } \quad \sum_{i=1}^{N} \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad \forall i$$

A convex quadratic programming problem with a global minimum is the optimization problem is therefore. In the non-linearly separable case does not allow to get a linear hyperplane for separation. In order to solve the problem, the margin maximization approach might be set not restricted by adding data points to fall on the incorrect margin by increasing degree of freedom for error. Representing $\boldsymbol{\xi}_i$ as a slack variable to extend the error degree for each input data point. The non-linearly separable case has one of three possibilities for data points: outside and correctly classified, with $\boldsymbol{\xi}_i = \mathbf{0}$, inside and correctly classified, with $\mathbf{0} < \xi_i < 1$, Outside and incorrectly classified, with $\boldsymbol{\xi}_i = \mathbf{1}$.

The data is linearly separable if all slack variables have a value of zero otherwise there

is a non-linearly separable case with nonzero values. The optimization problem then to minimize error $\boldsymbol{\xi}_i \neq \mathbf{0}$:

$$\text{find } \mathbf{w} \text{ and b that minimize:} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i^2 \qquad (2.5.1.7)$$

$$\text{subject to:} \quad y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i$$

, whereas *C* is an predefined variable to set the slack variable to the value closest to zero. The optimal choice of *C* is the challenging tasks deepening much on the dataset. The optimization problem can be expressed in a dual problem using conventional method:

$$\text{find } \alpha \text{ that maximizes} \quad \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \qquad (2.5.1.8)$$

$$\text{subject to} \quad \begin{array}{l} \sum_{i=1}^{N} \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq C, \quad \forall i \end{array}$$

In order to solve the non-linearly separable case, SVM uses of a nonlinear mapping function $\phi(x)$: $R^M \rightarrow F$ to transform the non-linear input into a higher dimension feature space where the data becomes linearly separable. The mapping function $\phi(x)$ applied to transform the training datasets. The dual problem is solved in feature space using:

$$\text{find } \alpha \text{ that maximizes} \quad \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \qquad (2.5.1.9)$$

$$\text{subject to} \quad \begin{array}{l} \sum_{i=1}^{N} \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq C, \quad \forall i \end{array}$$

the resulting SVM form:

$$f(\mathbf{x}) = \mathbf{w}^T\phi(\mathbf{x}_i) + b = \sum_{i=1}^{N} \alpha_i y_i \phi(\mathbf{x}_i)^T \phi(\mathbf{x}) + b \qquad (2.5.1.10)$$

Yu and Kim (2012) shows that SVMs does the mapping from input space to feature space to support nonlinear classification problems. The kernel trick is helpful for doing this by allowing the absence of the exact formulation of mapping function which could cause the issue of curse of dimensionality. This makes a linear classification in the new space equivalent to nonlinear classification in the original space. SVMs do these by mapping input vectors to a higher dimensional space where a maximal separating hyperplane is constructed.

With the intention of keeping the nonlinear properties for SVM classification algorithm efficient, the kernel trick is applied. The idea is to specify the dot product $\phi(x_i)^T \phi(x_j)$ as a function $K(x,y)$, instead of the function ϕ directly, where K is a corresponding kernel function as summarized in Table 8.

Table 8. *SVM Kernel functions*

Description	Definition
Linear kernel	$K(\mathbf{x},\mathbf{y}) = (\mathbf{x}^{\mathrm{T}}\mathbf{y})$
Polynomial kernel with degree d	$K(\mathbf{x},\mathbf{y}) = (\mathbf{x}^{\mathrm{T}}\mathbf{y} + \mathbf{1})^{\mathrm{d}}$
Radial basis function kernel with width σ	$K(\mathbf{x},\mathbf{y}) = e^{\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\sigma^2}}$

(Source: Author's adoption from Scholkopf and Smola (2001), Smola and Schölkopf (2004))

There is a *kernel trick* approach, where a kernel is a function with the original feature vectors returns the same value as the dot product of its corresponding mapped feature vectors. Thus, the dual problem then becomes to:

$$\text{find } \alpha \text{ that maximizes} \quad \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \qquad (2.5.1.11)$$

$$\text{subject to} \quad \begin{array}{l} \sum_{i=1}^{N} \alpha_i y_i = 0, \\ 0 \le \alpha_i \le C, \quad \forall i \end{array}$$

and the resulting SVM form:

$$f(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}_i) + b = \sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \qquad (2.5.1.12)$$

In statistical learning theory the problem of supervised learning is formulated as a given set of training data $\{(x_1, y_1) \ldots (x_n, y_n)\}$ in $R^n x R$ sampled according to unknown probability distribution $P(x, y)$, along with a loss function in form of $V(y, f(x))$. The loss function evaluates the error, for a given set $x, f(x)$, which is expected instead of the actual value y. The problem consists in finding a function f that minimizes the expectation of the error on new data that is, finding a function f that minimizes the expected error: $\int V(y, f(x)) P(x, y)\, dx\, dy$.

The concept of SVM is a useful technique for data classification. The typical application for supervised learning based on SVM is the classification assignment. Identified labeled classes assist to specify whether the model is acting correctly or not. In the beginning, SVM classification includes identification of the known classes. Feature selection is the important part of modelling. It brings to SVM classification notion of unknown classes, which starts processes to differentiate the classes. Nevertheless, researches point out that

potentially find that boosting mostly outperforms the autoregressive benchmark, and that *K*-fold cross-validation works much better as stopping criterion than the commonly used information criteria (Buchen and Wohlrabe (2014)). Gu and Han (2013) proposed a novel large margin classifier namely Clustered Support Vector Machine (CSVM) to divide the data into several clusters by *K*-means. The separation problem can also be formulated as maximization of the distance between two hyperplanes. Based on Scholkopf and Smola (2001), Smola and Schölkopf (2004) and following Drucker *et al.* (1997), Cristianini and Shawe-Taylor (2000), Scholkopf and Smola (2001), H.-C. Kim *et al.* (2002), H.-C. Kim *et al.* (2003), Smola and Schölkopf (2004), Wang (2005), Gu and Han (2013), Ma and Guo (2014), Witzany (2017) binary classification is the task of classifying the members of a given set of objects into groups on the basis of whether they have some property or not. In the linearly separable case, there is one or more hyperplanes that may separate the classes represented by the training data with absolute accuracy. The main question remains how to find the *optimal* hyperplane that would maximize the accuracy on the test data.

2.5.2. Artificial neural networks

There is given a brief background of the Artificial Neural Network (ANN) employed in ensemble technique. The ANN input layer represents a real-valued array. An ANN in which the input layer of source nodes projects into an output layer of neurons is identified as single feed-forward network. In single layer network, the layer refers to the output layer of computation nodes.

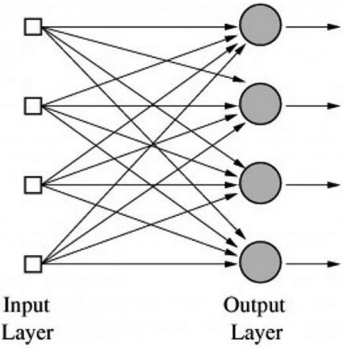


Figure 19. *Artificial Neural Network*
(Source: The Author's adoption from Graupe (2013))

The ANN consists of one or more hidden layers, whose computation nodes of hidden units. The function of hidden neurons is to interact between the external input and network output in some useful manner and to extract higher order statistics.

Graupe (2013) shows the theory in details, that any ANN consists of perceptron, which is elementary component of a network that takes multiple inputs and produces binary outputs. In machine learning terminology, the perceptron belongs to a supervised learning

technique, which can classify an input into binary class. It is a classification process than can do classification built on a linear predictor function joining weights of the feature vector. In machine learning, the perceptron is defined as a binary classifier function that maps its real-valued vector inputs $\boldsymbol{x}$ to an output value $\boldsymbol{f(x)}$:

$$f(x) = \begin{cases} 1, if\ w * x + b > 0 \\ 0\ otherwise \end{cases} \qquad (2.5.2.1)$$

where $\boldsymbol{w}$ is a vector of real-valued weights. The perceptron relation is given by $\boldsymbol{w} * \boldsymbol{x}$ is the dot product:

$$z_i = \sum_{i=0}^{m} w_i x_i \qquad (2.5.2.2)$$

where $\boldsymbol{m}$ is the number of real-valued vectors and $\boldsymbol{b}$ is the bias, which is independent of input values and helps correct the decision boundary.

The implication of ANN into ensemble techniques has significant advantages:

1. The method handles different data types
2. No glitches with outliers.
3. ANNs can estimate nonlinear functions with numerous inputs and layers entailed.
4. ANNs can incorporate stopping rules to prevent overfitting.

In particular, ANNs are most frequently used in previous studies since the prediction power is recognized to be better than the others techniques. Nevertheless, it has been regularly reported that models based on ANNs need a large amount of training data to evaluate the distribution of input pattern. There are difficulties of generalizing the results due to their overfitting bias. ANNs are t fully depends on feature set experience and understanding to preprocess data in order to select control parameters including appropriate input variables, hidden layers, learning rates, and momentum.

2.5.3. Random forest

Many studies have been motivated by the idea of combining complex models on the basis of set of other models. Researchers performing in the area of decision trees have offered a numerous solution in this context due to the nature of the decision trees. The concepts of bagging, boosting and other related approaches to merging decisions of sets of models are viewed by some sciences as the most significant success of data science and artificial intelligence research of the 1990s.

Averaging decisions theories of compound complex models can be justified in many researches. One of the most rational and practical path is the analysis grounded of Bayesian learning theory, where many models for given training data $\boldsymbol{D}$, the optimal choice of the class $\boldsymbol{c} \in \boldsymbol{C}$ for a data array $\boldsymbol{x}$ is defined by the Bayes Optimal Classifier:

$$BOC\ (x|D) = \arg\max P(c|x, D) = \arg\max \sum P(c|x, M)P(M|D) \qquad (2.5.3.1.)$$

Though, it is typically not possible to investigate the whole space ***M*** of probable models, the approximations of the models set is very reasonable approach. A single model is fitted on the basis of a criterion like *maximum a posteriori* or *maximum likelihood*:

$$M_{MAP} = \arg\max P(M|D) = \arg\max P(D|M)P(M) \qquad (2.5.3.2.)$$

$$M_{ML} = \arg\max P(M|D) \qquad (2.5.3.3.)$$

The formulation can be seen as extreme approximations of the ***M*** space by single element sets. Logically, the approximations have advantages over random collections of models, because the *maximum a posteriori* and *maximum likelihood* models fitted by parameters. The estimation of ***P(M|D)*** by some weights $\boldsymbol{w_m}$ leads to a classifier estimating class probabilities for given data array ***x*** as:

$$P(c|x) = \sum w_m \max P(c|x, M) \qquad (2.5.3.4.)$$

The conditional probabilities estimation of the ***P(c|x, M)*** of class ***c*** given decision tree models is straightforward tasks. The finding correct weights is $\boldsymbol{w_m}$ is more difficult task, because the definition of priors ***P(M)*** is not self-evident. In other implications, the weights are presumed to be equal in bagging and other unweighted situations or are determined by diverse algorithms to expose model accuracy and the power of its influence on final ensemble results.

Random Forest is one of the most popular decision tree-based ensemble techniques. The accuracy of these models is higher than most of the other competing decision trees. Breiman (2001) proposed a framework conception of Decision Tree Forests as a collection of tree-based classifiers built with respect to random vectors. In the study, given a random vector $\boldsymbol{\Theta_i}$ for each $\boldsymbol{i = 1, ..., s}$, a tree is grown for the training data and $\boldsymbol{\Theta_i}$. Thus a random forest is defined as the majority voting classifier based on the collection decision trees under the assumption that the random vectors are independent and identically distributed. Such characterization of Random Forest encompasses many different schemes of trees variation ensembles such as bagging, where the random vectors are directly corresponded to the number of elements in the training dataset. Boosted classifiers satisfy the definition of Random Forest, however the requirement of independent and identically distributed is not fulfilled. Therefore, the boosting algorithms are not in line with the main idea of random forest theory, despite the fact, that each particular model might be obtained within the random forest scheme.

The construction of Random Forest is defined in a way advocating the algorithm of growing forests, where for subsequent values of ***i***, makes the random vector $\boldsymbol{\Theta_i}$ and then uses it in the learning process to obtain model $\boldsymbol{M_\Theta}$.

The advantages of the method comprise:

1. The predictive power of Random Forest competes with the best supervised learning algorithms.
2. Random Forest provide a consistent feature importance assessement.

3. Random Forest offer effective estimates of the test error without encountering the cost of recursive model training related to cross-validations.

2.6. Research limitations

There are certain limitations for the proposed method, which have been discussed in the literature review and explicitly mention in the Section II:

1. Application of DEA requires handling a separate linear program for each DMU. Thus, the aggregation of DEA to problems that have many DMUs can be intensive in terms of gathering initial dataset.
2. Since DEA is an extreme point technique, errors in measurement can cause significant problems. DEA efficiencies are very sensitive to even small errors, making sensitivity analysis an important component of afterwards DEA procedure.
3. The proposed method is a non-parametric one and statistical hypothesis tests are difficult.
4. As efficiency scores in DEA are obtained after running a number of linear programming settings, it is not easy to explain intuitively the DEA for the case of more than two inputs and outputs.
5. From methodological point of view, the flexibility is needed to allow for one or more outputs or inputs for performance evaluation.
6. DEA approach is not able to handle environmental changes but only a number of assumptions.

Machine learning algorithms empowered by pre-defined kernel functions are good at coping with data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments over creating applicable yet easy interpretable knowledge about given issue. However, the Machine learning is also the subject of limitations:

1. Random Forest method does not give coefficient directly interpretable in economic terms.
2. Artificial Neural Networks are a *black box* approach, where merely not possible to interpret the processes within the networks estimation.
3. The choice of kernel and the determination of the parameters for a given value of the regularization. SVM shifts the problem from over-fitting to model selection, because kernel selection can be quite sensitive to over-fitting the model selection criteria.
4. The SVM loss function might not have an obvious statistical interpretation in many classification problems probability is missing.
5. Machine learning is highly susceptible to errors of the train datasets.

To overcome the limitation imposed by the DEA methods, the application of learning algorithms can help to process large volumes of data and discover specific trends and patterns that would not be apparent to prior models due to a regularization parameter, which makes possible avoiding over-fitting.

By Gareth *et al.* (2013) it is that in the absence of a very large designated test set that can be used to directly estimate the test error rate, a number of techniques can be selected

to evaluate this quantity by taking the entire available training datasets. Hackeling (2017) suggest that creating a large collection of supervised data can be costly in some domains. During development, and particularly when training data is scarce CV can be used to train and validate a model on the same data. In CV, the training data is partitioned and trained using the entire set but excluded one partition. The partitions are then rotated several times so that the model is trained and evaluated on all of the data. The mean of the model's scores on each of the partitions is a better estimate of performance in the real world than an evaluation using a single training split.

Ensemble methods in machine learning in this research assumes that data is stationary due to methods chosen. The accuracy of the result based on estimation and forecasting is affected significantly by how well the nature of the underlying process is identified. In particular, determining whether a process is stationary or not plays an important role in time series analysis. Stationarity in its weakest sense implies that the first and second moments of a time series remains constant over time. In such a situation, the future will behave very similar to the past and reliable forecasts based on past data can be easily obtained.

III. TESTING THE PROPOSED MODEL FOR THE ASSESSEMENT EFFICIENCY IN DECISION SUPPORT SYSTEMS UNDER UNCERTAINTY

3.1. Practical implementation steps of decision support systems

In this Section the empirical findings are presented. The model is tested on a sample of 63 DMUs listed on the Nasdaq Baltic[5], various datasets representing influence of the global environment on the regional scale. The empirical research is started with data mining process, assessing feature sets, non-parametric efficiency model with factors extension and is continued by determining the influence of uncertainty factors on the assessment efficiency using ensemble methods in machine learning. The general procedure consists of the following practical steps, which are needed to be integrated parts of the decision support systems:

1. Data model analysis to verify the credibility and integrity of datasets
2. Efficiency assessment and models estimation using various techniques
3. Based on the efficiency models estimation perform feature set selection
4. Apply ensemble machine learning methods
5. Verify the machine learning algorithms

At the end of the practical implementation, the robustness of the model should be verified as proposed in the Chapter II, Subsection *2.1 Model definition for the assessing efficiency under uncertainty factors*, *2.4 Two-stage DEA nonparametric efficiency analysis* and *2.5 Ensemble methods in machine learning*.

In Table 9 assessment efficiency under uncertainty is based on the proposed model in the previous Sections and the subsequently presented research methods. The assessment process consists of five processes in Column (1), each of them separated into running R script sequence in Column (2). The sequences are the R scripts, which executed manually via CLI interface on the server sider, initiated automatically chained by other sequence or via Web interface. Column (4) give the scientific method applied to each step. All methods are described in the Chapter II, Sections *1.5 Ensemble machine learning approach in decision-making process*, *2.4 Two-stage DEA nonparametric efficiency analysis*, *2.5 Ensemble methods in machine* learning. This Section provides technical details on applied methods.

5 Appendix 3. DMU list

Table 9. *Sequences of practical assessment using proposed methodological approaches*

Process[6]	Sequence[7]	Description	Scientific method
(I) Initial process		**Initiation**	**Association analysis**
	1	Initiation - DB Init	
	2	Initiation - Load libs	
(II) Data mining		**Data load**	**Clustering**
	1	Data load - Fit OMX_DT SPREAD and COEFF values	
	2	Data load - Fit OMX_DT values	
	3	Data load - Make graphs and charts for datasets	
	4	Data load - Descriptive statistics	
	5	Data load - Data array decomposition	
(III) Feature selection		**Feature selection**	**Sensitivity analysis, panel data analysis, regression analysis**
	1	Feature selection - Estimation model definition based on datasets	
	2	Feature selection - Tests on trustworthiness of feature sets	
	3	Feature selection - Empirical results of feature selection	
(IV) Two-stage nonparametric analysis		**DEA**	**Data Envelopment Analysis**
	1	DEA - DEA value test and correction for missing and NA values	
	2	DEA - Stage 1. CRS Input-orientated	
	3	DEA - Stage 1. VRS Input-orientated	
	4	DEA - Stage 1. CRS Output-orientated	
	5	DEA - Stage 1. VRS Output-orientated	
	6	DEA - Stage 2. VRS Output-orientated	
	7	DEA - Stage 2. VRS Input-orientated	
	8	DEA - DEA pairs - DEA pairs	
	9	DEA - Malmquist	

6 Appendix 7,8,9. The enumeration of Oracle procedures and functions for data mining process and R scripts definition for analytical framework.

7 Appendix 13. CLI automation

Process[6]	Sequence[7]	Description	Scientific method
(V) Ensemble machine learning		**Machine learning**	**Support Vector Machines, Random Forest, Neural Networks**
	1	Machine learning - Ensemble model	
	2	Machine learning – Classification	

(Source: The Author's representation)

First, data mining process is established by Oracle Data Mining (ODM)[8] representing algorithms implemented as SQL functions and leverage the strengths of the Oracle Database. The SQL data mining functions fetch normally structured data tables and data schemas including aggregations, some unstructured data and transactional data. The package offers wide range of possibilities to work with the analytics algorithms.

Large datasets are needed to be normalized to reduce the size by removing the highly correlated input or output factors. It is possible to measure if variables are independent when the distribution of one does not depend on the other by using the conditional distribution. Two variables are independent by the fixed probabilities of one variable in any case whether any condition given on another variable. In order to make appropriate assumption there is not much imbalance in the data sets is to have them at the same or similar normality. If not, then imbalance could cause problems in scaling and error problems may occur. A way of making sure the data is of the same or similar normality across and within datasets is to mean normalize the data relies on:

1. The mean of the data set for each input and output
2. Divide parameters by the mean for that specific factor.

The mean is defined by the mean equation (3.1) that sums up the value of each DMU input or output in that column and then divides the sum by the number of DMU.

$$\overline{V}_i = \frac{\sum_{n=1}^{N} V_{ni}}{N} \qquad (3.1)$$

, where $\overline{V}_1$ defined as the mean for i representing input or output parameter, N is the observation count and V_{ni} is the parameter n for a assumed input or output i parameter. All of the values of a given column divided by mean values:

$$VNorm_{ni} = \frac{V_{ni}}{\overline{V}_i} \qquad (3.2)$$

, where $VNorm_{ni}$ is the normalized value for the value associated with DMU n and input or output in column i as represented in the Annex contains the efficiency score for each DMU using various models.

8 Oracle Data Mining: https://www.oracle.com/technetwork/database/options/advanced-analytics/odm

It is expected that uncertainty represents significant influential factors on the efficiency but these fluctuations should have certain correlations. It is also expected that ensemble machine learning algorithms will outperform linear models. In order to evaluate the feasibility of the proposed approach, a set of baseline datasets using generalization and normalization of economic datasets is undertaken. The various datasets comprising uncertainty from different perspectives are selected for the model estimation due to its importance and inherent challenges. Both single class along with multi-class classification assignment performed and their respective outcome are examined. The initial performance results from conventional methods used as referenced to compare to those obtained using more complex approaches. These results demonstrate that the proposed architecture is capable of providing a reasonable solution to efficiency assessment tasks.

3.2. Datasets acquisition and analysis for decision support systems

3.2.1. Data model for decision support systems

As described in details in the Chapter II, Section *2.2 Taxonomy of datasets selection for decision support system* on page 80. The whole datasets are represented in form of panel data and times-series tables. The panel datasets with cross-sectional and time series feature, are the main data types used in this section for regression analysis fetched at a particular point in time and across several time periods for specific group.

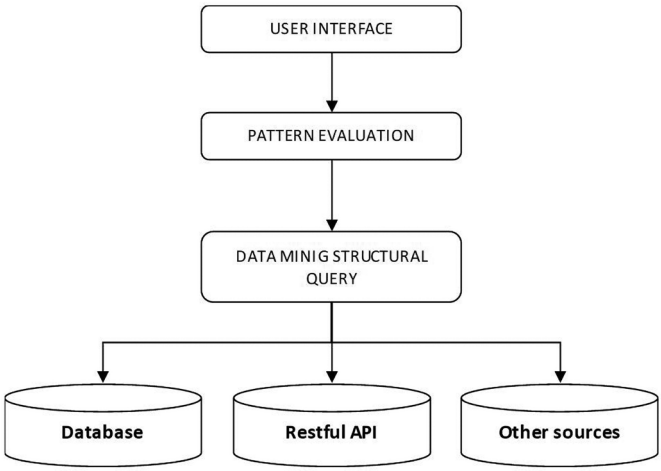


Figure 20. *Acquiring data for analysis*
(Source: Author's representation)

Figure 20 shows the entire data acquisition process, which can be run by user on-demand depends on the research targets:

1. **Web Interface.** The user interface is represented in graphical form written in Oracle APEX[9] as separate Web-application built upon the Oracle SQL scheme. The interface enables to evaluate data, combine data into groups, perform certain modification without explicit need for programming.
2. **Command-line interface (CLI)**. The CLI is designed to execute R scripts in batches triggered by various events and store results back into SQL.

After the processes have been set up, then the data can be obtained from:

1. **Database.** The data can be retrieved using Oracle PL/SQL queries, functions or procedures
2. **WebServices.** There is a number of services, provide data online via RESTful or SOAP protocol
3. **Other sources**. Data can be stored in files and the parsing script will translate CSV files into SQL statements in order to save them in the database.

For the research a high-level abstraction layer has been applied in order to provide structured datasets for the research.

A multiple-step process is designed to solve the problem of creating a robust decision support system. The presence of uncertainty factors makes any standard panel data model insufficient to characterize the dynamic nature of the economic process. Hence, dynamic models are advocated to provide more insights of the overall decision support system as in Chapter I, Section *1.3 Foundations of uncertainty in economics theories* on page 41.

The data model is divided into logical parts as represented in Table 10 for data-mining process and data acquisition. A PL/SQL packages are created for large volume data handling in order to enable SQL datasets will be used to do analytics with R. The data acquired from initial data sources via WebServices of via CSV parsing is stored in the Oracle 12c SE database in form of:

1. **Tables.** SQL collection of raw data with logical prefix DEA_%, EPU_%, OMX_%, organized in terms of rows and columns.
2. **Materialized view** (or snapshots) in Oracle with prefix MV_% is a database object that comprises the results of a select query. Materialized views are local copies of data fractions used to create summary tables based on aggregations of a table's data.
3. **Pipelined functions.** Pipelined functions in Oracle are designed to return data rows back to the calling query in the desired form. The advantage of using pipelined functions is to perform cross-validations upon function execution.

9 Oracle Application Express (Oracle APEX), is the low code web application development tool for the Oracle Database (https://apex.oracle.com/en/)

Table 10. *Data model's logical principle structured in data tables*

Description[10]	Source	Data table name	Records
Results data from two-stage DEA analysis	Model	DEA_DT	2 903
		DEA_MODEL	6
		DEA_SLACKS	21 832
		Total records count	***24 741***
Input-output parameters and results of two-stage analysis	Model	MV_DEA_STAGE1	573
		MV_DEA_STAGE2	23 082
		Total records count	***23 655***
Global uncertainty index data	External	EPU	241
		EPU_FIRM	66 448
		EPU_SETTINGS	143
		EPU_WUI	80
		EPU_WUI_COUNTRY	11 440
		Total records count	***78 352***
Country-specific datasets	External	MV_BUNDESBANKBBK01SU0202	22
		MV_BUNDESBANKBBXE1MI8WPRODNS	29
		MV_BUNDESBANKBBXL3AI6NUNEHTO	19
		MV_FREDWLEMUINDXD	21
		MV_NASDAQOMXVOLNDX	11
		MV_OECDKEIPRMNTO01ESTSTA	22
		MV_OECDKEIPRMNTO01LTUSTA	22
		MV_OECDKEIPRMNTO01LVASTA	20
		MV_OECDMEIBTSCOSBCBUTEESTBLS	26
		MV_OECDMEIBTSCOSBRBUTELTUBLS	36
		MV_OECDMEIBTSCOSBSSPFTESTBLS	29
		MV_OECDMEIBTSCOSBSSPFTLVABLS	28
		MV_OECDMEIBTSCOSBVDETELVABLS	19
		MV_OECDMEIBTSCOSBVEMFTLTUBLS	19
		MV_OECDPRICESCPIEU28CPHPTT01I	22
		MV_OECDSNATABLE1EU28P31S14VOB	19
		MV_WWDIESTBXKLTDINVWDGDZS	23
		MV_WWDILTUBXKLTDINVWDGDZS	25
		MV_WWDILVABXKLTDINVWDGDZS	26
		Total records count	***438***

10 Appendix 4. The structure of data tables. The MV_% represent Oracle "materialized views", which are selected from the aggregation table dynamically on-demand by prodecures.

Description[10]	Source	Data table name	Records
Nasdaq Baltics datasets. Financial performance records and daily stock activity	External	OMX	748
		OMX_FACTS	574
		OMX_M	470
		OMX_DT	166 911
		OMX_SETTINGS	40
		Total records count	***1 833***
		Total records count	***295 929***

(Source: Author's representation)

The analysis of efficiency of various stages are naturally linked. In environmental settings where multiple inputs are used to generate multiple outputs, aggregation methods are necessary to calculate aggregate input and output levels for a better decision support system. Therefore, panel data models are in focus because these might capture dynamic changes and used to evaluate fluctuations in efficiency over time. However, no true dynamics can be represented by a single nonparametric model but should assess multi-level approach.

The country-specific dataset. Country-specific factors should include market concentration, presence of foreign investments, fiscal indicators. In rapidly changing business settings evolving working environment, the ability to predict future trends and needs in terms of the knowledge and skills required to justify becomes critical for effective decision support system. These trends fluctuate by geography and industry, and so it is important to anticipate the industry and country-specific variables.

Country specific indicators in Table 11 used in the analysis were selected according to their statistical significance in at least one of the previous studies, with the exception of those indicators that contained biased data. Several proxy variables to comprise the most indicative variables are constructed in Column (2). There are various sources of information in Column (1). These indicators measure the annual evolution for EU Member States. The specific EU member states indicators are the subject developed to implement the *parallelism* principle ruled by EU Staff Regulations. The same OECD indicators derived from business and consumer survey are of key importance in assessing short-term economic developments. These indicators give fundamental information on and households assessments of the current economic situation and their intentions and expectations for the future. The OECD business survey indicators and composite indicators are collected for the industrial sector. Various strategies are used for the selection of time series to be included in a composite indicator. Standard country indicators could be used or an individual set of indicators per country. The use of a standard set of indicators across countries is a good approach for comparative analysis across countries.

To further expose the country-specific characteristic of uncertainty, the corresponding historical influence of uncertainty factors should be involved into efficiency assessment in decision support systems. Hence, the models should include multi-country but also

provide country-specific results focused on cross-county averages. Cooper *et al.* (2011) provides in-depth comparison between the results obtained from various nonparametric models revealed that the worse the quality of country-specific environmental datasets, the greater the changes in the efficiency scores. In the country-specific analyses, the mean efficiency of organizations suggests that the efficiency differences observed across countries are primarily attributable to country level effects. Hence, further analysis of efficiency and the Malmquist indices should be solved in models calibrated with time-varying and country-specific data.

Table 11. *Country specific variables composition with data sources and abbreviation*

Source	Dataset abbreviation	Implication	Short description
BUNDESBANK	BBXL3_A_I6_N_UNEH_TO	EU (17)	Unemployment
OECD	MEI_BTS_COS_BRBUTE_LTU_BLS	Lithuania	Business tendency
BUNDESBANK	BBK01_SU0202	EU	ECB Interest Rates
BUNDESBANK	BBXE1_M_I8_W_PROD_NS	EU	Industrial Production
OECD	MEI_BTS_COS_BVEMFT_LTU_BLS	Lithuania	Future Tendency
OECD	PRICES_CPI_EU28_CPHPTT01_I	EU (28)	HICP
OECD	KEI_PRMNTO01_EST_ST_A	Estonia	Total manufacturing
OECD	KEI_PRMNTO01_LTU_ST_A	Lithuania	Total manufacturing
OECD	MEI_BTS_COS_BCBUTE_EST_BLS	Estonia	Business tendency
NASDAQOMX	VOLNDX	Global	Volatility NASDAQ
OECD	MEI_BTS_COS_BVDETE_LVA_BLS	Latvia	Business tendency
WWDI	LVA_BX_KLT_DINV_WD_GD_ZS	Latvia	Foreign direct investment
OECD	MEI_BTS_COS_BSSPFT_EST_BLS	Estonia	Business tendency
WWDI	EST_BX_KLT_DINV_WD_GD_ZS	Estonia	Foreign direct investment
OECD	MEI_BTS_COS_BSSPFT_LVA_BLS	Latvia	Future Tendency
OECD	KEI_PRMNTO01_LVA_ST_A	Latvia	Total manufacturing
OECD	SNA_TABLE1_EU28_P31S14_VOB	EU (28)	Consumption of Households
WWDI	LTU_BX_KLT_DINV_WD_GD_ZS	Lithuania	Foreign direct investment

(Source: Author's representation)

The country-specific dataset includes macroeconomic time-series from 1990Q3 through 2018Q4 covering a wide range of economic activity relevant for policymakers. The series are obtained from multiple sources Federal Reserve Economic Database, Bundesbank, OECD, Nasdaq. The data sources for country-specific variables are enlisted in Table 12.

Table 12. *Country specific data sources*

Source	Description	Used
FRED	Federal Reserve Economic Data	1
BUNDESBANK	Deutsche Bundesbank Data Repository	3
OECD	Organization for Economic Co-operation and Development	11
NASDAQOMX	NASDAQ OMX Global Index Data	1
WWDI	World Bank World Development Indicators	3
	Total	***19***

(Source: Author's representation)

Uncertainty datasets. The country-specific datasets are enlarged with Baker *et al.* (2015) and Jurado *et al.* (2013), Jurado *et al.* (2015) developed index of Economic Policy Uncertainty, which is based on mass-media coverage frequency and also defined as the common volatility in the unforecastable component of a large number of economic indicators (Table 13). To investigate the impact of macroeconomic uncertainty on the efficiency model, the series are augmented with the economic uncertainty index. Most of the series enter the model in annualized log levels and multiply by factor. The global economic measures are taken reliably from the World Development Indicators.

The WUI contains the time series of the World Uncertainty Index (WUI) at the global level (simple average and GDP weighted average), income level (advanced, emerging, and low-income economies), and regional level (Africa, Asia and the Pacific, Europe, Middle East and Central Asia, and Western Hemisphere). All indices have been computed by counting the frequency of word uncertainty (or its variants) in the Economist Intelligence Unit (EIU) country reports. The indices are normalized by total number of words in each report, rescaled by multiplying by 1,000 and using the global average of 1996Q1 to 2010Q4 such that 1996Q1-2010Q4=100. A higher number means higher uncertainty and vice versa. The detailed WUI contains the time series of the World Uncertainty Index (WUI) for 143 countries from 1996Q1 to 2019Q1. All indices have been computed by counting the frequency of the world uncertainty (or its variant) in EIU country reports. The indices are normalized by total number of words and rescaled by multiplying by 1,000. A higher number means higher uncertainty and vice versa. Global Economic Policy Uncertainty (GEPU) Index includes normalized national EPU index to a mean of 100 from 1997 to 2018. The missing values is imputed for Australia, India, Greece, the Netherlands and Spain using a

regression-based method[11]. This phase requires balanced monthly panel datasets of EPU index of 18 countries starting from January 1997. Third, the GEPU Index value computed for each month as the GDP-weighted average of the 18 national EPU index values, using GDP data from the IMF's World Economic Outlook Database.

Table 13. *Datasets for World Uncertainty Index and Economic Policy Uncertainty*

Variable	Datasets	Description
WUI_GLOBAL	2	Global World Uncertainty Index
WUI_COUNTRY	140	Country specific index
EPU_GEPU	2	Global Economic Policy Uncertainty
EPU_COUNTRY	12	EPU uncertainty index by country

(Source: Economic Policy Uncertainty)

Firm level datasets. Practically seen, the choice and the number of parameters and the DMU determine what kind of discrimination exists among efficient and inefficient units. Considering amount of the data set there are some factors important for further analysis. The larger number of DMU included into the dataset there is a greater probability of capturing high performance units that would determine the efficient frontier and improve discriminatory power. The same time large datasets imply the homogeneity of the data set may get worse due to some random exogenous impacts or noise may affect the results.

3.2.2. Datasets structure analysis

Based on the results of estimating the linear models[12], there is possible to perform empirical results of feature selection presented in Table 14. The heterogeneity in responses is well recognized based on the Breusch-Pagan test with the significant p-value as causing statistical problems in experimental and non-experimental data. Allowing for heterogeneity in responses through standard methods, such as fixed or random effects models, there is not possible to identify efficiency and their factors according to the theory. Heterogeneity can be partially handled by accumulating and controlling information on observable characteristics in the statistical analysis. In the panel data analysis there is important to understand characteristics and the quality of datasets. It is also vital crucial to make decision between fixed or random effects based on a tests, where the null hypothesis is that the

11 Global Economic Policy Uncertainty Index: To construct a Global Economic Policy Uncertainty (GEPU) Index, we proceed as follows: First, we re-normalize each national EPU index to a mean of 100 from 1997 (or first year) to 2015. Second, we impute missing values for certain countries using a regression-based method. This step yields a balanced panel of monthly EPU index values for 21 countries from January 1997 onwards. Third, we compute the GEPU Index value for each month as the GDP-weighted average of the 21 national EPU index values, using GDP data from the IMF's World Economic Outlook Database (https://www.policyuncertainty.com/global_monthly.html)

12 Appendix 18. Linear models

preferred model is random effects or the alternative the fixed effects. It runs assessments whether the unique errors are correlated with the independent variable.

Table 14. *Results of estimating the datasets on random effects, cross-sectional dependences, serial correlation, stationary and heteroscedasticity*

Test[13]	Value	Result
Hausman Test	0.16276	Random effects
F test for individual effects	0.90384	No time-fixed effects
Breusch-Pagan test	0.22493	
Breusch-Pagan LM test	0.05723	No cross-sectional dependence
Pesaran CD test cross-sectional dependence	0.41003	
Breusch-Godfrey/Wooldridge test	0.07655	No serial correlation
Augmented Dickey-Fuller Test	0.01	No unit root. Series are stationary
Breusch-Pagan test	3.426e-08	Presence of heteroscedasticity

(Source: The Author's representation)

The null hypothesis accepted if is they are not. The p-value in Hausman tests does not give significant for fixed effects. The cross-sectional dependence can influence long time series. Base on the Breusch-Pagan LM and Pesaran CD test cross-sectional dependence there is no serial correlation. That means that future levels of the efficient cannot be reliably predictable based on the historical datasets. Hence it is almost not possible to find all these variables, which will have significant predictive power for efficiency. This result shows, that these variables in the forecasts do not poses any plausible predictive power for the future periods.

A potential drawback of using panel data in efficiency analysis is that it implicitly assumes that the market will not have any structural breaks over the panel, or that market conditions change in ways that can be readily accounted for time-fixed effects. Crises and economic turbulences in the overall market, will endanger this assumption. Finally, an important advance in recent years is the accessibility of panel data techniques to control for possibly omitted variables and dependences. The panel model should assess the latent uncertainty factors with serial dependence in terms of capturing the effects of cross-sectional dependences.

13 Appendix 19. Datasets testing

3.2.3. Preliminary results of data model analysis

The Figure 21 represents the breakdown of company dataset[14] count by stock marketplace. There are included the Baltic states stock markets, which are located geographically close and have a major common features. The proxy for other stock markets was selected Helsinki for the proxy reference. The organization selection is on purpose constructed unbalanced way in order to possibly prove the hypothesis of heterogeneous economic agents making decisions uncertainty conditions in modern economic settings, which consist of a large number of smaller complex subsets. The increased levels of complexity affected by uncertainty in many ways and thus increase risk factors across borders but not locally.

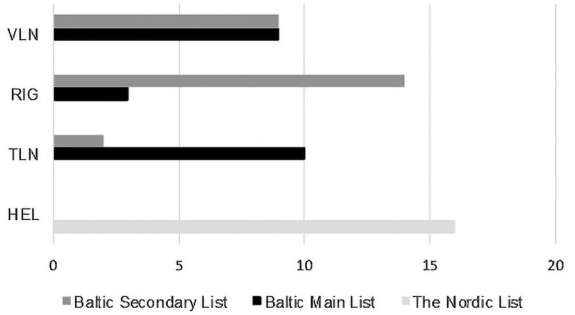


Figure 21. *Representative organizations datasets count by market place and lists (Source: The Author's representation)*

Each economic subset is modular in terms of being made up of a large number of functionally specific parts. The Figure 22 gives the breakdown by industry sectors.

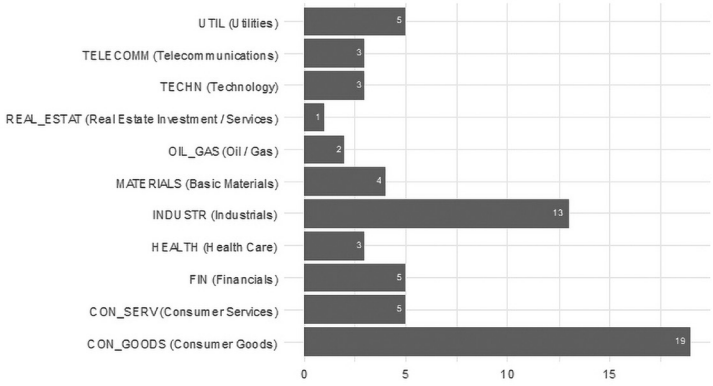


Figure 22. *Organizations count by industry and sectors (Source: The Author's representation)*

14 Organizations count is in Annex 3. DMU list

More specifically, following conditional plots by Murrell (2005) and Albert and Rizzo (2012) the two-stage efficiency ranks represented in Figure 23. It draws a separate figure for each level of the grouping factors. The data is given to this function as a formula of the form $y \sim x \mid g$, where g and h are factors. It includes first-stage frontier for the predictive revenues that the organizations could achieve and the estimation of the maximal market performance. The visualization represents one conditional variable of stock market. It means, that market place formula describing in the form of conditioning plot (3.2.1):

$$\text{Efficiency Stage 1} \sim \text{Efficiency Stage 2} \mid \text{Market Place} \qquad (3.2.1)$$

It indicates that plots of Efficiency Stage 1 versus Efficiency Stage 2 should be produced conditional on the market place. More specifically, the organizational efficiency is dependable on the volatility of the market place and market efficiency.

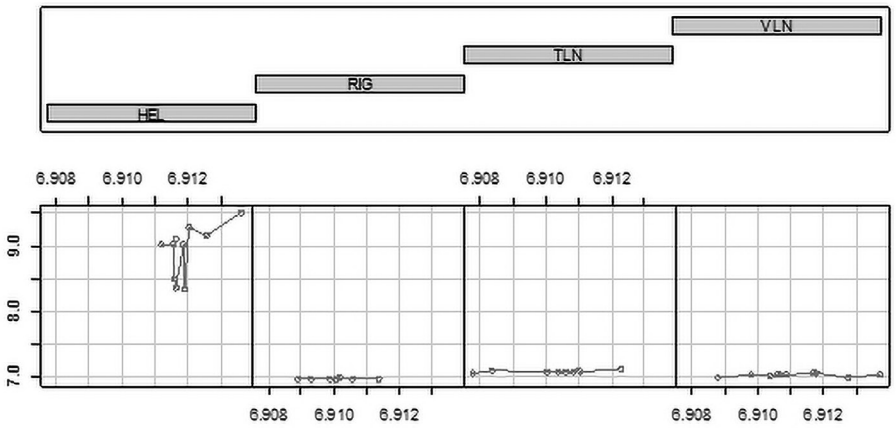


Figure 23. *Conditional two-stage efficiencies and market volatility (Source: The Author's representation)*

In Table 15 there is an estimation of stock market volatility. The stock markets in Column (1) differentiate mainly by regular activities of buying, selling, and issuance of shares of publicly-held companies take place. The expectations of stock uncertainty in Column (2) and volatility in Column (3) is calculated by the Nasdaq daily stock activities based on the OMX_DT values. This evidence is correlated with other empirical findings, which suggest that indexes are excessively volatile. Growth in the stock price volatility will cause the decline aggregate demand and generates a significant reduction in output.

Table 15. *Baltic stock market uncertainty and volatility (log)*

Stock market	Uncertainty	Market volatility
HEL	0,97430299	0,341187795
VLN	2,239761803	3,533308777
TLN	2,433921863	2,677679706
RIG	2,556601302	2,2303612

(Source: The Author's representation)

It underpins the assumption of various boundaries and limitations for the construction of conventional projection on the efficiency frontier, the lack of description of heterogeneous behavior in performance among many efficient agents. There is a shortcoming in a standard efficiency estimation model, where all efficient DMUs have the same estimation with no way to separate them. This has led to focused research to further discriminate between efficient DMUs, in order to arrive at a ranking, or even a numerical rating of these efficient DMUs, without affecting the results for the non-efficiency.

The further analysis of datasets incorporates the data from different layers in order to find out the nature of correlations. The Figure 24 corresponds with the economic theories of uncertainty described in the Chapter I, Section *1.3 Foundations of uncertainty in economics theories* on page 41 very well. The uncertainty has various non-linear economic effects on economic agents through various channels. It determines the role of firms is to take care of production decisions including production quantity of goods, means of production and price settings. Firms in general take disposable factors of production in order to sell own goods for consumption to the households or government. There is a number of proxies intended to measure effectively different layers of environmental uncertainty settings. Often in the empirical literature these proxies represented as a measure of the uncertainty impact on the economic activity in forms of industrial production, GDP, investment or consumption. Hence, variables on the macroeconomic level are strongly correlated to each other, whereas Macroeconomic uncertainty has no direct impact of the stock efficiency but the firm level is correlated to the by the mean of systemic uncertainty.

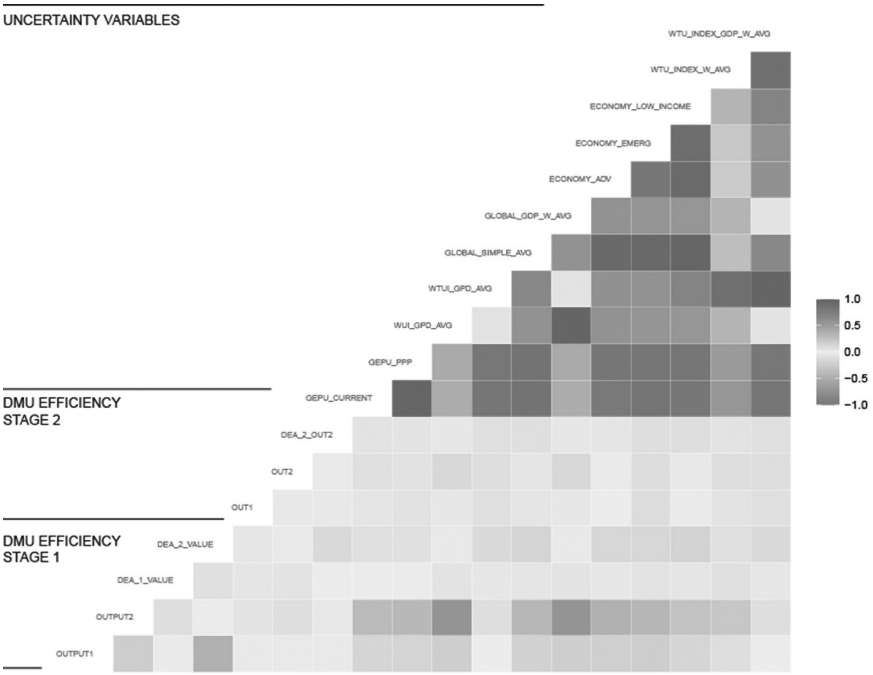


Figure 24. *Correlation map of models efficiency and environmental variables representing non-linear economic effects on economic agents through various channels (Source: The Author's representation)*

The practical measurement of uncertainty includes an encompassing set of datasets. Therefore, dynamic efficiency requires further fundamental analyses of economic environment, which becomes of great significance for policy implications.

3.3. Efficiency assessment by nonparametric model

3.3.1. Nonparametric model selection

The two-stage model per Chapter II, Section *2.4 Two-stage DEA nonparametric efficiency analysis* on page 91 respectively tested on the DMU datasets[15] with the input-oriented calculation of efficiency proposed to measure weighted output to weighted inputs and thus the efficiency score in range where output can never exceed input. To investigate effects of scale of operations both VRS and CRS approach of DEA models are chosen for comparative analysis and validation. The two-stage DEA developed ranks the performance of each organization comparative to each of the two frontiers calculated according to the parameters:

15 Appendix 3. DMU list

1. A first-stage frontier for the predictive information indicates the maximum revenues that the company would achieve:

$$Revenue \sim Financials + Leverage + Profitability \qquad (3.3.1)$$

2. The second-stage DEA for the estimation indicates the maximum market performance:

$$Capitalization \sim Efficiency\ 1 + Market\ coef. + Equity + Volatility \qquad (3.3.2)$$

The competitive DEA models are defined with the following features:

1. Model A1. Stage 1. CRS Input-orientated
2. Model A2. Stage 1. VRS Input-orientated
3. Model A3. Stage 1. CRS Output-orientated
4. Model A4. Stage 1. VRS Output-orientated
5. Model A5. Stage 2. VRS Output-orientated
6. Model A6. Stage 2. VRS Input-orientated

The entire process is illustrated in Figure 25. The crucial point is that the evaluating the performance of activities or organizations by DEA requires clear data. However, there are practical evidences that getting inputs and outputs data is difficult task. The datasets might consist of categorical and continuous variables. The available arrays of datasets are often vague and incomplete for further investigation. The sources from which data is acquired can be very different. The quality and usability of the data is directly affected by the manner in which it is generated. Since most models take data from multiple sources, this characteristic implies the need to integrate the corresponding data to some form normalization. Normalization and integration of data is not a straightforward assignment. The variables differ in range and type and therefore normalization of the values is needed.

Following Ravi Kumar and Ravi (2007) argue that in order to achieve an appropriate discriminatory power out of the CCR and BCC models the lower bound on the number of DMUs should be the multiple of the number of inputs and the number of outputs. Golany and Roll (1989) assert that the number of units should be at least twice the number of inputs and outputs in the analysis. Bowlin (1998) argues that the number of DMUs should exceed input and output variables. The study is derived from the issue that there is flexibility in the selection of weights to assign to parameter values in determining the efficiency of each DMU. Thus, in struggling to be efficient a DMU can assign all of its weight to a single parameter. The DMU with a single certain ratio as a measure of an output to an input as maximum will take all its weight to both specific efficient inputs and outputs. Requirements needed to ensure that productivity models are more discriminatory. However, in order to obtain a higher discriminatory power, it is still possible to reduce the number of parameters. DEA productivity models that can assistance discriminate among DMU more effectively regardless of the size of the dataset.

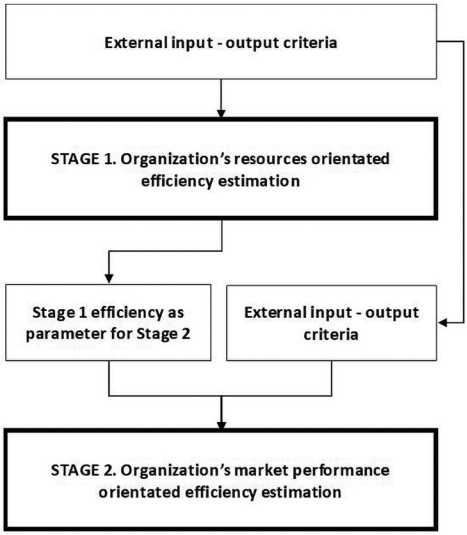


Figure 25. *Stepwise two-stage efficiency assessment (Source: The Author's representation)*

The DEA score measured on the year basis without taking into account environmental uncertainty factors but measured performance against the best peer. The two-stage models use data on outputs and inputs in the first stage, and use data on observable exogenous factors in the second stage, the objective being to determine the impact of the observable exogenous factors on initial evaluations. Especially in terms of analyzing financial statements, the relationship between financial information and organizational value is established through a two-stage fundamental process. In hypothesis testing, effect size, power, sample size, and critical significance level are related to each other. The chi-square test is used to analyze the contingency formed by categorical variables evaluating whether there is a significant association among variables.

A nonparametric study gives information about efficiencies of firms and among other peers. The verification of the robustness of the results plays an enormous role completed with sensitivity analysis. F-Statistics effect size is a concept that measures the strength of the relationship among variables. The greater the effect size, the greater the difference will be in determining if the difference is real or if it is due to a change of factors. Using structural equation modeling to investigate an efficiency effect on performance, the optimal strategy is constructing several models corresponding to the hypotheses, verify it against empirical datasets. The linear models evaluation involves comparing the *p-value* to the significance level, and rejecting the null hypothesis when the *p-value* is less than the significance level. If the estimated value is high, then this result is not statistically significant and the difference could have arisen by other variation between samples.

The models selection is rather complicated task, which can be solved two-fold approach. The best efficiency model selection should be done by:

1. **Method 1.** Using statistical techniques to fit the best models relied on the quantified results from the models.
2. **Method 2.** Using DEA analysis to find the most adequate models by explanatory approach.

Method 1. In Table 16 there are results of measuring by using data quantifies the relationship between a target performance variable and a set of covariate variables to determine the best possible efficiency model. The analysis also estimates a coefficient for each variable that corresponds to the difference in value. A high variance ***d(y, μ)*** denoted as the average of the squared distances from each point to the mean. It implies that the dataset points are extremely spread out from the mean, and from one another. Following the basic principles to use of the Akaike's Information Criteria (AIC) there is obvious that a lower AIC criterion indicates a more preferable model, relative to a model fit with a higher AIC. Bayesian information criterion (BIC) is another criterion for model selection among a set of models.

The sign of a regression coefficient tells about positive or negative correlation between each variable. A positive coefficient indicates that as the value of the independent variable increases, the mean of the dependent variable also tends to increase. A negative coefficient suggests that as the independent variable increases, the dependent variable tends to decrease.

Table 16. *Selection and validation of nonparametric models (A1, A2, A3, A4, A5, A6)*

Variables	Model A1	Model A2	Model A3	Model A4	Model A5	Model A6
Efficiency Stage	**1**	**1**	**1**	**1**	**2**	**2**
Orientation	Input	Input	Output	Output	Input	Output
Return to scale	Constant	Variable	Constant	Variable	Variable	Variable
Uncertainty levels	0.05	-0.03	0.05	0.12	-0.19	-0.16
Input levels	-1939.48	-1294.71	-1945.98	-1677.90	-1858.54	-1038.28
Efficiency 1	-18.11	-625.58	18,07	24,14		
Efficiency 2					0.26	2,44
AIC	*2,21*	*2.21*	*2,38*	*2.52*	*1,68*	*1,27*
BIC	*3,14*	*3.94*	*2,91*	*3.25*	*2,41*	*2.03*
LogLikelihood	*3,89*	*3,17*	*3,91*	*4,74*	*1,16*	*0,87*
d(y, μ)	*0.13*	*0.13*	*0.13*	*0,18*	*0,12*	*0,12*
Selected	-	*Yes*	-	-	*Yes*	-

(Source: Author's representation)

Based on the criteria described above the following best VRS IO models are selected for the assessment as described in the Chapter II, Subsection *2.4 Two-stage DEA nonparametric efficiency analysis* on page 91:

1. **Organizational efficiency measure** (Efficiency Stage 1)
2. **Market performance** (Efficiency Stage 1)

Method 2. Another prove on the model selection can be done using DEA approach. By verifying the models, the Model A2 for the Efficiency Stage 1 and the model A5 for the Efficiency Stage 2 give the most appropriate results. Table 17 reveals percentage of efficient organizations, average score and standard deviation amounts for slacks in inputs. The values differ in CRS and VRS models. As expected from the theoretical part, the values found in variable return to scale have advantage in ability to explain over the constant return to scale model. It is mainly because the variable in variable return to scale discloses comparability among organization in terms of inputs and outputs, what is actually defined as increased or decreased returns to scale. Nevertheless, the CRS based models are easier to interpret and they find own application in benchmarking based on CRS frontier. For this particular research CRS assumption does not fit the criteria of robustness. Only VRS models in its different form provide adequate explanatory power to efficiency assessment.

Table 17. *Validation of efficiency models (A1, A2, A3, A4, A5, A6)*

Statistics	**CRS IO**	**VRS IO**	**CRS OO**	**VRS OO**	**VRS IO**	**VRS OO**
	Model A1	**Model A2**	**Model A3**	**Model A4**	**Model A5**	**Model A6**
Efficient DMUs	0,84	0,94	0,84	0,94	0,83	0,88
Average Efficiency Score	0,93	0,95	1,01	1,01	12,61	1,01
Standard Deviation	31%	27%	32%	27%	28%	31%
Selected	-	*Yes*	-	-	*Yes*	-

(Source: The Author's representation)

The overall reductions in all inputs are not appropriate goals in real practical application. Then the slacks-based analysis provides the appropriate model structure to capture a DMU's performance measure. It means what kind of the additional improvement needed for a unit to become efficient presented in Table 18.

Table 18. *Mean and Standard deviation in inputs*

Variable	CRS IO		VRS IO		CRS OO		VRS OO	
	mean	std.dev	mean	std.dev	mean	std.dev	mean	std.dev
Gross margin	2,4110	18,9749	5,0768	21,2060	2,4123	19,0574	2,8970	17,1437
Op margin	2,3774	22,5355	5,1293	18,6937	2,3828	22,8184	2,8911	19,9614
ROA	1,5930	15,9726	2,0485	17,2716	1,6103	16,2556	1,6366	16,0252
ROE	1,5930	15,9726	3,4427	23,0267	1,6103	16,2556	1,6366	16,0252
Net margin	2,7963	47,9384	5,1363	24,9777	2,8023	48,7425	3,1173	27,8446
Leverage	-0,0850	8,7853	-0,9789	10,9122	-0,0959	8,8264	-0,1001	12,9620
Assets turnover	-0,5016	1,1613	-2,1877	10,2906	-0,4891	1,1734	-0,8131	10,5495

(Source: The Author's representation)

The further analysis represents the leftover fractions of inefficiencies in Figure 26 represented after log-transformation, where after proportional adjustments in inputs or respectively increases in outputs, a DMU attempt to reach efficiency frontier to its efficient target. In other words, the analysis indicates what portion of inputs factors can be optimized in order to achieve the optimum efficiency. The DEA slacks are related to the further changes in input that could be adjusted beyond that implied by the DEA projection of equal increase or decrease in all parameters. Depending on the model maximizing parameter, the efficient peers may have less of the input and correspondingly for input orientated models.

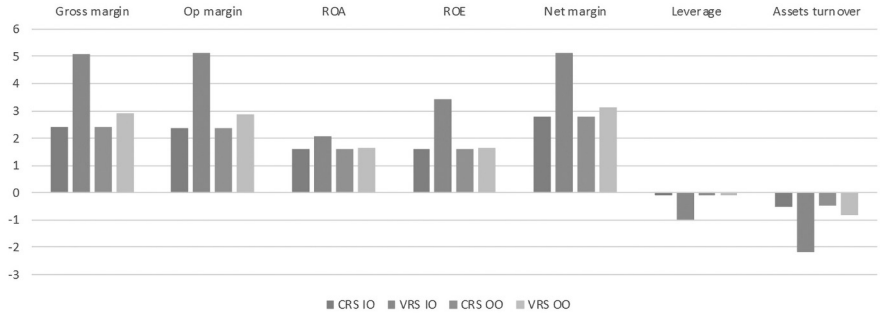


Figure 26. *Adjustments of input factor in nonparametric models (Log transformed)*
(Source: The Author's representation)

In order to gain appropriate result by applications of DEA methods, there is a lot of attention needed to be paid to important modeling issues. Some of these affect to clearly identifying the purpose of the analysis, deciding on inputs and outputs, choosing a model orientation, and giving more attention to the type of data involved.

3.3.2. Decomposition of factors influencing efficiency

The observation on the efficiency exhibits, that decomposition of the technical efficiency of into pure technical efficiency and scale efficiency so as identify the sources of inefficiency with emphasis on whether this is a result of managerial underperformance or caused by uncertainty. The further investigation of models is needed to allow to clarify full frontiers for efficiencies in CRS and VRS models and also to decompose the Total Factor Productivity (TFP) and its components: Technical Change (TC), Pure Efficiency Change (PECH), Scale Efficiency (SECH). To gain the main causes of changes, the TC index can be decomposed by Technical Efficiency (TE), Efficiency Change (EC).

The TC index is related with the changes in production technology, through innovations in resource optimization. The EC index, demonstrates the deviation of the performance of the DMU under consideration from the best efficient firms and is usually associated with decision making capabilities. On a second level the EC index can be disintegrated into PECH and the SECH. These indices represent the main source of changes in the TE.

The Figure 27 illustrates the PECH (the first graph) is associated with the changes in decision making and thus to the achievement of optimal allocation of resources in the production process. A progress of the PECH through a more efficient use of resources and the investigation of the possibility of one DMU to optimize its decision making, can reduce inefficiency. Hence, the SECH (the second graph) illustrates the extent to which one firm can improve its productivity by exploiting scale economies through the reduction of long run average cost by increasing production. Furthermore, it gives useful information to select the production scale that will achieve the required production level. Inappropriate size dimension of a firm causes technical inefficiency.

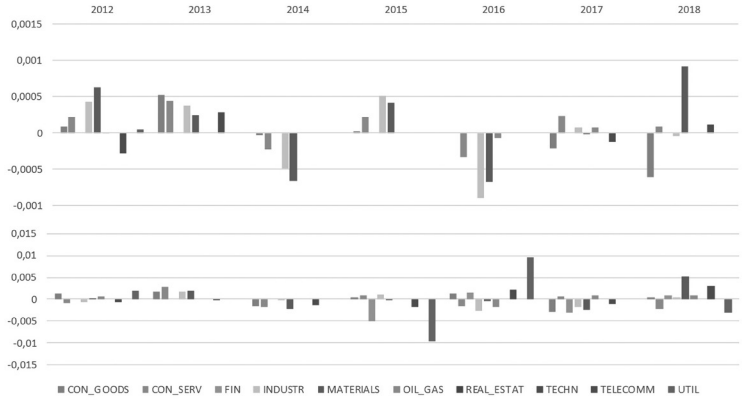


Figure 27. *Scale and Efficiency changes (Log transformed)*
(Source: The Author's representation)

There a number of evidences, that these outputs are appropriate for making relevant conclusions. The industries indicate the insignificant improvement of either the technical productivity or scale efficiency of enterprises. The average SECH is above average in tech-

nology, real estate and telecommunications indicating relatively significant economies of scale due to high demand in IT services and growing trend in real estate last years. Other industries stick to promote TE due to the fact that boundaries of SECH have been reached. The SECH relationship has a certain influence on the TFP of all sizes of enterprises. The critical view on the analysis based on PECH and SECH parameters is:

1. DEA method belongs to extreme point technique and therefore the efficiency frontier is formed by the concrete performance of the best selected peer.
2. DMU can obtain a better value of utility by advancing its performance by shifting focus on output parameter (=performance) ignoring others factors.
3. DMU can be measured as efficient if even it has not improved its performance in respect of all the outputs. Nevertheless, such DMU with unusual settings cannot be represented as a peer for many inefficient DMUs.

Using the productivity index also known as the Malmquist Index (MI) as described in the Chapter II, Section *2.4 Two-stage DEA nonparametric efficiency analysis* on page 91 illustrated Figure 28 there is no significant evidence in log-values, that firms are in pursuit of more explorative strategies towards new product and market developments except real estate and information technology sectors on the markets, which has a lower volatility and uncertainty. So they are able to cope better with the crisis. The increase in efficiency is normally result of an increase in both pure technical efficiency and scale efficiency. However, it is noticed that there is a disparity in the technical efficiency among firms during crises times. The process of significant innovations marked the end of the last century. Even if the inputs and outputs are physical factors, the data can still be different from the conventional precise values.

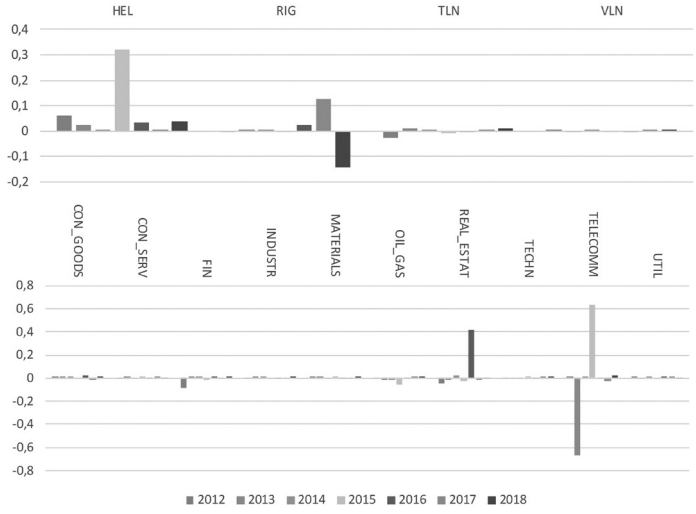


Figure 28. *Productivity index by stock market and industries (Log transformed, yearly) (Source: Author's representation)*

Due to the uncertain environment or imprecision in measurement, the data is not a precise value. Therefore, no doubt this is the toughest challenge for many researchers in the field is uncertainty. Such uncertainty introduces an unavoidable risk factor that is an integral part of theory. In order to handle the complex nature of the problem an interdisciplinary approach is strongly advocated.

The collapse of world trade since the end of 2008 was caused by the financial crisis, but the very unbalanced international trade also contributed to the creation of global problems. Numerous studies reported that short-term readiness of firms to invest in innovation has been reduced during 2008 economic crisis period. However, since 2000, the financial crisis started to appear in the financial market, and very quickly spread to the rest of the world. Followed crisis, of completely different background, hit the most developed markets in the world which made its effects and destructiveness even greater. At the beginning of the crisis no one could imagine its severity, extent and duration, which is caused by uncertainty. Archibugi *et al.* (2013) find that the crisis led to a concentration of innovative activities among fast growing and already innovative firms. The economic performance of innovations meant high rates of growth in both developed and less developed countries, followed by substantial trade and capital flows. The global economy recovered slowly in 2012, and all expectations are that the trend will continue in 2013. Recovery is reflected in the moderate growth rates for the most affected economies in 2010 and 2011, which resulted in the gradual intensification of trade flows in the world.

The decomposition of factors influencing efficiency is not a single step process due to the nature of the economic processes, which are nonlinear. In Chapter I, Section *1.1.4 The problematic of ergodicity and stability of stochastic processes* on page 35 explained that any stochastic process should model process with given parameters and interpretable prediction result. From statistical point of view, any stochastic process is meant to generate the infinite array of ensemble samples of all possible observed time series. Therefore, a statistical population can be formulated as an ensemble of stochastic processes in form of time series representing such processes.

In the Figure 29 using correlation coefficient to examine the strength and direction of the linear relationship between aggregated mean efficiencies, outputs and uncertainty variables, it shows, that Pearson's product-moment correlation has a weak correlation. The result shows that there is barely no association between aggregated mean efficiencies, output and uncertainty variable. Correlation and linear regression analysis are statistical techniques to quantify the dependencies. The correlation coefficient, denoted r, ranges between -1 and +1 and quantifies the direction and strength of the linear association between the two variables. The correlation between two variables can be positive where higher levels of one variable are associated with higher levels of the other or negative higher levels of one variable are associated with lower levels of the other.

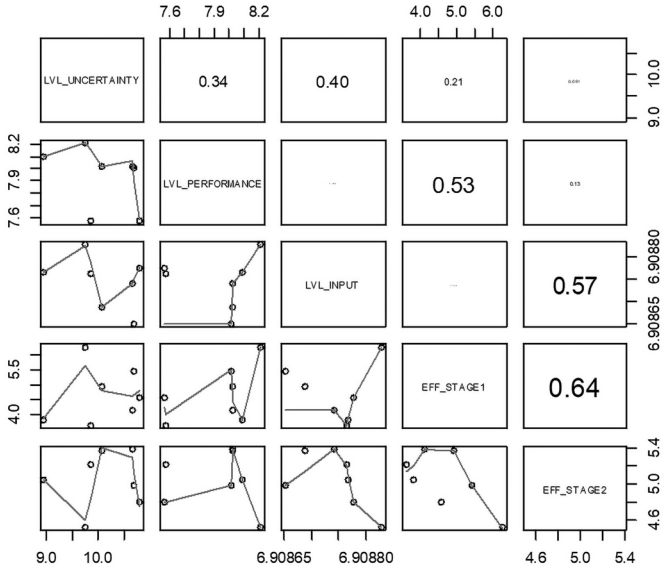


Figure 29. *Level variables correlation*
(Source: Author's representation)

The correlation coefficient between the two variables measures the strength and direction of the linear association. The correlation between two variables can be positive where higher levels of one variable are associated with higher levels of the other or negative higher levels of one variable are associated with lower levels of the other. The same result is achieved with linear regression that is appropriate to understand the association between one uncertainty variable and efficiency variable.

The findings in this subsection corresponds very well with the global development trends. First, using simple methodology is not possible to assess the efficiency under uncertainty condition using linear dependences. Second, there is substantial ongoing debate nowadays on the global innovation development has obviously slowed down. It can be seen that the total factor productivity growth rate has quite gradually decreased over the last decades. Another possible phenomenon of stalled technological progress and economic growth decrease in most countries is debt-driven characteristics of economies in order to diminish the demand gap. Due to uncertain nature of processes this is an unsustainable situation, which might cause another major global economic disturbance. The main reason for that is when access to credit markets will be closed for nonpayment borrowers or very high mark-ups on interest rates will be imposed on them.

3.4. Feature set selection and efficiency influencing factors

3.4.1. Sensitivity analysis of the factors influencing efficiency

The feature set selection is an important step in the decision support systems. The feature set selection should have sufficient explanatory power along with the economic theoretical background. Therefore, the sensitivity analysis is needed to understand the question to which extend performance can be estimated by which set of factors and weights. The estimation should provide the clear insight into factors influence the organizational performance.

An important issue to be evaluated in any practical research is whether the characteristics of ensemble methods in machine learning and a dataset facilitate assessing the suitability for this data set a priori adequately. The identification of feature sets can be carried out by regression modeling by gaining explanatory insight in that it provides an empirical fundament for further ensemble techniques. Gathering features of individual classifiers and characteristics of the benchmarking can be performed by using covariates in a regression framework to evaluate independent variable. Therefore, the entire process of feature selection should be broken down into two process:

1. **Datasets validation.** It is a common process in econometrics, where longitudinal multi-dimensional data is involved into measurements over time. These datasets are grouped into panels with observations of multiple phenomena acquired over time periods for the organizations. Hence, a number of analysis should be conducted in order to understand the datasets characteristics cross-sectional dependence, various effects, presence of heteroscedasticity, and stationary.
2. **Feature selection.** Based on the results of datasets validations there is possible to construct mixed methods of various statistical techniques to find dependencies taking into account individual-specific heterogeneity and effects.

Feature selection could help to expose the hidden meaning and increase the explanatory power. The model in the regression form is based on the data model based on the findings in Chapter III, Subsection *3.3 Efficiency assessment by nonparametric model* on page 122 and Subsection *3.2.1 Data model for decision support systems* on page 111:

$$\textit{Output performance} \sim \textit{Efficiency (1,2)} + \textit{Input factors} + \textit{Uncertainty factors} \qquad (3.4.1)$$

Following Banerjee *et al.* (1993) and Huynh *et al.* (2012) there are mixed methods of various statistical techniques has been applied to deal the correlation metrics between independent variables. In Table 19 there is the representation of the results of estimating the five linear models in Columns (2-5) for panel data model of output estimation and using the dataset. In Column (1) there are estimates using as explanatory variables of the efficiency, input factors and uncertainty variables. There are four models starting with the simple ordinary least squared (OLS) in Column (2) to examine the consistency of the coefficient of variables. More complicated models are involved into analysis with the Fixed

Effects in Column (3) and Random Effects in Column (4) respectively. Upon the analysis variables will continually be added to inspect whether they are significant in explaining depend variable. The best model will be preferred based on econometric reasons. Column (5) is Pooled ordinary least squares estimation technique applied on panel data. The analysis has a significant heterogeneity in the impact effect on firm-level output, depending on the level of uncertainty, and significant convexity in the response of input variables to market volatility. The indication of *Financial leverage* effect with the *Stock volatility* is explained by adjusting in the long run towards a target that is integrated with its overall uncertainty level. There is no evidence here that a permanent increase in the level of uncertainty would affect the level of the Stock volatility in the long run, but there is a clear evidence that rise in uncertainty decrease efficiency in the short run period by the means that are not fully explained by the qualitative additive interaction model.

Table 19. *Feature selection of efficiency and uncertainty variables using ordinary least squared, fixed effects, random effects and pooling models*

Variables	Model 1 OLS	Model 2 FE	Model 3 RE	Model 4 Pooling
Efficiency Stage 1	-3.83	-5.50	-5.03	-3.41
Efficiency Stage 2	0.02	0.03	0.01	0.02
Stock volatility	0.68***	0.66***	0.66***	0.68***
Stockholder expectation	0.24*	0.31**	0.29**	0.24*
Return on Assets (ROA)	0.93**	1.07**	1.04***	0.93**
Net margin	-0.22*	-0.29**	-0.27**	-0.22*
Financial leverage	-2.18***	-2.37***	-2.32***	-2.18***
Global Economic Policy Uncertainty	4.98**	4.90**	4.90***	4.98**
World Uncertainty Index	-88.80***	-87.28***	-87.72***	-88.80***
R^2	*0.78*	*0.80*	*0.80*	*0.78*
Adj.	*0.75*	*0.74*	*0.77*	*0.75*

(Source: The Author's representation)

Each component of the factor variable at the firm level is encompassing the effects particular to each firm. Since ordinary least squares regression does not consider heterogeneity across groups or time, predictors are not significant in the ordinary least squares model. Hence, application fixed-effects models are needed to be considered in analyzing the impact of time-variable parameters. The fixed-effects establish the relationship between predictors and outcome variables within groups. Fixed Effects are characterized by independent of time, whereas Random Effects include random disturbances.

Therefore, channels of uncertainty transmission can provide valuable insights regarding these interactions and raises the question of how the efficiency and effectiveness of the policies in achieving their objectives may be affected in influencing the economy through stock market.

3.4.2. Results of analysis of features selection

The main research question of this study set up by the Author is to shed light on the idea of economic decision process and its connection with investments from different perspective in terms of their ability to create or absorb technological innovations within on-going infinite technological progress. Although, the preliminary results of the study do not meet the announced research aim of improving decision-making process for policymakers, it is possible to derive the feature sets from the datasets represented in the Chapter II, Section *2.2 Taxonomy of datasets selection for decision support system* on page 83. The factors decomposing from the models in ranking shows, that firm-level outcome depends heavily on the decisions made by DMU in order to maximize performance under constraint of resources allocations. This assumptions corresponds with the theoretical background in the Chapter I on page 27, where economic science is defined as a study of human beings' activities and behavior related to exploiting scarce productive resources to satisfy their fundamentally unlimited needs. The concept of scarcity of resources is deeply rooted in the economic theory and fundamentally grounded in scientific field.

The analysis proved the assumption in the Chapter II, Section *2.4 Two-stage DEA non-parametric efficiency analysis* on page 91, where it is suggested to regard efficiencies separately due to the nature of the efficiency. In the Stage 1 efficiency, the input allocation and firm level efficiency is the key for the firm level efficiency assessment, where externalities do not have direct impact on the technology applied in production. The Stage 2 efficiency has output oriented approach. The Stage 2 efficiency is related with the market capitalization and market stock equity. Therefore, the input parameters are greatly influenced by uncertainties and market volatilities. stockholders are interested in profit maximization. Hence, they do not influence firm level efficiency directly.

The results of theoretical assumptions fully corresponds with the practical finding in Table 20, where the firms level output is influenced by input orientated decision making process for a better resources allocation within firms to gain best possible performance. The stock market output orientated decision making process supports the assumption of importance of stockholders in development strategy of stock listed firms. Pure uncertainty factors do still have significant strategic meaning, which considerably influence the decision-making process.

Table 20. *Feature set selection by factors for decision making process*

Factor	Rank	Percentage
Efficiency stage 1	1	49,45%
Efficiency stage 2	2	43,39%
Uncertainty	3	7,16%

(Source: Author's representation)

In Table 21 the uncertainty factors are scrutinized and ranked by its influential power. Not surprisingly, the top five ranks are related with the concept of the bounded rationality, which theoretically described in details in the Chapter I, Section *1.2.2 Context of heterogeneous economic agents* on page 40. The uncertainties often are cognitive with emotional constitution which influence occasionally making important rational decisions. Hence, there is obvious to observe that any decision making process consist two dimensions: environmental demands and bounds on adaptability in the given decision-making situation. Standard statistical techniques give the tools to distinguish systematic from random factors, so in principle it is possible to distinguish the rational, adaptive portion of a decision from bounds on rationality.

Table 21. *Influential power for externalities selection in ensemble machine learning*

Description	Rank	Value (log)	Percentage
The business tendency and consumer opinions	1	5,461378216	11,03%
Economic Policy Uncertainty Index	2	5,394624804	10,73%
Future Tendency	3	5,377094411	10,10%
Trade uncertainty	4	5,336047118	9,95%
World uncertainty index	5	5,272362216	9,77%
Industrial Production	6	5,227815053	7,82%
Equity Market Economic Uncertainty	7	4,985915285	6,89%
Nasdaq volatility	8	4,951239379	7,77%
Final Consumption Expenditure Of Households	9	4,932850058	5,31%
ECB Interest Rates Refinancing Operations	10	4,894053246	4,71%
Foreign direct investment	11	4,871783561	3,04%
Total manufacturing	12	4,867727747	2,81%
EU 28 Countries HICP	13	4,837812913	2,79%
Other parameters with importance (<2%)	14-18	4,816835951	7,28%

(Source: Author's representation)

The ranking of the most significant firm-level feature set in Table 22 corresponds with theoretical background in the Chapter II, Section *2.2 Taxonomy of datasets selection for decision support system* on page 83, where among direct profitability indicators, return on Assets (ROA) is the most widely established parameter of organizational performance. It is

a rational indicator of an organization's long-term financial perspective. It uses aggregated factors from both financial income statement and from the balance to evaluate profitability. It enables assess a better way the returns and risks related with an organization from operating decision's perspective and environmental effects. However, the uncertainty factor in terms of the *Stock volatility* plays an important role in the firm-level input orientated decision-making process. The background of the given influential factor is that managerial decision-making motivation is often linked with the stock performance of the organization through the economic profit channel. The economic profit represents the incremental difference in the rate of return over a company's cost of capital. If an economic profit is negative, it has a clear signal, that a company does not generate value from the invested funds. The investment allocation is a managerial decision-making processes, based on the perception of future business settings and market conditions. Thus, the economic profit concept in any form revolves around the assessment of the performance of a company though one single investment channel. Therefore, the understating of efficiency from economic point of view is narrowed to the perception is that only creation of wealth and returns for shareholders can improve the efficiency and performance.

Table 22. *Firm-level feature set definition for ensemble machine learning*

Variable	Rank	Value (log)	Percentage
ROA	1	6,937191539	39,38%
Stock volatility	2	6,503742633	14,52%
Gross margin	3	6,282605622	8,72%
Leverage	4	6,239561166	7,90%
ROE	5	6,222779822	7,60%
Revenue	6	6,090200391	5,60%
Net margin	7	6,061913013	5,25%
Assets turnover	8	6,030135106	4,88%
Dividends	9	5,910220681	3,70%
Operating margin	10	5,729964593	2,44%

(Source: Author's representation)

The market performance output orientated decision making process is decomposed in Table 23, where the most important role takes the *Stock equity* for the reasons of environmental demand and the decision-making expectations in terms of *Stockholder dividends*. The uncertainty factor is fully correlated with the findings around 9,77% as presented above. The market performance output orientated decision making process neglects the input orientated optimization.

Table 23. *Factors of market performance output orientated decision making process*

Variable	Rank	Value (log)	Percentage
Stock equity	1	7,109325591	66,72%
Stockholder dividends	2	6,379064768	12,42%
Stock volatility	3	6,330029443	11,09%
Efficiency stage 1	4	6,274884896	9,77%

(Source: Author's representation)

Only the structured approach using various methods in approaches both from Data Science and Economic studies might help researchers and policymaker rank the important factors and appreciate the factors underneath a better way. It is not possible to respect the results either from statistical nor theoretical point of view solely, but only as an integrated process with the fully qualified decision support system.

It underpins the assumption in Chapter I Section *1.3 Foundations of uncertainty in economics theories* on page *41* that uncertainties and efficiencies can hardly be modeled using error minimizing techniques. The modeling process takes place for large randomized subsets with agents having decision-making process limited by disposable information, considering the cognitive limitations and time constraint to make a decision. Therefore, the results might be vague due to the homoscedasticity and lack of correlations, which lead to the fundamental identifiability because of limited amount of observations and uncertainties underneath. Of course, with certain assumptions, it is not always required that the exogenous variables in the models should be fully observable. In this case, simulations should allow consider many sources of uncertainty simultaneously, but they can also lead to randomized combination of different effects. The most important factor is that, the lack of identifiability of the uncertainty factors makes identification of distributions rather difficult task from theoretical point of view and virtually impossible from practice implication. Ironically, the aspiration to find complete factors of uncertainty can lead to a reduction of the informational power of predictions and their usefulness in the decision support systems.

Selecting an appropriate set of features to represent the main information of original datasets is an important factor that influences the accuracy of efficiency and classification methods. Improving the classification accuracy and predictability ability, increasing the training process speed and decreasing the storage demands are some of the potential advantages of feature selections algorithms. Therefore, reducing the number of feature set, better understanding and interpretability of a figures can be achieved. To make a smaller feature set based on the initial feature space to obtaining more classification accuracy and precision, different kind of methods have been proposed (Miller and Forte (2017), Kuhn (2008), Kuhn and Johnson (2013)). When the original feature sets transformed by feature extraction and feature selection if none of transformation have been made.

3.5. Nonparametric efficiency models with ensemble machine learning

3.5.1. Results of ensemble machine learning techniques

The main application of machine learning algorithm is hidden pattern recognition. Therefore, the uncertainty feature is integrated into the set. The goal of normalization is to eliminate redundancy in the datasets, because balanced data attempts to give all attributes an equal weight. Normalization is particularly useful for classification algorithms. Without data normalization, the new machine learning techniques that were discussed in the previous section would simply not be possible. Therefore, the purpose of the next step is to select normalized inputs and outputs not depending on imbalanced data.

Incorporating results of nonparametric analysis of inputs and outputs allows to assess efficiency as classification problem. The DMU evaluation consists of performance and efficiency. Performance reflects the relationship between decisions made by DMU and environmental response. Efficiency reflects the competitiveness of the DMU in the given settings. The higher the efficiency, the stronger the competitiveness for DMU increase performance. The proposed efficiency and performance clusters concerning classifications are in Table 24. Column (1) describes the qualified term of the measure desired to achieve. The Column (2) and Colum (3) decompose the qualified characteristics attributed to a DMU. The *Category* represent barely only a combination of both *Efficiency* and *Performance* in order to present plausible result for policymaker. As described in the previous Section a hybrid method combining nonparametric analysis and machine learning techniques is proposed to model the classification problem in terms of efficiency and performance at different levels.

The initial feature set elaborated in the Chapter II, Section *2.2 Taxonomy of datasets selection for decision support system* on page 83 lead to ranking attributes feature set (FS*) described in the Chapter III, Section *3.4 Feature set selection and efficiency influencing factors* on page 131 should be able to classify the DMUs in efficient and inefficient in terms of the efficiency scores derived by nonparametric efficiency estimation under the externalities. Several methods have been proposed to improve the estimation of efficiency in the classification training according to the Chapter I, Section *1.3.3 Structural risk minimization* on page 50. The presence of noise increases the inefficiency bias by the simultaneous computational cost increase of the classifier. If the uncertainty of the samples, including the noisy and outlier samples, is identified and discarded before training the classifier, then useful samples are obtained, and the training samples are reduced.

Table 24. *Categorization of performance and efficiency*

Category	Performance	Efficiency	Characteristic
Excellent	Y	Y	Optimal resources allocation. Positive credit and development
Average	Y	N	Recourses consuming. Sub-optimal resources allocation.
Poor	N	Y	Might have issues in a long-term
Very poor	N	N	High risk of solvency

(Source: Author's representation)

Based on the adjusted feature set, ensemble methods in machine learning advocated to train the model that can be used evaluation of unknown DMUs. Figure 30 represents the entire sub-process of ensemble methods in machine learning multi-stage flow. The beginning of the process is to identify observed datasets and the results from previous models and procedures. Thereafter the entire dataset is divided into to the train dataset and test part in order to enable ensemble methods in machine learning algorithm to achieve the best possible predictive power. All machine learning algorithm are compared against a linear regression (=*benchmark algorithm*) in order to verify them statement whether or not machine learning approaches can considerably achieve a better result as exposed by sophisticated panel data analysis in the Chapter III, Section *3.2.2 Datasets structure analysis* on page 116. The analysis demonstrates rather weak correlation among variables due to the fundamental identifiability because of limited amount of observations and uncertainties underneath. As the result of the uncertain environment or imprecision in measurement, the data is not a precise value. Therefore, no doubt this is the toughest challenge for many researchers in the field is uncertainty. Such uncertainty introduces an unavoidable risk factor that is an integral part of theory. In order to handle the complex nature of the problem an interdisciplinary approach is strongly advocated.

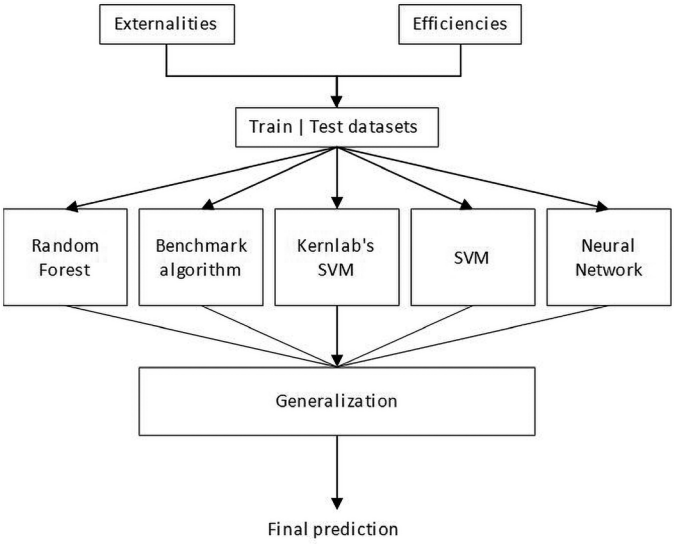


Figure 30. *The modeling processes of ensemble machine learning using various algorithms (Source: Author's representation)*

The ensemble method applied to train the model with the dataset using FS*. The suitable kernel function for SVM approach should be employed by experimenting with different options and parameters for each particular prediction model. The difference between SVM and KSVM is solely practical implementation of the same concept described in details in

the Chapter II, Section *2.5.1 Support vector machines* on page 100. The best is to start with a simple linear or low-degree polynomial kernel function and move to more complex kernel functions only if good performance cannot be achieved with this. Suitable selection of kernel function and the related parameters influence very much on the model accuracy to optimize the parameters C, γ associated with RBF kernel, and d associated with polynomial kernel based on 10-fold cross validation. However, tuning hyperparameters for classification machine learning algorithms might avoid limitations of extensive theorization of parameters. Deng *et al.* (2012) gives insights into the penalty parameter C determines the tradeoff between two conflicting goals: maximizing the margin and minimizing the training error. The larger C implies that more attention has been paid to minimizing the training error. From the practical point of view, it is critically lacking in quantitative meaning. Thus, the standard C-SVM is modified as ν-support vector classification. The significance of the parameter ν is practically hard to define as a training point to any plausible criteria, where whose inputs are separated *sufficient* correctly. The practical implementation in the decision support systems fails due to sensitivity to outliers and noise.

Many researchers and esteemed reviews refer to a typical approach in the practical realization of parameters method based on the widely-accepted notation of LIBSVM implementation Chang and Lin (2011), where SVM formulations supported in LIBSVM: C-Support Vector Classification (C-SVC), ν-Support Vector Classification (ν-SVC), distribution estimation (one-class SVM), ϵ-Support Vector Regression (ϵ-SVR), and ν-Support Vector Regression (ν-SVR). The Author's goes throughout the study with modeling approach, which is characterized by theoretical findings in the Chapter I, Subsection *1.5 Ensemble machine learning approach in decision-making process* on page 66 represented by data model definition and justified practically in the Chapter III, Subsection *3.4. Feature set selection and efficiency influencing factors* on the page 131. Therefore, tuning the hyperparameters in the algorithm over selection explicit parameters plays important role with the same practical result, that may improve the performance of the SVM in the SuperLearner[16] implementation. The advantage of tuning hyperparameters is that it has considerable benefit of tailoring the behavior of the algorithm to specific datasets. The Author's novelty is to propose the best possible approach to coupe with uncertainties underneath in terms of conjunction of the solid economic scientific literature body with novelty of data driven algorithms in the economic studies. From practical realization the hyperparameters approach differentiated from strict parameters methods, where coefficients along with weights found by the learning algorithm. Thus, hyperparameters are specified by configuring the data model as justified in details in the Chapter II, Subsection *2.2. Taxonomy of datasets selection for decision support system* on the page 83.

Kernel function parameter selection is one of the important parts of SVM modeling. Based on the calculation it is clear that the difference between SVM and other machine learning methods is significant. SVMs with different kernel functions outperforms other techniques. Important to mention, that some of the machine learning algorithms are not suitable for classification of large datasets. Worth to mention, that the training complexity

16 Appendix 10. R Ensemble Machine Learning

of machine learning algorithm especially SVM is highly dependent on the size of data set. The reduction of features that are foreseen as an input parameter to SVM is essential condition for obtaining reliable results.

The classification assignment takes input vectors at the first phase. Thereafter, based on training from exemplars of each class the classification algorithm should decide and identify the classes of each of ***N*** unknown classes. The most important point about the classification problem is that each example belongs to precisely one class, and the set of classes covers the whole possible output space.

The risk estimation qualifies the models accuracy and performance with intention to minimize the estimated risk. It means that the model should make low as possible mistakes in its prediction. The mean-squared error in a regression model has been chosen for the risk estimation.

Table 25. *Fitting machine learning algorithms*

Algorithm	Valid method	Coefficient	Average	Std. Err.	Min	Max
Random Forest	Yes	0.90787505	0,018829	0,0046548	0,0032914	0,056203
SVM	Yes	0.82235931	0,077622	0,0164955	0,0037596	0,165642
KSVM	Yes	0.78456423	0,087096	0,0166571	0,0118353	0,214417
Neural Networks	No	0.02448529	0,236482	0,0140787	0,1985978	0,290962
Mean	No	0	0,235147	0,0138448	0,1985978	0,290962

(Source: The Author's representation)

Table 25 gives the simultaneous model fitting based on the lowest risk parameter, which creates a weighted average of multiple models involved into analysis. It includes the mean of ***Y*** as a benchmark algorithm to compare very simple prediction against more complex algorithms in order to find the best single discrete algorithm with low weight in the weighted-average ensemble. The coefficient means the weight puts in the ensemble on the particular model in the weighted-average. In case of 0 or close to 0 coefficient, the model does not have any practical meaning. The Random Forest has the most weight, following by SVM and KSVM. The ANNs and the mean algorithms do not comply with any practical meaning in this research.

The failure of ANNs for this particular analysis is caused by the nature of the selected feature sets described in the Chapter III, Section *3.4 Feature set selection and efficiency influencing factors* on page 131. The ANNs learn from complex higher-order decision boundaries to fetch features to model phenomena. While the described feature set has a lot of characteristics of uncertainties and bounded rationality caused by unexpected behavior of the network, definition of proper network structure and difficulties to define the problem to the network.

The estimation of the performance of the ensemble algorithms should be done. At least 10fold cross-validated risk estimates is needed by incorporating a bunch of ensemble algorithms that might differ in bias and data-fitting degree. Therefore, the nested cross-validation prevents overfitting and selecting an extremely biased fit. The most common technique of 10fold cross-validation gives asymptotic optimality properties. The Figure 31 gives the graphical representation of the results of the application ensemble learning algorithms. It is the representation of the V-fold cross-validated risk estimates for each algorithm applied including an asymptotic 95% confidence interval.

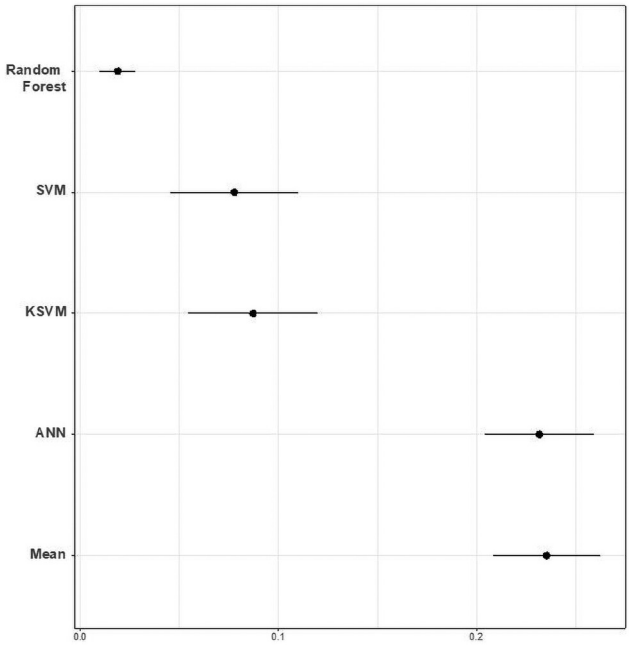


Figure 31. *Ensemble machine learning V-fold CV risk estimates with 95% CIs (Source: The Author's representation)*

It is clear, that the quality of the result depends to a high degree on the reduced feature set. Therefore, merely useful features should be considered left at the end from the entire set of features. To overcome the issue a random forest can be appreciated in the same terms. Random forest gives strong results on a variety of datasets, and is not extremely sensitive to tuning parameters. The practical results underpin the assertion that the Random Forest is primarily suited for multiclass problems definitions, while SVM is basically suited two-class problems.

In order to achieve a plausible and comparable results with Random Forest it is needed to reduce data into multiple binary classification problems. There are two typical methods for handling multiple class data sets:

1. A classifier algorithm is supplied with dataset with no amendments
2. Dividing the datasets into multiple binary sub-problems.

The first approach can be met relatively frequently in scientific literature and practical implication. The second approach requires generating binary problems from a multiple class data set joining the results of these binary classifiers produced by sum, sum with threshold, Hamming distance or loss-based function. The disadvantage of multiple binary classification problems is that the entire approach is grounded on data processing approach, which can hardly be justified by economic theories. Random Forest handles a mixture of numerical, categorical features *out-of-scale* values a better way. The SVM is designed to margins maximization. Therefore, the SVM relies on the distance theory between different points. From certain point of view, the argumentation of the distance concept might have more meaning in case of data with ***n*** points and ***m*** homogeneous features. In summary, for a decision making system based on the classification problem Random Forest produces probability of belonging to certain class. The SVM gives distance to the boundary, which is needed to be converted into probability or depending on the interpretation into other plausible result.

3.5.2. Discussion of models for the decision support systems

There is no clear one algorithm, which can be applicable in any situation. The Random Forest approach give slightly better result, but SVM can handle structured problems a better way. The analysis based on the confusion matrix analysis in Table 26 give insight into the ensemble methods in machine learning techniques. A confusion matrix is a powerful tool to compare algorithms on a classification problem, where the count of correct and incorrect predictions summarized and broken down by each class. The confusion matrix shows the parameters in which classification model is disordered while making predictions. It gives insight into the errors being made by a classifier and the types of errors.

Table 26. *Analysis of the confusion matrix*

Measure	Random Forest	Support Vector Machine
Accuracy	0,9722	0,9444
95% CI	(0,8547, 0,9993)	(0,8134, 0,9932)
No Information Rate	0,8166	0,7778
P-Value [Acc > NIR]	0,001328	0,007382
Kappa	0,9159	0,8393
Sensitivity	0,9737	0,9643
Specificity	0,8752	0,8751

Measure	Random Forest	Support Vector Machine
Pos Pred Value	0,9655	0,9643
Neg Pred Value	0,8963	0,875
Prevalence	0,7778	0,7778
Detection Rate	0,7778	0,7555
Detection Prevalence	0,8056	0,7778
Balanced Accuracy	0,9375	0,9196

(Source: Author's representation)

The models sensitivity and specificity are comparable and significantly higher than other methods used in the study in two-class classification tasks with reduced FS. That means that Random Forest and SVM has a better prediction power for sensitivity in terms of the proportion of observed positives that were predicted to be positive and specificity in terms of the proportion of observed negatives that were predicted to be negatives. Accuracy is then defined as the sum of the number of true positives and true negatives divided by the total number of examples. There is another complementary pair of measurements that can help to interpret the performance of a classifier, sensitivity and specificity. Sensitivity is the ratio of the number of correct positive examples to the number classified as positive, while specificity is the same ratio for negative examples. Moreover, at the 95% significance level for the accuracy *p*-values of the paired *t*-test results that the SVM with linear kernel (p=0.32) and logistic regression (p=0.281) are not statistically significant.

Table 27. *Results of SVM kernels*

Method	Accuracy	*p*-value	Sensitivity	Specificity
SVM - Linear	0,8968	0,32	0,9163	0,8356
SVM - RBF	0,9444	0,005	0,9643	0,8751
SVM - POLY	0,7243	0,001	0,8374	0,7118
Logistic regression	0,5848	0,29	0,645	0,3659
Naïve Bayes	0,6608	0,056	0,6941	0,4837

(Source: Author's representation)

Further details on SVM implication are in Table 27. However, the statistical power of SVM with RBF kernel (p=0.005) and SVM with polynomial kernel (p=0.001) provide statistically significant higher accuracy values. Non-linear SVM classification with RBF kernel with accuracy of 94,44% is better than SVM with linear 89,68% and polynomial 72,43% kernels. The standard deviation values of SVMs are also relatively lower comparing with

other methods, which means that SVM is a better choice for DMU classification. Logistic regression with 58,48% is a special case of linear regression where the target variable is categorical in nature to predict the probability of occurrence of a logit function. A naive Bayes classifier of 66,08% represents a family of machine learning algorithms, which treats feature set as independent one, where SVM looks at the interactions between them to a certain degree using a non-linear kernel. Validation results using 10-fold cross validation is performed to identify the class that the unknown DMU. The average accuracy and its standard deviation is computed to evaluate the performance of the proposed method. The data dispersion is characterized in terms of variance and standard deviation to indicate spreads of a data distribution. A low standard deviation means that the data observations tend to be very close to the mean, while a high standard deviation indicates that the data are spread out over a large range of values. At the 95% significance level for the accuracy p-values of the paired t-test results that the SVM with linear kernel (p=0.32) and logistic regression (p=0.281) are not statistically significant. However, the statistical power of SVM with RBF kernel (p=0.005) and SVM with polynomial kernel (p=0.001) provide statistically significant higher accuracy values.

The above analysis implies that Random Forest and SVM are suitable for the DMU classification task. In particular, the integration of SVM with the RBF kernel and DEA method achieved the best results. Proper method selection is necessary for the supplier evaluation which may guarantees DMU evaluation optimum solutions when compared with other artificial intelligence approaches. Especially for SVM, making an appropriate choice for kernel function is the key to construct a classification model which may enhance the prediction performance according to the above experimental results. Valid experiments using statistical test suggest that DEA score is a useful feature to improve the classification performance.

The results of the research have significant economic impact. First of all, the nonparametric models with reduced feature selection give extremely high prediction results. The findings underpin the main idea of the research, that hidden patterns can be recognized with higher probability. The importance of the proposed model is significant. Worth to mention, that many investors employ an investment model by selection process started with an economic environment drilled down to a single company performance. More favorable economic situation in a single country lead to a better financial activity on the market. But the current business settings are defined by complex global processes with uncertainties underneath. Economic turbulences affect all Baltic States really hard way. Then for the decision-makers the question of how a particular sector can overcome the crises is very important. Wrong unbalanced decisions might have serious long-run consequences. Thus, this research Section aims to explore the feasibility of the method associated with assessing efficiency in decision support systems under uncertainty condition with the real-world datasets. Using high-dimensional input space, efficiency assessment models and along with learning algorithm, the Author wants to deliver plausible result that optimize the predictable efforts. The proposed approach for efficiency assessment in decision support systems with learning algorithms is the one that fosters to avoid limitations of the conventional efficiency assessment methods. Taking common estimation techniques as referential, the

research will present that the learning algorithms methods associated with assessing efficiency in decision support systems under uncertainty condition can provide plausible yet clear unbiased results. The result should not provide some vague and unclear coefficients, which can be easily misinterpreted by decision maker, but give a clear answer estimated by the learning classification algorithm.

3.6. Empirical Research Results and Discussion

The objective of the empirical research part is to present generalized results on the efficiency assessment under uncertainty and to encourage scientific discussion on the research and the obtained results.

First, ensemble methods in machine learning are about to increase the models predictability by adapting parameters so that these actions get more accurate. The accuracy is measured by how well the chosen parameters reflect the correct ones. The recent studies show the necessity of multi-disciplinarily approach in decision making process. The ensemble methods in machine learning have been recognized as the promising ones. They merge ideas from statistics, mathematics and economics, to make models learn. Another thing that has made the change possible in direction of machine learning research is data mining, which looks at the extraction of useful information from massive datasets. The data mining requires efficient and fast algorithms, putting more of the importance onto data science. The relative complexity of the machine learning algorithms is also a part of interest for scientific research. It is particularly important because researchers might want to use some of the methods on very large datasets, so algorithms that have high degree polynomial complexity in the size of the dataset is a problem. The complexity consists from the complexity of training, and the complexity of applying the trained algorithm.

Important to *ex-post* evaluate each step of the empirical research for efficiency assessment under uncertainty:

1. **Data Collection and Preparation.** Throughout this study it is stressed out that he quantity of data needs to be considered. Machine learning algorithms need significant amounts of data, preferably without too much noise, but with increased dataset size comes increased computational errors.
2. **Feature Selection**. The identification of features that are important for modelling, it is also necessary that the features can be collected without significant errors, and that they are robust to noise and other corruption of the data that may arise in the collection process.
3. **Algorithm Choice**. With the dataset it is possible to make a choice of an appropriate algorithm, where the knowledge of the underlying principles of each algorithm and examples of their use is precisely what is required.
4. **Parameter and Model Selection**. There is an issue of handling various parameters that have to be adjusted manually. The process of adjustment for many models requires deep investigation to identify appropriate coefficients.
5. **Training**. With the dataset, algorithm, and parameters, training is a part of the process in order to build a model of the data to predict the outputs on new data. This

might be data consuming for supervised learning it all has to have target values attached and it is not always easy to get accurate labels.

6. **Evaluation**. Before a model can be deployed it needs to be tested and evaluated for accuracy on data that it was not trained on. This can often include a comparison with human experts in the field, and the selection of appropriate metrics for this comparison.

The empirical research shows, that established on the theory of the structural risk minimization principle to estimate a function SVM is shown to be very resistant to the overfitting problem, eventually achieving a high generalization performance. Due to the advantages of SVM algorithm in solving nonlinear problems, it can be used to capture and provide explanatory power of underlying uncertainties. It is proven, that the Author among other researchers investigated ensemble methods in machine learning approach to handle efficiency assessment under uncertainty feature sets and show how data uncertainty in feature sets can be treated in classification algorithms by employing robust feature selection.

However, in the Section 1 there are certain limitations for the assessment efficiency with nonparametric method are discussed. Application of DEA approach requires handling a separate linear program for each DMU. The aggregation of DEA to problems that have many DMUs can be intensive in terms of gathering initial dataset. The empirical research states clearly that this limitation can be solved by dataset normalization and careful feature selection. Since DEA is an extreme point technique, errors in measurement can cause significant problems. DEA efficiencies are very sensitive to even small errors, making sensitivity analysis an important component of afterwards DEA procedure. The Author argues, that machine learning approach can handle this limitation by employing various kernel specifications. The empirical research shows, that it is possible to overcome the limitation imposed by the non-parametric model, the application of machine learning can help to review large volumes of data and discover specific trends and patterns that would not be apparent to prior models due to a regularization parameter, which makes possible avoiding over-fitting.

It is confirmed that Machine Learning algorithms empowered with pre-defined kernel functions are good at coping with data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments over creating applicable yet easy interpretable knowledge about given issue.

CONCLUSIONS AND RECOMMENDATIONS

Definition and practical implementation of an effective decision support system under uncertainty is not a trivial task. As the result, the Author wants to express the idea that in the modern environment there is always space for new investigations and researches. The economics have been developing science adjusting methods and theories for the constantly mutable environment settings. The fact of the matter is that the dome has been opened for discussion, which theoretical and conceptual approach can embrace the processes itself and explain underneath causes a better way.

The study proved, that it is important to involve theoretical and empirical aspects of uncertainty, nonlinearities, complexity and bounded rationality as the major assumption of the framework, but not assumption of them in terms of other equilibria based theories. Analysis of the previous studies shows that theoretical part is detached from the statistical significant findings. It is obvious, that the economics itself from very early steps accepted equilibria concept and the study of generally balanced growing path. Thus, the statistical findings justified to established theories. Traditional studying equilibrium patterns of consistency required further behavioral adjustments.

The assessment of efficiency under uncertainty is defined by various sources of uncertainty, which cannot be quantified within other than hybrid model. From formal point of view, various uncertainties from missing data can be generalized with hypothesis of limited information. But there is to admit, that many real-world datasets may contain missing values for various reasons. Taking such data into a model with a lot of missing values can drastically impact the model's quality. The proposed model offers upfront how to deal with missing data using various machine learning techniques. The study underpins that the quality of the data is very important factor. The ensemble methods in machine learning require accurate measurement of both the inputs and outputs, construction of datasets for uncertainty. The DEA modelling should be regarded with cautious due to its subjective nature. There is no common approach while handling missing data. But within the proposed model in this study the treatment of missing data is one of the important tasks.

This study is one of the first attempts to assess efficiency within both classification and regression model. The Author among other researches investigate Random Forest, ANN and SVM classifiers in the face of uncertain knowledge sets and show how data uncertainty in knowledge sets can be treated in ensemble methods in machine learning by employing robust optimization. Consequently, various ensemble methods in machine learning can also be used as regression techniques, maintaining all the main features that characterize the algorithm of maximal margin. The Author is agreed with, that the future of the machine learning is in combination of different approaches, because fully supervised algorithms are a useful but perhaps an unnatural assumption due to latent variables in models.

The Author is the first who explicitly proposed to treat uncertainty not as a dummy variable, but phenomenon dissected within the proposed model on different layers: data-mining uncertainty, analytical framework uncertainty and uncertainty as a factor. Unlike the existing approaches, the combinations of machine learning techniques in this study do not require to think in terms of hypothetical assumption. Mathematically machine learn-

ing leads to the identification of implicit restrictions to weights, so there is a fundamental difference in these approaches, emerging from the way in which the data explicitly is gathered. In each process the uncertainty is emerging in different qualities and it should be assessed with respective techniques.

The proposed model deals with evaluating efficiencies in the absence of homogeneity gives rise to the issue of how to fairly compare a DMU to other units. A related problem, and one that has been examined extensively in the literature, is the missing data problem addressed directly to appropriate techniques of machine learning.

Depending on a number of factors, it is crucial to elaborate a theoretical framework, which can embrace as many factors dynamically. Therefore, any research on efficiency assessment under uncertainty should have a broader scope and should not be limited on country-specific parameters but include configurations in clusters. Uncertainty have been proven to be unstable factor, with the variations being most vividly seen during the crises, requiring researchers' attention. Due to the reasons mentioned here, the assessing efficiency under uncertainty is relevant in both theoretical and empirical aspects. This doctoral dissertation focuses on both aspects. The Author confirms that uncertainty is persistent phenomena in economics and it must be faced continually by policymakers. The measuring of macroeconomic uncertainty and understanding its impact on economic activity is thus crucial for assessing the current macroeconomic situation. From modern positions a robust and negative effect of uncertainty on economic growth is obvious and these consequences cannot be neglected by the theory. There are a vast number of studies arguing indicators of uncertainty which can be viewed as representative to the evidences of particular policy, involving a wide number of direct and indirect peers. The uncertainty factor is so large that the effects of policy decisions on the economy are thought to be ambiguous.

In this situation, any plausible expertise on the nature of uncertainty might be very useful. In order to understand how variations in uncertainty might affect the economic process, it is important to find its source. Uncertainty should not be oversimplified. The phenomena affect individual sectors of the economy in totally different manner with different impact and different degrees of persistence.

Bringing together the diverse characteristics of the economic activities into a framework, the analysis of networks appears. The networks occupy a significant place in a wide range of approaches in studying economic diversity. However, pretty frequently standard economic theory often ignores the economic networks characteristics explicitly in its analysis.

The relational nature of efficiency assessment under uncertainty implies that it can encounter certain challenges in statistical analysis out of standardized statistics problems. The simultaneous analysis of dependencies of the quantities and dimensionality is often containing a vast number of data collected into model. In researches of complex subsets, some forms of the structure and characteristics can represent dynamics with a particular pattern. The issues involving the transfer of information, knowledge or commodities viewed in terms of efficiency assessment under uncertainty and progress along those paths.

This study presents a model for efficiency assessment under uncertainty, where proposed applications for efficiency assessment under uncertainty are described with its gen-

eral features of complexity or simplicity of further analysis. The next development of the study will be deeper experiments, where applications will be tested in a much larger dataset.

The ensemble method has been proposed to provide a good generalization performance. The classification result of the practically implemented Random Forest and Support Vector Machines are often far from the theoretically expected level. Each practical implication has own initial condition and characterized by approximations in datasets and various degree of complexity algorithms.

Machine learning and data science require more than just getting more data into a model. Data scientists need to actually understand the real processes behind the data to be able to implement a successful model. One promising methodology to implementation is knowing when a model might benefit from utilizing ensemble models. In this case future researches will take advantage of complex combinations of the predictors taken from multiple machine learning methods in order to achieve more accurate forecasts than any initial individual model.

REFERENCES

1. Abdou, H., and Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent Systems in Accounting, Finance and Management, 18*(2-3), 59-88.
2. Abe, S. (2010). *Support Vector Machines for Pattern Classification*: Springer London.
3. Abel, A., Mankiw, G., Summers, L., and Zeckhauser, R. (1989). Assessing dynamic efficiency: Theory and evidence. *The review of economic studies, 56*(1), 1-19.
4. Acemoglu, D. (2008). *Introduction to modern economic growth*: Princeton University Press.
5. Addo, P., Guegan, D., and Hassani, B. (2018). Credit Risk Analysis Using Machine and Deep Learning Models. *Risks, 6*(2), 38.
6. Adler, J. (2010). *R in a Nutshell: A Desktop Quick Reference*: O'Reilly Media.
7. Aggarwal, C. (2014). *Data classification: algorithms and applications*: CRC press.
8. Aggarwal, C., and Yu, P. (2009). A survey of uncertain data algorithms and applications. *IEEE Transactions on Knowledge and Data Engineering, 21*(5), 609-623.
9. Ahmad, A., Hassan, M., Abdullah, M., Rahman, H., Hussin, F., Abdullah, H., and Saidur, R. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews, 33*(1), 102-109.
10. Aigner, D. J., and Chu, S. F. (1968). On Estimating the Industry Production Function. *The American Economic Review, 58*(4), 826-839.
11. Aigner, D. J., Lovell, C., and Schmidt, P. (1977). *Formation and Estimation of Stochastic Frontier Production Function Models* (Vol. 6).
12. Akkizidis, I., and Stagars, M. (2015). *Marketplace Lending, Financial Analysis, and the Future of Credit: Integration, Profitability, and Risk Management*: Wiley.
13. Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., and Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications, 94*, 164-184.
14. Alavi, M., and Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly, 25*(1), 107-136. doi:10.2307/3250961
15. Albert, J. (2007). *Bayesian Computation with R*: Springer.
16. Albert, J., and Rizzo, M. (2012). *R by Example*: Springer New York.
17. Aldamak, A., and Zolfaghari, S. (2017). Review of efficiency ranking methods in data envelopment analysis. *Measurement, 106*, 161-172. doi: https://doi.org/10.1016/j.measurement.2017.04.028
18. Alejo, R., García, V., Marqués, A., Sánchez, J., and Antonio-Velázquez, J. (2013). Making accurate credit risk predictions with cost-sensitive mlp neural networks. In *Management Intelligent Systems* (pp. 1-8): Springer.
19. Ali, A., and Lerme, C. (1997). Comparative advantage and disadvantage in DEA. *Annals of Operations Research, 73*, 215-232.

20. Ali, A., Lerme, C., and Seiford, L. (1995). Components of efficiency evaluation in data envelopment analysis. *European Journal of Operational Research, 80*(3), 462-473.
21. Ali, A., and Seiford, L. (1990). Translation invariance in data envelopment analysis. *Operations research letters, 9*(6), 403-405.
22. Allen, T. T. (2011). *Introduction to Discrete Event Simulation and Agent-based Modeling: Voting Systems, Health Care, Military, and Manufacturing*: Springer London.
23. Anadol, B., Joseph, P., Simak, P., and Yang, X. (2014). Valuing private companies: a DEA approach. *International Journal of Business and Management, 9*(12), 16.
24. Andersen, E. (1998). *The Evolution of the Organisation of Industry*.
25. Anderson, P. W. (2018). *The Economy As An Evolving Complex System*: CRC Press.
26. Archibugi, D., Filippetti, A., and Frenz, M. (2013). Economic crisis and innovation: is destruction prevailing over accumulation? *Research Policy, 42*(2), 303-314.
27. Arthur, W. B. (2018). *The Economy As An Evolving Complex System II*: CRC Press.
28. Asheim, B. (2007). *Differentiated Knowledge Bases and Varieties of Regional Innovation System* (Vol. 20).
29. Attigeri, G., Pai, M., and Pai, R. (2017). Credit Risk Assessment Using Machine Learning Algorithms. *Advanced Science Letters, 23*(4), 3649-3653.
30. Awad, M., and Khanna, R. (2015). Machine learning in action: examples. In *Efficient Learning Machines* (pp. 209-240): Springer.
31. Ayyadevara, V. K. (2018). *Pro Machine Learning Algorithms: A Hands-On Approach to Implementing Algorithms in Python and R*: Apress.
32. Baker, S., Bloom, N., and Davis, S. (2015). *Measuring Economic Policy Uncertainty*. Retrieved from https://EconPapers.repec.org/RePEc:nbr:nberwo:21633
33. Banerjee, A., Dolado, J. J., Galbraith, J. W., Press, O. U., and Hendry, D. (1993). *Co-integration, Error Correction, and the Econometric Analysis of Non-stationary Data*: Oxford University Press.
34. Banker, R., Chang, H., and Cooper, W. (1996). *Simulation Studies of Efficiency, Returns to Scale and Misspecification with Nonlinear Functions in DEA* (Vol. 66).
35. Banker, R., Charnes, A., and Cooper, W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science, 30*(9), 1078-1092.
36. Banker, R., Emrouznejad, A., Lopes, A., and de Almeida, M. (2012). *Data Envelopment Analysis: Theory and Applications.* Paper presented at the Proceedings of the 10th International Conference on DEA.
37. Banker, R., and Morey, R. C. (1986). The use of categorical variables in data envelopment analysis. *Management Science, 32*(12), 1613-1627.
38. Bartodziej, C. J. (2016). *The Concept Industry 4.0: An Empirical Analysis of Technologies and Applications in Production Logistics*: Springer Fachmedien Wiesbaden.
39. Bekaert, G., Hoerova, M., and Lo Duca, M. (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics, 60*(7), 771-788.
40. Benhabib, J., Wang, P., and Wen, Y. (2013). *Uncertainty and Sentiment-Driven Equilibria*.

41. Bensusan, H., and Kalousis, A. (2001). *Estimating the predictive accuracy of a classifier.* Paper presented at the European Conference on Machine Learning.
42. Bernal, O., Gnabo, J.-Y., and Guilmin, G. (2016). Economic policy uncertainty and risk spillovers in the Eurozone. *Journal of International Money and Finance, 65*(C), 24-45.
43. Berument, H., Yalcin, Y., and Yildirim, J. (2009). The effect of inflation uncertainty on inflation: Stochastic volatility in mean model within a dynamic framework. *Economic Modelling, 26*(6), 1201-1207.
44. Beyer, A., Nicoletti, G., Papadopoulou, N., Papsdorf, P., Rünstler, G., Schwarz, C., . . . Vergote, O. (2017). *The transmission channels of monetary, macro-and microprudential policies and their interrelations*: ECB Occasional Paper.
45. Beyersmann, J., Allignol, A., and Schumacher, M. (2011). *Competing Risks and Multistate Models with R*: Springer New York.
46. Bhavsar, H., and Ganatra, A. (2012). A comparative study of training algorithms for supervised machine learning. *International Journal of Soft Computing and Engineering (IJSCE), 2*(4), 2231-2307.
47. Binder, M., Lieberknecht, P., Quintana, J., and Wieland, V. (2017). Model uncertainty in macroeconomics: On the implications of financial frictions.
48. Bird, R., Choi, D. F. S., and Yeung, D. (2013). Market uncertainty, market sentiment, and the post-earnings announcement drift. *Review of Quantitative Finance and Accounting, 43*(1), 45-73. doi:DOI:10.1007/s11156-013-0364-x
49. Bishop, C. M., and Tipping, M. E. (2000). *Variational relevance vector machines.* Paper presented at the Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence.
50. Bivand, R. S., Pebesma, E., and Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R*: Springer New York.
51. Black, J., Hashimzade, N., and Myles, G. (2018). The impact of uncertainty on activity in the euro area. *German Institute for Economic Research (DIW Berlin Deutsches Institut für Wirtschaftsforschung e.V.)*.
52. Blackburn, V., Brennan, S., and Ruggiero, J. (2014). *Nonparametric Estimation of Educational Production and Costs using Data Envelopment Analysis*: Springer US.
53. Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica, 77*(3), 623-685.
54. Bloom, N. (2014). Fluctuations in uncertainty. *Journal of Economic perspectives, 28*(2), 153-176.
55. Bloom, N., Bond, S., and Reenen, J. (2007). Uncertainty and investment dynamics. *The review of economic studies, 74*(2), 391-415.
56. Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I., and Terry, S. (2018). Really uncertain business cycles. *Econometrica, 86*(3), 1031-1065.
57. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. a., and Hwang, D.-U. (2006). *Complex networks: Structure and dynamics* (Vol. 424).
58. Bogetoft, P. (2013). *Performance Benchmarking: Measuring and Managing Performance*: Springer US.

59. Bogetoft, P., and Otto, L. (2010). *Benchmarking with DEA, SFA, and R*: Springer New York.
60. Bolker, B. M. (2008). *Ecological Models and Data in R*: Princeton University Press.
61. Bomberger, W. A. (1996). Disagreement as a Measure of Uncertainty. *Journal of Money, Credit and Banking, 28*(3), 381-392.
62. Born, B., Breuer, S., and Elstner, S. (2018). Uncertainty and the great recession. *Oxford Bulletin of Economics and Statistics, 80*(5), 951-971.
63. Borovkov, A. A. (1998). *Ergodicity and Stability of Stochastic Processes*: Wiley.
64. Bowlin, W. F. (1998). Measuring performance: An introduction to data envelopment analysis (DEA). *The Journal of Cost Analysis, 15*(2), 3-27.
65. Breiman, L. (2001). Random Forests. *Mach. Learn., 45*(1), 5-32. doi:10.1023/a:1010933404324
66. Brian, A. W. (2006). Out-of-Equilibrium Economics and Agent-Based Modeling. In (Vol. 2, pp. 1551-1564): Elsevier.
67. Broek, J. v. d., Førsund, F. R., Hjalmarsson, L., and Meeusen, W. (1980). On the estimation of deterministic and stochastic frontier production functions: A comparison. *Journal of Econometrics, 13*(1), 117-138. doi: https://doi.org/10.1016/0304-4076(80)90046-9
68. Brose, M. S., Flood, M. D., Krishna, D., and Nichols, B. (2014a). *Handbook of Financial Data and Risk Information I*: Cambridge University Press.
69. Brose, M. S., Flood, M. D., Krishna, D., and Nichols, B. (2014b). *Handbook of Financial Data and Risk Information II*: Cambridge University Press.
70. Brown, R. G. (1983). *Introduction to random signal analysis and Kalman filtering*: Wiley.
71. Brown, R. G., and Hwang, P. Y. C. (2012). *Introduction to Random Signals and Applied Kalman Filtering with Matlab Exercises*: Wiley.
72. Bruun, C. (2004). *Agent-based computational economics-An introduction*.
73. Bryant, W. (2014). The Microeconomics of Choice under Risk and Uncertainty: Where Are We? *Vikalpa, 39*(1), 21-40.
74. Bryant, W., Ankit, K., Mahajan, P., Dangi, M., D'Aguiar, B., and Damodaran, U. (2014). Uncertainty, Ambiguity, and Financial Decision-Making. *Vikalpa, 39*(1), 103-128. doi:10.1177/0256090920140107
75. Buchen, T., and Wohlrabe, K. (2014). *Assessing the Macroeconomic Forecasting Performance of Boosting - Evidence for the United States, the Euro Area, and Germany* (Vol. 33).
76. Buckley, R., Arner, D., and Barberis, J. (2016). The Evolution of Fintech: A New Post-Crisis Paradigm? *Georgetown Journal of International Law, 47*, 1271-1319. doi:10.2139/ssrn.2676553
77. Burkov, A. (2019). *The Hundred-page Machine Learning Book*: Andriy Burkov.
78. Campbell, C., and Ying, Y. (2011). *Learning with Support Vector Machines*: Morgan & Claypool.

79. Cao, L.-J., and Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on neural networks, 14*(6), 1506-1518.
80. Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of Economics, 118*(1), 269-298.
81. Cass, D. (1965). Optimum growth in an aggregative model of capital accumulation. *The review of economic studies, 32*(3), 233-240.
82. Caves, D. W., Christensen, L. R., and Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica: Journal of the Econometric Society*, 1393-1414.
83. Celebi, M. E., and Aydin, K. (2016). *Unsupervised Learning Algorithms*: Springer International Publishing.
84. Chan, C., and Steiglitz, K. (2009). *An Agent-Based Model of a Minimal Economy*.
85. Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST), 2*(3), 1-27.
86. Charemza, W., Diaz, C., and Makarova, S. (2014). *Term Structure Of Inflation Forecast Uncertainties And Skew Normal Distributions*. Retrieved from https://ideas.repec.org/p/lec/leecon/14-01.html
87. Charnes, A., Cooper, W., Lewin, A., and Seiford, L. (2013). *Data Envelopment Analysis: Theory, Methodology, and Applications*: Springer Netherlands.
88. Charnes, A., Cooper, W., and Rhodes, E. (1979). *Measuring The Efficiency of Decision Making Units* (Vol. 2).
89. Chen, D., Batra, D., and Freeman, W. T. (2013). *Group norm for learning structured SVMs with unstructured latent variables.* Paper presented at the Proceedings of the IEEE International Conference on Computer Vision.
90. Chen, H.-H. (2008). Stock selection using data envelopment analysis. *Industrial Management and Data Systems, 108*, 1255-1268. doi:10.1108/02635570810914928
91. Chen, J. (2005). Information Theory and Market Behavior. *SSRN Electronic Journal*. doi:10.2139/ssrn.622901
92. Chen, T., and Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*.
93. Chuen, L. D. K., and Linda, L. (2018). *Inclusive FinTech: Blockchain, Cryptocurrency and ICO*: World Scientific Publishing Company.
94. Clayton, T., and Radcliffe, N. (2018). *Sustainability: a systems approach*: taylorfrancis.com.
95. Clements, M. (2014). Forecast Uncertainty- Ex Ante and Ex Post: U.S. Inflation and Output Growth. *Journal of Business & Economic Statistics, 32*(2), 206-216.
96. Clements, M., and Harvey, D. (2011). Combining probability forecasts. *International Journal of Forecasting, 27*(2), 208-223.
97. Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., and Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*: Springer US.
98. Cohen, Y., and Cohen, J. Y. (2008). *Statistics and Data with R: An Applied Approach Through Examples*: Wiley.

99. Colander, D., Howitt, P., Kirman, A., Leijonhufvud, A., and Mehrling, P. (2008). *Beyond DSGE Models: Toward an Empirically Based Macroeconomics* (Vol. 98).
100. Cook, W., and Seiford, L. (2009). *Data envelopment analysis (DEA)-Thirty years on* (Vol. 192).
101. Cook, W., Tone, K., and Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega, 44*(C). doi:10.1016/j.omega.2013.09.0
102. Cook, W., and Zhu, J. (2006). Rank order data in DEA: A general framework. *European Journal of Operational Research, 174*(2), 1021-1038.
103. Cooke, P. (2008). *Regional innovation systems: origin of the species* (Vol. 1).
104. Cooper, W., Seiford, L., and Tone, K. (2007). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*: Springer US.
105. Cooper, W., Seiford, L., and Zhu, J. (2011). *Handbook on Data Envelopment Analysis*: Springer US.
106. Cooper, W., and Tone, K. (1997). *Measuring inefficiency in data envelopment analysis and stochastic frontier estimation* (Vol. 99).
107. Cornett, M. M., and Saunders, A. (2003). *Financial institutions management: A risk management approach*: McGraw-Hill/Irwin.
108. Coudène, Y., and Erné, R. (2016). *Ergodic Theory and Dynamical Systems*: Springer London.
109. Cowpertwait, P. S. P., and Metcalfe, A. V. (2009). *Introductory Time Series with R*: Springer New York.
110. Cristianini, N., and Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*: Cambridge university press.
111. D'Amico, S., and Orphanides, A. (2008). *Uncertainty and Disagreement in Economic Forecasting*.
112. da Silva, I. N., Spatti, D. H., Flauzino, R. A., Liboni, L. H. B., and dos Reis Alves, S. F. (2016). *Artificial Neural Networks: A Practical Course*: Springer International Publishing.
113. Daal, E., Naka, A., and Sanchez, B. (2005). Re-examining inflation and inflation uncertainty in developed and emerging countries. *Economics Letters, 89*(2), 180-186.
114. Dangeti, P. (2017). *Statistics for Machine Learning*: Packt Publishing.
115. Del Negro, M., and Schorfheide, F. (2004). Priors from general equilibrium models for VARs. *International Economic Review, 45*(2), 643-673.
116. Deng, N., Tian, Y., and Zhang, C. (2012). *Support vector machines: optimization based theory, algorithms, and extensions*: Chapman and Hall/CRC.
117. Dequech, D. (2004). *Uncertainty: Individuals, Institutions and Technology* (Vol. 28).
118. Devezas, T., Leitão, J., and Sarygulov, A. (2017). *Industry 4.0: Entrepreneurship and Structural Change in the New Digital Landscape*: Springer International Publishing.
119. Diamond, P. A. (1965). National debt in a neoclassical growth model. *The American Economic Review, 55*(5), 1126-1150.

120. Diebold, F. X., Tay, A. S., and Wallis, K. F. (1997). Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters. *National Bureau of Economic Research Working Paper Series, No. 6228*. doi:10.3386/w6228
121. Diether, K. B., Malloy, C., and Scherbina, A. (2002). Differences of Opinion and the Cross Section of Stock Returns. *Journal of Finance, 57*(5), 2113-2141.
122. Donders, A. R. T., Van Der Heijden, G. J., Stijnen, T., and Moons, K. G. (2006). A gentle introduction to imputation of missing values. *Journal of clinical epidemiology, 59*(10), 1087-1091.
123. Doumpos, M., Lemonakis, C., Niklis, D., and Zopounidis, C. (2018). *Analytical Techniques in the Assessment of Credit Risk: An Overview of Methodologies and Applications*: Springer International Publishing.
124. Drake, L., Hall, M. J., and Simper, R. (2006). The impact of macroeconomic and regulatory factors on bank efficiency: A non-parametric analysis of Hong Kong's banking system. *Journal of Banking & Finance, 30*(5), 1443-1466.
125. Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., and Vapnik, V. (1997). *Support vector regression machines.* Paper presented at the Advances in neural information processing systems.
126. Dumitrescu, E., Hue, S., Hurlin, C., and Tokpavi, S. (2018). Machine Learning for Credit Scoring: Improving Logistic Regression with Non Linear Decision Tree Effects.
127. Dunis, C. L., Middleton, P. W., Karathanasopolous, A., and Theofilatos, K. (2016). *Artificial Intelligence in Financial Markets: Cutting Edge Applications for Risk Management, Portfolio Optimization and Economics*: Palgrave Macmillan UK.
128. Dutt, L. S., and Kurian, M. (2013). Handling of uncertainty-A Survey. *International Journal of Scientific and Research Publications, 3*(1), 1-4.
129. Easley, D., and Kleinberg, J. (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*: Cambridge University Press.
130. Eberlein, E., and Kallsen, J. (2019). Optimal Investment. In *Mathematical Finance* (pp. 461-535). Cham: Springer International Publishing.
131. Eftekhary, M., Gholami, P., Safari, S., and Shojaee, M. (2012). Ranking normalization methods for improving the accuracy of SVM algorithm by DEA method. *Modern Applied Science, 6*(10), 26.
132. Ehrgott, M., Holder, A., and Nohadani, O. (2018). Uncertain Data Envelopment Analysis. *European Journal of Operational Research, 268*(1), 231-242. doi: https://doi.org/10.1016/j.ejor.2018.01.005
133. Elder, J. (2004). Another Perspective on the Effects of Inflation Uncertainty. *Journal of Money, Credit and Banking, 36*(5), 911-928.
134. Ellison, G., and Glaeser, E. (1997). *Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach* (Vol. 105).
135. Emrouznejad, A. (2006). *Back-propagation DEA*.
136. Emrouznejad, A., and Anouze, A. (2010). *Data envelopment analysis with classification and regression tree - A case of banking efficiency* (Vol. 27).

137. Emrouznejad, A., and Shale, E. (2009). A combined neural network and DEA for measuring efficiency of large scale datasets. *Computers & Industrial Engineering, 56*(1), 249-254.
138. Emrouznejad, A., and Tavana, M. (2013). *Performance Measurement with Fuzzy Data Envelopment Analysis*: Springer Berlin Heidelberg.
139. Engel, J. S., and del-Palacio, I. (2009). Global networks of clusters of innovation: Accelerating the innovation process. *Business Horizons, 52*(5), 493-503.
140. Ericsson, R. N., Jansen, E., Kerbeshian, N., and N., R. (1999). *Interpreting a Monetary Conditions Index in Economic Policy*.
141. Erkan, B., and Yildirimci, E. (2015). Economic Complexity and Export Competitiveness: The Case of Turkey. *Procedia-Social and Behavioral Sciences, 195*, 524-533.
142. Ernst, E., and Viegelahn, C. (2014). *Hiring uncertainty: a new labour market indicator*.
143. Evans, G. W., and Honkapohja, S. (2012). *Learning and expectations in macroeconomics*: Princeton University Press.
144. Everitt, B., and Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*: Springer New York.
145. Fama, E. F., and French, K. R. (2002). The equity premium. *The Journal of Finance, 57*(2), 637-659.
146. Fan, A., and Palaniswami, M. (2000). *Selecting bankruptcy predictors using a support vector machine approach.* Paper presented at the Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium.
147. Färe, R., and Grosskopf, S. (2012). *Intertemporal production frontiers: with dynamic DEA*: Springer Science & Business Media.
148. Färe, R., Grosskopf, S., and Lovell, C. A. K. (1994). *Production Frontiers*: Cambridge University Press.
149. Färe, R., Grosskopf, S., and Lovell, C. A. K. (2013). *The Measurement of Efficiency of Production*: Springer Netherlands.
150. Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General), 120*(3), 253-290. doi:10.2307/2343100
151. Faust, J., and Wright, J. H. (2007). Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset. *National Bureau of Economic Research Working Paper Series, No. 13397*. doi:10.3386/w13397
152. Fethi, M. D., and Pasiouras, F. (2010). Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. *European Journal of Operational Research, 204*(2), 189-198. doi:https://doi.org/10.1016/j.ejor.2009.08.003
153. Fildes, R., and Stekler, H. (2002). The state of macroeconomic forecasting. *Journal of Macroeconomics, 24*(4), 435-468.
154. Firoozye, N., and Ariff, F. (2016). *Managing Uncertainty, Mitigating Risk: Tackling the Unknown in Financial Risk Assessment and Decision Making*: Palgrave Macmillan UK.

155. Fischer, S. (1993). The role of macroeconomic factors in growth. *Journal of Monetary Economics, 32*(3), 485-512.
156. Fountas, S. (2010). Inflation, inflation uncertainty and growth: Are they related? *Economic Modelling, 27*(5), 896-899.
157. Franceschini, F., Galetto, M., and Maisano, D. (2007). *Management by Measurement: Designing Key Indicators and Performance Measurement Systems*: Springer Berlin Heidelberg.
158. Freund, A. (2018). Automated, Decentralized Trust: A Path to Financial Inclusion. In D. Lee Kuo Chuen & R. Deng (Eds.), *Handbook of Blockchain, Digital Finance, and Inclusion, Volume 1* (pp. 431-450): Academic Press.
159. Fried, H. O., Lovell, C. A. K., and Schmidt, S. S. (2008). *The Measurement of Productive Efficiency and Productivity Growth*: Oxford University Press.
160. Gabbay, D. M., and Smets, P. (2013). *Quantified Representation of Uncertainty and Imprecision* (Vol. 1): Springer Science & Business Media.
161. Galati, G., and Moessner, R. (2013). Macroprudential policy–a literature review. *Journal of Economic Surveys, 27*(5), 846-878.
162. García, V., Marqués, A. I., and Sánchez, J. S. (2012). *Improving risk predictions by preprocessing imbalanced credit data.* Paper presented at the International Conference on Neural Information Processing.
163. Gareth, J., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112): Springer.
164. Gattoufi, S., Oral, M., and Reisman, A. (2004). *A taxonomy for data envelopment analysis* (Vol. 38).
165. Gauch, S., and Blind, K. (2014). Technological convergence and the absorptive capacity of standardisation. *Technological Forecasting and Social Change, 91*. doi:10.1016/j.techfore.2014.02.022
166. Gelman, A., and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*: Cambridge University Press.
167. Geum, Y.-J., Kim, C., Lee, S., and Kim, M.-S. (2012). Technological Convergence of IT and BT: Evidence from Patent Analysis. *ETRI Journal, 34*. doi:10.4218/etrij.12.1711.0010
168. Gilboa, I., Postlewaite, A., and Schmeidler, D. (2012). Rationality of belief or: why savage's axioms are neither necessary nor sufficient for rationality. *Synthese, 187*(1), 11-31.
169. Gilboa, I., and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of mathematical economics, 18*(2), 141-153.
170. Giordani, P., and Söderlind, P. (2003). Inflation forecast uncertainty. *European Economic Review, 47*(6), 1037-1059.
171. Glass, K., and Fritsche, U. (2015). *Real-time Macroeconomic Data and Uncertainty*. Retrieved from https://EconPapers.repec.org/RePEc:hep:macppr:201406
172. Golany, B., and Roll, Y. (1989). An application procedure for DEA. *Omega, 17*(3), 237-250.

173. Gollier, C. (2018). *The economics of risk and uncertainty*: Edward Elgar Publishing Limited.
174. Gonzalez-Soto, M., Sucar, L. E., and Escalante, H. J. (2019). von Neumann-Morgenstern and Savage Theorems for Causal Decision Making. *arXiv preprint arXiv:1907.11752*.
175. Grąbczewski, K. (2013). *Meta-Learning in Decision Tree Induction*: Springer International Publishing.
176. Graupe, D. (2013). *Principles of Artificial Neural Networks*: World Scientific.
177. Gray, R. M. (2013). *Entropy and Information Theory*: Springer New York.
178. Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., and Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337-350. doi:10.1007/s10654-016-0149-3
179. Gu, Q., and Han, J. (2013). *Clustered support vector machines*.
180. Guiso, L., Sapienza, P., and Zingales, L. (2018). Time varying risk aversion. *Journal of Financial Economics*.
181. Guvenen, F. (2011). *Macroeconomics with heterogeneity: A practical guide*. Retrieved from
182. Hackeling, G. (2017). *Mastering Machine Learning with scikit-learn*: Packt Publishing.
183. Haddow, A., Hare, C., Hooley, J., and Shakir, T. (2013). Macroeconomic uncertainty: what is it, how can we measure it and why does it matter? *Bank of England Quarterly Bulletin, 53*(2), 100-109.
184. Hahn, F., and Huang, P. (1999). *Markets, Information and Uncertainty: Essays in Economic Theory in Honor of Kenneth J. Arrow*: Cambridge University Press.
185. Han, J., Pei, J., and Kamber, M. (2011). *Data Mining: Concepts and Techniques*: Elsevier Science.
186. Hansen, L., and Singleton, K. (1983). Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy, 91*(2), 249-265.
187. Hansen, L. P. (2013). Challenges in identifying and measuring systemic risk. In *Risk topography: Systemic risk and macro modeling* (pp. 15-30): University of Chicago Press.
188. Hansen, L. P., and Sargent, T. J. (2001). Robust control and model uncertainty. *American Economic Review, 91*(2), 60-66.
189. Hansson, P., Juslin, P., and Winman, A. (2008). *The Role of Short-Term Memory Capacity and Task Experience for Overconfidence in Judgment Under Uncertainty* (Vol. 34).
190. Härdle, W., Moro, R., and Schäfer, D. (2006). Bankruptcy Analysis with Support Vector Machines. *SSRN Electronic Journal*. doi:10.2139/ssrn.912511
191. Harrison, G. W., and Rutström, E. E. (2009). Expected utility theory and prospect theory: One wedding and a decent funeral. *Experimental economics, 12*(2), 133.
192. Hartley, J. E. (1996). Retrospectives: The Origins of the Representative Agent. *Journal of Economic perspectives, 10*(2), 169-177. doi:doi: 10.1257/jep.10.2.169

193. Hartley, J. E. (2002). *The representative agent in macroeconomics*: Routledge.
194. Hartmann, M., and Herwartz, H. (2012). Causal relations between inflation and inflation uncertainty—Cross sectional evidence in favour of the Friedman–Ball hypothesis. *Economics Letters, 115*(2), 144-147.
195. Hatami-Marbini, A., Emrouznejad, A., and Tavana, M. (2011). *A Taxonomy and Review of the Fuzzy Data Envelopment Analysis Literature: Two Decades in the Making* (Vol. 214).
196. He, H., and Garcia, E. A. (2008). Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*(9), 1263-1284.
197. Heathfield, D. F. (1995). Production frontiers: Rolf Fare, Shawna Grosskopf and C.A. Knox Lovell, (Camridge University Press, Cambridge, 1994) pp. xv + 296, L. 30, ISBN 0 521 42033 4. *Structural Change and Economic Dynamics, 6*(4), 508-510.
198. Helbekkmo, H., Kshirsagar, A., Schlosser, A., Selandari, F., Stegemann, U., and Vorholt, J. (2013). Enterprise Risk Management—Shaping the Risk Revolution. *New York: McKinsey & Co.*
199. Henry, Ó., Olekalns, N., and Suardi, S. (2007). Testing for rate dependence and asymmetry in inflation uncertainty: Evidence from the G7 economies. *Economics Letters, 94*(3), 383-388.
200. Heylighen, F. (2008). *Complexity and Self-organization*.
201. Hidalgo, C., and Hausmann, R. (2009). *The Building Blocks of Economic Complexity* (Vol. 106).
202. Hirshleifer, J. (1965). Investment Decision under Uncertainty: Choice—Theoretic Approaches. *The Quarterly Journal of Economics, 79*(4), 509-536. doi:10.2307/1880650
203. Højsgaard, S., Edwards, D., and Lauritzen, S. (2012). *Graphical Models with R*: Springer New York.
204. Hooker, M. A. (2002). Are Oil Shocks Inflationary? Asymmetric and Nonlinear Specifications versus Changes in Regime. *Journal of Money, Credit and Banking, 34*(2), 540-561.
205. Huang, W., Nakamori, Y., and Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research, 32*(10), 2513-2522.
206. Huynh, V. N., Kreinovich, V., Sriboonchitta, S., and Suriya, K. (2012). *Uncertainty Analysis in Econometrics with Applications*: Springer Berlin Heidelberg.
207. Hwang, S. N., Lee, H. S., and Zhu, J. (2016). *Handbook of Operations Analytics Using Data Envelopment Analysis*: Springer US.
208. Inigo, E., and Albareda, L. (2016). *Understanding sustainable innovation as a complex adaptive system: A systemic approach to the firm* (Vol. 126).
209. Inigo, E., Albareda, L., and Ritala, P. (2017). Business model innovation for sustainability: exploring evolutionary and radical approaches through dynamic capabilities. 1-28. doi:10.1080/13662716.2017.1310034
210. Istrefi, K., and Piloiu, A. (2014). *Economic Policy Uncertainty and Inflation Expectations*. Retrieved from https://EconPapers.repec.org/RePEc:bfr:banfra:511

211. Jaccard, J., Sage Publications, i., and Jaccard, J. (2001). *Interaction Effects in Logistic Regression*: SAGE Publications.
212. Jeyakumar, V., Li, G., and Suthaharan, S. (2014). Support vector machine classifiers with uncertain knowledge sets via robust optimization. *Optimization, 63*(7), 1099-1116. doi:10.1080/02331934.2012.703667
213. Jo, S., and Sekkel, R. (2016). *Macroeconomic Uncertainty Through the Lens of Professional Forecasters*. Retrieved from https://EconPapers.repec.org/RePEc:bca:bocawp:16-5
214. Jofré, A., Rockafellar, R. T., and Wets, R. J. (2007). Variational inequalities and economic equilibrium. *Mathematics of Operations Research, 32*(1), 32-50.
215. Jones, B. D. (2003). *Bounded Rationality* (Vol. 2).
216. Jordà, Ò., Knüppel, M., and Marcellino, M. (2013). Empirical simultaneous prediction regions for path-forecasts. *International Journal of Forecasting, 29*(3), 456-468. doi: https://doi.org/10.1016/j.ijforecast.2012.12.002
217. Jorion, P. (2000). Value at risk.
218. Joro, T., and Korhonen, P. J. (2015). *Extension of Data Envelopment Analysis with Preference Information: Value Efficiency*: Springer US.
219. Jurado, K., Ludvigson, S. C., and Ng, S. (2013). Measuring Uncertainty. *National Bureau of Economic Research Working Paper Series, No. 19456*. doi:10.3386/w19456
220. Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *American Economic Review, 105*(3), 1177-1216.
221. Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R*: Springer Berlin Heidelberg.
222. Kahneman, D., and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-291. doi:10.2307/1914185
223. Kao, C. (2016). *Network Data Envelopment Analysis: Foundations and Extensions*: Springer International Publishing.
224. Kao, H.-Y., Chang, T.-K., and Chang, Y.-C. (2013). *Classification of Hospital Web Security Efficiency Using Data Envelopment Analysis and Support Vector Machine* (Vol. 2013).
225. Kara, Y., Boyacioglu, M., and Baykan, O. (2011). *Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange* (Vol. 38).
226. Karaa, A., and Krichene, A. (2012). Credit-risk assessment using support vectors machine and multilayer neural network models: a comparative study case of a Tunisian bank. *Accounting and Management Information Systems, 11*(4), 587.
227. Karadgi, S. (2014). *A Reference Architecture for Real-Time Performance Measurement: An Approach to Monitor and Control Manufacturing Processes*: Springer International Publishing.
228. Karian, Z. A., and Dudewicz, E. J. (2016). *Handbook of Fitting Statistical Distributions with R*: CRC Press.
229. Kassambara, A. (2013). *ggplot2: Guide to Create Beautiful Graphics in R*: Alboukadel KASSAMBARA.

230. Kassambara, A. (2017a). *Network Analysis and Visualization in R: Quick Start Guide*: CreateSpace Independent Publishing Platform.
231. Kassambara, A. (2017b). *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning (Part I)*: CreateSpace Independent Publishing Platform.
232. Kelleher, J. D., Namee, B. M., and D'Arcy, A. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*: MIT Press.
233. Kelly, M. G. (1998). *Tackling change and uncertainty in credit scoring.* The Open University,
234. Khan, S. H., Hayat, M., Bennamoun, M., Sohel, F. A., and Togneri, R. (2018). Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE transactions on neural networks and learning systems, 29*(8), 3573-3587.
235. Khandani, A. E., Kim, A. J., and Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance, 34*(11), 2767-2787.
236. Khemakhem, S., and Boujelbene, Y. (2017). Artificial Intelligence for Credit Risk Assessment: Artificial Neural Network and Support Vector Machines. *Oxford Journal of Finance and Risk Perspectives, 6.2*.
237. Khetrapal, P., and Thakur, T. (2014). A review of benchmarking approaches for productivity and efficiency measurement in electricity distribution sector. *International Journal of Electronics and Electrical Engineering, 2*(3), 214-221.
238. Khezrimotlagh, D., and Chen, Y. (2018). *Decision Making and Performance Evaluation Using Data Envelopment Analysis*: Springer International Publishing.
239. Kim, H.-C., Pang, S., Je, H.-M., Kim, D., and Bang, S.-Y. (2002). *Support vector machine ensemble with bagging.* Paper presented at the International Workshop on Support Vector Machines.
240. Kim, H.-C., Pang, S., Je, H.-M., Kim, D., and Bang, S. Y. (2003). Constructing support vector machine ensemble. *Pattern recognition, 36*(12), 2757-2767.
241. Kim, K.-j. (2003). Financial time series forecasting using support vector machines. *Neurocomputing, 55*(1-2), 307-319.
242. Kirman, A. (2004). *Economics and complexity* (Vol. 07).
243. Kirman, A. (2010). *Complex Economics: Individual and Collective Rationality*: Taylor & Francis.
244. Kirman, A., and Zimmermann, J. B. (2012). *Economics with Heterogeneous Interacting Agents*: Springer Berlin Heidelberg.
245. Kjellberg, D., and Post, E. (2007). *A Critical Look at Measures of Macroeconomic Uncertainty*. Retrieved from https://ideas.repec.org/p/hhs/uunewp/2007_014.html
246. Kleiber, C., and Zeileis, A. (2008). *Applied Econometrics with R*: Springer New York.
247. Knight, F. H. (2012). *Risk, Uncertainty and Profit*: Dover Publications.
248. Knüppel, M. (2014). Efficient estimation of forecast uncertainty based on recent forecast errors. *International Journal of Forecasting, 30*(2), 257-267. doi:https://doi.org/10.1016/j.ijforecast.2013.08.004
249. Kolaczyk, E. D., and Csárdi, G. (2014). *Statistical Analysis of Network Data with R*: Springer New York.

250. Konnov, I. (2007). *Equilibrium Models and Variational Inequalities*: Elsevier Science.
251. Kontonikas, A. (2004). Inflation and inflation uncertainty in the United Kingdom, evidence from GARCH modelling. *Economic Modelling, 21*.
252. Koopmans, T. (1952). Activity analysis of production and allocation. *The Economic Journal, 62*.
253. Koopmans, T. (1965). *On the concept of optimal economic growth*. Retrieved from
254. Kornilov, S., and Polajeva, T. (2016). Economic Complexity: The Future of the Knowledge-Based Regional Development. *SGEM2016 Conference Proceedings, 3*. doi:10.5593/SGEMSOCIAL2016/B13/S03.032
255. Kornilov, S., Ridala, S., and Aasma, A. (2016). Dynamics of Economic Adjustment Under Uncertainty: MS-VAR Model for Baltic Dry Index. *3rd International Multidisciplinary Scientific Conference on Social Sciences and Arts SGEM, 5*, 189-196. doi:10.5593/SGEMSOCIAL2016/B25/S07.025
256. Kowalczyk, H. (2013). *Inflation Fan Charts and Different Dimensions of Uncertainty: What if Macroeconomic Uncertainty is High?*
257. Kowalczyk, H., Lyziak, T., and Stanisławska, E. (2013). *A new approach to probabilistic surveys of professional forecasters and its application in the monetary policy context*. Retrieved from https://ideas.repec.org/p/nbp/nbpmis/142.html
258. Krayer von Krauss, M. P., Walker, W., van der Sluijs, J. P., Janssen, P., van Asselt, M. B. A., and Rotmans, J. (2019). *Response to "To what extent, and how, might uncertainty be defined" by Norton, Brown, and Mysiak*.
259. Kreienkamp, T., and Kateshov, A. (2014). Credit risk modeling: combining classification and regression algorithms to predict expected loss. *Journal of Corporate Finance Research*(4 (32)), 4-10.
260. Kreps, D. (2018). *Notes on the Theory of Choice*: content.taylorfrancis.com.
261. Kreps, D. M., and Porteus, E. L. (1979). Dynamic choice theory and dynamic programming. *Econometrica: Journal of the Econometric Society*, 91-100.
262. Krüger, F., Clark, T. E., and Ravazzolo, F. (2017). Using Entropic Tilting to Combine BVAR Forecasts With External Nowcasts. *Journal of Business & Economic Statistics, 35*(3), 470-485. doi:10.1080/07350015.2015.1087856
263. Kruppa, J., Ziegler, A., and König, I. R. (2012). Risk estimation and risk prediction using machine-learning methods. *Human genetics, 131*(10), 1639-1654.
264. Kuhn, M. (2008). *Building Predictive Models in R Using the caret Package* (Vol. 28).
265. Kuhn, M., and Johnson, K. (2013). *Applied Predictive Modeling*: Springer New York.
266. Lahiri, K., Peng, H., and Sheng, X. (2015). *Measuring Uncertainty of a Combined Forecast and Some Tests for Forecaster Heterogeneity*. Retrieved from https://ideas.repec.org/p/ces/ceswps/_5468.html
267. Lahiri, K., and Sheng, X. (2010). Measuring forecast uncertainty by disagreement: The missing link. *Journal of Applied Econometrics, 25*(4), 514-538.
268. Lane, D., and Maxfield, R. (2004). *Ontological uncertainty and innovation* (Vol. 15).
269. Le Bellac, M., and Viricel, A. (2017). *Deep Dive Into Financial Models: Modeling Risk and Uncertainty*: World Scientific.

270. Lee, E. A., and Seshia, S. A. (2016). *Introduction to Embedded Systems: A Cyber-Physical Systems Approach*: MIT Press.
271. Leipzig, J., and Li, X. Y. (2011). *Data Mashups in R: A Case Study in Real-World Data Analysis*: O'Reilly Media.
272. Lensink, R., Bo, H., and Sterken, E. (1999). Does uncertainty affect economic growth? An empirical analysis. *Weltwirtschaftliches Archiv, 135*(3), 379-396.
273. Lertworasirikul, S., Fang, S. C., Nuttle, H. L. W., and Joines, J. (2002). *Fuzzy data envelopment analysis*.
274. Levin, A. T., Onatski, A., Williams, J. C., and Williams, N. (2005). Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models. *National Bureau of Economic Research Working Paper Series, No. 11523*. doi:10.3386/w11523
275. Lewis, H. F., and Sexton, T. R. (2004). Network DEA: efficiency analysis of organizations with complex internal structure. *Computers & Operations Research, 31*(9), 1365-1410.
276. Lewis, N. D. (2015). *92 Applied Predictive Modeling Techniques in R: With Step by Step Instructions on How to Build Them Fast!* : CreateSpace Independent Publishing Platform.
277. Li, S., Tsang, I. W., and Chaudhari, N. S. (2012). Relevance vector machine based infinite decision agent ensemble learning for credit risk analysis. *Expert Systems with Applications, 39*(5), 4947-4953.
278. Li, Z., Crook, J., and Andreeva, G. (2017). Dynamic prediction of financial distress using Malmquist DEA. *Expert Systems with Applications, 80*, 94-106.
279. Liaw, A., and Wiener, M. (2001). *Classification and Regression by RandomForest* (Vol. 23).
280. Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association, 83*(404), 1198-1202.
281. Liu, J. S., Lu, L. Y. Y., Lu, W.-M., and Lin, B. J. Y. (2013). Data envelopment analysis 1978–2010: A citation-based literature survey. *Omega, 41*(1), 3-15. doi:https://doi.org/10.1016/j.omega.2010.12.006
282. Ljungqvist, L., and Sargent, T. J. (2012). *Recursive Macroeconomic Theory*: MIT Press.
283. Loasby, B. (1999). *Institutions and Evolution in Economics*.
284. Lovell, C. K., Pastor, J. T., and Turner, J. A. (1995). Measuring macroeconomic performance in the OECD: A comparison of European and non-European countries. *European Journal of Operational Research, 87*(3), 507-518.
285. Ma, Y., and Guo, G. (2014). *Support Vector Machines Applications*: Springer International Publishing.
286. Macal, C., and North, M. (2010). *Tutorial on agent-based modelling and simulation* (Vol. 4).
287. Malinvaud, E. (1953). Capital accumulation and efficient allocation of resources. *Econometrica, Journal of the Econometric Society*, 233-268.
288. Mankiw, N. G. (2008). *Principles of Economics*: Cengage Learning.

289. Mankiw, N. G., Reis, R., and Wolfers, J. (2003). Disagreement about Inflation Expectations. *National Bureau of Economic Research Working Paper Series, No. 9796*. doi:10.3386/w9796
290. Markowitz, H. M., Todd, G. P., and Sharpe, W. F. (2000). *Mean-Variance Analysis in Portfolio Choice and Capital Markets*: Wiley.
291. Marsland, S. (2011). *Machine Learning: An Algorithmic Perspective*: CRC Press.
292. Mårtensson, M. (2000). *A Critical Review of Knowledge Management as a Management Tool* (Vol. 4).
293. Martin, R., and Sunley, P. (2011). *Conceptualizing Cluster Evolution: Beyond the Life Cycle Model?* (Vol. 45).
294. Marwala, T., and Hurwitz, E. (2017). *Artificial Intelligence and Economic Theory: Skynet in the Market*: Springer International Publishing.
295. Mason, L., Baxter, J., Bartlett, P. L., and Frean, M. R. (2000). *Boosting algorithms as gradient descent.* Paper presented at the Advances in neural information processing systems.
296. Massari, M., Gianfrate, G., and Zanetti, L. (2016). *Corporate Valuation: Measuring the Value of Companies in Turbulent Times*: Wiley.
297. Masuch, M., and Huang, Z. (1996). *A case study in logical deconstruction: Formalizing J.D. Thompson's Organizations in Action in a multi-agent action logic* (Vol. 2).
298. Matsushima, H. (1997). *Bounded Rationality in Economics: A Game Theorist's View* (Vol. 48).
299. Meinen, P., and Röhe, O. (2017). On measuring uncertainty and its impact on investment: Cross-country evidence from the euro area. *European Economic Review, 92*, 161-179.
300. Miller, J. D., and Forte, R. M. (2017). *Mastering Predictive Analytics with R*: Packt Publishing.
301. Min, J.-H., and Lee, Y.-C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications, 28*(4), 603-614.
302. Mittal, S., Mirza, S., and Zaman, M. (2016). *A Review of Data Mining Literature* (Vol. 14).
303. Monti, F. (2010). Combining Judgment and Models. *Journal of Money, Credit and Banking, 42*(8), 1641-1662. doi:10.1111/j.1538-4616.2010.00357.x
304. Mosini, V. (2008). *Equilibrium in Economics: Scope and Limits*: Taylor & Francis.
305. Mullainathan, S., and Spiess, J. (2017). *Machine Learning: An Applied Econometric Approach* (Vol. 31).
306. Müller, V. C. (2016). *Fundamental Issues of Artificial Intelligence*: Springer International Publishing.
307. Murrell, P. (2005). *R Graphics*: CRC Press.
308. Murty, M. N., and Raghava, R. (2016). *Support Vector Machines and Perceptrons: Learning, Optimization, Classification, and Application to Social Networks*: Springer International Publishing.

309. Nava, N., Di Matteo, T., and Aste, T. (2018). Financial time series forecasting using empirical mode decomposition and support vector regression. *Risks, 6*(1), 7.
310. Neanidis, K., and Savva, C. (2011). Nominal uncertainty and inflation: The role of European Union membership. *Economics Letters, 112*(1), 26-30.
311. Newman, M. (2010). *Networks: An Introduction*: OUP Oxford.
312. Niaf, E., Flamary, R., Lartizien, C., and Canu, S. (2011). *Handling uncertainties in SVM classification*.
313. Nishimura, K. G., and Ozaki, H. (2017). Decision-Theoretic Foundations of Knightian Uncertainty. In *Economics of Pessimism and Optimism* (pp. 51-75): Springer.
314. Niu, D.-X., Wang, X.-M., and Gao, C. (2008). *The combination evaluation method based on DEA and SVM*. Paper presented at the 2008 International Conference on Machine Learning and Cybernetics.
315. Norton, J. P. B., J. B.; Mysiak, J. (2006). To what extent, and how, might uncertainty be defined? Comments engendered by 'Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support: Walker et al.'. *The Integrated AssessmentJournal, 6*(1).
316. O'Neil, C., and Schutt, R. (2013). *Doing Data Science: Straight Talk from the Frontline*: O'Reilly Media.
317. Oberkampf, W., Helton, J., and Sentz, K. (2001). *Mathematical representation of uncertainty*.
318. Onatski, A., and Williams, N. (2003). Modeling model uncertainty. *Journal of the European Economic Association, 1*(5), 1087-1122.
319. Onozaki, T. (2018). *Nonlinearity, Bounded Rationality, and Heterogeneity: Some Aspects of Market Economies as Complex Systems*: Springer Japan.
320. Orlik, A., and Veldkamp, L. (2014). Understanding Uncertainty Shocks and the Role of Black Swans. *National Bureau of Economic Research Working Paper Series, No. 20445*. doi:10.3386/w20445
321. Ouellette, P., and Yan, L. (2008). Investment and dynamic DEA. *Journal of Productivity Analysis, 29*(3), 235-247.
322. Paradi, J. C., Sherman, H. D., and Tam, F. K. (2017). *Data Envelopment Analysis in the Financial Services Industry: A Guide for Practitioners and Analysts Working in Operations Research Using DEA*: Springer International Publishing.
323. Pareek, M. (2006). Optimizing controls to test as part of a risk-based audit strategy. *Information Systems Control Journal, 2*, 39-41.
324. Pfaff, B. (2008). *Analysis of Integrated and Cointegrated Time Series with R*: Springer New York.
325. Pfaff, B. (2016). *Financial Risk Modelling and Portfolio Optimization with R*: Wiley.
326. Porter, M. E. (2008). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*: Free Press.
327. Pozzi, L. (2005). Income uncertainty and aggregate consumption. *National Bank of Belgium Working Paper*(77).
328. Praveena, M., and Jaiganesh, V. (2017). *A Literature Review on Supervised Machine Learning Algorithms and Boosting Process* (Vol. 169).

329. Preiser, R., Biggs, R., De Vos, A., and C., F. (2018). *Social-ecological systems as complex adaptive systems: organizing principles for advancing research methods and approaches*: ecologyandsociety.org.
330. Provost, F., and Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*: O'Reilly Media.
331. Puu, T. (2010). Nonlinear economic dynamics and global analysis. In (pp. 3-12).
332. Pyka, A., and Hanusch, H. (2007). *The Elgar-Companion to Neo-Schumpeterian Economics*.
333. Rachev, S. T., Stoyanov, S. V., and Fabozzi, F. J. (2008). *Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measures*: Wiley.
334. Rakotomamonjy, A. (2004). Support vector machines and area under ROC curve. *PSI-INSA de Rouen: Technical Report*.
335. Ramsey, F. P. (1928). A mathematical theory of saving. *The Economic Journal, 38*(152), 543-559.
336. Rappaport, A. (1986). *Creating shareholder value: the new standard for business performance*: Free Press.
337. Ravi Kumar, P., and Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review. *European Journal of Operational Research, 180*(1), 1-28.
338. Rawley, T., and Benton, G. (2009). *The Valuation Handbook, (Custom Chapter 14): Valuation Techniques from Today's Top Practitioners*: John Wiley & Sons.
339. Rich, R., and Butler, J. (1998). Disagreement as a Measure of Uncertainty: A Comment on Bomberger. *Journal of Money, Credit and Banking, 30*(3), 411-419.
340. Rich, R., and Tracy, J. (2010). The Relationships among Expected Inflation, Disagreement, and Uncertainty: Evidence from Matched Point and Density Forecasts. *The Review of Economics and Statistics, 92*(1), 200-207.
341. Richardson, G. B. (1972). The Organisation of Industry. *Economic Journal, 82*(327), 883-896.
342. Rigby, D., and Bilodeau, B. (2011). Management tools & trends 2013. *London: Bain & Company*.
343. Romer, D. (2018). *Advanced Macroeconomics*: McGraw-Hill Education.
344. Röthig, A. (2009). *Microeconomic Risk Management and Macroeconomic Stability*: Springer Berlin Heidelberg.
345. Ruggiero, J. (1998). Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research, 111*(3), 461-469.
346. Sakouvogui, K. (2019). *Banks performance evaluation: A hybrid DEA-SVM- The case of U.S. agricultural banks*.
347. Samoilenko, S., and Osei-Bryson, K.-M. (2008). *Increasing the discriminatory power of DEA in the presence of the sample heterogeneity with cluster analysis and decision trees* (Vol. 34).
348. Samuelson, P. A. (1958). An exact consumption-loan model of interest with or without the social contrivance of money. *Journal of Political Economy, 66*(6), 467-482.

349. Samuelson, P. A. (1983). *Foundations of Economic Analysis*: Harvard University Press.
350. Samuelson, P. A. (2010). *Economics*: Tata McGraw Hill.
351. Sarel, M. (1996). Nonlinear Effects of Inflation on Economic Growth. *IMF Staff Papers, 43*(1), 199-215.
352. Saunders, A., Cornett, M. M., and McGraw, P. A. (2006). *Financial Institutions Management : a Risk Management Approach*: McGraw-Hill Ryerson, Limited.
353. Schapire, R. E., Freund, Y., Bartlett, P., and Lee, W. S. (1998). Boosting the margin: A new explanation for the effectiveness of voting methods. *The annals of statistics, 26*(5), 1651-1686.
354. Schilirò, D. (2012). *Bounded rationality: psychology, economics and the financial crisis*. Retrieved from https://ideas.repec.org/p/pra/mprapa/40280.html
355. Scholkopf, B., and Smola, A. J. (2001). *Learning with kernels: support vector machines, regularization, optimization, and beyond*: MIT press.
356. Schorfheide, F. (2011). *Estimation and evaluation of DSGE models: progress and challenges*. Retrieved from
357. Schwab, K. (2016). *The fourth industrial revolution*: World Economic Forum.
358. Segal, U. (1987). The Ellsberg paradox and risk aversion: An anticipated utility approach. *International Economic Review*, 175-202.
359. Seiford, L. (1996). *Data Envelopment Analysis: The Evolution of the State of the Art (1978-1995)* (Vol. 7).
360. Seiford, L. (1997). A bibliography for data envelopment analysis (1978-1996). *Annals of Operations Research, 73*, 393-438.
361. Seiford, L. (2005). A Cyber-Bibliography for Data Envelopment Analysis (1978-2005). *University of Michigan, Michigan*.
362. Seol, H., Choi, J., Park, G., and Park, Y. (2007). *A framework for benchmarking service process using data envelopment analysis and decision tree* (Vol. 32).
363. Sexton, T., H. Silkman, R., and J. Hogan, A. (1986). *Data Envelopment Analysis: Critique and Extensions* (Vol. 1986).
364. Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal, 27*(3), 379-423. doi:10.1002/j.1538-7305.1948.tb01338.x
365. Shawe-Taylor, J., Bartlett, P. L., Williamson, R. C., and Anthony, M. (1998). Structural risk minimization over data-dependent hierarchies. *IEEE transactions on Information Theory, 44*(5), 1926-1940.
366. Shin, K.-S., Lee, T. S., and Kim, H.-j. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications, 28*(1), 127-135.
367. Shumway, R. H., and Stoffer, D. S. (2014). *Time Series Analysis and Its Applications*: Springer.
368. Siguenza-Guzman, L., Saquicela, V., Avila-Ordoñez, E., Vandewalle, J., and Cattrysse, D. (2015). *Literature Review of Data Mining Applications in Academic Libraries* (Vol. 41).

369. Siklos, P. L. (2013). Sources of disagreement in inflation forecasts: An international empirical investigation. *Journal of International Economics, 90*(1), 218-231. doi:10.1016/j.jinteco.2012.11
370. Simon, H. (1997). *Models of Bounded Rationality: Empirically grounded economic reason*: MIT Press.
371. Simon, H., Egidi, M., Viale, R., and Marris, R. (1992). *Economics, Bounded Rationality and the Cognitive Revolution*: Edward Elgar Publishing.
372. Smith, P. (1997). Model misspecification in data envelopment analysis. *Annals of Operations Research, 73*, 233-252.
373. Smola, A. J., and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing, 14*(3), 199-222.
374. So, E. C. (2013). A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts? *Journal of Financial Economics, 108*(3), 615-640. doi:https://doi.org/10.1016/j.jfineco.2013.02.002
375. Soll, J., and Klayman, J. (2004). *Overconfidence in Interval Estimates* (Vol. 30).
376. Stolp, C. (1990). *Strengths and Weaknesses of Data Envelopment Analysis: An Urban and Regional Perspective* (Vol. 14).
377. Tattar, P. N. (2018). *Hands-On Ensemble Learning with R: A beginner's guide to combining the power of machine learning algorithms using ensemble techniques*: Packt Publishing.
378. Tesfatsion, L. (2003). *Agent-based computational economics: Modelling economies as complex adaptive systems* (Vol. 149).
379. Thanassoulis, E. (1993). A comparison of regression analysis and data envelopment analysis as alternative methods for performance assessments. *Journal of the Operational Research Society, 44*(11), 1129-1144.
380. Thomas, L., Crook, J., and Edelman, D. (2017). *Credit scoring and its applications* (Vol. 2): Siam.
381. Thompson, J. D., Scott, W. R., and Zald, M. N. (2011). *Organizations in Action: Social Science Bases of Administrative Theory*: Transaction Publishers.
382. Tobback, E., Naudts, H., Daelemans, W., de Fortuny, E. J., and Martens, D. (2018). Belgian economic policy uncertainty index: Improvement through text mining. *International Journal of Forecasting, 34*(2), 355-365.
383. Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research, 130*, 498-509. doi:10.1016/S0377-2217(99)00407-5
384. Tone, K., and Tsutsui, M. (2009). Network DEA: A slacks-based measure approach. *European Journal of Operational Research, 197*(1), 243-252.
385. Tone, K., and Tsutsui, M. (2010). Dynamic DEA: A slacks-based measure approach. *Omega, 38*(3-4), 145-156.
386. Torgo, L. (2016). *Data Mining with R: Learning with Case Studies, Second Edition*: CRC Press.
387. Tsang, S., Kao, B., Yip, K. Y., Ho, W.-S., and Lee, S. D. (2011). Decision trees for uncertain data. *IEEE Transactions on Knowledge and Data Engineering, 23*(1), 64-78.

388. Van Gestel, T., Baesens, B., Suykens, J., Espinoza, M., Baestaens, D., Vanthienen, J., and De Moor, B. (2003). *Bankruptcy prediction with least squares support vector machine classifiers*. Paper presented at the 2003 IEEE International Conference on Computational Intelligence for Financial Engineering, 2003. Proceedings.
389. Van Liebergen, B. (2017). Machine learning: A revolution in risk management and compliance? *Journal of Financial Transformation, 45*, 60-67.
390. Vapnik, V. (1992). *Principles of risk minimization for learning theory.* Paper presented at the Advances in neural information processing systems.
391. Vapnik, V. (2013). *The Nature of Statistical Learning Theory*: Springer New York.
392. Viana, M., and Oliveira, K. (2016). *Foundations of Ergodic Theory*: Cambridge University Press.
393. Viebig, J., Poddig, T., and Varmaz, A. (2008). *Equity Valuation: Models from Leading Investment Banks*: Wiley.
394. Vilela, L. F., Leme, R. C., Pinheiro, C. A., and Carpinteiro, O. A. (2018). Forecasting financial series using clustering methods and support vector regression. *Artificial Intelligence Review*, 1-31.
395. von Neumann, J., Morgenstern, O., Kuhn, H. W., and Rubinstein, A. (2007). *Theory of Games and Economic Behavior: 60th Anniversary Commemorative Edition*: Princeton University Press.
396. Wagner, J. M., and Shimshak, D. G. (2007). Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. *European Journal of Operational Research, 180*(1), 57-67.
397. Walker, W., E. Harremoës, P., Rotmans, J., P. van der Sluijs, J., van Asselt, M. B. A., Janssen, P., and Krayer von Kraus, M. P. (2003). *Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support* (Vol. 4).
398. Wang, L. (2005). *Support Vector Machines: Theory and Applications*: Springer Berlin Heidelberg.
399. Welling, M. (2004). Support vector regression. *Department of Computer Science, University of Toronto, Toronto (Kanada)*.
400. Wen, M. (2014). *Uncertain Data Envelopment Analysis*: Springer Berlin Heidelberg.
401. Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.
402. Witzany, J. (2017). *Credit Risk Management: Pricing, Measurement, and Modeling*: Springer International Publishing.
403. Xie, J., Girshick, R., and Farhadi, A. (2016). *Unsupervised deep embedding for clustering analysis.* Paper presented at the International conference on machine learning.
404. Xu, X., and Wang, Y. (2009). Financial failure prediction using efficiency as a predictor. *Expert Systems with Applications, 36*(1), 366-373.
405. Yang, X., and Dimitrov, S. (2017). *Data envelopment analysis may obfuscate corporate financial data: Using support vector machine and data envelopment analysis to predict corporate failure for nonmanufacturing firms* (Vol. 55).
406. Yang, Y. (2016). *Temporal Data Mining via Unsupervised Ensemble Learning*: Elsevier Science.

407. Yeh, C.-C., Chi, D.-J., and Hsu, M.-F. (2010). A hybrid approach of DEA, rough set and support vector machines for business failure prediction. *Expert Systems with Applications, 37*, 1535-1541. doi:10.1016/j.eswa.2009.06.088
408. Young Sohn, S., and Hee Moon, T. (2004). *Decision Tree based on data envelopment analysis for effective technology commercialization* (Vol. 26).
409. Yu, H., and Kim, S. (2012). SVM tutorial—classification, regression and ranking. In *Handbook of Natural computing* (pp. 479-506): Springer.
410. Zarnowitz, V., and Lambros, L. A. (1987). Consensus and Uncertainty in Economic Prediction. *Journal of Political Economy, 95*(3), 591-621.
411. Zelenkov, Y., Fedorova, E., and Chekrizov, D. (2017). Two-step classification method based on genetic algorithm for bankruptcy forecasting. *Expert Systems with Applications, 88*, 393-401.
412. Zhang, M. (2008). *Artificial higher order neural networks for economics and business*: IGI Global.
413. Zhang, Q., and Wang, C. (2018). DEA efficiency prediction based on IG–SVM. *Neural Computing and Applications*, 1-10.
414. Zhou, K., Liu, T., and Zhou, L. (2015). *Industry 4.0: Towards future industrial opportunities and challenges*.
415. Zhou, L., Lai, K. K., and Yen, J. (2014). Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation. *International Journal of Systems Science, 45*(3), 241-253.
416. Zhu, J. (2016a). *Data Envelopment Analysis: A Handbook of Empirical Studies and Applications (Volume 238)*: Springer US.
417. Zhu, J. (2016b). *Data Envelopment Analysis: A Handbook of Empirical Studies and Applications (Volume 221)*: Springer US.
418. Zolbanin, H., and Delen, D. (2018). *The analytics paradigm in business research* (Vol. 90).

APPENDICES

1. MODELS IN LITERATURE BY TECHNIQUE AND ITS APPLICATION

Adopted form Thomas *et al.* (2017)

Modeling technique	Application	Reference
Discriminant analysis	General	Durand (1941), Myers & Forgy (1963), Boggess (1967), Morrison (1969), Lane (1972), Bates (1973), Chang & Affifi (1974), Apilado et al. (1974), Eisenbeis (1977, 1978), Grablowsky & Talley (1981), Taffler (1982), Moses & Liao (1987), Falbo (1991), Overstreet & Bradley (1994), Rosenberg & Geit (1994), Trevino & Daniels (1995), Hand et al. (1996), Lee, Jo & Han (1997), Kim, Kim, Kim, Ye & Lee (2000), Kim & Oommen (2008)
	Financial data	Churchill (1941), Merwin (1947), Myers & Forgy (1963), Hills (1967), Altman (1968, 1988 and 1993), Hoskins (1968), Altman et al. (1974), Altman et al. (1977), Altman & Eisenbeis (1978), Deakin (1972), Edmister (1972), Bates (1973), Apilado et al. (1974), Blum (1974), Eisenbeis (1977), Taffler & Tisshaw (1977), Bidelbeek (1979), Misha (1984), Gombola et al. (1987), Piesse & Wood (1992), Lussier (1995), Altman et al. (1995).
	Credit Scoring	Bardos (1998), Desai et al. (1996), Martell and Fits (1981), Overstreet, Bradley & Kemp (1992), Reichert et al. (1983), Titterington (1992), Lee et al (2002).
Logistic regression	General	Lachenbruch (1975), Orgler (1970), Orgler (1971), Cox (1972), Breslow (1974), Dawes (1974, 1979), Dawid (1976), Fitzpatrick (1976), Wainer (1976, 1978), Gunst & Mason (1977), Laughlin (1978), Darroch et al. (1980), Haggstrom (1983), Garthwaite & Diskey (1988), Lucas (1992), Lai & Ying (1994), Henley (1995), Djebarni & Al-Abed (1998), Flagg, Giroux &Wiggins (1991), Kay, Warde & Martens (2000), Laitinen & Laitinen (2000), Lau (1987), Suh, Noh & Suh (1999), Vellido, Lisboa & Vaughan (1999), Wong, Bodnovich & Selvi (1997), Zavgen (1983), Gentry et al. (1985), Keasey & Watson (1987), Aziz et al. (1988), Cox & Snell (1989), Hosmer & Lemeshow (1989), Platt & Platt (1990), Ooghe et al. (1995), Crook (1996), Mossman et al. (1998), Charitou & Trigeorpis (2002), Becchetti and Sierra (2002).
	Credit Scoring	Banasik (1996), Berkowitz & Hynes (1999), Henley (1995), Joanes (1993), Laitinen (1999), Westgaard & van der Wijst (2001).

Modeling technique	Application	Reference
Probit regression		Badu & Daniels (1997), Badu et al (2002), Boyes et al. (1989), Crook (2001), Banasik, Crook & Thomas (2003); Greene (1998); Guillen & Artis (1992), Loviscek & Crowley (1990), Tsaih et al (2004), Wallace (1978; 1981).
Neural Networks	General	Jacobs (1988), Tang et al. (1991), Kuan & White (1992), Lee et al. (1993), Coats & Fant (1993), Cheng & Titterington (1994), Ripley (1994), Hill et al. (1994), Kuan & Liu (1995), Lachtermacher & Fuller (1995), Drossu & Obradovic (1996), Boussabaine & Duff (1996), Wong et al. (1997), Zhang, Patuwo & Hu (1998), Gruca & Klemz (1998), Vellido et al. (1999), Lau et al. (2001), Tkacz (2001), Papadas & Hutchison (2002), Heravi et al. (2004), Santin et al. (2004), Delgado (2005), Nakamura (2005), Hippert et al. (2005), Longhi et al. (2005), Longhi et al. (2005), Erbas & Stefanou (2008), Anderson & Rosenfeld (1988), Cheng & Titterington (1994), Haykin (1994), Stern (199), Vellido et al. (1999), Zhang, Patuwo, & Hu (1998).
	Credit Scoring	Gallant(1988), Nelson & Illingworth (1990), Eberhart & Dobbins (1990), Kim & Scott (1991), Davis et al. (1992), Jensen (1992), Salchenberger, Cinar & Lash (1992), Tam & Kiang (1992), Deng (1993), Robins (1993), Rosenberg & Gleit (1994), Altman et al. (1994), Kerling & Poddig (1994), Podding (1994), Piramuthu, Shaw & Gentry (1994), Richeson, Zimmermann & Barnett (1994), Borrowsky (1995), Lacher et al. (1995), Williamson (1995), Sharda & Wilson (1996), Torsun (1996), Desai et al. (1996), Glorfeld (1996), Jagielska & Jaworski (1996), Glorfeld & Hardgrave (1996), Hand & Henley (1997), Desai et al. (1997), Arminger, Enache & Bonne (1997), Brill (1998), Piramuthu (1999), Barney, Graves & Johnson (1999), Zhang, Hu, Patuwo, & Indro (1999), Yang et al. (1999), West (2000), Malhotra & Malhotra (2003), Lee et al. (2002), Kim & Sohn (2004), Lee & Chen (2005), Blochlinger & Leippold (2006)
Time varying model	General	Anderson & Goodman (1957), Cyert et al. (1962), Bierman & Hausman (1970), Metha (1970), Dirickx & Wakeman (1976), Long (1976), Corcoran (1978), Van Kuelen et al. (1981), Frydman (1984), Frydman et al. (1985), Srinivasan & Kim (1987b), Edelman (1992), Clemen et al. (1995).
	Credit Scoring	Fix and Hodges (1952), Cover & Hart (1967), Chatterjee & Barcun (1970), Hand (1986), Henley & Hand (1996), Tam & Kiang (1992).

Modeling technique	Application	Reference
Recursive partitioning		Raiffa and Schlaifer (1961), Metha (1968), Sparks (1972), Breiman et al. (1984), Frydman, Altman & Kao (1985), Makowski (1985), Coffman (1986), Carter & Catlett (1987), Safavian & Landgrebe (1991), Boyle et al. (1992), Davis, Edelman & Gammerman (1992), Altman et al. (1994), Zakrzewska (2007).
Mathematical programming	General	Kendall (1966), Rao (1971), Pye & Tezel (1974), Hand (1981), Showers and Chakrin (1981), Kolesar and Showers (1985), Hardy and Adrian (1985), Joachimsthlaer & Stam (1990), Glover (1990), Ziari et al. (1997); Gehrlein and Wagner (1997), Hamsici & Martinez (2008).
	Credit Scoring	Hardy & Adrian (1985) , Gehrlein & Wagner (1997).
Genetic algorithms	General	Efron (1977), Fogarthy & Ireson (1993), Desai et al. (1997), Yobas et al. (2000).
	Credit Scoring	Ong et al. (2005).
Support Vector Machine		Vapnik (1995, 2000), Burges & Schölkopf (1997), Schölkopf et al. (1996, 1998), Vapnik et al. (1997), Joachims (1998), Pontil & Verri (1998), Baudat et al. (2000), Scholkopf & Smola (2000), Cristianini & Shawe-Taylor (2000), Zhang (2000), Kecman (2001), Weston et al. (2001), Guyon et al. (2002), Yu et al. (2003), Frohlich & Chapelle (2003), Gestel et al. (2003), Baesens et al. (2003), Huang et al. (2004), Mao (2004), Schebesch & Stecking (2005), Schebesch (2005), Somol et al. (2005), Lai et al. (2006), Huang et al. (2007), Zhou et al. (2009).
Comparison	Traditionnal vs. Modern ones	Lee & Chen (2005), Lee et al. (2002), Zekic-Suzac et al. (2004), Malhotra & Malhotra (2003), Ong, Huang, & Tzeng (2005), Abdou et al. (2008), Arminger, Enache & Bonne (1997), Gilbert et al. (1990).

2. NORMALIZATION

Vector normalization

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}$$

Linear normalization (I)

$$r_{ij} = \frac{x_{ij}}{x_j^*}$$

$$i = 1,\ldots.m; j = 1,\ldots.n; x_j^* = \max\left\{x_{ij}\right\}$$

Linear normalization (II)

$$r_{ij} = \frac{x_{ij} - x_j^{\sim}}{x_j^* - x_j^{\sim}}$$

$$i = 1,\ldots,m; j = 1,\ldots,n; x_j^{\sim} = \min\{x_{ij}\}$$

Linear normalization (III)

$$r_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}$$

$$i = 1,\ldots,m; j = 1,\ldots,n$$

3. DMU LIST

Download: https://nasdaqbaltic.com/
Verified by May 2020

Nr	Name	Abbr	Market	Stock	Industry	Sector
1	AUGA group	AUG1L	VLN	BALT_M	CON_GOODS	FOOD
2	Amber Grid	AMG1L	VLN	BALT_S	OIL_GAS	OIL_GAS
3	Apranga	APG1L	VLN	BALT_M	CON_SERV	RETAIL
4	Arco Vara	ARC1T	TLN	BALT_M	FIN	REAL_EST
5	Aspo Oyj	ASPO	HEL	NORD	INDUSTR	
6	Atria Oyj A	ATRAV	HEL	NORD	CON_GOODS	
7	Baltika	BLT1T	TLN	BALT_M	CON_GOODS	HOUS_PROD
8	Bittium Oyj	BITTI	HEL	NORD	TECHN	
9	Citycon Oyj	CTY1S	HEL	NORD	FIN	
10	Ditton pievadu rpnca	DPK1R	RIG	BALT_S	INDUSTR	IND_GOODS
11	Ekspress Grupp	EEG1T	TLN	BALT_M	CON_SERV	MEDIA
12	Elisa Oyj	ELISA	HEL	NORD	TELECOMM	
13	F-Secure Oyj	FSC1V	HEL	NORD	TECHN	
14	Fortum Oyj	FORTUM	HEL	NORD	UTIL	
15	Grigeo	GRG1L	VLN	BALT_M	MATERIALS	RESOURCE
16	Grindeks	GRD1R	RIG	BALT_M	HEALTH	HEALTH
17	Harju Elekter	HAE1T	TLN	BALT_M	INDUSTR	IND_GOODS
18	Invalda INVL	IVL1L	VLN	BALT_S	FIN	FIN
19	KONE Oyj	KNEBV	HEL	NORD	INDUSTR	
20	Kauno energija	KNR1L	VLN	BALT_S	UTIL	UTIL
21	Kemira Oyj	KEMIRA	HEL	NORD	MATERIALS	CHEM
22	Kurzemes atslga 1	KA11R	RIG	BALT_S	CON_GOODS	HOUS_PROD
23	LITGRID	LGD1L	VLN	BALT_S	UTIL	UTIL
24	Latvijas Gze	GZE1R	RIG	BALT_S	UTIL	UTIL

Nr	Name	Abbr	Market	Stock	Industry	Sector
25	Latvijas Jras medicnas centrs	LJM1R	RIG	BALT_S	HEALTH	HEALTH
26	Latvijas balzams	BAL1R	RIG	BALT_S	CON_GOODS	FOOD
27	Linas	LNS1L	VLN	BALT_S	CON_GOODS	HOUS_PROD
28	Linas Agro Group	LNA1L	VLN	BALT_M	CON_GOODS	FOOD
29	Merko Ehitus	MRK1T	TLN	BALT_M	INDUSTR	CONSTR
30	Metso Oyj	METSO	HEL	NORD	INDUSTR	
31	Neste Oyj	NESTE	HEL	NORD	OIL_GAS	
32	Nordecon	NCN1T	TLN	BALT_M	INDUSTR	CONSTR
33	Nordic Fibreboard	SKN1T	TLN	BALT_S	CON_GOODS	HOUS_PROD
34	Olainfarm	OLF1R	RIG	BALT_M	HEALTH	HEALTH
35	PATA Saldus	SMA1R	RIG	BALT_S	MATERIALS	RESOURCE
36	Panevio statybos trestas	PTR1L	VLN	BALT_M	INDUSTR	CONSTR
37	Pieno vaigds	PZV1L	VLN	BALT_M	CON_GOODS	FOOD
38	Rgas autoe-lektroapartu rpnca	RAR1R	RIG	BALT_S	REAL_ESTAT	RE_DEV
39	Rgas elek-tromanbves rpnca	RER1R	RIG	BALT_S	INDUSTR	IND_GOODS
40	Rgas juvelie-rizstrdjumu rpnca	RJR1R	RIG	BALT_S	CON_GOODS	HOUS_PROD
41	Rgas kuu bvtava	RKB1R	RIG	BALT_S	INDUSTR	IND_GOODS
42	Rokikio sris	RSU1L	VLN	BALT_M	CON_GOODS	FOOD
43	SAF Tehnika	SAF1R	RIG	BALT_M	TECHN	TECHN
44	Sanoma Oyj	SAA1V	HEL	NORD	CON_SERV	
45	Siguldas ciltslietu un mkslgs apskloanas stacija	SCM1R	RIG	BALT_S	CON_GOODS	FOOD

Nr	Name	Abbr	Market	Stock	Industry	Sector
46	Silvano Fashion Group	SFG1T	TLN	BALT_M	CON_GOODS	HOUS_PROD
47	Snaig	SNG1L	VLN	BALT_S	CON_GOODS	HOUS_PROD
48	Tallink Grupp	TAL1T	TLN	BALT_M	CON_SERV	TRAVEL
49	Tallinna Kaubamaja Grupp	TKM1T	TLN	BALT_M	CON_SERV	RETAIL
50	Tallinna Vesi	TVEAT	TLN	BALT_M	UTIL	UTIL
51	Telia Company AB	TELIA1	HEL	NORD	TELECOMM	
52	Telia Lietuva	TEL1L	VLN	BALT_M	TELECOMM	TELECOMM
53	Trigon Property Development	TPD1T	TLN	BALT_S	FIN	REAL_EST
54	Utenos trikotaas	UTR1L	VLN	BALT_S	CON_GOODS	HOUS_PROD
55	VEF	VEF1R	RIG	BALT_S	FIN	REAL_EST
56	VEF Radiotehnika RRR	RRR1R	RIG	BALT_S	CON_GOODS	HOUS_PROD
57	Valmet Corporation	VALMT	HEL	NORD	INDUSTR	
58	Valmieras stikla iedra	VSS1R	RIG	BALT_S	MATERIALS	CHEM
59	Vilkyki pienin	VLP1L	VLN	BALT_M	CON_GOODS	FOOD
60	Vilniaus baldai	VBL1L	VLN	BALT_S	CON_GOODS	HOUS_PROD
61	Wrtsil Oyj Abp	WRT1V	HEL	NORD	INDUSTR	
62	YIT Oyj	YIT	HEL	NORD	INDUSTR	
63	Žemaitijos pienas	ZMP1L	VLN	BALT_S	CON_GOODS	FOOD

Source: The Author's representation

```
CREATE TABLE DEASVM.DEA_DT (
  ID             NUMBER,
  NAME_OMX       VARCHAR2(10 BYTE),
  ORIG_DATE      VARCHAR2(15 BYTE),
  DEA_VALUE      NUMBER,
  DEA_MODEL_ID   NUMBER,
  IS_VALID       VARCHAR2(1 BYTE),
  RECORD_DATE    DATE,
  YEAR           NUMBER)

CREATE TABLE DEASVM.DEA_MODEL (
  ID            NUMBER,
  NAME          VARCHAR2(25 BYTE),
  ORIENTATION   VARCHAR2(15 BYTE),
  RTS           VARCHAR2(15 BYTE),
  RSCRIPT_ID    NUMBER)

CREATE TABLE DEASVM.DEA_SLACKS (
  ID             NUMBER,
  OMX_NAME       VARCHAR2(15 BYTE),
  ORIG_DATE      VARCHAR2(15 BYTE),
  DEA_MODEL_ID   NUMBER,
  VARIABLE       VARCHAR2(15 BYTE),
  VALUE          NUMBER)

CREATE TABLE DEASVM.DT (
  ID            NUMBER,
  ORIG_DATE     VARCHAR2(12 BYTE),
  VALUE         NUMBER,
  SOURCE        VARCHAR2(50 BYTE),
  RECORD_DATE   DATE,
  YEAR          NUMBER,
  TBL           VARCHAR2(50 BYTE))

CREATE TABLE DEASVM.EPU (
  ID             NUMBER,
  ORIG_DATE      VARCHAR2(12 BYTE),
  GEPU_CURRENT   NUMBER,
  GEPU_PPP       NUMBER,
  RECORD_DATE    DATE)

CREATE TABLE DEASVM.EPU_FIRM (
  ID                    NUMBER,
  GVKEY                 VARCHAR2(10 BYTE),
  ORIG_DATE             VARCHAR2(25 BYTE),
  PRISK                 NUMBER,
  NPRISK                NUMBER,
  RISK                  NUMBER,
  PSENTIMENT            NUMBER,
  NPSENTIMENT           NUMBER,
  SENTIMENT             NUMBER,
```

```
  PRISKT_ECONOMIC      NUMBER,
  PRISKT_ENVIRONMENT   NUMBER,
  PRISKT_TRADE         NUMBER,
  PRISKT_INSTITUTIONS  NUMBER,
  PRISKT_HEALTH        NUMBER,
  PRISKT_SECURITY      NUMBER,
  PRISKT_TAX           NUMBER,
  PRISKT_TECHNOLOGY    NUMBER,
  COMPANY_NAME         VARCHAR2(50 BYTE),
  RECORD_DATE          DATE)

CREATE TABLE DEASVM.EPU_SETTINGS (
  ID            NUMBER,
  ISO_CODE      VARCHAR2(3 BYTE),
  WEO_CODE      VARCHAR2(3 BYTE),
  COUNTRY_NAME  VARCHAR2(50 BYTE))

CREATE TABLE DEASVM.EPU_WUI (
  ID                   NUMBER,
  ORIG_DATE            VARCHAR2(12 BYTE),
  WUI_GPD_AVG          NUMBER,
  WTUI_GPD_AVG         NUMBER,
  GLOBAL_SIMPLE_AVG    NUMBER,
  GLOBAL_GDP_W_AVG     NUMBER,
  ECONOMY_ADV          NUMBER,
  ECONOMY_EMERG        NUMBER,
  ECONOMY_LOW_INCOME   NUMBER,
  WTU_INDEX_W_AVG      NUMBER,
  WTU_INDEX_GDP_W_AVG  NUMBER,
  RECORD_DATE          DATE)

CREATE TABLE DEASVM.EPU_WUI_COUNTRY (
  ID           NUMBER,
  ORIG_DATE    VARCHAR2(12 BYTE),
  WUI          NUMBER,
  COUNTRY      VARCHAR2(3 BYTE),
  RECORD_DATE  DATE)

CREATE TABLE DEASVM.OMX (
  ID              NUMBER,
  NAME_FULL       VARCHAR2(100 BYTE),
  NAME_OMX        VARCHAR2(25 BYTE),
  OMX_INDEX_FULL  VARCHAR2(25 BYTE),
  SECTOR_ID       NUMBER,
  INDUSTRY_ID     NUMBER,
  TICKER          VARCHAR2(15 BYTE),
  CURRENCY        VARCHAR2(3 BYTE),
  MARKET_PL       VARCHAR2(3 BYTE),
  LIST_ID         NUMBER,
  LIST            VARCHAR2(50 BYTE),
  INDUSTRY        VARCHAR2(50 BYTE),
  SECTOR          VARCHAR2(50 BYTE),
  IS_DT           VARCHAR2(1 BYTE))
```

```
CREATE TABLE DEASVM.OMX_DT (
  ID                NUMBER                        NOT NULL,
  OMX_ID            NUMBER,
  RECORD_DATE       DATE,
  OMX_BID           NUMBER,
  OMX_ASK           NUMBER,
  OMX_OPEN_PRICE    NUMBER,
  OMX_HIGH_PRICE    NUMBER,
  OMX_LOW_PRICE     NUMBER,
  OMX_CLOSE_PRICE   NUMBER,
  OMX_AVG_PRICE     NUMBER,
  OMX_TOTAL_VOL     NUMBER,
  OMX_TURNOVER      NUMBER,
  OMX_TRADES        NUMBER,
  OMX_NAME          VARCHAR2(10 BYTE),
  CURRENCY          VARCHAR2(3 BYTE),
  COEFF             FLOAT(126),
  SPREAD            FLOAT(126),
  SPREAD_EXP        NUMBER)

CREATE TABLE DEASVM.OMX_FACTS (
  ID                    NUMBER,
  OMX_ID                NUMBER,
  ORIG_DATE             VARCHAR2(12 BYTE),
  RECORD_PERIOD_ID      NUMBER,
  OMX_EQUITY            NUMBER,
  OMX_INDEX             NUMBER,
  DIVIDEND_YIELD        NUMBER,
  MARKET_CAP            NUMBER,
  REVENUE               NUMBER,
  GROSS_MARGIN          NUMBER,
  OP_INCOME             NUMBER,
  OP_MARGIN             NUMBER,
  EBIT                  NUMBER,
  NET_INCOME            NUMBER,
  BASIC_EARN_PER_SHARE  NUMBER,
  DIVIDEND_PER_SHARE    NUMBER,
  AVG_DIL_SHARES        NUMBER,
  OP_CF                 NUMBER,
  CAP_EXPENDITURE       NUMBER,
  FREE_CF               NUMBER,
  ROA                   NUMBER,
  ROE                   NUMBER,
  NET_MARGIN            NUMBER,
  ASSET_TURNOVER        NUMBER,
  LEVERAGE              NUMBER,
  WORKING_CAPITAL       NUMBER,
  LONG_TERM_DEBT        NUMBER,
  TOTAL_EQUITY          NUMBER,
  CASH                  NUMBER,
  INVENTORY             NUMBER,
  ACCOUNTS_RECEIVABLE   NUMBER,
  CURRENT_ASSETS        NUMBER,
  NET_PPE               NUMBER,
```

```
  INTANGIBLES           NUMBER,
  ACCOUNTS_PAYABLE      NUMBER,
  CURRENT_DEBT          NUMBER,
  CURRENT_LIABILITIES   NUMBER,
  RPT_ID                NUMBER,
  CURRENCY              VARCHAR2(3 BYTE),
  NAME_OMX              VARCHAR2(10 BYTE),
  RECORD_DATE           DATE,
  AVG_SPREAD            NUMBER,
  AVG_COEFF             NUMBER,
  Y_RCD                 NUMBER,
  IS_VALID              VARCHAR2(1 BYTE),
  IS_TEST               VARCHAR2(1 BYTE))

CREATE TABLE DEASVM.OMX_RAW (
  ID          NUMBER,
  OMX_ID      NUMBER,
  CREATED_AT  DATE,
  IS_PARSED   VARCHAR2(1 BYTE),
  TXT         CLOB,
  BTXT        BLOB,
  MIME_TYPE   VARCHAR2(50 BYTE),
  FILENAME    VARCHAR2(50 BYTE),
  CHARSET     VARCHAR2(50 BYTE),
  UPDATED_AT  DATE,
  MD5         VARCHAR2(50 BYTE))

CREATE TABLE DEASVM.OMX_SETTINGS (
  ID           NUMBER,
  NAME         VARCHAR2(50 BYTE),
  ABBR         VARCHAR2(10 BYTE),
  IS_RPT       VARCHAR2(1 BYTE),
  IS_SECTOR    VARCHAR2(1 BYTE),
  IS_INDUSTRY  VARCHAR2(1 BYTE),
  IS_LIST      VARCHAR2(1 BYTE))

CREATE TABLE DEASVM.RLOG (
  ID          NUMBER,
  CREATED_AT  TIMESTAMP(6),
  BATCH       VARCHAR2(30 BYTE),
  USERNAME    VARCHAR2(50 BYTE),
  APPUSER     VARCHAR2(50 BYTE),
  RSCRIPT_ID  NUMBER,
  VALUE       VARCHAR2(255 BYTE),
  SCRIPT      VARCHAR2(25 BYTE))

CREATE TABLE DEASVM.RPACKS (
  ID               NUMBER,
  PACKAGE          VARCHAR2(25 BYTE),
  IS_VALID         VARCHAR2(1 BYTE),
  IS_DEFAULT_LOAD  VARCHAR2(1 BYTE),
  IS_DELETE        VARCHAR2(1 BYTE))
```

```
CREATE TABLE DEASVM.RSCRIPT (
  ID          NUMBER,
  PID         NUMBER,
  SEQ         NUMBER,
  NAME        VARCHAR2(75 BYTE),
  CODE        CLOB,
  IS_RUN      VARCHAR2(1 BYTE),
  UPDATED_AT  DATE,
  UPDATED_BY  VARCHAR2(50 BYTE),
  VERSION     NUMBER DEFAULT 1)
```

5. ORACLE OBJECTS DEFINITIONS

Name	Type	Description
DEA_SAVE	PROCEDURE	Save DEA values from R script
DEA_SAVE_SLACK	PROCEDURE	
DT_MV	PROCEDURE	Update MV_% aggregated tables
DT_YEAR_AVG	PROCEDURE	
GET_DEA_MODEL_ID	FUNCTION	Retrieve DEA model parameters
GET_DEA_MODEL_ORIENTATION	FUNCTION	
GET_DEA_MODEL_RTS	FUNCTION	
GET_DEA_SLACK	FUNCTION	
GET_DT	FUNCTION	Fetch OMX data value per time period and data type
GET_OMX_FACT	FUNCTION	
GET_OMX_NAME	FUNCTION	
GET_OMX_NAME_PR	PROCEDURE	
OMX_DT_COEFF	PROCEDURE	Verify Nasdaq Baltics data consistency
OMX_DT_FACTS	PROCEDURE	
OMX_DT_SPREAD_EXP	PROCEDURE	
OMX_FIX_ABBR	PROCEDURE	
OMX_IS_DT	PROCEDURE	
OMX_M_CALC	PROCEDURE	
OMX_SETTINGS_CREATE	PROCEDURE	
PARSE	PACKAGE	Parse CSV files from Nasdaq Baltics
RLOGGER	PROCEDURE	Log excecution of R scripts
STRING_TO_DATE	PROCEDURE	Format string to date
TBL	PACKAGE	Package for data aggregation
GET_USERNAME	FUNCTION	Get username for launching R scripts
IS_NUMBER	FUNCTION	Verify numeric value
MD5	FUNCTION	Calculate MD5 hash value for string
GET_APPUSER	FUNCTION	Get application username
GET_BATCH	FUNCTION	Create unique batch number for logging
CLEAN	FUNCTION	Clean all data
RLOG	TABLE	R Scripts functions
RPACKS	TABLE	
RSCRIPT	TABLE	

Source: The Author's representation

Source	Abbreviation	Table name	Area	Description
FRED	WLEMUINDXD	MV_FREDWLEMUINDXD	Global	Equity Market-related Economic Uncertainty
BUNDESBANK	BBXL3_A_I6_N_UNEH_TO	MV_BUNDESBANKBBXL3AI6NUNEHTO	Europe Union	Euro Area 17 Unemployment
OECD	MEI_BTS_COS_BRBUTE_LTU_BLS	MV_OECDMEIBTSCOSBRBUTELTUBLS	Lithuania	The business tendency and consumer opinion
BUNDESBANK	BBK01_SU0202	MV_BUNDESBANKBBK01SU0202	Europe Union	ECB Interest Rates For Main Refinancing
BUNDESBANK	BBXE1_M_I8_W_PROD_NS	MV_BUNDESBANKBBXE1MI8WPRODNS	Europe Union	Industrial Production / Total Industry
OECD	MEI_BTS_COS_BVEMFT_LTU_BLS	MV_OECDMEIBTSCOSBVEMFTLTUBLS	Lithuania	Future Tendency
OECD	PRICES_CPI_EU28_CPHPTT01_I	MV_OECDPRICESCPIEU28CPHPTT01I	Europe Union	EU 28 Countries HICP
OECD	KEI_PRMNTO01_EST_ST_A	MV_OECDKEIPRMNTO01ESTSTA	Estonia	Total manufacturing
OECD	KEI_PRMNTO01_LTU_ST_A	MV_OECDKEIPRMNTO01LTUSTA	Lithuania	Total manufacturing
OECD	MEI_BTS_COS_BCBUTE_EST_BLS	MV_OECDMEIBTSCOSBCBUTEESTBLS	Estonia	The business tendency and consumer opinion
NASDAQOMX	VOLNDX	MV_NASDAQOMXVOLNDX	Global	Volatility NASDAQ – 100
OECD	MEI_BTS_COS_BVDETE_LVA_BLS	MV_OECDMEIBTSCOSBVDETELVABLS	Latvia	The business tendency and consumer opinion
WWDI	LVA_BX_KLT_DINV_WD_GD_ZS	MV_WWDILVABXKLTDINVWDGDZS	Latvia	Foreign direct investment
OECD	MEI_BTS_COS_BSSPFT_EST_BLS	MV_OECDMEIBTSCOSBSSPFTESTBLS	Estonia	The business tendency and consumer opinion
WWDI	EST_BX_KLT_DINV_WD_GD_ZS	MV_WWDIESTBXKLTDINVWDGDZS	Estonia	Foreign direct investment
OECD	MEI_BTS_COS_BSSPFT_LVA_BLS	MV_OECDMEIBTSCOSBSSPFTLVABLS	Latvia	Future Tendency
OECD	KEI_PRMNTO01_LVA_ST_A	MV_OECDKEIPRMNTO01LVASTA	Latvia	Total manufacturing
OECD	SNA_TABLE1_EU28_P31S14_VOB	MV_OECDSNATABLE1EU28P31S14VOB	Europe Union	EU 28 final Consumption of Households
WWDI	LTU_BX_KLT_DINV_WD_GD_ZS	MV_WWDILTUBXKLTDINVWDGDZS	Lithuania	Foreign direct investment

Source: The Author's representation

```
#
# Initial script initialization to run in Windows environment
# without arguments used in Linux settings
#
rm(list = ls())
batch <- toString(as.numeric(Sys.time()) * 10000)
print(paste("Generate a new batch: ", batch))
Sys.setenv(OCI_LIB64 = "C:/ora19")
Sys.setenv(OCI_INC = "C:/ora19/sdk/include")
Sys.setenv(ORACLE_HOME = "C:/ora19")
setwd("C:/WIP")

#
# Database connectivity
#
drv <- dbDriver("Oracle")
host <- "127.0.0.1"
port <- 1521
svc <- "inscrio"
connect.string <- paste("(DESCRIPTION=","(ADDRESS=(PROTOCOL=tcp)
(HOST=",host,") (PORT=",port,"))", "(CONNECT_DATA= (SERVICE_
NAME=",svc,")))",sep = "")
con <- dbConnect(drv, username = " inscrio", password = "000", dbname
= connect.string)
stmt <-paste("begin rlogger(p_msg=>'INIT: ",script_name,"',p_
batch=>'", batch, "'); end;")
rs <- dbSendQuery(con,statement = stmt)

#
# Packages load
#
stmt<-paste("SELECT PACKAGE FROM RPACKS")
rs <- dbSendQuery(con,statement = stmt)
list_packs_sql <- fetch(rs, n = -1)

for (i in 1:nrow(list_packs_sql))
{
  list_packs_r <- c(list_packs_sql[i, "PACKAGE"])
  if(!require(list_packs_r, character.only=TRUE))
  {
    install.packages(list_packs_r)
    library(list_packs_r, character.only=TRUE)
  }
}
```

8. R SCRIPT FOR DEA

```
#
# DATA LOAD
#
stmt <-paste   (“
               SELECT DISTINCT ORIG_DATE
               FROM OMX_FACTS
               WHERE IS_VALID=’Y’
               ORDER BY ORIG_DATE ASC
               “)
rs <-dbSendQuery(con, statement = stmt)

omx_data <- fetch(rs, n = -1)
print(“Loop”)

for (i in 1:nrow(omx_data))
{
omx_date <- c(omx_data[i, “ORIG_DATE”])
print(paste(
“ YEAR: “,omx_data[i, “ORIG_DATE”],
“ DEA orienzation: “,z_orientation,
“ RTS: “,z_rts,
“ MODEL: “,z_model_name,
“ ID: “,z_model))

stmt <- paste  (“
                     SELECT *
                     FROM MV_DEA_STAGE1 WHERE
                     ORIG_DATE=CLEAN(‘”,omx_data[i, “ORIG_DATE”],”’)
                “)
  rs <- dbSendQuery(con,statement = stmt)
  omx_dt <- fetch(rs, n = -1)
  omx_dt <- read_data  (
                       omx_dt,
                       outputs = c(“OUT1”, “OUT2”),
                       inputs = c(“IN1”,”IN2”,”IN3”,”IN4”,”IN5”,”IN6”,
                       ”IN7”)
                       )

  fp_dea_m <- model_basic     (
                              omx_dt,
                              orientation = z_orientation,
                              rts = z_rts
                              )

  eff <- efficiencies(fp_dea_m)
  s <- slacks(fp_dea_m)
  lamb <- lambdas(fp_dea_m)
  tar <- targets(fp_dea_m)
  ref <- references(fp_dea_m)
  returns <- rts(fp_dea_m)
  eff <- data.frame(as.list(eff))
```

```
  print("DONE DMU ESTIMATION. SAVING...")

#
# DEA SLACK SAVE INPUTS
#
  s_d<-as.data.frame(s$slack_input)

  for (j in 1:nrow(s_d))
  {
    print(paste("SLACK SAVE INPUTS ROW: ",j," OF ",nrow(s_d)))
    for (k in 1:ncol(s_d))
    {
      slack_col_name <- names(s_d)[k]
      slack_row_name <- rownames(s_d)[j]
      slack_val <- s_d[j, k]
      if(!is.na(slack_val))
      {
        stmt<-paste("   BEGIN DEA_SAVE_SLACK (
                        P_OMX_NAME =>'",slack_row_name,"',
                        P_ORIG_DATE =>'",omx_date,"',
                        P_DEA_METHOD =>'",z_model,"',
                        P_VARIABLE =>'",slack_col_name,"',
                        P_VALUE =>'",slack_val,"');
                        END;")
        print(stmt)
        rs <- dbSendQuery(con,statement = stmt)
      }

    }
  }

#
# DEA SLACK SAVE OUTPUTS
#
  s_d2<-as.data.frame(s$slack_output)
  for (j in 1:nrow(s_d2))
  {

    for (k in 1:ncol(s_d2))
    {

      slack_col_name <- names(s_d2)[k]
      slack_row_name <- rownames(s_d2)[j]
      slack_val <- s_d2[j, k]

      if(!is.na(slack_val))
      {
        stmt<-paste("   BEGIN
                        DEA_SAVE_SLACK (
                        P_OMX_NAME =>'",slack_row_name,"',
                        P_ORIG_DATE =>'",omx_date,"',
                        P_DEA_METHOD =>'",z_model,"',
                        P_VARIABLE =>'",slack_col_name,"',
                        P_VALUE =>'",slack_val,"'
```

```
                        );
                        END;
                        ")
        print(stmt)
        rs <- dbSendQuery(con,statement = stmt)
      }
    }
  }

#
# DEA SAVE
#
  for (j in 1:nrow(eff))
  {
    for (k in 1:ncol(eff))
    {
      eff_dea_val <- eff[j, k]
      eff_dea_name <- names(eff)[k]
      if(!is.na(eff_dea_val))
      {
      stmt<-paste("
                  BEGIN
                  DEA_SAVE(   P_OMX_NAME   =>'",eff_dea_name,"',
                  P_DEA_VALUE  =>'",eff_dea_val,"',
                  P_ORIG_DATE  =>' ",omx_date,"',
                  P_DEA_METHOD =>'",z_model,"'
                  );
                  END;
                  ")
      print(stmt)
      rs <- dbSendQuery(con,statement =stmt)
      }
    }
  }
}
```

```
#
# HEATMAP DATA ANALYSIS
#
dt.cor = cor(dt[,-1], method = c("spearman"))
palette = colorRampPalette(c("green", "white", "red")) (20)
heatmap(x = dt.cor, col = palette, symm = TRUE)

#
# PANEL DATA ANALYSIS
#
pn.mod1.ols <-lm(mod1, data=dt)
summary(pn.mod1.ols)
pn.mod1.fe <- plm(mod1, data = dt, index=c("OMX_NAME", "PERIOD"),
model = "within")
summary(pn.mod1.fe)
pFtest(pn.mod1.fe, pn.mod1.ols)
pool1 <- plm(mod1, data=dt, index=c("OMX_NAME", "PERIOD"),
model="pooling")
plmtest(pool1, type=c("bp"))
pcdtest(pn.mod1.fe, test = c("lm"))
pbgtest(pn.mod1.fe)
pFtest(fixed.time2, pn.mod2.fe)

#
# CORRELATION MATRIX
#
corstars <-function(x, method=c("pearson", "spearman"),
removeTriangle=c("upper", "lower"),
result=c("none", "html", "latex")){
correlation_matrix<-rcorr(x, type=method[1])
corr_R <- correlation_matrix$r
corr_p <- correlation_matrix$P
SIGstars <- ifelse(p < .0001, "****", ifelse(p < .001, "*** ",
ifelse(p < .01, "**  ", ifelse(p < .05, "*   ", "    "))))
corstars(dt[,-1], result="none")
cor(dt[,-1], use="complete.obs", method = c("pearson", "kendall",
"spearman"))
ggcorr(dt[,2:20], method = c("all.obs", "spearman"))

#
# LM REGRESSION
#
lmfit <- lm(dt$EFFICIENCY_2 ~ dt$EFFICIENCY_1+dt$UNCERTAINY_
GLOBAL+dt$GLOBAL_AVG+dt$TRADE_UNCERTAINTY)
summary(lmfit)
par(mfrow = c(2, 2))
plot(lmfit)
```

10. R ENSEMBLE MACHINE LEARNING

https://cran.r-project.org/web/packages/SuperLearner/SuperLearner.pdf

```
Package         SuperLearner
December        10, 2019
Type            Package
Title           Super Learner Prediction
Version         2.0-26
Date            2019-10-27
Maintainer      Eric Polley <polley.eric@mayo.edu>
Description     Implements the super learner prediction method and
                contains a
                library of prediction algorithms to be used in the
                superlearner.
License         GPL-3
URL             https://github.com/ecpolley/SuperLearner
```

```
#
# DATA LOAD
#
stmt            <- paste("SELECT [...] FROM MV_SVM_1 [...]")
rs              <- dbSendQuery(con,statement = stmt)
dt              <- fetch(rs, n = -1)

#
# DATA SEPARATION
#
dt_r <- createDataPartition(y = dt$CLASSFICATOR, p= 0.7, list = FALSE)
train <- dt[dt_r,]
test <- dt[-dt_r,]
dim(train);
dim(test);
anyNA(dt)

#
# CONFUSIONMATRIX
# Calculates a cross-tabulation of observed and
# predicted classes with associated statistics.
#
test_pred <- predict(svm_Linear, newdata = testing)
confusionMatrix(test_pred,as.factor(testing$ CLASSFICATOR))
```

```
#
# RPART ANALYSIS
#
rpart_model =  rpart(CLASSFICATOR ~ ., data = train)
rpart.plot(rpart_model, type=2, under = TRUE)

#
# FIT MODELS
# SuperLearner fits the super learner prediction algorithm. The weights
for
# each algorithm in SL.library is estimated, along with the fit of each
# algorithm. All prediction algorithm wrappers in SuperLearner:
#
[1] "SL.bartMachine"
[2] "SL.bayesglm"
[3] "SL.biglasso"
[4] "SL.caret"
[5] "SL.caret.rpart"
[6] "SL.cforest"
[7] "SL.earth"
[8] "SL.extraTrees"
[9] "SL.gam"
[10] "SL.gbm"
[11] "SL.glm"
[12] "SL.glm.interaction"
[13] "SL.glmnet"
[14] "SL.ipredbagg"
[15] "SL.kernelKnn"
[16] "SL.knn"
[17] "SL.ksvm"
[18] "SL.lda"
[19] "SL.leekasso"
[20] "SL.lm"
[21] "SL.loess"
[22] "SL.logreg"
[23] "SL.mean"
[24] "SL.nnet"
[25] "SL.nnls"
[26] "SL.polymars"
[27] "SL.qda"
[28] "SL.randomForest"
[29] "SL.ranger"
[30] "SL.ridge"
[31] "SL.rpart"
[32] "SL.rpartPrune"
[33] "SL.speedglm"
[34] "SL.speedlm"
[35] "SL.step"
[36] "SL.step.forward"
[37] "SL.step.interaction"
[38] "SL.stepAIC"
[39] "SL.svm"
[40] "SL.template"
[41] "SL.xgboost"
```

```
All screening algorithm wrappers in SuperLearner:
[1] "All"
[1] "screen.corP"
[2] "screen.corRank"
[3] "screen.glmnet"
[4] "screen.randomForest"
[5] "screen.SIS"
[6] "screen.template"
[7] "screen.ttest"
[8] "write.screen.template"
sl_l = SuperLearner(Y = y_train, X = x_train, family = binomial(),
SL.library = ALGORITHM)

#
# V-FOLD CV
# Function to get V-fold cross-validated risk estimate for super
learner.
# This function simply splits the data into V folds and then
# calls SuperLearner.
# Most of the arguments are passed directly to SuperLearner.
#
cv_sl = CV.SuperLearner(Y = y_train,
                        X = x_train,
                        family = binomial(),
                        V = 10,
                        parallel = "multicore",
                        SL.library = my_ml_3
                        )
```

11. R PACKAGES USED

R package	Description
deaR	Set of functions for Data Envelopment Analysis. It runs both classic and fuzzy DEA models.See: Banker, R.; Charnes, A.; Cooper, W.W. (1984). , Charnes, A.; Cooper, W.W.; Rhodes, E. (1978). and Charnes, A.; Cooper, W.W.; Rhodes, E. (1981).
plm	plm is a package for R which intends to make the estimation of linear panel models straightforward.
tseries	Title Time Series Analysis and Computational Finance
lmtest	Testing Linear Regression Models. A collection of tests, data sets, and examples for diagnostic checking in linear regression models.
caTools	Contains several basic utility functions including: moving (rolling, running) window statistic functions, read/write for GIF and ENVI binary files
LiblineaR	A wrapper around the 'LIBLINEAR' C/C++ library for machine learning
pastecs	Regularisation, decomposition and analysis of space-time series. The pastecs R package is a PNEC-Art4 and IFREMER
fastAdaboost	Implements Adaboost based on C++ backend code. This is blazingly fast and especially useful for large, in memory data sets. The package uses decision trees as weak classifiers.
xgboost	Extreme Gradient Boosting, which is an efficient implementation of the gradient boosting framework from Chen & Guestrin (2016) . This package is its R interface. The package includes efficient linear model solver and tree learning algorithms.
ggplot2	ggplot2 is a system for declaratively creating graphics, based on The Grammar of Graphics.
corrplot	A graphical display of a correlation matrix or general matrix. It also contains some algorithms to do matrix reordering
Hmisc	Contains many functions useful for data analysis, high-level graphics, utility operations, functions for computing sample size and power, importing and annotating datasets, imputing missing values, advanced table making, variable clustering, character string manipulation, conversion of R objects to LaTeX and html code, and recoding variables.
caret	Misc functions for training and plotting classification and regression models.
tidyverse	The 'tidyverse' is a set of packages that work in harmony because they share common data representations and 'API' design.

R package	Description
e1071	Functions for latent class analysis, short time Fourier transform, fuzzy clustering, support vector machines, shortest path computation, bagged clustering, naive Bayes classifier etc.
RColorBrewer	Provides color schemes for maps (and other graphics) designed by Cynthia Brewer as described at http://colorbrewer2.org
SuperLearner	Implements the super learner prediction method and contains a library of prediction algorithms to be used in the super learner.
glmnet	Extremely efficient procedures for fitting the entire lasso or elastic-net regularization path for linear regression, logistic and multinomial regression models, Poisson regression and the Cox model. Two recent additions are the multiple-response Gaussian, and the grouped multinomial regression.
randomForest	Classification and regression based on a forest of trees using random inputs, based on Breiman (2001)
RhpcBLASctl	Control the number of threads on 'BLAS' (Aka 'GotoBLAS', 'OpenBLAS', 'ACML', 'BLIS' and 'MKL'). And possible to control the number of threads in 'OpenMP'. Get a number of logical cores and physical cores if feasible.
rpart	Recursive partitioning for classification, regression and survival trees. An implementation of most of the functionality of the 1984 book by Breiman, Friedman, Olshen and Stone.
rpart.plot	Plot 'rpart' models. Extends plot.rpart() and text.rpart() in the 'rpart' package.
psych	A general purpose toolbox for personality, psychometric theory and experimental psychology. Functions are primarily for multivariate analysis and scale construction using factor analysis, principal component analysis, cluster analysis and reliability analysis, although others provide basic descriptive statistics.
Ggally	The R package 'ggplot2' is a plotting system based on the grammar of graphics. 'GGally' extends 'ggplot2' by adding several functions to reduce the complexity of combining geometric objects with transformed data.

Source: The Author's adoption from respective libraries

12. ABBREVIATIONS USED IN DATASETS

Name	Abbreviation	Sector	Industry	OMX List
Nonequity Investment Instruments	NONEQ_INV		Y	
Utilities	UTIL		Y	
Industrial Transportation	IND_TRANSP		Y	
Technology	TECHN		Y	
Consumer Services	CON_SERV		Y	
Consumer Goods	CON_GOODS		Y	
Real Estate Investment / Services	REAL_ESTAT		Y	
Industrials	INDUSTR		Y	
Basic Materials	MATERIALS		Y	
Personal Goods	PER_GOODS		Y	
Health Care	HEALTH		Y	
Oil / Gas	OIL_GAS		Y	
Telecommunications	TELECOMM		Y	
Financials	FIN		Y	
Banks	BANKS		Y	
Nonequity Investment Instruments	NONEQ_INV	Y		
Travel / Leisure	TRAVEL	Y		
Personal Products	PER_PROD	Y		
Personal / Household Goods	HOUS_PROD	Y		
Retail	RETAIL	Y		
Utilities	UTIL	Y		
Basic Resources	RESOURCE	Y		
Transportation Services	TRANSP	Y		
Technology	TECHN	Y		
Chemicals	CHEM	Y		

Name	Abbreviation	Sector	Industry	OMX List
Real Estate	REAL_EST	Y		
Media	MEDIA	Y		
Industrial Goods / Services	IND_GOODS	Y		
Food / Beverage	FOOD	Y		
Real Estate Holding / Development	RE_DEV	Y		
Construction / Materials	CONSTR	Y		
Health Care	HEALTH	Y		
Banks	BANKS	Y		
Telecommunications	TELECOMM	Y		
Financial Services	FIN	Y		
Oil / Gas	OIL_GAS	Y		
The Nordic List	NORD			Y
Baltic Main List	BALT_M			Y
First North Baltic Share List	1NORTH_BAL			Y
Baltic Secondary List	BALT_S			Y

Source: The Author's representation

13. CLI AUTOMATION FOR DATA LOADING

Part 1. List of scripts needed for a massive data loading (CSV, API) into database:

```
[sergei@INSCORIO toolbox]$ ls -la phd*
-rw-rw-rw- 1 sergei sergei   1926 Mar 10 10:44 phd_data_pump.php
-rw-rw-rw- 1 sergei sergei   1265 Feb 17 13:56 phd_epu_all.php
-rw-rw-rw- 1 root   sergei   1399 Feb 17 13:56 phd_epu_firm.php
-rw-rw-rw- 1 sergei sergei   2105 Feb 18 17:48 phd_omx_facts_loader.
php
-rw-rw-rw- 1 sergei sergei   1591 Feb 20 11:16 phd_omx_facts.php
-rw-rw-rw- 1 sergei sergei   1398 Feb 19 18:52 phd_omx.php
-rw-rw-rw- 1 sergei sergei   1976 Feb 21 17:19 phd_omx_ticker_all.php
-rwxrwxrwx 1 sergei sergei     38 Feb 20 17:15 phd_run
-rw-rw-rw- 1 oracle sergei   2989 Feb 23 19:43 phd_run.php
-rw-rw-rw- 1 sergei sergei   1066 Feb 21 18:52 phd_util_fix.php
-rw-rw-rw- 1 sergei sergei   1703 Feb 17 18:20 phd_wui_country.php
-rw-rw-rw- 1 sergei sergei   1390 Feb 17 13:55 phd_wui.php
-rw-rw-rw- 1 sergei sergei   1277 Feb 17 13:55 phd_wui_settings.php
```

Part 2. Execution of R scripts saved in database on servers

```
[sergei@INSCORIO toolbox]$ ./phd_run
```

```
================================================================================
R SCRIPTS WRAPPER FOR LINUX ENVIROMENT
================================================================================

 ORACLE_CONNECT
 Schema...........: INSCORIO
 IP...............: 127.0.0.1/ALPHA
 Charset..........: AL32UTF8
 Connection.......: CHECK_OK
 * Load all scripts IDs

 ID      NAME
 --      ----
 42       DEA >>> rDEA test
 62       DEA >>> Stage 1. CRS Input-orientated
 82       DEA >>> Stage 1. VRS Input-orientated
 83       DEA >>> Stage 1. CRS Output-orientated
 84       DEA >>> Stage 1. VRS Output-orientated
 85       DEA >>> Stage 2. VRS Input-orientated
 101      DEA >>> Stage 2. VRS Input-orientated / DT2
 122      DEA >>> DEA pairs
 123      DEA >>> Malmquist
 43       Initiation >>> Init
 44       Initiation >>> Load libs
 41       Data load >>> Calculate OMX_DT SPREAD and COEFF per each row
 61       Data load >>> Fix OMX_DT to NULL values
 102      Data load >>> Make barcharts
```

```
121     Data load >>> Descriptive statistics
127     Data load >>> MV update
125     Machine learning >>> Ensemble model
126     Machine learning >>> Classification
* Please enter script ID to run: _
```

14. DATA ENVELOPMENT ANALYSIS

result_malmquist$mi

Log transformed for graphical representation

PERIOD	HEL			RIG			TLN			VLN		
	MI	TC	SECH	MI	TC	SECH	MI	TC	SECH	MI	TC	SECH
2012	0,06175	0,06175	0	-0,002052004	0,00062	-0,0027	-0,0277	-0,0277	0	0,00079	0,00079	0
2013	0,02492	-0,4696	0	0,001177407	-0,0015	0,00267	0,00871	0,00871	0	-0,0033	-0,0028	-0,0004
2014	0,00372	0,00372	0	0,005227015	0,00523	0	0,00284	0,00284	0	0,00475	0,00425	0,00036
2015	0,32195	0,32195	0	-0,001464472	-0,0015	0	-0,0081	-0,0081	0	-0,0041	-0,0041	0
2016	0,03494	0,03494	0	0,024509669	0,02451	0	-0,0007	-0,0007	0	-0,0003	-0,0003	0
2017	0,00088	0,00088	0	0,125769195	0,12577	0	1,4E-05	0,0006	-0,0004	0,00102	0,00102	0
2018	0,03999	0,03999	0	-0,141542251	-0,1415	0	0,00816	0,00757	0,00043	0,00104	0,00104	0

result_malmquist$mi

	CON_GOODS	CON_SERV	FIN	INDUSTR	MATERIALS	OIL_GAS	REAL_ESTAT	TECHN	TELECOMM	UTIL
2012	1,001567	0,9997616	0,9194461	0,9987866	1,0003085	0,9996652	0,9590016	0,9979846	1,0006087	1,0124008
2013	1,0015429	1,003993	1,0138742	1,0007583	1,0018096	0,9865221	0,981893	0,9994109	0,5133639	0,9895074
2014	1,00171	0,9974189	1,0084983	1,0015401	0,9990099	0,9831652	1,0242914	0,9980284	1,0005807	1,0014416
2015	0,9981213	1,0021431	0,9830451	0,9999418	1,0004333	0,948607	0,977332	1,000821	1,882543	0,9935968
2016	1,0256422	0,9984751	1,0005572	0,9975037	0,9991055	0,9898974	1,517805	0,9999751	0,9964546	1,0034392
2017	0,979249	1,0019482	0,9984484	0,9997195	0,9989158	1,0018264	0,9830451	1,0000831	0,974272	1,0014181
2018	1,0019318	0,9993915	1,0020017	1,0018216	1,0140247	1,0030789	0,9964546	1,0050171	1,0243008	0,9983632

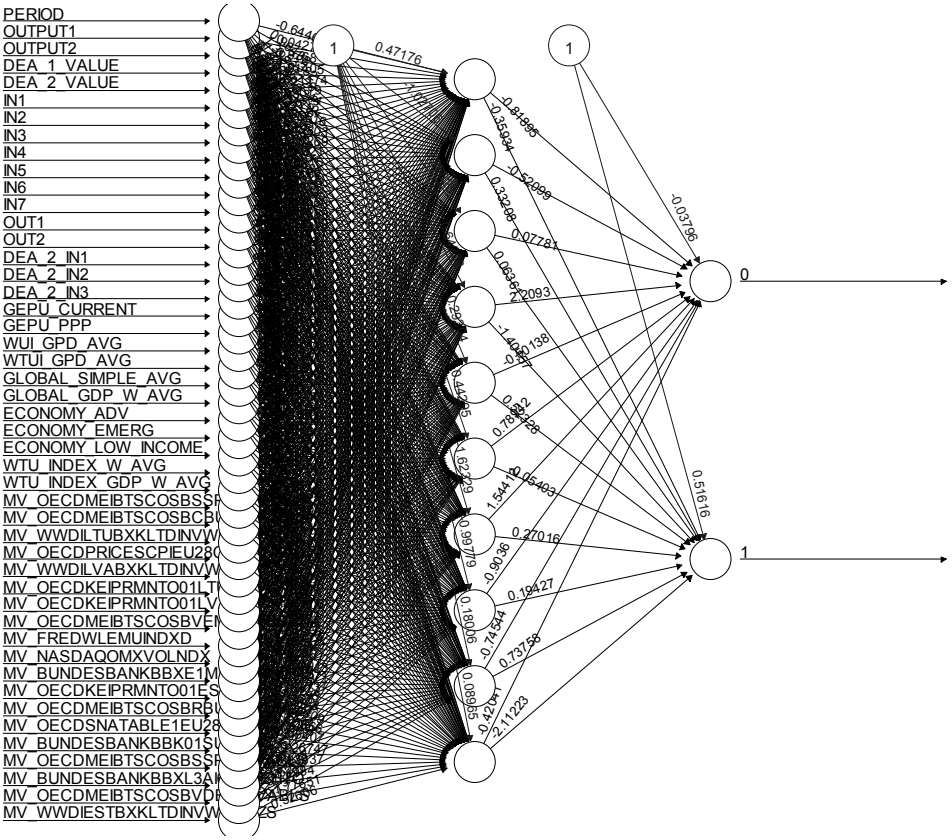
result_malmquist$pech

	CON_GOODS	CON_SERV	FIN	INDUSTR	MATERIALS	OIL_GAS	REAL_ESTAT	TECHN	TELECOMM	UTIL
2012	1,0000918	1,0002255	1	1,0004308	1,0006291	1,0000111	1	0,9997182	1	1,000048
2013	1,0005276	1,0004458	1	1,0003826	1,0002451	1	1	1,0002819	1	1
2014	0,9999702	0,9997765	1	0,9995082	0,999344	1	1	1	1	1
2015	1,0000298	1,0002235	1	1,0005085	1,0004197	1	1	1	1	1
2016	1	0,9996661	1	0,9991035	0,9993236	0,9999262	1	1	1	1
2017	0,9997888	1,0002302	1	1,0000786	0,9999878	1,0000738	1	0,9998767	1	1
2018	0,9993938	1,000095	1	0,9999545	1,0009259	1	1	1,0001233	1	1

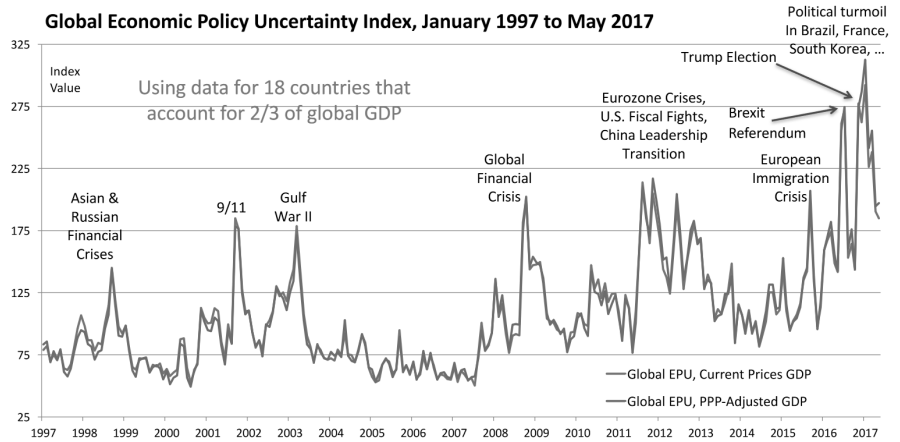
result_malmquist$sech

	CON_GOODS	CON_SERV	FIN	INDUSTR	MATERIALS	OIL_GAS	REAL_ESTAT	TECHN	TELECOMM	UTIL
2012	1,0012065	0,9991184	1	0,9994081	1,0000104	1,0006564	1	0,9992519	1	1,0019245
2013	1,0017023	1,0028447	1	1,0017351	1,0020467	1	1	0,9998135	1	1
2014	0,9984505	0,9983283	1	0,9998523	0,9979158	1	1	0,9986979	1	1
2015	1,0003364	1,0009791	0,9950247	1,0010914	0,9997438	1	1	0,9981954	1	0,9903555
2016	1,0012151	0,9985636	1,001623	0,9974939	0,9996263	0,9983171	1	1,002109	1	1,0097384
2017	0,9971107	1,0007103	0,996989	0,9983101	0,9975433	1,0009153	1	0,9989825	1	1
2018	1,0004507	0,9978008	1,0007966	1,0005057	1,0052381	1,0007697	1	1,0029598	1	0,9968522

15. ANN VISUALIZATION



16. WORLD UNCERTAINTY INDEX

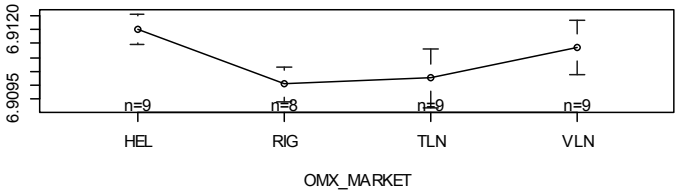


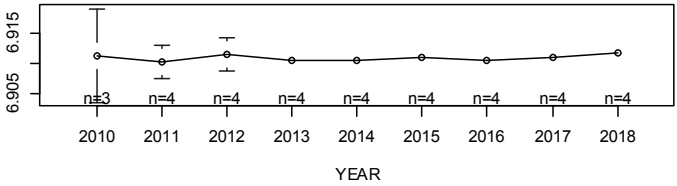
Notes: Global EPU calculated as the GDP-weighted average of monthly EPU index values for US, Canada, Brazil, Chile, UK, Germany, Italy, Spain, France, Netherlands, Russia, India, China, South Korea, Japan, Ireland, Sweden, and Australia, using GDP data from the IMF's World Economic Outlook Database. National EPU index values are from www.PolicyUncertainty.com and Baker, Bloom and Davis (2016). Each national EPU Index is renormalized to a mean of 100 from 1997 to 2015 before calculating the Global EPU Index.

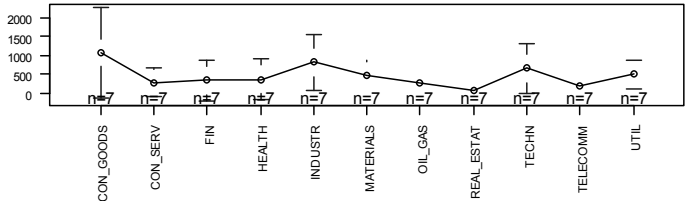
A monthly index of Global Economic Policy Uncertainty (GEPU) that runs from January 1997 to the present. The GEPU Index is a GDP-weighted average of national EPU indices for 20 countries: Australia, Brazil, Canada, Chile, China, France, Germany, Greece, India, Ireland, Italy, Japan, Mexico, the Netherlands, Russia, South Korea, Spain, Sweden, the United Kingdom, and the United States.

Each national EPU index reflects the relative frequency of own-country newspaper articles that contain a trio of terms pertaining to the economy (E), policy (P) and uncertainty (U). In other words, each monthly national EPU index value is proportional to the share of own-country newspaper articles that discuss economic policy uncertainty in that month.

17. MEANS AND CONFIDENCE INTERVALS PER MARKETS, PERIODS AND INDUSTRIES







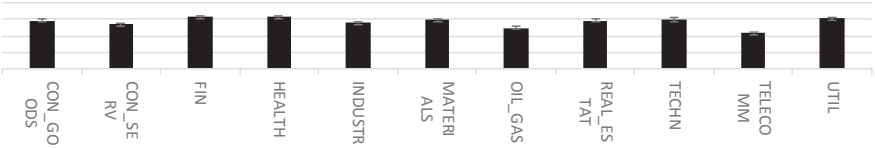
```
# Models 1|2 specifications:

OUTPUT1|2   ~ DEA_STAGE_1 +
              DEA_STAGE_2 +
              INPUTS +
              UNCERTANTY +
              EXTERNALITIES +
              factors (OMX_MARKET | OMX_NAME | PERIOD)

# Fitted with significance codes:
# 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


# GENERALIZED LINEAR MODELS Model1

                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -4.624e+02  2.317e+02  -1.996  0.05045 .
   DEA_1_VALUE   -8.631e+00  5.530e+00  -1.561  0.12373
   DEA_2_VALUE    3.191e-01  5.166e-01   0.618  0.53900
   DEA_2_IN1     -1.050e+04  9.422e+03  -1.114  0.26956
   DEA_2_IN2     -1.679e-01  1.076e+00  -0.156  0.87646
   DEA_2_IN3      6.646e-01  1.363e-01   4.877 8.06e-06 ***
   IN1            8.465e-02  9.400e-02   0.901  0.37138
   IN2            2.280e-01  1.176e-01   1.938  0.05722 .
   IN3            9.569e-01  3.680e-01   2.600  0.01168 *
   IN4            1.100e-01  8.078e-02   1.362  0.17834
   IN5           -2.758e-01  1.224e-01  -2.254  0.02782 *
   IN6           -1.831e+00  6.599e-01  -2.775  0.00731 **
   IN7            2.027e+00  3.662e+00   0.554  0.58191
   GEPU_PPP       6.159e+00  2.527e+00   2.438  0.01771 *
   WUI_GPD_AVG   -1.066e+00  1.628e+00  -0.655  0.51501
   WTUI_GPD_AVG  -1.012e+02  4.271e+01  -2.370  0.02096 *

# GENERALIZED LINEAR MODELS Model2
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -7.877e+02  1.453e+02  -5.420 1.07e-06 ***
   DEA_1_VALUE    9.535e+00  3.468e+00   2.749  0.00785 **
   DEA_2_VALUE   -3.749e-01  3.240e-01  -1.157  0.25181
   DEA_2_IN1      1.118e+04  5.910e+03   1.892  0.06320 .
   DEA_2_IN2      5.978e-01  6.747e-01   0.886  0.37914
   DEA_2_IN3     -1.472e-01  8.548e-02  -1.722  0.09007 .
   IN1           -3.654e-02  5.896e-02  -0.620  0.53777
   IN2            2.124e-01  7.378e-02   2.879  0.00549 **
   IN3            5.082e-01  2.309e-01   2.201  0.03150 *
   IN4            1.348e-01  5.067e-02   2.661  0.00995 **
   IN5           -1.609e-01  7.677e-02  -2.096  0.04020 *
   IN6           -2.230e+00  4.139e-01  -5.387 1.21e-06 ***
   IN7            4.541e+00  2.297e+00   1.977  0.05259 .
   GEPU_PPP       4.717e+00  1.585e+00   2.976  0.00418 **
   WUI_GPD_AVG    9.708e-01  1.021e+00   0.951  0.34555
   WTUI_GPD_AVG  -7.145e+01  2.679e+01  -2.667  0.00979 **
```

 # Oneway (individual) effect Within Model

	Estimate	Std. Error	t-value	Pr(>|t|)	
DEA_1_VALUE	-5.495162	4.836737	-1.1361	0.2606564	
DEA_2_VALUE	0.025091	0.344884	0.0728	0.9422581	
DEA_2_IN3	0.655854	0.126771	5.1735	3.093e-06	***
IN2	0.310969	0.112495	2.7643	0.0076707	**
IN3	1.074531	0.320106	3.3568	0.0014091	**
IN5	-0.292833	0.107424	-2.7260	0.0085024	**
IN6	-2.374053	0.662759	-3.5821	0.0007060	***
GEPU_PPP	4.901567	1.561246	3.1395	0.0026805	**
WTUI_GPD_AVG	-87.275088	23.430421	-3.7249	0.0004502	***

F-statistic: 25.7881, p-value: < 2.22e-16

Fixed-effects estimates:



 # No time-fixed effects
 # plmtest(fixed, c("time"), type=("bp")):
Lagrange Multiplier Test - time effects (Breusch-Pagan) for balanced panels
chisq = 1.4451, df = 1, p-value = 0.2293

 # No significant differences across market places (plmtest)
 # plmtest(pool, type=c("bp"))
Lagrange Multiplier Test - (Breusch-Pagan) for balanced panels
chisq = 0.060505, df = 1, p-value = 0.8057

	Fixed time	Fixed	Random	OLS	Pool
DEA_1_VALUE	-4.36	-7.27	-5.03	-8.63	-8.63
DEA_2_VALUE	-0.06	0.18	0.01	0.32	0.32
DEA_2_IN3	0.66 ***	0.65 ***	0.66 ***	0.66 ***	0.66 ***
IN2	0.31 *	0.29 *	0.29 **	0.23	0.23
IN3	1.05 **	1.00 *	1.04 ***	0.96 *	0.96 *
IN5	-0.30 *	-0.29 *	-0.27 **	-0.28 *	-0.28 *
IN6	-2.56 ***	-2.20 **	-2.32 ***	-1.83 **	-1.83 **
GEPU_PPP	4.77 **	5.48 *	4.90 ***	6.16 *	6.16 *
WTUI_GPD_AVG	-85.88 **	-91.79 *	-87.72 ***	-101.23 *	-101.23 *
factor(PERIOD)3	31.65				
factor(PERIOD)4	130.19				
factor(PERIOD)5	72.78				
factor(PERIOD)6	-15.01				
factor(PERIOD)7	-54.13				
factor(PERIOD)8	27.49				
DEA_2_IN1		-4355.02		-10497.90	-10497.90
DEA_2_IN2		-0.28		-0.17	-0.17
IN1		0.03		0.08	0.08
IN4		0.10		0.11	0.11
IN7		1.57		2.03	2.03
WUI_GPD_AVG		-0.50		-1.07	-1.07
(Intercept)			-473.94 *	-462.36	-462.36

***** p < 0.001, ** p < 0.01, * p < 0.05**

```
# Test for random or fixed effects.
# phtest(fixed, random)

Hausman Test
chisq = 0.48851, df = 9, p-value = 0.16276
alternative hypothesis: one model is inconsistent

# Testing time-fixed effects.
# pFtest(fixed.time, fixed)

F test for individual effects
F = 0.34705, df1 = 6, df2 = 51, p-value = 0.9084
alternative hypothesis: significant effects

Lagrange Multiplier Test - time effects (Breusch-Pagan) for balanced
panels
chisq = 1.4451, df = 1, p-value = 0.2293
alternative hypothesis: significant effects

# Testing cross-sectional dependence
# H0) The null is that there is not cross-sectional dependence
# pcdtest(fixed, test = c("lm")), pcdtest(fixed, test = c("cd"))

Breusch-Pagan LM test for cross-sectional dependence in panels
chisq = 72.472, df = 55, p-value = 0.05723
alternative hypothesis: cross-sectional dependence

Pesaran CD test for cross-sectional dependence in panels
z = -0.82331, p-value = 0.4103
alternative hypothesis: cross-sectional dependence

# Testing serial correlations
# pbgtest(fixed)

Breusch-Godfrey/Wooldridge test for serial correlation in panel models
chisq = 12.823, df = 7, p-value = 0.07655
alternative hypothesis: serial correlation in idiosyncratic errors

# Unit root tests adf.test
Augmented Dickey-Fuller Test
Dickey-Fuller = -4.0489, Lag order = 4, p-value = 0.012
alternative hypothesis: stationary

# Testing heteroscedasticity with bptest
Breusch-Pagan test
BP = 72.567, df = 19, p-value = 3.426e-08
```

20. DATASETS SOURCES

Valid by May 2020

Source	Description	Data Source API
Federal Reserve Economic Data	Equity Market-related Economic Uncertainty Index	https://www.quandl.com/data/FRED/WLEMUINDXD-Equity-Market-related-Economic-Uncertainty-Index
Deutsche Bundesbank Data Repository	Euro Area 17 (fixed composition) / Unemployment	https://www.quandl.com/data/BUNDESBANK/BBXL3_A_I6_N_UNEH_TOTAL0_ILO_TOEC_RAT_I00-Euro-Area-17-fixed-composition-Unemployment-Standardised-ILO-Total-Total-economy-Rate-Neither-seasonally-nor-working-day-adjusted
		https://www.quandl.com/data/BUNDESBANK/BBXL3_M_I6_S_UNEH_TOTAL0_ILO_TOEC_RAT_I00-Euro-area-17-fixed-composition-Unemployment-Standardised-ILO-Total-Total-economy-Rate-Seasonally-adjusted
	ECB Interest Rates For Main Refinancing Operations	https://www.quandl.com/data/BUNDESBANK/BBK01_SU0202-Ecb-Interest-Rates-For-Main-Refinancing-Operations-End-Of-Month
	Industrial Production EU Area	https://www.quandl.com/data/BUNDESBANK/BBXE1_M_I8_W_PROD_NS0020_IND_I00-Industrial-Production-Total-Industry-excluding-Construction-Index-Calendar-Adjusted-Only-Euro-Area-19
		https://www.quandl.com/data/BUNDESBANK/BBXE1_M_I6_W_PROD_NS0020_IND_I00-Industrial-production-Total-industry-excluding-construction-Index-Calendar-adjusted-only-Euro-area-17-fixed-composition
OECD	Future Tendency, Lithuania, Balance, S.A.	https://www.quandl.com/data/OECD/MEI_BTS_COS_BVEMFT_LTU_BLSA_M-Future-Tendency-Lithuania-Balance-S-A
	European Union (28 Countries), Hicp	https://www.quandl.com/data/OECD/PRICES_CPI_EU28_CPHPTT01_IXOB_A-European-Union-28-Countries-Hicp-All-Items-Index-Annual
	Total manufacturing, s.a., Estonia	https://www.quandl.com/data/OECD/KEI_PRMNTO01_EST_ST_A-Total-manufacturing-s-a-Estonia-Level-ratio-or-index-Annual
	Total manufacturing, s.a., Lithuania	https://www.quandl.com/data/OECD/KEI_PRMNTO01_LTU_ST_A-Total-manufacturing-s-a-Lithuania-Level-ratio-or-index-Annual
	Tendency, Estonia, Balance, S.A.	https://www.quandl.com/data/OECD/MEI_BTS_COS_BCBUTE_EST_BLSA_M-Tendency-Estonia-Balance-S-A

Source	Description	Data Source API
OECD	Tendency, Latvia, Balance, S.A.	https://www.quandl.com/data/OECD/MEI_BTS_COS_BVDETE_LVA_BLSA_M-Tendency-Latvia-Balance-S-A
	Future Tendency, Estonia, Balance, S.A.	https://www.quandl.com/data/OECD/MEI_BTS_COS_BSSPFT_EST_BLSA_M-Future-Tendency-Estonia-Balance-S-A
	Future Tendency, Latvia, Balance, S.A.	https://www.quandl.com/data/OECD/MEI_BTS_COS_BSSPFT_LVA_BLSA_M-Future-Tendency-Latvia-Balance-S-A
	Total manufacturing, s.a., Latvia	https://www.quandl.com/data/OECD/KEI_PRMNTO01_LVA_ST_A-Total-manufacturing-s-a-Latvia-Level-ratio-or-index-Annual
	European Union (28 Countries), Final Consumption Expenditure Of Households, Constant Prices, Oecd Base Year	https://www.quandl.com/data/OECD/SNA_TABLE1_EU28_P31S14_VOB-European-Union-28-Countries-Final-Consumption-Expenditure-Of-Households-Constant-Prices-Oecd-Base-Year
	Tendency, Lithuania, Balance, S.A.	https://www.quandl.com/data/OECD/MEI_BTS_COS_BRBUTE_LTU_BLSA_M-Tendency-Lithuania-Balance-S-A
NASDAQ OMX Global Index Data	Volatility NASDAQ - 100 (VOLNDX)	https://www.quandl.com/data/NASDAQOMX/VOLNDX-Volatility-NASDAQ-100-VOLNDX
World Bank World Development Indicators	Foreign direct investment, net inflows (% of GDP) - Estonia	https://www.quandl.com/data/WWDI/EST_BX_KLT_DINV_WD_GD_ZS-Foreign-direct-investment-net-inflows-of-GDP-Estonia
	Foreign direct investment, net inflows (% of GDP) - Latvia	https://www.quandl.com/data/WWDI/LVA_BX_KLT_DINV_WD_GD_ZS-Foreign-direct-investment-net-inflows-of-GDP-Latvia
	Foreign direct investment, net inflows (% of GDP) - Lithuania	https://www.quandl.com/data/WWDI/LTU_BX_KLT_DINV_WD_GD_ZS-Foreign-direct-investment-net-inflows-of-GDP-Lithuania
Uncertainty data	Monthly EPU Indices	https://www.policyuncertainty.com/media/All_Country_Data.xlsx
	Global Economic Policy Uncertainty Index	https://www.policyuncertainty.com/media/Global_Policy_Uncertainty_Data.xlsx
		https://www.policyuncertainty.com/media/Global_Annotated_Series.pdf

Source	Description	Data Source API
Uncertainty data	US Policy Categories	https://www.policyuncertainty.com/media/Categorical_EPU_Data.xlsx
	World Uncertainty Index (WUI)	https://worlduncertaintyindex.com/data/
	Financial Stress Indicator	https://www.policyuncertainty.com/media/Financial_Stress.xlsx
	Firm-Level Political Risk	http://www.firmlevelrisk.com/download
	Geopolitical Risk Index	https://www.matteoiacoviello.com/gpr_files/gpr_web_latest.xlsx
	Economic Uncertainty Related Queries (EURQ)	https://www.dropbox.com/s/c6yp22weychlobn/EURQ_data.xlsx?dl=0
		https://www.dropbox.com/s/ie7d8frwdtv01za/EURQ_paper.pdf?dl=0
	Global Economic Policy Uncertainty Index	https://www.policyuncertainty.com/media/Global_Policy_Uncertainty_Data.xlsx

Source: The Author's representation

MYKOLO ROMERIO UNIVERSITETAS

Sergei Kornilov

SISTEMŲ, PADEDANČIŲ PRIIMTI SPRENDIMUS, ORGANIZACINIO EFEKTYVUMO VERTINIMAS ESANT MAKROEKONOMINIAM NEAPIBRĖŽTUMUI: APIBENDRINTI MOKYMOSI METODŲ ALGORITMAI TAIKANT DVIEJŲ PAKOPŲ NEPARAMETRINIUS EFEKTYVUMO MODELIUS

Daktaro disertacijos santrauka
Socialiniai mokslai, ekonomika (S 004)

Vilnius, 2020

Mokslo daktaro disertacija rengta 2011-2017 metais Talino technologijos universitete (Estijos Respublika) ir 2019-2020 metais Mykolo Romerio universitete pagal Vytauto Didžiojo universitetui su ISM Vadybos ir ekonomikos universitetu, Mykolo Romerio universitetu, Šiaulių universitetu Lietuvos Respublikos švietimo, mokslo ir sporto ministro 2019 m. vasario 22 d. įsakymu Nr. V-160 suteiktą doktorantūros teisę.

Disertacija ginama eksternu.

Mokslinė konsultantė:

Prof. dr. Asta Vasiliauskaitė (Mykolo Romerio universitetas, socialiniai mokslai, ekonomika S 004).

Mokslinė vadovė Talino technologijos universitete 2011-2017 metais:

Prof. dr. Tatjana Põlajeva (Talino technologijos universitetas, Estijos Respublika, socialiniai mokslai, ekonomika S 004).

Mokslo daktaro disertacija ginama Vytauto Didžiojo universiteto, ISM Vadybos ir ekonomikos universiteto, Mykolo Romerio universiteto, Šiaulių universiteto Ekonomikos mokslo krypties taryboje:

Pirmininkė:

prof. dr. Violeta Pukelienė (Vytauto Didžiojo universitetas, socialiniai mokslai, ekonomika S 004).

Nariai:

prof. dr. Natalja Lace (Rygos technikos universitetas, Latvijos Respublika, socialiniai mokslai, ekonomika S 004);

prof. habil. dr. Žaneta Simanavičienė (Mykolo Romerio universitetas, socialiniai mokslai, ekonomika S 004);

doc. dr. Sigita Urbonienė (Vytauto Didžiojo universitetas, gamtos mokslai, matematika N 001);

doc. dr. Inga Žilinskienė (Mykolo Romerio universitetas, technologijos mokslai, informatikos inžinerija T 007).

Mokslo daktaro disertacija bus ginama viešame Ekonomikos mokslo krypties tarybos posėdyje 2020 m. birželio 30 d. 12 val. Mykolo Romerio universiteto I-414 aud.

Adresas: Ateities g. 20, 08303 Vilnius.

Temos aktualumas. Šiuolaikinei ekonomikai būdingas didėjantis informacijos, renkamos sprendimų priėmimo procesui, srautas, auganti pasaulinė makroekonominio lygmens konkurencija ir riboti fiziniai ištekliai. Sunku paneigti, kad XXI a. informacijos ir žinių vaidmuo tampa vis svarbesnis. Laikui bėgant vis sudėtingėjančiame sprendimų priėmimo procese gali kilti vienas svarbus susirūpinimą keliantis klausimas. Sprendimų priėmimo proceso rezultatas turėtų būti pats efektyviausias sprendimas, reiškiantis, kad prireiks minimaliai paskirstyti išteklius ir bus gauta didžiausia išeiga *ceteris paribus*. Taigi, sprendimo priėmimo procese itin svarbus vaidmuo tenka efektyvumo vertinimui. Pasauliniu mastu daugėjant konkurentų, blogo sprendimo kaina tampa didžiulė. Šiuolaikinėje ekonomikoje technologijų pažanga tampa svarbiu konkurencinį pranašumą lemiančiu veiksniu. Plėtojantis technologijoms, labai sparčiai didėja alternatyviųjų sąnaudų svarba. Pavyzdžiui, nagrinėjant besiformuojančios rinkos ekonomikos šalies, galinčios moderniųjų technologijų srityje pasauliniu mastu pasiekti technologinę pažangą situaciją, sunku padaryti kokią nors patikimą prielaidą neturint pagrindinės informacijos ir ekspertinės patirties. Siedami išvengti galimų alternatyvių išlaidų sprendimų priėmėjai ieško sudėtingesnių, bet kartu patikimų prognozavimo metodų. Pagrindine mokslinio tyrimo problematika šiame kontekste tampa įvairiarūšių ūkio subjektų sprendimų priėmimas esant neapibrėžtumo veiksniams.

Autorius visoje disertacijoje teigia, kad nė vienas iš pirmiau paminėtų klausimų neturėtų būti nagrinėjamas atskirai. Šios prielaidos pagrindimą ekonomikos srities mokslinėje literatūroje galima rasti jau seniai. Bet tik per praėjusius dešimtmečius moksliniai metodai, sustiprinti algoritmų mokymosi technologijomis, galėjo padėti rastis lanksčiai analitinei neapibrėžtumo vertinimo įvairiais lygmenimis struktūrai ir nelinijiškumą procesuose nagrinėti taikant ne tik bendruosius mokslinius metodus, pagrįstus bendrosiomis lūkesčių prielaidomis. Dėl finansinių technologijų inovacijų tampa įmanoma ekonomikos augimo idėją ir jos ryšį su inovacijomis paaiškinti iš kitos perspektyvos, t. y. pagal jų gebėjimą sukurti arba panaudoti technologines inovacijas esant neribotai technologinei pažangai.

Vertinant teoriškai, šiuolaikinės ekonomikos visumą sudaro daugybė mažesnių sudėtingų poaibių. Ankstesniame straipsnyje Kornilov and Polajeva (2016) jau nagrinėjo sudėtingą ekonominių procesų prigimtį. Tyrimai parodė, kad padidėjęs ekonominių procesų sudėtingumas esant neapibrėžtumui padidina rizikos veiksnius. Kiekvienas ekonominis poaibis yra modulinis, t. y. jį sudaro daug funkciniu požiūriu savitų dalių. Jis yra atviras, nes toms dalims būdinga tam tikro laipsnio laisvė.

Bet koks pasirinktas mokslinis požiūris turėtų padėti atpažinti ekonominių veiksnių prigimtį, kuri yra nevienalytė, o tai padidina ekonominių procesų sudėtingumą. Tam reikia sudėtingesnių mokslinių metodų, kurie turėtų paaiškinti kolektyvinių žinių veiksnių individualumą, kuriant bendrą rezultatą ir individualią reakciją į bendrą rezultatą. Tai reikalauja daugiau sudėtingų mokslinių metodų, kurie turėtų paaiškinti agentų individualumą kolektyvinių žinių atžvilgiu, kuriant bendrą rezultatą ir individualią reakciją į bendrąjį rezultatą. Šiuolaikinėje mokslinėje literatūroje atkreipiamas dėmesys į svarbius ekonominių tyrimų klausimus, įskaitant erdvinę integraciją ir ekonominį kompleksiškumą. Ekonominiai poaibiai tapo vis labiau siejami su plačiai žinoma žinių ekonomikos sąvoka.

Šiuolaikiniam verslui būdingi greiti ir radikalūs pokyčiai, nulemti informacijos pasiekiamumo. Šiuolaikinį ekonomikos mokslą dabar krečia neramumai (Chuen and Linda (2018),

Dunis *et al.* (2016)). Pastarojo meto finansinio automatizavimo tendencijomis galima paaiškinti tuo, kad vis labiau pereinama prie kompiuterinių programų taikymo prognozuojant, modeliuojant finansų rinkas bei informaciją ir prekiaujant rinkose. Ateityje mokslas turės analizuoti naujas kriptovaliutų ir skaitmeninių finansų tendencijas, bet jau iš dabar turimų įrodymų matyti, kad tai tampa esminiu reiškiniu. Tai jau tapo rezultatu, kurį galima pastebėti, kai pelno motyvai susilieja su socialiniais tikslais ir formuojasi didelių įmonių, susijusių su finansų technologijomis, grupė. Nūdienoj tarp sektorių vyksta intensyvūs technologiniai mainai, taigi pagrindinės inovacijos vienu metu gali būti taikomos labai įvairiose pramonės šakose. *Technologinė konvergencija* – tai dar vienas svarbus veiksnys, palyginti naujas ekonomijos moksle, ir neabejotinai ateityje tai taps išsamių tyrimų dalyku.

Temos aktualumą patvirtina ir ES valstybių narių, siekiančių užtikrinti didesnį šalių ekonomikos efektyvumą ir pašalinti ekonominius skirtumus, integracijos procesai. Integracijos plėtrą sudaro pagrindinė globalizacijos dinamika rinkų ir kapitalo požiūriu, taip pat judėjimas link glaudesnio tarptautinio bendradarbiavimo toliau plėtojant profesines sąjungas ir koordinuojant politiką. Tai rodo skirtingus ekonominės integracijos procesų režimus, kai tarp valstybių narių pašalinamos bet kokios rūšies sienos ir prekiaujant su kitomis valstybėmis, nesančiomis narėmis, laikomasi bendros ekonominės politikos bei struktūros. Todėl efektyvumo vertinimas esant neapibrėžtumui politikos formuotojams yra didžiausio prioriteto užduotis.

Efektyvumo įvertinimas daugiausia priklauso nuo patikimų įrodymų, kuriems įtakos turi netikrumas. Yra daugybė skirtingų metodų, kaip įvertinti ekonominių procesų ekonominį sudėtingumą, neapibrėžtumą kaip ekonominių procesų efektyvumo veiksnį. Tačiau tyrėjų tikslai šioje srityje yra skirtingi ir nėra apibendrinti vieno metodo rėmuose. Tyrimai rodo, kad netikrumas yra ekonominių procesų nestabilumo veiksnys, be to, jis ryškiausiai pasireiškia krizių metu. Dėl čia nurodytų priežasčių efektyvumo įvertinimas esant neapibrėžtumui yra svarbus tiek teoriniu, tiek empiriniu aspektais. Šioje daktaro disertacijoje nagrinėjami abu aspektai.

Naujausiuose Onatski and Williams (2003) tyrimuose pritariama, kad neapibrėžtumas yra nuolatinis ekonomikos reiškinys, ir politikai turi į jį nuolat atkreipti dėmesį. Black *et al.* (2018), Meinen and Röhe (2017) pritaria, kad vertinti makroekonominį neapibrėžtumą ir suprasti jo poveikį ekonominei veiklai yra svarbu vertinant dabartinę makroekonominę situaciją. Vertinant iš šiuolaikinės pozicijos, stiprus ir neigiamas neapibrėžtumo poveikis ekonomikos augimui yra akivaizdus, ir teorijoje šių padarinių negalima nevertinti (Lensink *et al.* (1999), Levin *et al.* (2005), Ljungqvist and Sargent (2012)). Atlikta daugybė tyrimų, kuriuose įrodinėjami neapibrėžtumo rodikliai, kuriuos galima laikyti tipiškais konkrečios politikos įrodymais, įtraukiant daug tiesioginių ir netiesioginių lygių dalyvių (Ericsson *et al.* (1999), Benhabib *et al.* (2013), Bird *et al.* (2013), Jurado *et al.* (2013), Ernst and Viegelahn (2014), Baker *et al.* (2015), Jurado *et al.* (2015)). Neapibrėžtumo veiksnys yra toks platus, kad manoma, jog politikos sprendimų poveikis ekonomikai yra nevienareikšmiškas. Šioje situacijoje bet kokia patikima kompetencija neapibrėžtumo prigimties klausimais būtų labai naudinga. Siekiant suprasti, kaip neapibrėžtumo lygio kitimas gali daryti poveikį ekonomikos procesui, svarbu rasti neapibrėžtumo šaltinį.

Iš daugybės tyrimų matyti, kad efektyvumo analizė tapo svarbia operacijų tyrimų, vie-

šosios politikos, energetikos bei aplinkos valdymo ir regioninės plėtros tema. Taigi akivaizdu, kad dviejų pakopų neparametriniai metodai pastarojo meto mokslinėje literatūroje, susijusiose su efektyvumo vertinimu, buvo plačiai taikomi. Empiriniuose tyrimuose pasirenkama viena vertinimo metodų grupė.

Todėl temos aktualumas pagrindžiamas tuo, kad disertacijos teoriniai ir praktiniai rezultatai atskleidžia algoritmų mokymosi metodų taikymo atliekant efektyvumo vertinimą esant neapibrėžtumui galimybes, ir būsimi politikos formuotojai gali tuo naudotis.

Antra, esama akivaizdaus poreikio prognozuoti heteroskedastiškumo įtaką pasauliniu mastu už ES ribų ir ES viduje. Sprendžiant tokias problemas reikia taikyti mechanizmą, kuriame persipynę tarpvalstybiniai komponentai, turintys daug laisvės, susiejami į vieną sistemą, kurios parametrai yra ekonominiai įvesties duomenys ir išvesties duomenys ir kurioje gebama suvaldyti neapibrėžtumą kitu lygmeniu: ribotos informacijos, apriboto racionalumo ir jų lūkesčių, taip pat chaotiškumo.

Trečia, ir, svarbiausia, efektyvumo įvertinimas esant neapibrėžtumo veiksniams vaidina pagrindinį vaidmenį šiuolaikinėje ekonomikoje, kurią lemia naujovės ir žinios su augančiu sudėtingumo lygiu. Atsižvelgiant į veiksnius, svarbu sukurti teorinį požiūrį, kuris apimtų kuo daugiau parametrų. Taigi, atliekant bet kokius neapibrėžties veiksnių efektyvumo įvertinimo tyrimus turėtų būti suteikta daugiau galimybių parametruoti modelį ir jie neturėtų apsiriboti tikslinėmis šalimis, bet į tyrimą turėtų būti analizuojami klasteriai peržengiantys nustatytas ribas. Ekonominė raida turėtų būti nagrinėjama ne tik ekonomine prasme, bet taip pat turėtų apimti ekonominių subjektų keitimąsi žiniomis. Prie analizės pridėjus neapibrėžtumo veiksnius, iškyla specifinis klausimas apie socialinių procesų vaidmenį tyrimuose.

Mokslinė problematika ir problemos ištyrimo lygis. Ekonomikos moksle esama daug ir labai įvairių puikų darbų, kuriuose vertinamas neapibrėžtumas. Eksperimentais grindžiamų modelių, parengtų remiantis naujausiais stebėjimais, priešakinėje linijoje rikiuojasi Elder (2004), Kontonikas (2004), Daal *et al.* (2005), Fountas (2010), Fountas (2010), Henry *et al.* (2007), Neanidis and Savva (2011). Tyrimuose neapibrėžtumui sukurti paprastai naudojamos deterministinės paradigmos. Šioje kategorijoje dominuoja autoregresinių sąlyginio heteroskedastiškumo modelių, tiek su klaidų variacija, tiek bendrąja forma nustatančių sąlygines autoregresines klaidas, siejamas su neapibrėžtumu, šeima. Metodologinius klausimus, susijusius su neapibrėžtumo vertinimu, kėlė Giordani and Söderlind (2003), Diebold *et al.* (1997), Clements and Harvey (2011). Walker *et al.* (2003) pasiūlė klasifikacijos fundamentines nuostatas. Berument *et al.* (2009) ir Hartmann and Herwartz (2012) tyrimuose standartinę prielaidą išplėtė įtraukdami atsitiktinio kintamumo modelius. Orlik and Veldkamp (2014), taip pat Glass and Fritsche (2015) tvirtina, kad neapibrėžtumas yra neciklinių neapibrėžtumo pokyčių, esant sukrėtimams, rezultato vertė. Zarnowitz and Lambros (1987), Bomberger (1996), Rich and Butler (1998), taip pat D'Amico and Orphanides (2008) episteminį neapibrėžtumą įrodinėja tiesiogiai apskaičiuodami parametrinį pasiskirstymą tarp individų. Lahiri and Sheng (2010), Siklos (2013), Lahiri *et al.* (2015) modelį išplečia daugelį dalykų pataisydami ir pakeisdami. Walker *et al.* (2003), Dequech (2004) nagrinėja episteminį neapibrėžtumą, kurį sukelia neišsamios ekspertų žinios ir

kintamumo neapibrėžtumas, priskirtinas atsitiktiniams atsitiktinai atsirandantiems veiksniams. Lane and Maxfield (2004) kintamumo neapibrėžtumą išplečia įtraukdami ontologinį neapibrėžtumą. Diskusijoje, pradėtoje dėl Walker *et al.* (2003) ekonominių neapibrėžtumų klasifikacijos, pereinama prie infliacijos neapibrėžtumo, nagrinėjamo Norton (2006), Kowalczyk (2013), Krayer von Krauss *et al.* (2019). Infliacijos neapibrėžtumas daro didelę įtaką ekonominių modelių sudarymui. Tai ypač aktualu modeliuojant sprendimus, paremtus tokia analize.

Gelman and Hill (2007) pradeda naudoti daugialygį ir apibendrintą linijinį modelį, kuriame parametrams taikomas tikimybės modelis. Jordà *et al.* (2013), Knüppel (2014) siūlo apsvarstyti neapibrėžtumo įtraukimą į racionalaus ekspertų prognozavimo modelius ar jų derinius su tokiais modeliais, kurie nebūtinai turi matematinį ir ekonometrinį prognozių pagrindimą, tačiau remiasi rizikos veiksnių vertinimu, pagrįstu postprognozavimo klaidų paskirstymu.

Mokslinių tyrimų gausa patvirtina neparametrinio efektyvumo vertinimo svarbą. Seiford (1997) paskelbtoje apžvalgoje kalbama apie 800 publikacijų, o jau vėlesnėje paskelbtoje Seiford (2005) publikacijoje, nurodoma jau apie 2800 paskelbtų straipsnių apie DEA metodikos naudojimą. Pradedant pagrindiniais Farrell (1957), Koopmans (1952), Aigner ir Chu (1968), Aigner ir kt. (1977), Broek ir kt. (1980) darbais efektyvumo metodikos koncepcija vertinant gamybos funkcijas tapo plačiai paplitusi. Šis klausimas buvo kruopščiai ištirtas iš visų pusių. Todėl yra daugybė kritinių darbų, nurodančių pagrindinius tradicinių efektyvumo vertinimo metodų trūkumus. Sexton ir kt. (1986), o po to Smith (1997) atskleidė neteisingų specifikacijų įtaką; Pillar (1990) apibendrino, kad technologijų nevienalytiškumas organizacijose, neapibrėžtumas pasirenkant įvestis ir išvestis gali paveikti veiklos vertinimo objektyvumą.

Tačiau, Tobback *et al.* (2018) tvirtina, kad vertinant neapibrėžtumą taikomi įprasti metodai, sukurti Baker *et al.* (2015), negali prognozuoti jokių kintamųjų, o algoritmų mokymosi metodas lenkia tradicinius ARCH grindžiamus metodus. Brose *et al.* (2014a) ir Brose *et al.* (2014b) tvirtina, kad rizikos ir neapibrėžtumo valdymas labai priklauso nuo informacijos. Pastarąjį dešimtmetį buvo atlikta keletas tyrimų, išsamiau nagrinėjusių optimizavimo algoritmų, pagrįstų linijinio programavimo modeliu, naudojimą siekiant nustatyti kontrolės priemones, kurias reikia testuoti, siekiant valdyti riziką, ir tai gali būti plėtojama kaip hibridiniai efektyvumo klasifikavimo metodai (Pareek (2006), H.-Y. Kao *et al.* (2013)). Įvairūs linijinio optimizavimo būdai buvo sėkmingai taikyti prognozuojant laiko eilutes ir jų bendrą judėjimą (Kara *et al.* (2011), Karaa and Krichene (2012)).

Todėl naujausiuose tyrimuose algoritminis mokymasis naudojamas vertinant ir neapibrėžtumą, ir efektyvumą. Įvairių algoritmų mokymosi būdų, pavyzdžiui, neuroninių tinklų, prognozavimo galia literatūroje plačiai patvirtinta, o jų praktinį poveikį nustatė Alejo *et al.* (2013). Attigeri *et al.* (2017) tvirtina, kad empirinio požiūrio laikomasi kuriant modelius, taikomus vertinant finansinę riziką, kai naudojami mokymosi algoritmai. Iš Kruppa *et al.* (2012), Kreienkamp and Kateshov (2014), Addo *et al.* (2018) rezultatų matyti, kad nelinijiniai būdai itin pasiteisina, kai modeliuojama tikėtina vertė. Per kelerius praėjusius metus daug tyrėjų naudojo algoritminio mokymosi būdą ir neparametrinį būdą kurdami naujus metodus efektyvumui prognozuoti atliekant duomenų analizę (Xu and Wang

(2009), L. Zhou *et al.* (2014), X. Yang and Dimitrov (2017), Zelenkov *et al.* (2017), Alaka *et al.* (2018)). Q. Zhang and Wang (2018) pasiūlė efektyvumo prognozavimo modelį, kuris pirmą kartą apima prižiūrimą mokymąsi analizuojant informaciją pagal neparametrinį modelį, siekiant įvertinti sprendimų priėmimo padalinio būsimą efektyvumą.

Mokslinė problema. Šio mokslinio tyrimo tikslas yra efektyvumo vertinimo modelio, kuris grįstas ekonomikos mokslu, ir kuriame išnaudoti algoritminio mokymosi pranašumai siekiant gauti teisingą ir patikimą rezultatą, sukūrimas. Po nuodugnios mokslinės literatūros apžvalgos ir praktinio modelių taikymo analizės, efektyvumo įvertinimo klausimas buvo vienareikšmiškai apibrėžtas kaip tolesnių tyrimų užduotis, atskleidžiant neapibrėžtumų, atsirandančių dėl ekonominių procesų sudėtingumo ir netiesiškumo, įtraukimo į sprendimų priėmimo procesą galimybes. Anksčiau daugelis mokslinių tyrimų suteikė vertingų ir išsamių žinių ekonomikos mokslui, kad galėtų atskleisti ekonominius procesus ir ekonominių agentų dalyvavimo vaidmenį. Įtakingiausi mokslininkai buvo apdovanoti Nobelio ekonomikos premija už jų reikšmingą indėlį į elgesnos ekonomiką esant ribotam racionalumui. Tačiau vis dar yra spragos tarp teorinių išvadų ir praktinio ekonominių agentų ekonominio efektyvumo įvertinimo priimant sprendimus netikrumo sąlygomis.

Tyrimo objektas - neapibrėžtumo ir efektyvumo veiksniai, taikomi sprendimų priėmimo sistemose, vertinant organizacijų efektyvumą neapibrėžtumo sąlygomis, ir naudojant mokymo algoritminio mokymosi metodus.

Tyrimo tikslas - atskleidus tarpdisciplinį požiūrį į ekonominio kompleksiškumo, neapibrėžtumo ir efektyvumo veiksnius, parengti efektyvumo vertinimo esant neapibrėžtumui metodiką ir išbandyti ją naudojant finansų srities duomenų rinkinius.

Tyrimo uždaviniai:

1. Atlikti empirinį efektyvumo vertinimo tyrimą, taikant bendrus algoritmų mokymosi būdus, grindžiamus hibridiniu modeliu.
2. Atskleisti neapibrėžtumo esmę ir šaltinius ir išanalizuoti juos siūlomuose efektyvumo vertinimo metoduose.
3. Išnagrinėti efektyvumą kaip ekonominę koncepciją ir išanalizuoti siūlomus efektyvumo įvertinimo metodus, naudojant algoritminio mokymosi metodus.
4. Pasiūlyti konceptualų efektyvumo vertinimo esant neapibrėžtumui modelį, taikant linijinio optimizavimo metodus ir algoritmų mokymąsi.
5. Atlikti empirinį efektyvumo vertinimo tyrimą, taikant bendrus algoritmų mokymosi būdus, grindžiamus hibridiniu modeliu.
6. Apibūdinti efektyvumo vertinimo esant neapibrėžtumo sąlygoms empirinio tyrimo rezultatus, kad būtų galima pasiūlyti jų taikymo rekomendacijas.

Tyrimo metodai. Autoriaus taikomi tyrimo metodai – tai mokslinės literatūros analizė, sintezė ir lyginimas siekiant apibūdinti neapibrėžtumą ir efektyvumą. Analizei naudojama praktinė programinė įranga R (The R Project for Statistical Computing). Dideliam duo-

menų kiekiui tvarkyti naudota duomenų bazė „Oracle 12c“, kad būtų galima naudoti SQL duomenų rinkinius analizei programinėje įrangoje R atlikti. Duomenys buvo gauti iš pirminių duomenų šaltinių per „WebServices“ arba per CSV analizę, esančią duomenų bazėje.

Tyrimo duomenys ir jų šaltiniai. Tyrime nagrinėjamas atrinktų bendrovių, įtrauktų į *Nasdaq* Baltijos biržos rinkos indeksus, efektyvumas. Neapibrėžtumo duomenų rinkiniai yra iš daugelio šaltinių:

1. Ekonominės politikos neapibrėžtumo indeksas – jis siejamas su žiniasklaidos komentarų dažnumu, taip pat apibrėžiamas, kaip daugelio ekonominių rodiklių neprognozuojamos sudedamosios dalies bendras kintamumas.
2. Konkrečiai šaliai būdingi veiksniai turėtų apimti rinkos koncentraciją, užsienio investicijų buvimą, fiskalinius rodiklius. Sparčiai kintančioje verslo aplinkoje besiformuojanti darbo aplinka, gebėjimas prognozuoti ateities tendencijas ir poreikius, susijusius su žiniomis ir įgūdžiais, tampa labai svarbūs, norint užtikrinti veiksmingą sistemą, padedančią priimti sprendimus. Šios tendencijos kinta atsižvelgiant į geografiją ir pramonės šakas, taigi svarbu numatyti pramonės šaliai ir konkrečiai šaliai būdingus kintamuosius. Šaltiniai:
   – „Federal Reserve Economic Data“
   – „Deutsche Bundesbank Data Repository“
   – „Organization for Economic Co-operation and Development“
   – „NASDAQ OMX Global Index Data“
   – Pasaulio banko Pasauliniai išsivystymo rodikliai
3. Organizaciniai duomenų rinkiniai, kuriuos pateikė NASDAQ OMX.

Tyrimo apribojimai. Keli tyrimo ribotumo atvejai rodo, kad tyrimo metodikos naudojimo rezultatai labai priklauso nuo duomenų kokybės. Vis dėlto, rezultatų svarba dėl to nesumažėja nei teoriniu, nei praktiniu lygmeniu. Teoriniu lygmeniu ši tyrimo metodika, parengta atsižvelgiant į nustatytą tyrimų spragą, yra vienas pirmųjų bandymų išsamiai įvertinti ne tik neapibrėžtumo ir efektyvumo veiksnius atskirai, bet ir rasti moderniausią būdą suprasti juos kaip visumą. Empiriniu lygmeniu šiame tyrime aprėpiama dauguma klausimų, susijusių su efektyvumo vertinimu esant neapibrėžtumui, atsižvelgiant į apribojimus, kurie tokiai analizei nustatomi pagal kiekvieną ūkio subjektų branduolį. Į siūlomą modelį įtraukti ne visi galimi veiksniai. Siūlomas modelis yra tik vienas galimas būdas patikimiems rezultatams pasiekti esant neapibrėžtumui. Tačiau tam tikti empiriniai duomenų rinkiniai aprėpia 10 metų laikotarpį. Makroekonominių veiksnių skaičius yra ribotas ir naudojami tik pagrindiniai rodikliai. Duomenų kaupimo lygmenyje taikyti pradiniai filtrai. Tai reiškia, kad duomenų rinkinyje nėra atsitiktinių dydžių. Realiomis aplinkybėmis, kur duomenų rinkiniai gaunami automatiškai, yra tikimybė, kad bus trūkstamų arba netinkamų dydžių.

Mokslinis tyrimo naujumas:

1. Šiuo tyrimu siekiama įtraukti teorinius ir empirinius neapibrėžtumo, netiesiškumo, sudėtingumo ir riboto racionalumo aspektus kaip pagrindinę prielaidą, kurios peržengia pusiausvyros teorijas. Ankstesnių tyrimų, pagrįstų pusiausvyros teo-

rija, analizė rodo, kad teorinė dalis dažnai yra atskirta nuo statistiškai reikšmingų rezultatų. Visuotinai žinoma, kad ekonomikos teorija remiasi pusiausvyros samprata, o tyrimai vykdomi subalansuoto augimo srityje. Todėl statistiniais duomenimis pagrindžiamos nusistovėjusios teorijos. Tačiau jau Brianas (2006) konceptualiai pabrėžė, kad į tradicinius pusiausvyros modelius reikia įtraukti elgsenos aspektus.

2. Efektyvumo vertinimas esant neapibrėžtumui apibrėžiamas įvairiais neapibrėžtumo šaltiniais, ir ne hibridiniuose modeliuose to negalima įvertinti kiekybiškai. Formaliu požiūriu, įvairus neapibrėžtumas, atsirandantis dėl trūkstamų duomenų, gali būti apibendrintas taikant ribotos informacijos hipotezę. Tačiau reikia pripažinti, kad daugelyje realiomis aplinkybėmis esančių duomenų rinkinių dėl įvairių priežasčių gali būti trūkstamų dydžių. Tokius duomenis įkeliant į modelį, kuriame yra trūkstamų dydžių, gali būti daromas didžiulis poveikis modelio kokybei. Siūlomame modelyje tiesiogiai nurodoma, ką daryti, kai yra trūkstamų duomenų, ir tam naudojami įvairūs algoritmų mokymosi būdai. Keletas tyrėjų primygtinai akcentuoja duomenų kokybę, o Lertworasirikul *et al.* (2002) parodo, kad neparametrinio modeliavimo metodams reikia tiksliai apskaičiuoti ir įvesties, ir išvesties duomenis. Visose mokslininkų ir neparametrinio modeliavimo specialistų apibūdintose situacijose tebesilaikoma palyginti subjektyvaus požiūrio į spragos, atsirandančios dėl trūkstamų duomenų, užpildymą. Bet šiame tyrime siūlomame modelyje trūkstamų duomenų traktavimas yra viena iš svarbių užduočių.
3. Šis tyrimas yra vienas iš nedaugelio, kuriame neapibrėžtumui apskaičiuoti taikomas struktūrinis maksimalaus rizikos sumažinimo principas, o algoritmų mokymosi būdais, užuot maksimaliai mažinus pastebėtas mokymosi klaidas, siekiama maksimaliai sumažinti apibendrinimo klaidos apribojimą, kad būtų pagerintas bendras veikimas.
4. Šis tyrimas yra vienas pirmųjų bandymų įvertinti efektyvumą taikant tiek klasifikavimo, tiek regresijos modelį. Autorius, kaip ir kiti tyrėjai, nagrinėja bendrus metodus, taikomus algoritmų mokymosi klasifikatoriuose, kai esama neapibrėžtų žinių rinkinių, ir parodo, kaip duomenų neapibrėžtumas žinių rinkiniuose gali būti traktuojamas taikant bendrus algoritmų mokymosi klasifikavimo metodus, panaudojant optimizavimą. Vadinasi, bendri algoritmų mokymosi metodai taip pat gali būti naudojami kaip regresijos metodas, išlaikant visus pagrindinius bruožus, kuriais apibūdinamas maksimalios maržos algoritmas. Autorius sutinka, kad algoritminio mokymosi ateitis priklauso nuo įvairių požiūrių derinio, nes visiškai kontroliuojami algoritmai yra naudingi, tačiau dėl paslėptų modelių kintamųjų gali būti apribojimų (Chen ir kt. (2013)).
5. Siūlomas modelis yra susijęs su efektyvumo įvertinimu nesant homogeniškumo, iškyla klausimą, kaip teisingai palyginti organizacijas tarpusavyje. Susijusi problema – ji plačiai nagrinėjama literatūroje – yra trūkstamų duomenų problema, sprendžiama tiesiogiai naudojant tinkamus algoritmų mokymosi būdus (Zhu (2016b)).
6. Autorius pirmas aiškiai pasiūlė neapibrėžtumą traktuoti ne kaip fiktyvų kintamąjį, bet kaip reiškinį, smulkiai analizuojamą siūlomame modelyje skirtingais lygmenimis: duomenų kaupimo neapibrėžtumo, analitinės struktūros neapibrėžtumo

ir neapibrėžtumo, kaip veiksnio. Skirtingai nei esami metodai, algoritmų mokymosi būdai, įtraukti į šį tyrimą, nereikalauja remtis hipotetine prielaida. Matematiniu požiūriu, taikant algoritmų mokymosi būdus nustatomi numanomi svorinių koeficientų apribojimai, taigi esama esminio skirtumo tarp šių metodų, ir tas skirtumas atsiranda iš to, kaip duomenys yra renkami. Kiekviename procese neapibrėžtumo mastas yra skirtingas, ir jis turėtų būti vertinamas atitinkamais būdais.

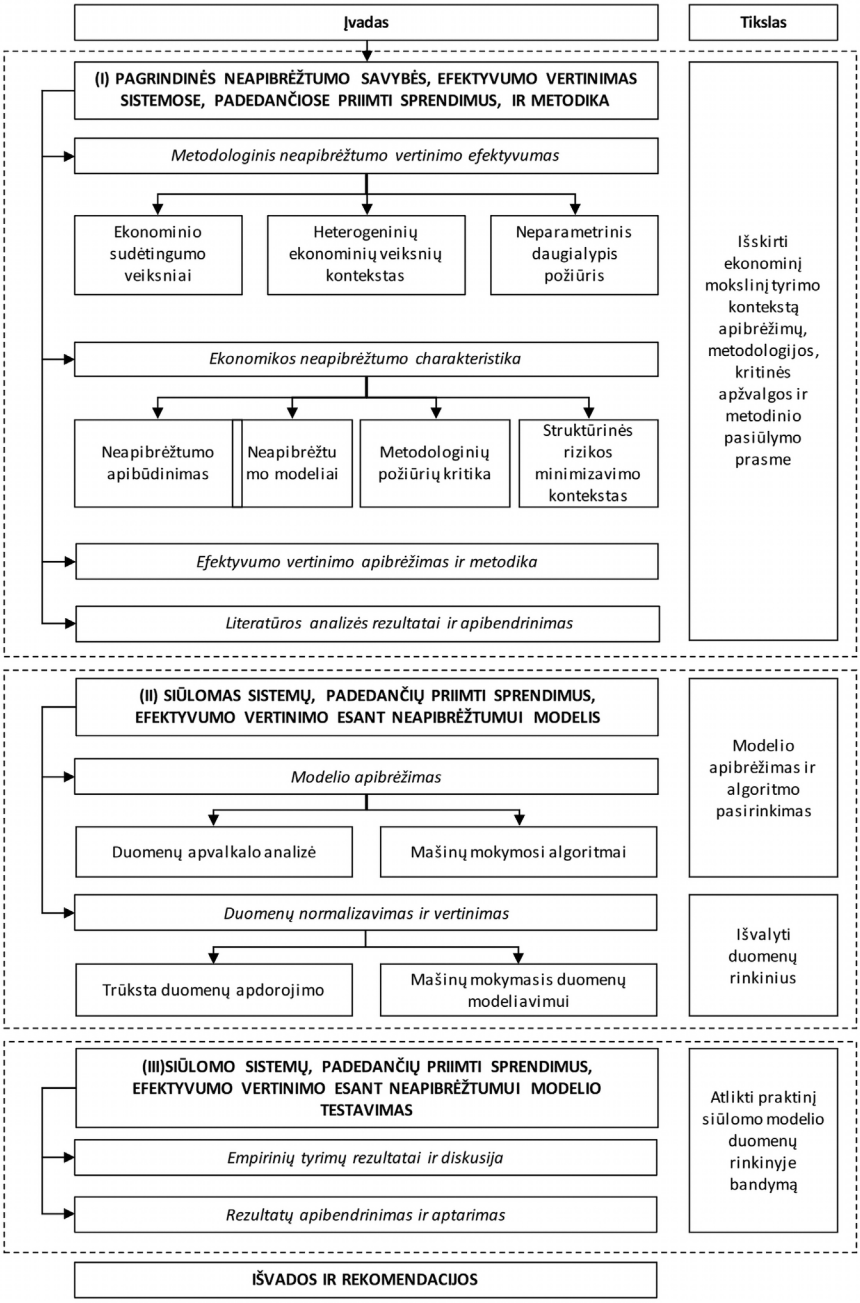
Praktinė tyrimo rezultatų svarba. Atliekant tyrimą gauti aiškūs rezultatai, susiję su automatizavimo procedūrų integravimu į bet kurias sprendimo priėmimo sistemas. Matyti, kad įmanoma kurti sistemas naudojant modulines funkcijas. Esama keleto pritaikymo sričių, kuriose galima taikyti sistemas, padedančias priimti sprendimus:

1. Verslo analitikos sistemos, sukurtos taip, kad duomenis galima greitai stebėti ir filtruoti pagal kelis skirtingus matmenis, siekiant iškart suvokti naujausius organizacinių padalinių veiklos rezultatus.
2. Duomenų kaupimo programos veikia naudodamos milžiniškus duomenų ir faktų rinkinius, sujungtus ir sukauptus vykstant sąveikai su kitomis sandorių šalimis ir aplinka. Šiems duomenų rinkiniams tenka svarbus vaidmuo atliekant statistinę analizę, kai svarbiausia yra gauti metainformaciją ir paslėptus duomenis, susijusius su naudingumu, tuo, kam teikiama pirmenybė, tendencijomis arba kitų susijusių subjektų elgsena.
3. Plataus masto įmonės išteklių planavimo programa suteikia galimybę formuoti organizacinę darbo eigą pagal efektyvumą, labiau sutelkiant dėmesį į kapitalo investicijas, atsargas, produkciją ir logistiką.

Tik struktūrizuota metodika, kai taikomi įvairūs metodai ir iš duomenų mokslo, ir iš ekonomikos tyrimų, gali padėti tyrėjams ir politikos formuotojams surikiuoti svarbius veiksnius ir geriau įvertinti giluminius veiksnius. Rezultatai negali būti vertinami tik statistiniu arba tik teoriniu požiūriu, juos galima vertinti tik taikant integruotą procesą, esant tinkamai struktūrizuotai sprendimų priėmimo sistemai.

Tyrimas yra grindžiamas istoriniais duomenų šaltiniais iš pirmiau nurodytų šaltinių, naudotasi API sąsaja, taip pat atsisiųsta iš atitinkamų šaltinių.

Disertacijos loginė struktūra. Disertaciją sudaro įvadas, trys dalys, išvados ir rekomendacijos, priedai. Disertacijos apimtis – 166 puslapiai. Joje yra 31 paveikslų, 27 lentelės, 418 literatūros šaltinių ir 20 priedų.



1 pav. *Disertacijos loginė struktūra*

DISERTACIJOS TURINIO APŽVALGA

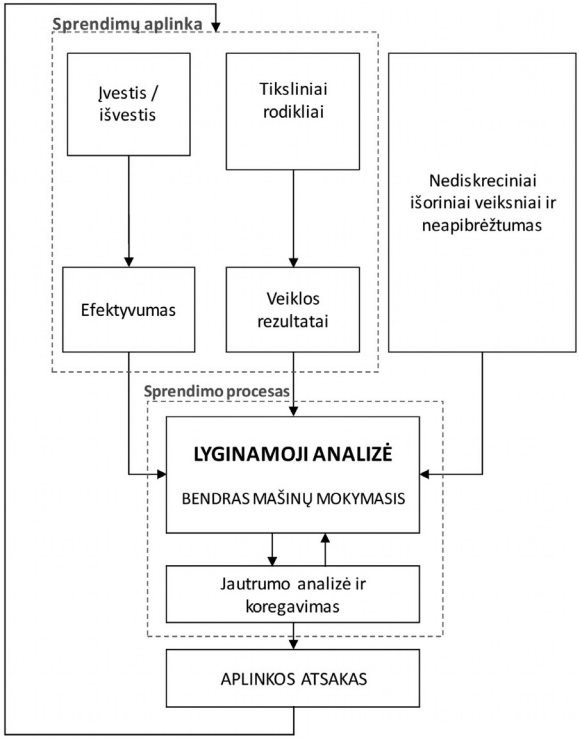
I. PAGRINDINĖS NEAPIBRĖŽTUMO SAVYBĖS, EFEKTYVUMO VERTINIMAS SISTEMOSE, PADEDANČIOSE PRIIMTI SPRENDIMUS, IR METODIKA

Šiame skirsnyje aprašoma rekursinio spendimų priėmimo proceso, kaip ekonomikos procesų stuburo, sąvokos plėtra. Pačioje ekonomikoje nuo labai ankstyvų etapų sutikta su pusiausvyros sąvoka ir iš esmės subalansuoto augimo trajektorijos tyrimu. Remiantis *ex post* įrodymais, ekonomistai pirmenybę teikė išgaubtosioms struktūroms, galinčiomis užtikrinti unikalią pusiausvyros būklę, kurią galima sieti su optimalaus augimo trajektorija. Bet ekonomikos ir informacijos mokslui darant pažangą, pusiausvyros sąvokos, esant linijiniam išteklių paskirstymui, yra neįmanomos dėl ūkio subjektų riboto racionalumo, viena vertus, ir ekonominio kompleksiškumo reiškinių, kartu su makroekonominiu neapibrėžtumu, kita vertus. Ūkio subjektų efektyvumo vertinimas, esant makroekonominiam neapibrėžtumui, yra dalis pasaulinio masto ekonomikos proceso. Sprendimų priėmimo efektyvumo prognozavimo esant neapibrėžtumui problematika kelia kompleksinės dinaminės sistemos klausimų, kuriais turėtų būti paaiškinti pagrindiniai visų šių komponentų sąveikos ir tarpusavio susietumo šablonai. Atliekant literatūros apžvalgą, apimančią senus ir naujesnius tyrimus, matomi ontologiniai veiksniai, kurie daro poveikį rezultatui esant neapibrėžtumui.

1.1. Sprendimų priėmimo kontekstas vykstant įvairiarūšiams ekonomikos procesams

Atlikus mokslinės literatūros analizę nustatyta, kad riboto racionalumo teorija yra solidus analitinis metodas, taikomas daugelyje įvairių sričių. Esant ekonominiam kompleksiškumui didėja komponentų sąveikos tvarka arba dėsningumas ir net susidaro nauja tvarka ir konfigūracija, įvairiems komponentams elgiantis autonomiškai. Taigi, susiformuoja radikaliai nauji procesai ir naujos komponentų sąveikos. Išanalizavus tyrimo literatūrą nustatyta, kad efektyvumo vertinimas esant neapibrėžtumo sąlygoms yra ekonominio sudėtingumo ir sprendimų priėmimo procesų nelinijiškumo rezultatas.

Figure 10 pateiktas *sprendimų priėmimo modelis, kurį galima dinamiškai koreguoti*. Modelis gerai dera su kitais rekursiniais modeliais, kuriuos taikant ūkio subjektai savo sprendimus priima reaguodami į aplinkos atsaką. Kiekvienoje organizacijoje yra tam tikrų disponuojamųjų įvesties duomenų, pageidaujamų išvesties duomenų ir numanomų tikslinių rodiklių, ir dėl jų pobūdžio į juos reikia atsižvelgti atskirai. Taigi, labai rekomenduojama taikyti dviejų pakopų efektyvumo neparametrinę analizę. Analizės rezultatai yra integruota sprendimų proceso dalis, ją galima sustiprinti taikant bendrus algoritmų mokymosi metodus.



2 pav. *Sprendimų priėmimo modelis*
(Šaltinis: sudaryta autoriaus)

Todėl *sprendimų priėmimo sistemų* apibrėžtis tinka organizacijoms, taikančioms skirtingus statistinio modeliavimo ir rezultatų lentelių modeliavimo verslo analitikos būdus, neuroninius tinklus, įvairias ekspertines sistemas, ūkio subjektais grindžiamas sistemas, neuroapytiksles sistemas, įvairiais atvejais grindžiamas sistemas arba paprasčiausiai gaires, parengtas remiantis duomenimis pagrįsta patirtimi.

Efektyvumo vertinimo, atliekamo sistemoje, padedančioje priimti sprendimus, procesą sudaro keletas procesų, ir tai reiškia, kad imamasi veiksmų skirtingais lygmenimis. Tyrime siūloma neapibrėžtumą traktuoti kaip reiškinį, smulkiai analizuojamą skirtingais lygmenimis: duomenų gavybos neapibrėžtumo, analitinės struktūros neapibrėžtumo ir neapibrėžtumo, kaip veiksnio. Dauguma metodų, į kuriuos buvo įtrauktas neapibrėžtumas, yra paremti tuo, kad daugiklio modelyje ribojami svoriniai koeficientai. Skirtingai nei esami metodai, algoritmų mokymosi būdai, įtraukti į šį tyrimą, nereikalauja remtis hipotetine prielaida. Matematiniu požiūriu, taikant algoritmų mokymosi būdus yra nustatomi numanomi svorinių koeficientų apribojimai, taigi esama esminio skirtumo tarp šių metodų, ir tas skirtumas atsiranda iš to, kaip duomenys aiškiai renkami. Kiekviename procese neapibrėžtumo mastas yra skirtingas, ir jis turėtų būti vertinamas atitinkamais būdais.

1.2. Organizacinio efektyvumo vertinimo metodinis kontekstas

Šiame skirsnyje išsamiai nagrinėjamas organizacinis efektyvumas ir sprendimų palaikymo sistemos, kurios įtraukia ekonominio sudėtingumo sąvoką. Ši sąvoka yra būtina norint išvengti klaidinančių apibendrinimų. Sprendimų palaikymo sistemose naudojami esami efektyvumo įvertinimo modeliai, pagrįsti įvesties-išvesties parametrais, kurie dabar plačiai naudojami ekonominės veiklos rezultatams įvertinti. Šie modeliai paprastai vertina išorinius kintamuosius ir išteklius, tačiau neanalizuodami pasikartojančio šalutinio poveikio. Ekonominės veiklos agentų požiūriu, tokia analizė labiau domina būtent tas organizacijas, kurios naudojasi materialiaisiais ištekliais, nes jų veikla gali būti nuolat keičiama ir veikiama įvairių veiksnių. Šiame tyrime siūloma pagerinti išorinių veiksnių poveikio vertinimo metodiką. Metodologiniame kontekste siūloma praktinė metodika. Taigi siūlomas metodas leis pagerins išorinio poveikio ekonominės veiklos agentams vertinimo modelį.

Standartiniai statistiniai būdai yra priemonės, padedančios atskirti sisteminius ir atsitiktinius veiksnius, taigi iš esmės turėtų būti įmanoma atskirti sprendimo racionalią adaptyvią dalį nuo riboto racionalumo. Todėl bet kuris prasmingas makroekonomikos modelis turėtų analizuoti ne tik individų charakteristikas, bet ir jų sąveikos struktūrą. Ūkio subjektais grindžiamo modeliavimo ir makroekonomikoje taikomo modeliavimo metodo pranašumas yra tas, kad nelieka išsprendžiamumo apribojimų, labai ribojančių analitinę mikroekonomiką. Ūkio subjektais grindžiamas modeliavimas ir modeliavimo metodas tyrėjams suteikia galimybę pasirinkti mikroekonomikos formą, tinkamą konkrečiu atveju, įskaitant ūkio subjektų rūšių platumą, kiekvienos rūšies ūkio subjektų skaičių ir ūkio subjektų hierarchinę struktūrą. Taip pat tyrėjams suteikiama galimybė atsižvelgti į ūkio subjektų sąveiką tuo pat metu kaip ir į ūkio subjektų sprendimus bei nagrinėti tarp ūkio subjektų vykstančią dinaminę makrosąveiką.

1.3. Neapibrėžtumo pagrindai ekonomikos teorijose

Ekonomikos teorijoje jau seniai atliekami tyrimai, susiję su tiek mikrolygmens, tiek makrolygmens reiškinių rizika bei neapibrėžtumu. Tyrime apibendrinta, kad neapibrėžtumas gali turėti įvairų ekonominį poveikį ūkio subjektams, ir to poveikio kanalai gali būti įvairūs. Visoje ekonomikos teorijoje išskiriami trys pagrindiniai jungtiniai ūkio subjektai, vykdantys įvairią ekonominę veiklą, įskaitant gamybą, vartojimą ir mainus, bet neapsiribojant tik tuo. Todėl nūdienoje ekonomikos teorija gali padėti spręsti sudėtingus klausimus, kaip ūkio subjektai, pavyzdžiui, gamintojai, namų ūkiai ir investuotojai, elgtųsi konkrečiomis aplinkybėmis. Tokiu pat būdu įmanoma patikrinti, kaip būtų paskirstomi ištekliai ir atsirastų rinkos netobulumo atvejų. Ekonominiu požiūriu esama diskusijų dėl neapibrėžtumo sąvokų apibūdinimo. Ekonominiuose tyrimuose dažniausia taikoma neapibrėžtumo sąvokos koncepcija pagal F. Knight. Vadovaujantis šia koncepcija neapibrėžtumas vyrauja ekonomikos tyrimuose kaip hipotezė, susijusi su visuma, ir jo negalima tiesiogiai išmatuoti. Neapibrėžtumas yra nestebimas, jo negalima tiesiogiai išmatuoti. Tačiau taikant modelius galima naudoti pakaitinius kintamuosius, kad būtų galima įvertinti jo pokyčius laikui bėgant.

1.4. Veiklos rezultatų, veiksmingumo ir efektyvumo klasifikavimas

Efektyvumo, veiklos rezultatų ir veiksmingumo skirtumas yra didžiulis, nors visais šiais terminais turėtų būti lyginami įvesties ir išvesties veiksniai pagal sunaudotus išteklius ir gautą rezultatą. Verslo veiklos rezultatai tyrime suprantami kaip verslo veiklos rezultatų vertinimas pagal Rappaport (1986), teigusį, kad akcininko vertė turėtų tapti pasauliniu standartu, taikomu vertinant verslo veiklos rezultatus. Po diskusijos išvardyti apskaitinės investicijų grąžos ir apskaitinės nuosavo kapitalo grąžos, kaip standartų, taikomų verslo veiklos rezultatams vertinti, trūkumai. Efektyvumą galima išmatuoti kaip veiksmingumo ir veiklos rezultatų derinį. Šiame tyrime nenagrinėjamas tiesioginis efektyvumo matavimas. Taigi, tyrime pirmiausia siekiama nustatyti verslo veiklos rezultatų ryšį finansiniu, ekonomikos augimo veiksnių ir sprendimų priėmimo proceso požiūriu.

Ekonometrinių metodų taikymas reikalauja konkrečios prielaidos dėl įvesties duomenų ir išvesties duomenų ryšio, ir apskaičiuoja tos funkcinės formos parametrus. Ekonometriniai būdai neturėtų būti vertinami deterministiniu arba tikimybiniu požiūriu. Taikant deterministinį ribų metodą tariama, kad visas nukrypimas nuo apskaičiuotų ribų daugiausia atsiranda dėl techninio neefektyvumo, o atsitiktiniai veiksniai jokio vaidmens neatlieka. Tačiau, skirtingai nei taikant deterministinį ribų metodą, taikant tikimybinį produkcijos ribų metodą į modelio specifikacijas įtraukiamas ir triukšmo, ir neefektyvumo komponentas.

1.5. Bendras algoritmų mokymosi metodas, taikomas sprendimų priėmimo procese

Tyrime nagrinėjami bendri metodai, susiję su algoritmų mokymusi, kai įvairūs būdai yra taikomi kartu, siekiant gauti kuo geresnį įvertį. Algoritmų mokymosi pranašumas yra tas, kad juo galima nustatyti apibendrintus pavydžius, nes esama galimybės atskleisti sudėtingas ir iš anksto nenumatytas struktūras. Algoritmų mokymasis gali tikti lanksčioms, bet sudėtingoms duomenų struktūroms, ir tada neprireikia persimokymo. Bendri algoritmų mokymosi metodai gali būti laikomi neparametriniu metodu, vertinant pagal parametrus, priklausančius nuo modelio, grindžiamo duomenimis, kad modelio pajėgumas atitiktų duomenų sudėtingumą.

Dėl savo principų algoritmų mokymosi būdai reikalauja didesnės abstrakcijos, palyginti su įprastais statistiniais būdais. Tačiau, vertinant ekonomikos mokslo požiūriu, algoritmų mokymuisi tebebūdinga galimybės apibendrinti stoka, nes jo prigimtis yra susijusi su duomenų apdorojimu ir šablono atpažinimu. Bendriausiu atveju, pavyzdžiui, kai taikoma aprašomoji analitika, labiausiai tradiciniai metodai yra sustiprinami naujausiomis algoritmų mokymosi pasiekimais, o galimybės atrasti geresnių žinių ir užtikrinti geresnį sprendimų priėmimą išsiplėtė. Prognozuojamoji analitika vis dažniau koncentruojasi į modelių, kuriais siekiama teikti empirines prognozes, net ir esant silpniau išvystytai teorinei struktūrai, kūrimą ir vertinimą.

1.6. Metodų integravimas į sistemas, padedančias priimti sprendimus

Sistemos, padedančios priimti sprendimus, šiandien yra patrauklus tyrimų dalykas ir praktiniu, ir moksliniu požiūriu. Geresnis sprendimo priėmimo procesas padeda didinti bendrą efektyvumą ir gerinti veiklos rezultatus, nes suformuluojama strateginė informacija apie esamas operacijas ir aplinkos struktūrą. Platesnio vaizdo matymas gali daryti poveikį vadovybės sprendimų priėmimo procesui. Įvairiarūšė aplinkos struktūra ir procesų nelinijiškumas taip pat savo ruožtu daro poveikį akcijų rinkai, tam, kam vartotojai teikia pirmenybę, ir net konkurentų politikai. Atsižvelgus į šiuos svarstymus ir prielaidas, nuspręsta, kad tyrimas turi apimti ir verslą, ir mokslą.

1.7. Literatūros analizės rezultatai ir apibendrinimas

Nuo literatūros apžvalgos ir metodikos aiškiai pereinama prie išmanesnių sistemų, padedančių priimti sprendimus. Įmonių sprendimų priėmimo procesuose taikyta daug inovatyvių būdų, ir remtasi labai įvairiais šaltiniais – nuo finansinių koeficientų, finansinių ataskaitų iki matematinio modeliavimo ir vertinimų. Autorius, išanalizavęs tyrimo literatūrą, nustatė, kad efektyvumo vertinimas esant neapibrėžtumo sąlygoms yra ekonominio kompleksiškumo ir sprendimų priėmimo procesų nelinijiškumo rezultatas. Sistemos, padedančios priimti sprendimus, šiandien yra patrauklus tyrimų dalykas ir praktiniu, ir moksliniu požiūriu. Geresnis sprendimo priėmimo procesas padeda didinti bendrą efektyvumą ir gerinti veiklos rezultatus, nes suformuluojama strateginė informacija apie esamas operacijas ir aplinkos struktūrą.

II. SISTEMŲ, PADEDANČIŲ PRIIMTI SPRENDIMUS, EFEKTYVUMO VERTINIMO ESANT NEAPIBRĖŽTUMUI MODELIS

Šiame skyriuje aprašomi modelio tyrimo metodai, įžvalgos, tyrimo priemonės ir vertinimas. Daugiamatės analizės pagrindą sudaro algoritminio mokymosi metodai, leidžiantys atlikti tokią analizę.

2.1. Efektyvumo vertinimo esant neapibrėžtumo veiksniams modelio apibrėžimas

Į modelį turėtų būti įtraukta aplinkos poaibių, iš kurių sudaryta visa imtis, vertinimo įtaka. Sistemoje, padedančioje priimti sprendimus, kiekvienas ekonominis poaibis atitinka modulinį principą, t. y. yra sudarytas iš didelio skaičiaus sudėtingų, bet funkciniu požiūriu tikslių dalių. Atvirumas imtyje turėtų reikšti, kad šios dalys turi tam tikro laipsnio laisvę. Išsamus ūkio subjektų vertinimas yra svarbi priemonė, reikalinga norint užtikrinti

objektyvų ir efektyvų išteklių paskirstymą. Ši priemonė gali padėti pagerinti sprendimų priėmimo procesą, siekiant sustiprinti valdymą ir suteikti sprendimų priėmimo pagrindą. Apžvelgus literatūrą, galima rasti daugybę vertinimo metodų. Daugelis šių metodų nustatomi apskaičiuojant indeksų svorį ir gali būti vertinami šališkai ir subjektyviai. Akivaizdu, kad šie ekspertų metodai yra gana teisingi pagrindimuose ir skaičiavimuose, tačiau vertinimo rezultatai dažnai nėra akivaizdūs. Algoritminio mokymosi sistema, pagrįsta bendrais būdais – jų taikymas pastaraisiais dešimtmečiais išaugo, – šias problemas gali tinkamai išspręsti. Dvejetainiais klasifikatoriais, atraminių vektorių algoritmais ir dirbtiniais tinklais pagrįsti algoritmai turi pagrįstą pranašumą klasifikavimo ir regresijos problemose.

2.2. Duomenų rinkinių taksonomijos pasirinkimas sistemai, padedančiai priimti sprendimus

Tyrime nagrinėjamas atrinktų į biržos sąrašus įtrauktų bendrovių, efektyvumas. Tyrimas apima daug daugiau nei bendrovių veiklos rezultatus, vertinamus finansiniu požiūriu. Efektyvumo vertinimas labai priklauso nuo duomenų rinkinio, naudojamo kaip produktyvumo modelio įvestis, kokybės. Santykinių duomenų kaupimas apima įvairius duomenų gavimo būdus, naudojant daugybę duomenų rinkinių, iš kurių siekiama gauti žinių. Duomenų rinkinių sluoksnių bendrąją struktūrą sudaro:

1. makrolygis,
2. konkrečiai šaliai būdingi duomenų rinkiniai,
3. organizaciniai duomenų rinkiniai.

Neapibrėžtumo duomenų rinkiniai pateikiami keliuose šaltiniuose, kurie remiasi tam tikrų meta duomenų žiniasklaidos aprėpties dažnumu. Šie duomenys yra apibūdinami kaip bendras procesų nepastovumas, kurie ekonominiu požiūriu nėra stebimi. Kiekvienai šaliai būdingi duomenų rinkiniai parodo konkrečios šalies ekonominę aplinką. Organizaciniai duomenų rinkiniai yra rengiami į biržos sąrašus įtrauktoms bendrovėms. Mokslinėje literatūroje, taip pat pramonės sektoriuje, nuolat vykta diskusijos dėl tinkamos įvesties duomenų ir išvesties duomenų apibrėžties ir atrankos. Laikantis finansinio požiūrio į efektyvumą, skaičiavimas turi būti atliekamas taikant su įvestimi susijusį būdą, nes esama pagrindinės prielaidos, kad finansų įstaiga paprastai gali labiau kontroliuoti įvesties, o ne išvesties duomenis. Technologijų susiliejimo laikais konkurencinis pranašumas yra apibūdinamas geresniu įvesties išteklių valdymu, o ne masto poveikiu.

2.3. Duomenų kaupimo būdai

Masto sumažinimui tenka pagrindinis vaidmuo daugelyje duomenimis grindžiamų sričių, ir jis buvo plačiai nagrinėtas atstumo funkcijų ir grupavimo algoritmų požiūriu. Parametrai apskaičiuojami optimizuojant duomenų ir modelio tarpusavio tinkamumą, išreikštą tikėtinumu.

1. *Asociacijos analizė*. Taikant šį metodą siekiama rasti ryšius tarp subjektų remiantis sandoriais arba įvykiais, su kuriais tie subjektai buvo susiję.
2. *Klasterizavimas*. Gerai žinomas objektų grupavimo būdas.

2.4. Dviejų pakopų duomenų gaubtinės analizės neparametrinio efektyvumo analizė

Duomenų gaubtinės analizės neparametrinis modeliavimas yra matematinis programavimo metodas, taikomas analizuojant našumo ribas. Todėl efektyvumo matavimas yra susijęs su tomis ribomis, ir kiekvienai poaibio organizacijai suteikiamas efektyvumo balas.

Dviejų pakopų modeliai duomenų įvestį ir išvestį naudoja pirmoje pakopoje, o antroje pakopoje jie naudoja rezultatus ir išorinius veiksnius, kuriuos galima pastebėti. Visų pirma kalbant apie finansinių ataskaitų analizavimą, finansinės informacijos ir organizacinės vertės ryšys nustatomas vykstant dviejų pakopų pagrindiniam procesui:

1. Prognozuojamasis organizacinis efektyvumo vertinimas siekiant susieti esamus veiklos rezultatų duomenis su išteklių paskirstymu.
2. Vertinimo sąsaja, pagal kurią organizacinis efektyvumas perkeliamas į rinkos rezultatus.

Tikslas – nustatyti stebimųjų išorės veiksnių poveikį pirminiams vertinimams.

2.5. Bendri algoritmų mokymosi metodai

Remiantis literatūros apžvalga, šiame tyrime aprėpiama 80,19 proc. algoritmų mokymosi metodų, taikytinų konstruojant sistemą, padedančią priimti sprendimus, kai atliekamas efektyvumo vertinimas. Algoritmų mokymosi modelių sukūrimui ir diegimui reikia atlikti tam tikrus veiksmus, ir šie veiksmai gana panašūs į statistinio modeliavimo procesą, kai siekiama surinkti, patikrinti ir išmokyti modelį, naudojant hiperparametrus. Metodinė klaida dažnai padaroma atsižvelgus į prognozavimo funkcijos parametrus ir testuojant tame pačiame duomenų rinkinyje. Siekiant išvengti persimokymo, algoritme reikia naudoti didesnį kiekį mokymo šablonų. Bendrus algoritmų mokymosi metodus sudaro šie metodai:

1. *Atraminių vektorių algoritmų* modelis, pasiūlytas Vapniko (1992 m.), yra prižiūrimo mokymosi matematinis metodas, naudojamas ir klasifikavimo užduotims, ir regresijoms.
2. *Dirbtiniai neuroniniai tinklai* yra įvairūs duomenų gavybos būdai, kuriuos sudaro keli apdorojimo elementai, kai gaunami įvesties duomenys ir pateikiami išvesties duomenys, paremti jų iš anksto nustatytomis aktyvavimo funkcijomis.
3. *Atsitiktinis miškas* yra klasifikavimo algoritmas, susidedantis iš daugelio medžių pavidalo sprendimų schemų.

Šiame tyrime taikant algoritmų mokymosi metodus laikoma, kad dėl pasirinkto metodo duomenys yra stacionarūs. Skaičiavimais ir prognozėmis grindžiamo rezultato tikslumui didelį poveikį daro tai, kaip gerai nustatytas pagrindinio proceso pobūdis.

III. SIŪLOMO SISTEMŲ, PADEDANČIŲ PRIIMTI SPRENDIMUS, EFEKTYVUMO VERTINIMO ESANT NEAPIBRĖŽTUMUI MODELIO TESTAVIMAS

Šioje dalyje pateikiamas empirinis tyrimas. Siūlomo modelio svarba yra didžiulė. Verta paminėti, kad daug investuotojų naudoja investicinį modelį, taikydami atrankos procesą, kuris prasideda nuo ekonominės aplinkos ir pereina prie pavienės bendrovės veiklos rezultatų. Siūlomas efektyvumo vertinimo metodas, taikomas sistemose, padedančiose priimti sprendimus, yra metodas, kuriuo siekiama išvengti įprastų efektyvumo vertinimo metodų ribotumo.

3.1. Sistemų, padedančių priimti sprendimus, praktinio įgyvendinimo etapai

Bendrą procedūrą sudaro šie praktiniai veiksmai, ir jie turi būti integruota sistemų, padedančių priimti sprendimus, dalis:

1. Duomenų modelio analizė, siekiant patikrinti duomenų rinkinių patikimumą ir vientisumą.
2. Efektyvumo vertinimas ir modelių apskaičiavimas taikant įvairius būdus.
3. Remiantis apskaičiuotais efektyvumo modelių įverčiais atliekama savybių rinkinio atranka.
4. Taikomi bendro algoritmų mokymosi metodai.
5. Tikrinami algoritmų mokymosi algoritmai.

3.2. Sistemoms, padedančioms priimti sprendimus, skirtų duomenų rinkinių gavimas ir analizė

Įvairių pakopų efektyvumo analizė yra natūraliai susijusi. Aplinkos struktūroje, kur naudojama daug įvesties duomenų siekiant gauti daug išvesties duomenų, sumavimo metodai yra būtini siekiant apskaičiuoti bendrą įvesties duomenų ir išvesties duomenų lygį, kad sistema, padedanti priimti sprendimą, būtų geresnė. Atliekant tyrimą taikytas aukšto lygio abstrakcijos lygmuo, siekiant tyrimui pateikti struktūrizuotus duomenų rinkinius. Daugelio veiksmų procesas yra skirtas spręsti patikimos sprendimų priėmimo sistemos sukūrimo problemą. Kadangi esama neapibrėžtumo veiksnių, bet kuris standartinis grupės duomenų modelis yra nepakankamas ekonominio proceso dinaminiam pobūdžiui apibūdinti.

Grupės duomenų modeliams skiriamas dėmesys, nes jie gali užfiksuoti dinaminius pokyčius ir yra naudojami vertinant efektyvumo svyravimus laikui bėgant. Tačiau jokia tikra dinamika negali būti parodyta taikant vieną neparametrinį modelį, veikiau turėtų būti vertinamas daugiapakopis metodas.

3.3. Efektyvumo vertinimas taikant neparametrinį modelį

Operacijų masto poveikiui nagrinėti pasirinkti duomenų gaubtinės analizės (DEA) metodo kintamojo pelno pagal mastą (VRS) ir pastoviojo pelno pagal mastą (CRS) būdai, taikyti lyginamajai analizei ir patvirtinimui atlikti. Atliekant dviejų pakopų duomenų gaubtinę analizę surikiuoti kiekvienos organizacijos veiklos rezultatai, ir jie palyginti su kiekviena iš dviejų ribų, apskaičiuotų pagal parametrus. Norint gauti tinkamą rezultatą taikant duomenų gaubtinės analizės metodus, daug dėmesio reikia skirti svarbiems modeliavimo klausimams. Keli jų daro poveikį aiškiam analizės paskirties nustatymui, sprendimui dėl įvesties ir išvesties duomenų, modelio krypties pasirinkimui ir didesniam dėmesiui naudojamų duomenų rūšiai. Modeliavimo procesas apima didelius atsitiktinai sudarytus duomenų rinkinius, kai ūkio subjektų sprendimų priėmimo procesas yra ribotas dėl disponuojamos informacijos, atsižvelgiant į kognityvinius apribojimus ir laiko apribojimus priimant sprendimą. Todėl rezultatai gali būti neaiškūs dėl homoskedastiškumo ir per menkos koreliacijos, o tai nulemia pagrindinę galimybę identifikuoti, nes pastebėtų faktų ir neapibrėžtumo už to yra nedaug. Žinoma, darant tam tikras prielaidas, ne visada reikalaujama, kad būtų stebimi visi modelyje naudojami išorės kintamieji. Svarbiausias veiksnys yra tai, kad dėl per menkos galimybės identifikuoti neapibrėžtumo veiksnius pasiskirstymo identifikavimas tampa gana sunki užduotis teoriniu požiūriu, o praktiškai tai išgyvendinti yra beveik neįmanoma. Ironiška, siekis rasti išsamius neapibrėžtumo veiksnius gali nulemti prognozių informacinės galios sumažėjimą, taip pat jų naudingumo sistemose, padedančiose priimti sprendimus, sumažėjimą.

3.4. Savybių rinkinio atranka ir efektyvumui poveikį darantys veiksniai

Tinkamo savybių, kurios atitiktų pagrindinę originalių duomenų rinkinių informaciją, rinkinio atranka yra labai svarbus veiksnys, darantis įtaką efektyvumui ir klasifikavimo metodams. Klasifikavimo tikslumo didinimas ir gebėjimo prognozuoti gerinimas, mokymosi proceso spartinimas ir atminties poreikio mažinimas – tai keletas pranašumų, kurių teikia savybių atrankos algoritmai. Todėl, sumažinus savybių rinkinį, gali tapti lengviau suprasti ir aiškinti skaičius.

Iš jautrumo analizės matyti, kad savybių atrankos rezultatų poveikis įmonės lygmens išvesties duomenims yra labai įvairialytiškas ir priklauso nuo neapibrėžtumo lygio, taip pat matyti, kad įvesties kintamųjų atsakui į rinkos kintamumą būdingas didelis išgaubtumas. *Finansinio sverto* poveikis *akcijų kintamumui* paaiškintas ilgoje perspektyvoje koreguojant tikslinį rodiklį, įtrauktą taikant jo bendrą neapibrėžtumo lygį. Todėl neapibrėžtumo perdavimo kanalai gali būti vertingas įžvalgų, susijusių su šia sąveika, šaltinis, ir kyla klausimas, kaip gali būti daromas poveikis politikos veiksmingumui ir efektyvumui, siekiant jos tikslų, kai poveikis ekonomikai daromas per akcijų rinką. Analizė įrodė II skyriaus *2.4* skirsnyje „*Two-stage DEA nonparametric efficiency analysis*“ pateiktą prielaidą, kad dėl efektyvumo pobūdžio veiksmingumą reikėtų vertinti atskirai. Esant 1 pakopos efektyvumui, įvesties

duomenų paskirstymas ir įmonės lygmens efektyvumas yra pagrindiniai rodikliai vertinant įmonės lygmens efektyvumą, kai išorinis poveikis neturi tiesioginio poveikio gamyboje taikomoms technologijoms. Esant 2 pakopos efektyvumui, laikomasi į išvesties rodiklius orientuoto požiūrio, susijusio su rinkos kapitalizacija ir rinkos akcijomis. Todėl 2 pakopos įvesties duomenų parametrams didelę įtaką daro neapibrėžtumas.

Teorinių prielaidų rezultatai visiškai atitinka praktiškai nustatytus faktus. Įmonių lygmens išvesties rezultatų efektyvumui daromas 49,45 proc. poveikis dėl į įvesties duomenis orientuoto sprendimų priėmimo proceso. Į akcijų rinką orientuotas sprendimų priėmimo procesas patvirtina prielaidą, kad akcininkai yra svarbūs plėtojant į biržos sąrašus įtrauktų įmonių strategiją, ir tiems veiksniams tenka 43,39 proc. O 7,16 proc. gryno neapibrėžtumo veiksnių tikrai turi didelės strateginės reikšmės, ir tai daro didelę įtaką sprendimų priėmimo procese.

3.5. Neparametriniai efektyvumo modeliai, taikant bendrą algoritmų mokymąsi

Pagrindinė algoritmų mokymosi algoritmo taikymo paskirtis yra atpažinti paslėptus šablonus. Todėl į rinkinį yra integruota neapibrėžtumo savybė. Normalizavimo tikslas yra pašalinti perteklių duomenų rinkiniuose, nes esant subalansuotiems duomenims stengiamasi visiems požymiams teikti vienodą svarbą. Įtraukus įvesties ir išvesties duomenų neparametrinės analizės rezultatus galima įvertinti efektyvumą kaip klasifikavimo problemą. Atsitiktiniam miškui tenka didžiausias svorinis koeficientas (0,9078), po jo eina atraminių vektorių algoritmai (0,822). Dirbtiniai neuroniniai tinklai (0,024) ir vidurkio algoritmai šiame tyrime praktinės reikšmės neturi.

Nėra vieno aiškaus algoritmo, kurį būtų galima taikyti bet kurioje situacijoje. Taikant atsitiktinio miško metodą rezultatai gaunami kiek geresni, bet taikant atraminių vektorių algoritmus galima geriau spręsti struktūrines problemas.

Iš pirmiau pateiktos analizės matyti, kad atsitiktinio miško metodas ir atraminių vektorių algoritmų metodas yra tinkami, kai sprendimų priėmimo sistemose vykdoma klasifikavimo užduotis, taikant neparametrinius efektyvumo vertinimo modelius. Visų pirma, atraminių vektorių algoritmų modelių integravimas į įvairias pagrindines funkcijas ir duomenų gaubtinės analizės metodas davė geriausius rezultatus. Tinkamas metodo pasirinkimas yra būtinas atliekant tiekėjo vertinimą, nes tuo gali būti užtikrinti optimalūs įmonės vertinimo sprendimai, palyginti su kitais dirbtinio intelekto metodais. Ypač tada, kai taikomas atraminių vektorių algoritmų metodas, labai svarbus yra tinkamas pagrindinės funkcijos pasirinkimas sudarant klasifikavimo modelį, nes dėl to gali pagerėti prognozavimas, remiantis pirmiau pateiktais eksperimentiniais rezultatais. Iš pagrįstų eksperimentų, atliekant statistinį testavimą, matyti, kad duomenų gaubtumo analizės rezultatas yra naudinga savybė siekiant padidinti klasifikavimo veiksmingumą.

3.6. Empirinio tyrimo rezultatai ir diskusija

Empirinio tyrimo dalies tikslas yra pateikti apibendrintus efektyvumo vertinimo, esant neapibrėžtumui, rezultatus ir paskatinti mokslinę diskusiją dėl tyrimo bei gautų rezultatų. Dėl atraminių vektorių algoritmo pranašumų sprendžiant nelinijines problemas, jį galima taikyti siekiant užfiksuoti ir paaiškinti pagrindinį neapibrėžtumą.

Patvirtinta, kad algoritmų mokymosi algoritmai, sustiprinti iš anksto apibrėžtomis pagrindinėmis funkcijomis, gerai susitvarko su daugiamačiais ir didelės įvairovės duomenimis, ir jie tai gali padaryti dinaminėje arba neapibrėžtumo aplinkoje, nes apie tam tikrą klausimą gaunama žinių, kurias galima taikyti ir lengvai aiškinti.

IŠVADOS

Apibrėžti ir praktiškai įdiegti veiksmingą sistemą, padedančią priimti sprendimus, yra nelengva užduotis. Taigi, tyrime teigiama, kad šiuolaikinėje aplinkoje visada esama erdvės naujiems tyrimams.

Atlikti tyrimai parodė, kad svarbu įtraukti teorinius ir empirinius neapibėžtumo, nelinijiškumo, kompleksiškumo ir riboto racionalumo aspektus, kaip pagrindinę struktūros prielaidą, bet ne pusiausvyra pagrįstas teorijas. Šis efektyvumo vertinimas esant neapibrėžtumui apibrėžiamas įvairiais neapibrėžtumo šaltiniais, ir to negalima įvertinti kiekybiškai ne hibridiniuose modeliuose. Tyrime pagrindžiama, kad duomenų kokybė yra labai svarbus veiksnys. Norint taikyti bendrus algoritmų mokymosi metodus, reikia labai tiksliai įvertinti ir įvesties, ir išvesties duomenis, taip pat sukurti neapibrėžtumo duomenų rinkinius. Neparametrinio modeliavimo turėtų būti imamasi atsargiai, nes jam būdingas subjektyvumas. Nėra bendro metodo, taikytino sprendžiant trūkstamų duomenų problemą. Bet šiame tyrime siūlomame modelyje trūkstamų duomenų traktavimas yra viena iš svarbių užduočių.

Šis tyrimas yra vienas pirmųjų bandymų įvertinti efektyvumą taikant tiek klasifikavimo, tiek regresijos modelį. Tyrime, be kitų dalykų, tiriami atsitiktinio miško, dirbtinių neuroninių tinklų ir atraminių vektorių algoritmų klasifikatoriai, sprendžiant duomenų rinkiniuose esančių neapibrėžtų žinių problemą. Svarbu parodyti, kaip žinių duomenų rinkinių neapibrėžtumo duomenys gali būti traktuojami taikant bendrus algoritmų mokymosi metodus, panaudojant optimizavimą. Algoritmų mokymosi ateitis siejama su įvairių metodų taikymu, nes visiškai prižiūrimi algoritmai yra naudingi, bet praktiniu požiūriu juos sunku taikyti.

Tyrime aiškiai pasiūlyta neapibrėžtumą traktuoti ne kaip fiktyvų kintamąjį, bet kaip reiškinį, smulkiai analizuojamą siūlomame modelyje skirtingais lygmenimis: duomenų kaupimo neapibrėžtumo, analitinės struktūros neapibrėžtumo ir neapibrėžtumo, kaip veiksnio. Skirtingai nei esami metodai, algoritmų mokymosi būdai, įtraukti į šį tyrimą, nereikalauja remtis hipotetine prielaida. Matematiniu požiūriu, taikant algoritmų mokymosi būdus nustatomi numanomi svorinių koeficientų apribojimai, taigi esama esminio skirtumo tarp

šių metodų, ir tas skirtumas atsiranda iš to, kaip duomenys renkami. Kiekviename procese neapibrėžtumo mastas yra skirtingas, ir jis turėtų būti vertinamas atitinkamais būdais.

Atsižvelgiant į keletą veiksnių, svarbu detaliai parengti teorinį pagrindą, galintį dinamiškai apimti kuo daugiau veiksnių. Todėl bet kurio efektyvumo vertinimo esant neapibrėžtumui tyrimo aprėptis turi būti platesnė, taip pat nereikėtų apsiriboti konkrečiai šaliai būdingais parametrais – veikiau reikėtų įtraukti konfigūracijas klasteriuose. Įrodyta, kad neapibrėžtumas yra nestabilus veiksnys, o nuokrypiai ryškiausiai matomi per krizes, ir tai reikalauja tyrėjų dėmesio. Dėl šiame darbe paminėtų priežasčių efektyvumo vertinimas esant neapibrėžtumui yra aktualus ir teoriniu, ir empiriniu aspektu. Šiame tyrime nagrinėjami abu aspektai. Tyrime patvirtinta, kad neapibrėžtumas yra nuolatinis ekonomikos reiškinys, ir politikai turi į jį nuolat atkreipti dėmesį. Vertinti makroekonominį neapibrėžtumą ir suprasti jo poveikį ekonominei veiklai yra svarbu vertinant dabartinę makroekonominę situaciją. Neapibrėžtumas neturėtų būti pernelyg supaprastintas. Reiškiniai pavieniams ekonomikos sektoriams poveikį daro visiškai skirtingais būdais, skirtingas būna poveikis ir atsparumas jam. Ekonominės veiklos rūšių skirtingas savybes sutelkus į sistemą atliekama klasterių analizė.

Algoritmų mokymosi ir duomenų mokslas reikalauja daugiau nei tik į modelį įtraukti daugiau duomenų. Kad būtų galima įdiegti sėkmingą metodą, mokslininkai turi iš tikrųjų suprasti tikruosius procesus, slypinčius už duomenų. Viena perspektyvi įgyvendinimo metodika – žinojimas, kai modeliui gali būti naudingas bendrų modelių taikymas. Šiuo atveju atliekant būsimus tyrimus bus pasinaudota kompleksiniais prediktorių deriniais, gautais taikant daug algoritmų mokymosi metodų, siekiant gauti tikslesnes prognozes nei tai būtų įmanoma taikant bet kurį vieną atskirą modelį.

Tolesnių tyrimų kryptys. Šio darbo rezultatai parodė, kad reikia atlikti papildomus tyrimus elgsenos ekonomikos ir algoritmų mokymosi srityse. Atlikti tyrimai parodė, kad nepaisant visų problemų, su kuriomis susiduria elgsenos ekonomika ir algoritmų mokymasis ekonomikoje, atlikti tyrimai suteikia naujų žinių ekonomistams, padeda tiriant ekonominio žmogaus elgesio anomalijas ir suteikia atsakymus į klausimus apie nukrypimus nuo racionalaus elgesio, plėtojant ekonomiką, priimant ekonominius sprendimus. Reikėtų pažymėti, kad elgsenos ekonomika ir algoritmų mokymasis vis dar formuojasi kaip savarankiška ekonomikos teorijos sritis.

Todėl tolimesniuose tyrimuose būtina:

1. Įvertinti ekonominių neapibrėžtumų įtaką ūkio subjektų elgesio modeliams. Šis tyrimas turėtų padėti išsiaiškinti, kaip ir kuriais kanalais netikrumas daro įtaką agentų ekonominiams sprendimams.
2. Reikia kritiškai įvertinti elgsenos modelių konstravimą ir analizuoti veiksnius, susijusius su žmonių sprendimų priėmimo procesu elgsenos ekonomikoje.

MOKSLINIŲ PUBLIKACIJŲ DISERTACIJOS TEMA SĄRAŠAS

1. S. Kornilov, T. Polajeva (2016). Economic Complexity: the future of the knowledge-based regional development. SGEM 2016. ISBN 978-619-7105-72-8 / ISSN 2367-5659. Book 1, Vol. 3, P. 239-246, DOI: 10.5593/SGEMSOCIAL2016/B13/S03.032.
2. S. Kornilov, S. Ridala, A. Aasma (2016). Dynamics of economic adjustment under uncertainty: MS-VAR model for Baltic Dry Index. SGEM 2016. ISBN 978-619-7105-76-6 / ISSN 2367-5659. Book 2, Vol. 5, P. 189-196, DOI: 10.5593/SGEMSOCIAL2016/B25/S07.025.
3. Sergei Kornilov, Tatjana Põlajeva (2012). Infrastructure development: Economic Growth Effects // The 7th international scientific conference «Business and Management 2012», May 10-11, 2012, Vilnius, Lithuania: selected papers. Vilnius: Technika, 2012. ISBN 9786094571169. ISSN 2029-4441. eISSN 2029-929X. P. 156–161.

Su publikacijomis galima susipažinti čia: https://bit.ly/SergeiKornilov.

KONFERENCIJOS

1. EAEPE (European Association for Evolutionary Political Economy) 2013, Italija. Impact of innovation and infrastructure development on growth. Estimation of agent-based model approach.
2. VGTU Scientific Conference "Business and Management", 2012. Vilnius. Infrastructure development: the risks assessments and economic growth effects.
3. International PhD Summer School 2011, Vinistu, Estija. Adopting Value Creation Policy in Economic Shocks.
4. ICEM 2011. Economics and management, Kaunas. Strategic approach in risk-adjusted tariff policy.

GYVENIMO APRAŠYMAS

SERGEI KORNILOV
Tel. +372 56 215 980
El. paštas sergei.kornilov@gmail.com

IŠSILAVINIMAS

2019–2020	**Mykolo Romerio universitetas** Doktorantūros studijos	*Lietuva*
2011–2017	**Talino technologijos universitetas** Doktorantūros studijos	*Estija*
2008–2010	**Talino technologijos universitetas** Magistro studijos	*Estija*
1999–2003	**Taikomųjų mokslų universitetas Furtwangen** Bakalauro laipsnis	*Vokietija*

DARBO PATIRTIS

Nuo 2019	**EVER SOLUTIONS OÜ** Duomenų analitikas	*Estija*
Nuo 2018	**MTÜ KOOL 21. SAJANDIL** Tarybos narys, mokslinis vadovas	*Estija*
2007–2018	**E.R.S. AS** Informacinių technologijų departamento vadovas	*Estija*
2003–2005	**GOMAB MÖÖBEL AS** Projekto vadovas	*Estija*
2002–2003	**ROBERT BOSCH GMBH** Stažuotė	*Vokietija*
2001–2002	**SIEMENS LTD.** Stažuotė	*Kinija*

KALBŲ MOKĖJIMAS

Rusų	Gimtoji
Anglų	Puikiai
Vokiečių	Puikiai
Estų	Puikiai

MYKOLAS ROMERIS UNIVERSITY

Sergei Kornilov

ASSESSING ORGANIZATIONAL EFFICIENCY UNDER MACROECONOMIC UNCERTAINTY IN DECISION SUPPORT SYSTEMS: ENSEMBLE METHODS IN MACHINE LEARNING WITH TWO-STAGE NONPARAMETRIC EFFICIENCY MODELS

Summary of Doctoral Dissertation
Social Sciences, Economics (S 004)

Vilnius, 2020

This dissertation was prepared during the period 2011-2017 at Tallinn University of Technology (Estonia) and during the period 2019-2020 at Mykolas Romeris University under the doctoral program right conferred to Vytautas Magnus University, ISM University of Management and Economics, Mykolas Romeris University and Šiauliai University on 22 February 2019 by the Order No. V-160 of the Minister of Education, Science and Sport of the Republic of Lithuania.

Dissertation is defended on a non-resident basis.

Scientific consultant:

Prof. Dr. Asta Vasiliauskaitė (Mykolas Romeris University, Social Sciences, Economics S 004).

Scientific supervisor at Tallinn University of Technology in 2011-2017:

Prof. Dr. Tatjana Põlajeva (Tallinn University of Technology, Estonia, Social Sciences, Economics S 004).

The doctoral dissertation is defended at the Council of Economics of Vytautas Magnus University, ISM University of Management and Economics, Mykolas Romeris University and Šiauliai University:

Chairperson:

Prof. Dr. Violeta Pukelienė (Vytautas Magnus University, Social Sciences, Economics S 004)

Members:

Prof. Dr. Natalja Lace (Riga Technical University, Latvia, Social Sciences, Economics S 004);

Prof. Habil. Dr. Žaneta Simanavičienė (Mykolas Romeris University, Social Sciences, Economics S 004);

Assoc. Prof. Dr. Sigita Urbonienė (Vytautas Magnus University, Natural Sciences, Mathematics N 001);

Assoc. Prof. Dr. Inga Žilinskienė (Mykolas Romeris University, Technological Sciences, Informatics Engineering T 007).

The doctoral dissertation will be defended at a public meeting of the Council of Economics on 30 June 2020 at 12:00 at Mykolas Romeris University, auditorium I-414.

Address: Ateities str. 20, 08303 Vilnius, Lithuania.

Relevance of the topic. The modern environment, where we all live in, is the subject of constant changes over time. There are evidences both from mass-media and science, that the modern economic setting is characterized by increasing information flow gathered for decision making process, growing global competition on the macroeconomic level and limited physical resources. It is hard to deny the increasing role of information and knowledge in the XXI century. One particular concern could emerge from decision-making process, which is getting more and more complicated with time passing by. The decision-making process has the result to make the most efficient decision, which implies the minimal allocation of resources and maximum output ceteris paribus. Thus, the assessing of effectivity plays enormous role in the decision-making process. The cost of wrong decision is enormous due to increased competitors on the global scale. However, a clear effect of convergence in economics generally is observed on global scale, but the technological advance is the factor, which determines competitive advantage over decades. With developing of technologies, the opportunity cost is getting higher at the explosive scale. For example, it is hard to make any credible assumption in rise of a hypothetical emerging economy, which is able to achieve the technological advance on global scale in a hi-tech area without having prior fundamental knowledge and expertise. In order to avoid possible opportunity costs, the decision-makers are demanding more sophisticated yet reliable prediction techniques. In this context, the issue of handling decisions of generally heterogeneous economic agents under factors of uncertainties emerges as the front-line problematic of scientific research.

The Author asserts throughout the thesis, that none of the above mentioned issues should be regarded isolated manner. This assumption finds its justification in the literature body of economic science virtually since the very beginning. But only in the past decades the scientific methodology reinforced with technologies of machine learning could bring a feasible analytical framework for assessment uncertainty on different levels and capture nonlinearities in processes not only within generalized scientific techniques prevailing with generalized expectation assumptions underneath. Through innovations caused by financial technologies it becomes possible to shed light on the idea of economic growth and its connection with investments from different perspective in terms of their ability to create or absorb technological innovations within on-going infinite technological progress.

Theoretically, the modern economy as a whole is made up of many smaller complex subsets. In the preceding paper, Kornilov and Polajeva (2016) have already investigated a complex nature of economic processes. The study shows, that increased levels of complexity affected by uncertainty in many ways and thus increase risk factors. Each economic subset is modular in terms of being made up of a large number of functionally specific parts. It is open in the sense that these parts deal with a certain degrees of freedom.

Any scientific approach should be able to recognize that agents are naturally heterogeneous, what rose complexity of economic process demanded more sophisticated methods, which should explain individual set of knowledge collectively, creation of an aggregated outcome and their reaction to this outcome. This differs from other approaches tend itself to expression in equation form, whereas by definition a general pattern that

does not change. Modern scientific literature draws attention to important issues in economic studies, including spatial integration and economic complexity. The economic subsets have become increasingly associated with the widely known concept as knowledge economy.

The modern business settings are defined by rapid and radical changes caused by information accessibility. The modern economic science is being in turbulence nowadays (Chuen and Linda (2018), Dunis *et al.* (2016)). The recent trends of financial automation explain increasing progress to have computational applications for forecasting, modelling and trading financial markets and information. New trends of cryptocurrency and digital finance has to be analyzed in the future science but today's evidences have already exhibited that its phenomena. It has been already forged as the result we might see as the convergence of profit motives with social objectives creating a class of large companies in financial technologies. Technological exchange among sectors is intense nowadays, so the underlying innovations may be applied to a wide range of industries simultaneously. The *technological convergence* is another important factor, which is a relative new to the economic science and it is definitely the subject of future in-depth investigation.

The relevance of the topic is supported by the integration processes among EU member states focused of increasing economy efficiency and eliminating economic disparities among nations. The integration development consists of the underlying dynamics of globalization in terms of markets and capital as well as the move towards closer international co-operation through the further development of trade unions and policy co-ordination. Such arrangements represent different modes of economic integration processes by eliminating borders of any kind among member states and applying a common policy and structure the economies to trade with other non-members. Therefore, the assessment of efficiency under uncertainty for the policy-makers belongs to the tasks with the highest priority.

Assessing efficiency is highly dependent on reliable evidences, which are the subjects influenced by uncertainty. A number of various methods exist to assess complexity of economic, uncertainty as the factor of economic processes and assessing efficiency. But the attitudes of researchers in the field are detached. Uncertainty have been proven to be unstable factor, with the variations being most vividly seen during the crises. Due to the reasons mentioned here, the assessing efficiency under uncertainty is relevant in both theoretical and empirical aspects. This doctoral dissertation focuses on both aspects.

The recent studies of Onatski and Williams (2003) argues that uncertainty is persistent phenomena in economics and it must be faced continually by policymakers. Black *et al.* (2018), Meinen and Röhe (2017) supports that measuring macroeconomic uncertainty and understanding its impact on economic activity is thus crucial for assessing the current macroeconomic situation. From modern positions a robust and negative effect of uncertainty on economic growth is obvious and these consequences cannot be neglected by the theory (Lensink *et al.* (1999), Levin *et al.* (2005), Ljungqvist and Sargent (2012)). There are a vast number of studies arguing indicators of uncertainty which can be viewed as representative to the evidences of particular policy, involving a wide number of direct and indirect peers (Ericsson *et al.* (1999), Benhabib *et al.* (2013), Bird *et al.* (2013),Jurado *et al.* (2013), Ernst

and Viegelahn (2014), Baker *et al.* (2015), Jurado *et al.* (2015)). The uncertainty factor is so large that the effects of policy decisions on the economy are thought to be ambiguous. In this situation, any plausible expertise on the nature of uncertainty might be very useful. In order to understand how variations in uncertainty might affect the economic process, it is important to find its source.

The large number of studies shows that assessment of efficiency analysis has become an important topic in operational research, public policy, energy-environment management, and regional development. So it is obvious, that two-stage nonparametric methods have been widely used in the recent literature on productive efficiency measurement and in a large literature of studies. Empirical applications choose one group of measurement techniques.

Therefore, the relevance of the topic shows, that first of all the theoretical and practical findings of the thesis underpin the idea of applying machine learning methods for efficiency assessment under uncertainty, which can be utilized for the future policy-makers.

Second, there is a clear need to predict the influence of the heteroscedasticity on the global scale beyond and within the EU. It contributes to the mechanism linking intertwined cross-border components with high degree of freedom into a system, which has economic inputs and outputs as parameters and is able to handle uncertainties on different layer: limited information, bounded rationality and their expectations, and randomness.

Third, but the most important, to argue that assessment of efficiency under uncertainty plays the leading role in a knowledge-driven economic system with continuously increasing complexity. Depending on a number of factors, it is crucial to elaborate a theoretical framework, which can embrace as many factors. Therefore, any research on assessment of efficiency under uncertainty should have a broader scope and should not be limited on country-specific parameters but include configurations in clusters over the borders. The economic development should be captured not solely in economic terms, but should be shaped for knowledge exchange among economic agents. Adding a factor of uncertainty into analysis opens a specific question of incorporating social processes aspects into study.

Research problematic and the level of its investigation. The economic science represents a huge variety of perfect works assessing uncertainty. At the frontier line of experiment-based models derived from recent observations are Elder (2004), Kontonikas (2004), Daal *et al.* (2005), Fountas (2010), Fountas (2010), Henry *et al.* (2007), Neanidis and Savva (2011). The studies are keen to follow deterministic paradigms to cause uncertainties. In this category prevail a wide family of autoregressive conditional heteroscedasticity models both with error variance or in its general form imposing conditionally-autoregressive errors associated with uncertainties. Methodological questions on measure of the uncertainty raised by Giordani and Söderlind (2003), Diebold *et al.* (1997), Clements and Harvey (2011). Classification of Walker *et al.* (2003) gives fundamental notion on it. Berument *et al.* (2009) and Hartmann and Herwartz (2012) extend the standard assumption with stochastic volatility models. Orlik and Veldkamp (2014) and Glass and Fritsche (2015) argue that uncertainty is an outcome value of acyclical changes in uncertainty while shocks. Zarnowitz and Lambros (1987), Bomberger (1996), Rich and Butler (1998) and D'Amico

and Orphanides (2008) argue the epistemic uncertainty by direct estimation of parametric distributions across individuals. Lahiri and Sheng (2010), Siklos (2013), Lahiri *et al.* (2015) extend the model by to numerous improvements and modifications. Walker *et al.* (2003), Dequech (2004) look into epistemic uncertainty caused by experts incomplete knowledge and the variability uncertainty attributed to accidental factors randomly appeared. Lane and Maxfield (2004) extends the variability uncertainty with the ontological uncertainty. Discussion raised by Walker *et al.* (2003) classification goes into inflation uncertainty by Norton (2006), Kowalczyk (2013), Krayer von Krauss *et al.* (2019). Inflation uncertainty has a major impact on economic modeling. This is especially evident when modeling solutions based on such an analysis.

Gelman and Hill (2007) introduces multilevel linear and generalized linear model in which the parameters are given a probability model. Jordà et al. (2013), Knüppel (2014) suggest considering the inclusion of uncertainty in rational expert forecasting models or combinations of various models that do not necessarily have a mathematical and econometric basis for forecasting, but rely on risk factor assessment based on the distribution of *ex-post* forecast errors.

The same level of scientific investigation exhibit nonparametric efficiency assessment. Originated from Seiford (1997) with 800 publications, the more recent overview by Seiford (2005) mentions some 2800 published articles on DEA. Since fundamental contributions by Farrell (1957), Koopmans (1952), Aigner and Chu (1968), Aigner *et al.* (1977), Broek *et al.* (1980) concept of efficiency methodology in frontier production function estimation has been rapid developed. There are a number of critical reviews emerged by principle weakness of the conventional methods to assess efficiency. Sexton *et al.* (1986) and followed by Smith (1997) identified the impact of misspecification, Stolp (1990) generalized that homogeneity of technology across DMU, uncertainty over the choice of inputs and outputs can affect the performance assessment.

However, Tobback *et al.* (2018) argues that common methods of measuring uncertainty developed by Baker *et al.* (2015) does not have any predictive power for any of its variables but the machine learning approach outperform the traditional ARCH-based models. Brose *et al.* (2014a) and Brose *et al.* (2014b) argue that managing risks and uncertainty depends critically on information. Past decade, a number of research look deep into usage of an optimization algorithms based on a linear programming model to identify controls that need to be tested to address the risks, which can be developed as hybrid approaches for efficiency classification (Pareek (2006), H.-Y. Kao *et al.* (2013)). Various linear optimization techniques has been successfully applied to predict time series and their co-movements (Kara *et al.* (2011), Karaa and Krichene (2012)).

Therefore, the recent studies employ machine learning both for assessment uncertainty and efficiency measurement. Predictive power of various machine learning techniques like neural networks widely confirmed in the literature and found practical implications as by Alejo *et al.* (2013). Attigeri *et al.* (2017) argue empirical approach is used to build models for financial risk assessment with supervised machine learning algorithms. Kruppa *et al.* (2012), Kreienkamp and Kateshov (2014), Addo *et al.* (2018) results indicate that nonlinear techniques work especially well to model expected value. Past a few years many

researchers exploit machine learning technique and nonparametric technique to provide a new method for predicting efficiency by using data envelopment analysis (Xu and Wang (2009), L. Zhou *et al.* (2014), X. Yang and Dimitrov (2017), Zelenkov *et al.* (2017), Alaka *et al.* (2018)). Q. Zhang and Wang (2018) proposed efficiency prediction model which for the first time combines supervised learning for information analysis with nonparametric model, to evaluate the future efficiency of decision making unit.

Scientific problem. The main focus of the research is the development of an efficiency evaluation model that is supported on the one hand by economic science and on the other hand uses the advantages of algorithmic learning to obtain a correct and reliable result. After an in-depth review of the scientific literature and the practical application of models, the issue of performance evaluation is unambiguously defined as a task for further research to face factors of uncertainties arising from economic processes and nonlinearities in decision-making processes. In the past, much research has provided valuable and in-depth knowledge of economics to uncover economic processes and the role of economic agents in it. The most influential scholars are awarded the Nobel Prize in Economics for their significant contribution to behavioral economics with bounded rationality. However, there are still gaps between theoretical findings and the practical assessment of the economic efficiency of economic agents in decision-making process under uncertainty.

The object of the research are the factors of uncertainty and efficiency applied to the decision-making system, assessing the effectiveness of organizations in conditions of uncertainty, using the methods of training algorithms.

The aim of the research is to develop a methodology for assessing the effectiveness of uncertainty and to test it using financial data sets, after revealing the factors of multidisciplinary approach, economic complexity, uncertainty and efficiency.

Research tasks:

1. Conduct an empirical study of efficiency evaluation using common algorithmic learning methods based on a hybrid model.
2. Reveal the essence and sources of uncertainty and analyze them in the proposed efficiency assessment methods.
3. Examine efficiency as an economic concept and to analyze the proposed methods of efficiency evaluation using algorithmic learning methods.
4. Propose a conceptual model of performance evaluation under uncertainty using linear optimization methods and algorithm learning.
5. Carry out an empirical study of efficiency evaluation using common algorithmic learning methods based on a hybrid model.
6. Describe the results of the empirical study of the evaluation of efficacy under conditions of uncertainty in order to offer recommendations for their application.

Research methods. The research methods used by the Author are the analysis, synthesis and comparison of scientific literature in order to describe the uncertainty and efficiency. The analysis uses practical software R (The R Project for Statistical Computing). Oracle 12c database is used to handle large amounts of data using SQL datasets for analysis in R software. Data were obtained from primary data sources through WebServices or through CSV parsing to database.

Research data and their sources. The study examines the performance of selected companies listed in the Nasdaq Baltic Exchange Market Indices. Uncertainty datasets come from many sources:

1. The Economic Policy Uncertainty Index is based on the frequency of media coverage and is defined as the overall variability of the unpredictable component of many economic indicators.
2. Country-specific factors should include market concentration, the presence of foreign investment, and fiscal indicators. In a rapidly changing business environment, the evolving work environment, the ability to anticipate future trends, and the needs for knowledge and skills, are becoming critical to providing an effective decision support system. These trends vary by geography and industry, so it is important to anticipate country-specific and country-specific variables. Sources:
   - Federal Reserve Economic Data
   - Deutsche Bundesbank Data Repository
   - Organization for Economic Co-operation and Development
   - NASDAQ OMX Global Index Data
   - World Bank Global Development Indicators
3. Organizational datasets provided by NASDAQ OMX.

Limitations of the study. Several cases of study limitation show that the results of using the research methodology are highly dependent on the quality of the data. However, the importance of the results does not diminish as a result, either theoretically or practically. On a theoretical level, this research methodology, developed in response to an identified research gap, is one of the first attempts to assess in detail not only the factors of uncertainty and efficiency in isolation, but also to find the most modern way to understand them as a whole. At the empirical level, this study covers most of the issues related to the assessment of effectiveness in the presence of uncertainty, given the limitations that are imposed on such an analysis by each nucleus of economic operators. Not all possible factors are included in the proposed model. The proposed model is only one possible way to achieve reliable results under uncertainty. However, certain empirical data sets cover a period of 10 years. The number of macroeconomic factors is limited and only key indicators are used. At the data mining level, there are several assumptions that primary filters have been applied to the data. This means that there are no random variables in the data set. In real-world circumstances where data sets are retrieved automatically, there is a chance that there will be missing or incorrect sizes.

Scientific novelty:

1. This research is designed to involve the theoretical and empirical aspects of uncertainty, nonlinearities, complexity and bounded rationality as the major assumption of the framework, but not assumption of them in terms of other equilibria based theories. Analysis of the previous studies shows that theoretical part is detached from the statistical significant findings. It is obvious, that the economics itself from very early steps accepted equilibria concept and the study of generally balanced growing path. Thus, the statistical findings justified to established theories. Even though, Brian (2006) pointed conceptually out that traditional studying equilibrium patterns of consistency required further behavioral adjustments.
2. The assessment of efficiency under uncertainty is defined by various sources of uncertainty, which cannot be quantified within other than hybrid model. From formal point of view, various uncertainties from missing data can be generalized with hypothesis of limited information. But there is to admit, that many real-world datasets may contain missing values for various reasons. Taking such data into a model with a lot of missing values can drastically impact the model's quality. The proposed model offers upfront how to deal with missing data using various machine learning techniques. A number of researchers insist on the quality of the data, whereas Lertworasirikul *et al.* (2002) shows that the nonparametric modelling methods require accurate measurement of both the inputs and outputs. In all the situations presented by researchers and practitioners of nonparametric modelling, it is still a relatively subjective approach in filling a gap from missing data. But within the proposed model in this study the treatment of missing data is one of the important tasks.
3. This research is one of the few that employ structural risk minimization principle to estimate uncertainties, whereas instead of minimizing the observed training error proposed machine learning techniques attempts to minimize the generalization error bound so as to achieve generalized performance.
4. This study is one of the first attempts to assess efficiency within both classification and regression model. The Author among other researches investigate ensemble methods in machine learning classifiers in the face of uncertain knowledge sets and show how data uncertainty in knowledge sets can be treated in ensemble methods in machine learning classification by employing robust optimization. Consequently, ensemble methods in machine learning can also be used as a regression method, maintaining all the main features that characterize the algorithm of maximal margin. The Author is agreed with, that the future of the machine learning is in combination of different approaches, because fully supervised algorithms are a useful but perhaps an unnatural assumption due to latent variables in models (D. Chen *et al.* (2013)).
5. The proposed model deals with evaluating efficiencies in the absence of homogeneity gives rise to the issue of how to fairly compare a DMU to other units. A related problem, and one that has been examined extensively in the literature, is the missing data problem addressed directly to appropriate techniques of machine learning (Zhu (2016b)).

6. The Author is the first who explicitly proposed to treat uncertainty not as a dummy variable, but phenomenon dissected within the proposed model on different layers: data-mining uncertainty, analytical framework uncertainty and uncertainty as a factor. Unlike the existing approaches, the combinations of machine learning techniques in this study do not require to think in terms of hypothetical assumption. Mathematically machine learning leads to the identification of implicit restrictions to weights, so there is a fundamental difference in these approaches, emerging from the way in which the data explicitly is gathered. In each process the uncertainty is emerging in different qualities and it should be assessed with respective techniques.

Practical importance of research results. The study gives clear results on integration of an automated core for any decision support systems. It shows that it is possible to design systems by using modular functions. There are a number of areas of applications, where decision support systems can be applied in:

1. Business Intelligence systems designed the way where data can be promptly observed and filtered by a number of different dimensions in order to obtain immediately insights into recent performance of organizational units.
2. Data mining applications operate with enormous sets of data and facts which have been combined and accumulated through ongoing interaction with counter-parties and environment. These datasets play important role for statistical analysis focused on acquiring meta-information and hidden patterns on utility, preferences, trends, or other associated agents' behavior.
3. Full-scale Enterprise Resource Planning application gives opportunity to conduct organizational workflow based on efficiency a better way with focus on capital investment, inventory, production and logistics.

Only the structured methodology using various methods in approaches both from Data Science and Economic studies might help researchers and policymaker rank the important factors and appreciate the factors underneath a better way. It is not possible to respect the results either from statistical nor theoretical point of view solely, but only as an integrated process with the fully qualified decision support system.

The logical structure of the dissertation. Dissertation consists of introduction, three parts, conclusions and recommendations, references and appendices. The volume of the dissertation is 166 pages. It contains 31 figures, 27 tables, 418 references and 20 appendices.

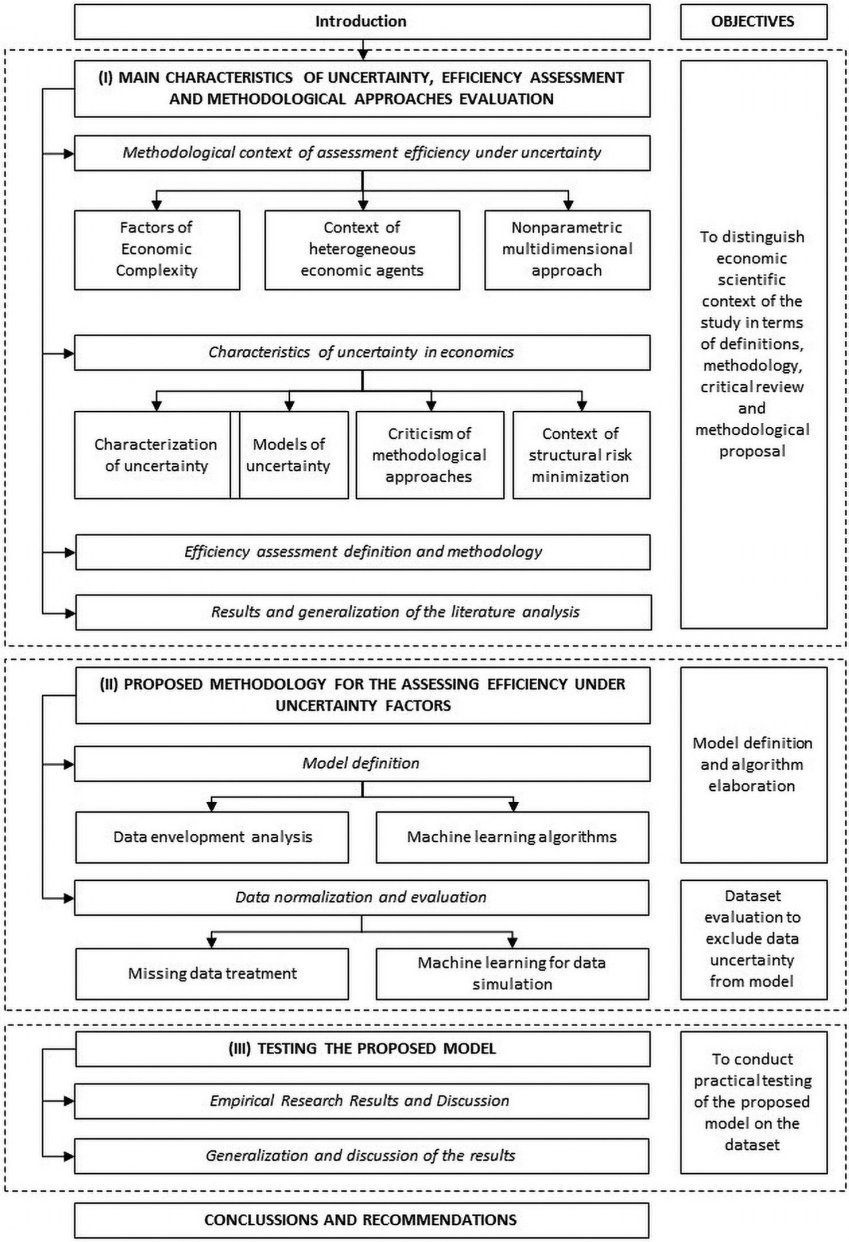


Figure 1. *Dissertation structure*

OVERVIEW OF THE DISSERTATION CONTENT

I. MAIN CHARACTERISTICS OF UNCERTAINTY, EFFICIENCY ASSESSMENT IN DECISION SUPPORT SYSTEMS AND METHODOLOGICAL APPROACHES

In this subsection, the development of the concept of recursive decision-making process as a backbone of economic processes is represented. The economics itself from very early steps accepted equilibria concept and the study of generally balanced growing path. Based on ex-post evidences economists preferred convex structures providing for unique equilibrium states, which could be associated with optimal growth path. But with advances in economic and information science, the equilibrium concepts with the linear distribution of resources is not possible due to the bounded rationality of economic agents from one hand and the economic complexity phenomena along with macroeconomic uncertainties from another. The efficiency assessing of economic agents under macroeconomic uncertainties is a part of economic process on the global scale. The problematic of efficiency forecasting for decision-making under uncertainty yields complex dynamic system questions, which should explain main patterns of how all these components interact and being interconnected. The literature review from the past and recent researches exhibit ontological factors, which influence outcome under uncertainty.

1.1. The decision-making context in heterogeneous economic processes

The analysis of the scientific literature body reports that the theory of bounded rationality is a solid analytical approach, which found its application in many diverse areas. Economic complexity increases the order or regularity between the components interaction and even generates a new order and configuration with the different components behaving autonomously. Thus, radically new processes and component interactions emerge. Analysis of research literature enabled to determine that the efficiency assessment under uncertainty conditions is the result of economic complexity and nonlinearities of the decision-making processes.

The *Dynamically adjustable decision-making model* is presented in Figure 10. The model corresponds well with other recursive models, where agents make their decision in respect to the environmental response. Each organization has certain disposable inputs, desired outputs and assumed targets, which should be regarded separately due to its nature. Hence, a two-stage efficiency nonparametric analysis is strongly advocated. The results of the analysis are integrated part of decision process, which can be reinforced with ensemble methods in machine learning.

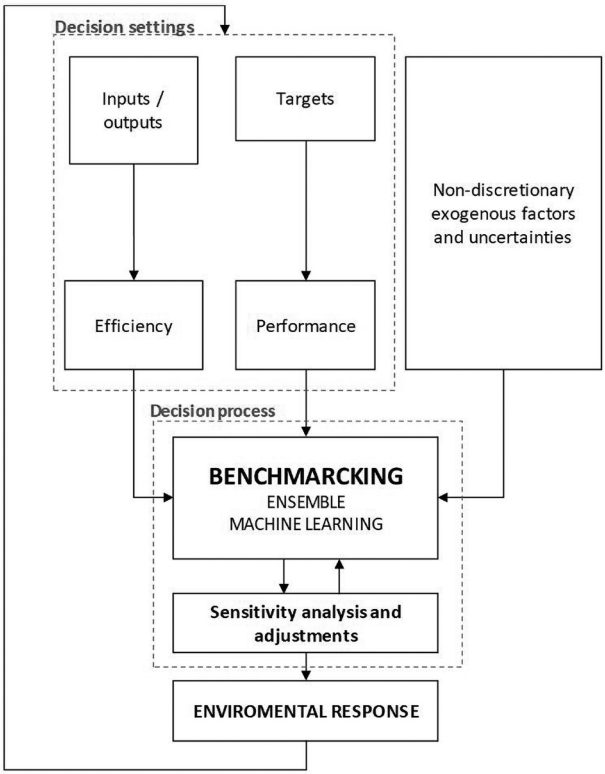


Figure 2. *Dynamically adjustable decision-making model (Source: The Author's representation)*

Therefore, the definition of decision support systems emerges for organizations applying different business intelligence techniques of statistical and scoring modeling, neural networks, various expert systems, agent-based systems, neuro-fuzzy systems, various case-based systems, or simply guidelines that have been developed through data-driven experience.

The process of the efficiency assessment in decision support system has a number of processes, which imply actions on different layers. The study proposes to treat uncertainty as phenomenon dissected on different layers: data-mining uncertainty, analytical framework uncertainty and uncertainty as a factor. Most approaches which used to incorporate uncertainty are based on restricting the weights in the multiplier model. Unlike the existing approaches, the combinations of machine learning techniques in this study do not require to think in terms of hypothetical assumption. Mathematically machine learning leads to the identification of implicit restrictions to weights, so there is a fundamental difference in these approaches, emerging from the way in which the data explicitly is gathered. In each process the uncertainty is emerging in different qualities and it should be assessed with respective techniques.

1.2. Methodological context of assessment organizational efficiency

This subsection goes into details of the organizational efficiency and decision support systems, which involve the concept of the economic complexity upfront in order to avoid misleading generalizations. Decision support systems employ efficiency assessment models based on input-output parameters are commonly used to evaluate economic impacts. These models typically evaluate exogenous variables in resource demanding elements with no look at associated effects of recursive simultaneous connections. An analysis from the economic agent perspective is of greater interest to economic that exploit natural resources because their activity is subject to variations or various factors beyond, what formal approach estimates. Herewith proposes a methodology to improve the estimation of the impacts of these variations. Within the methodological context of economic context analysis, a practical methodology is introduced. Hence, the proposed method will improve impact assessments derived from economic agents to environmental events.

Standard statistical techniques give the tools to distinguish systematic from random factors, so in principle it should be possible to distinguish the rational, adaptive portion of a decision from bounds on rationality. So any meaningful model of the macro economy should analyze not only the characteristics of the individuals but also the structure of their interactions. The advantage of the agent-based modelling and simulation approach for macroeconomics in particular is that it removes the tractability limitations that so limit analytic macroeconomics. Agent-based modelling and simulation modelling allows researchers to choose a form of microeconomics appropriate for the issues at hand, including breadth of agent types, number of agents of each type, and nested hierarchical arrangements of agents. It also allows re-searchers to consider the interactions among agents simultaneously with agent decisions, and to study the dynamic macro interplay among agents.

1.3. Foundations of uncertainty in economics theories

In economic theory there is already a long history of studies attributed to risk and uncertainty both macro and microeconomic phenomenon. The study summarizes, that the uncertainty might have various economic effects on economic agents through various channels. The whole economic theory distinguishes three basic aggregated economic agents, who undertake various economic actives including but not limited to production, consumption and exchange. Therefore, nowadays the theory of economic is able to enlighten sophisticated issues of how economic agents, for instance, manufacturers, households and investors would behave under given circumstances. The same way it is possible to inspects how resources would be allocated and market imperfections arise. From the economic point of view there are discussion between two main concepts of uncertainties *Knightian* and non-*Knightian*. The *Knightian* uncertainty predominantly seen in economic studies as a hypothesis related to an aggregate and is not directly measurable. Uncertainty is unobservable and not directly measurable. However, models might rely on proxies in order to evaluate its changes in time.

1.4. Classification of performance, effectiveness and efficiency

The difference among efficiency, performance and effectiveness is enormous despite the fact all of these terms supposed to benchmark the input and output factors in terms of resources consumed and output produced. Under the business performance the study understands the measurement of business performance following Rappaport (1986), who stated that the shareholder value should become the global standard for measuring business performance. The discussion is followed by an enumeration of the shortcomings of the accounting return on investment and accounting return on equity as standards for measuring business performance. Effectiveness can be measured as a combination of efficiency and performance. The explicit measurement of effectiveness is out of the scope of this research. Thus, particular interest of the research is to establish the link between business performance in financial terms, economic growth factors and the decision-making process.

Econometric techniques require specific assumption about the relationship between the inputs and outputs, and estimate the parameters of a functional form representing this. Econometric techniques should be seen from deterministic or stochastic point of view. The deterministic frontier approach assumes that all the deviation from an estimated frontier is mainly due to technical inefficiency, with no role played by random factors. Unlike the deterministic frontier approach, a stochastic production frontier approach, however, incorporates both noise and inefficiency component into the model specification.

1.5. Ensemble machine learning approach in decision-making process

The study explores ensemble methods in machine learning approach, where various techniques are combined in order to deliver the best possible estimation. The advantage of machine learning is that it can expose generalizable patterns by ability to uncover complex structures that was not stipulated in advance. Machine learning can fit flexible yet complex data settings without overfitting. Ensemble methods in machine learning can be seen as a nonparametric approach in terms of parameters defined by the capacity of the model, which is data-driven to match the model capacity to data complexity.

Machine learning techniques require more abstraction than common statistical techniques due to its principles. However, from economic science point of view, machine learning still suffers from lack of generalization due to its nature arise from data processing and pattern recognitions. In general-class approaches such as descriptive analytics, the most traditional methods are reinforced with recent developments in machine learning and it got extended its capabilities for enhanced knowledge discovery and improved decision-making. The growing class is the predictive analytics focused on the building and assessment of models that seek to make empirical predictions with weaker theoretical framework.

1.6. Integration of methods into decision support systems

Decision support systems are nowadays an attractive research issue in the practical field and from scientific point of view. A better decision-making process contributes to overall efficiency and performance by articulating strategic information about the current operations and environmental settings. Awareness of a bigger picture it may affect a management decision-making process. Heterogeneous environmental settings and nonlinearities of processes might also in turn affect the stock market, consumers' preferences, and even competitors' policy. All of these considerations and presumptions lead to concentrate specific research efforts in both business and science.

From the literature review and methodology there is a clear shift to more intelligent decision support systems. A considerable number of innovative approaches have been used in integrated decision-making processes, most of which employ a wide of information sources from financial ratios, financial statements to mathematical modeling and environmental evaluations of externalities.

1.7. Results and generalization of the literature analysis

From the literature review and methodology there is a clear shift to more intelligent decision support systems. A considerable number of innovative approaches have been used in corporate decision-making processes, most of which employ a wide of information sources from financial ratios, financial statements to mathematical modeling and evaluations. Analysis of research literature enabled the Author to determine that the efficiency assessment under uncertainty conditions is the result of economic complexity and nonlinearities of the decision-making processes. Decision support systems are nowadays an attractive research issue in the practical field and from scientific point of view. A better decision-making process contributes to overall efficiency and performance by articulating strategic information about the current operations and environmental settings.

II. PROPOSED MODEL FOR THE ASSESSING EFFICIENCY IN DECISION SUPPORT SYSTEMS UNDER UNCERTAINTY

This section describes methods, technological approach and tools for research, insights and evaluation, such as framework for the model. A framework for multidimensional analysis based on machine learning techniques, which enables analysis from different point of views.

2.1. Model definition for the assessing efficiency under uncertainty factors

The model should include influence of the environmental subsets evaluation, which the entire sample is made of. For decision support system each economic subset exhibits modularity in terms of being created from a large number of complex yet functionally

specific parts. The openness within sample should mean in the sense that these parts deal with degrees of freedom. The comprehensive evaluation for the economic agents is an important tool to achieve the objective effective resource allocation. It can help to improve decision-making process in order to strengthen the management and provide basis for decision-making. The literature review exhibits many evaluation methods. But most of the methods need to decide the weight of the indices first, scores weighted indices and biased estimations. These methods obviously have a certain subjective fairness of evaluation results are not obvious.

The machine learning system based on ensemble techniques which application is growing in past decades can solve the given problematic appropriately. Random Forest, Support Vector Machines and Artificial Networks have the convenient superiority in the classification and regression.

2.2. Taxonomy of datasets selection for decision support system

The research investigates the efficiency of the selected stock-listed companies. The research goes far beyond estimation of the companies' performances from the financial point of view. Efficiency assessment are heavily dependent on the dataset quality that is used as an input to the productivity model. Relational data mining combines various data mining techniques with multiple dataset for extracting the knowledge from it. The general structure of the datasets layers consists of:

1. Macrolevel
2. Country-specific datasets
3. Organizational datasets

The datasets for uncertainty are represented by multiple sources, which are based on mass-media coverage frequency and also defined as the common volatility in the unforecastable component of a large number of economic indicators. Country-specific datasets represent the economic environment of a given country. Organizational datasets are shaped for stock listed companies. In the scientific literature as well as in industry there is a continuous discussion regarding the proper definition and selection of inputs and outputs. The financial point of view at efficiency require estimation under the input-oriented approach due to underlying assumption that financial institution poses higher control over inputs as a general rule rather than outputs. In the time of technological convergence, the competitive advantage is characterized by a better input resource management rather than scale effects.

2.3. Data mining techniques

The dimensionality reduction has a central role to many data-driven application domains and has been studied extensively in terms of distance functions and grouping algorithms. Parameters are estimated by optimizing the fit, expressed by the likelihood, between the data and the model:

1. *Association analysis*. This method attempts to find the relations between entities based on transactions or events that involve them.
2. *Clustering*. Well-known method of group a set of objects.

2.4. Two-stage DEA nonparametric efficiency analysis

The nonparametric modelling of Data Envelopment Analysis is a mathematical programming method for the analyzing of production frontiers. The measurement of efficiency is therefore relative to these frontiers, where each organization in subset is assigned an efficiency score.

The two-stage models use data outputs and inputs in the first stage, and employ results with observable exogenous factors in the second stage. Especially in terms of analyzing financial statements, the relationship between financial information and organizational value is established through a two-stage fundamental process:

1. A predictive organizational efficiency measure to bind current performance data to resources allocation
2. Valuation connection that projects organizational efficiency to market performance.

The objective is to determine the impact of the observable exogenous factors on initial evaluations.

2.5. Ensemble methods in machine learning

From the literature review, the current research covers 80,19% of applicable methods in machine learning for constructing decision support system for efficiency assessment. The introduction and deployment of machine learning models involves a series of steps that are almost similar to the statistical modeling process, in order to collect, validate and train model with hyper-parameters. An often methodological mistake occurs by taking into account the parameters of a prediction function and testing it on the same dataset. In order to avoid overfitting, algorithm requires a larger number of training patterns. The ensemble machine learning techniques consists of the following approaches:

1. *Support Vector Machines* (SVM) proposed by Vapnik (1992) is a mathematical approach in a supervised learning used for both classification assignments and regressions.
2. *Artificial Neural Networks* (ANNs) are various data mining techniques that consists of several processing elements that receive inputs and deliver outputs based on their predefined activation functions.
3. *Random Forest* (RF) is a classification algorithm consisting of many decisions trees.

Ensemble methods in machine learning in this research assumes that data is stationary due to methods chosen. The accuracy of the result based on estimation and forecasting is affected significantly by how well the nature of the underlying process is identified.

III. TESTING THE PROPOSED MODEL FOR THE ASSESSEMENT EFFICIENCY IN DECISION SUPPORT SYSTEMS UNDER UNCERTAINTY

In this Section the empirical findings are presented. The importance of the proposed model is significant. Worth to mention, that many investors employ an investment model by selection process started with an economic environment drilled down to a single company performance. The proposed approach for efficiency assessment in decision support systems with learning algorithms is the one that fosters to avoid limitations of the conventional efficiency assessment methods.

3.1. Practical implementation steps of decision support systems

The general procedure consists of the following practical steps, which are needed to be integrated parts of the decision support systems:

1. Data model analysis to verify the credibility and integrity of datasets
2. Efficiency assessment and models estimation using various techniques
3. Based on the efficiency models estimation perform feature set selection
4. Apply ensemble machine learning methods
5. Verify the machine learning algorithms

3.2. Datasets acquisition and analysis for decision support systems

The analysis of efficiency of various stages are naturally linked. In environmental settings where multiple inputs are used to generate multiple outputs, aggregation methods are necessary to calculate aggregate input and output levels for a better decision support system. For the research a high-level abstraction layer has been applied in order to provide structured datasets for the research. A multiple-step process is designed to solve the problem of creating a robust decision support system. The presence of uncertainty factors makes any standard panel data model insufficient to characterize the dynamic nature of the economic process.

Therefore, panel data models are in focus because these might capture dynamic changes and used to evaluate fluctuations in efficiency over time. However, no true dynamics can be represented by a single nonparametric model but should assess multi-level approach.

3.3. Efficiency assessment by nonparametric model

To investigate effects of scale of operations both VRS and CRS approach of DEA models are chosen for comparative analysis and validation. The two-stage DEA developed ranks the performance of each organization comparative to each of the two frontiers calculated according to the parameters. In order to gain appropriate result by applications of DEA methods, there is a lot of attention needed to be paid to important modeling issues. Some

of these affect to clearly identifying the purpose of the analysis, deciding on inputs and outputs, choosing a model orientation, and giving more attention to the type of data involved. The modeling process takes place for large randomized subsets with agents having decision-making process limited by disposable information, considering the cognitive limitations and time constraint to make a decision. Therefore, the results might be vague due to the homoscedasticity and lack of correlations, which lead to the fundamental identifiability because of limited amount of observations and uncertainties underneath. Of course, with certain assumptions, it is not always required that the exogenous variables in the models should be fully observable. The most important factor is that, the lack of identifiability of the uncertainty factors makes identification of distributions rather difficult task from theoretical point of view and virtually impossible from practice implication. Ironically, the aspiration to find complete factors of uncertainty can lead to a reduction of the informational power of predictions and their usefulness in the decision support systems.

3.4. Feature set selection and efficiency influencing factors

Selecting an appropriate set of features to represent the main information of original datasets is an important factor that influences the accuracy of efficiency and classification methods. Improving the classification accuracy and predictability ability, increasing the training process speed and decreasing the storage demands are some of the potential advantages of feature selections algorithms. Therefore, reducing the number of feature set, better understanding and interpretability of figures can be achieved.

Sensitivity analysis shows that the results of Feature selection have a significant heterogeneity in the impact effect on firm-level output, depending on the level of uncertainty, and significant convexity in the response of input variables to market volatility. The indication of *Financial leverage* effect with the *Stock volatility* is explained by adjusting in the long run towards a target that is integrated with its overall uncertainty level. Therefore, channels of uncertainty transmission can provide valuable insights regarding these interactions and raises the question of how the efficiency and effectiveness of the policies in achieving their objectives may be affected in influencing the economy through stock market. The analysis proved the assumption in the Chapter II, Section *2.4 Two-stage DEA nonparametric efficiency analysis*, where it is suggested to regard efficiencies separately due to the nature of the efficiency. In the Stage 1 efficiency, the input allocation and firm level efficiency is the key for the firm level efficiency assessment, where externalities do not have direct impact on the technology applied in production. The Stage 2 efficiency has output oriented approach, which is related with the market capitalization and market stock equity. Therefore, the Stage 2 input parameters are greatly influenced by uncertainties.

The results of theoretical assumptions fully correspond with the practical finding. The firms level output efficiency is influenced by 49,45% with input orientated decision making process. The stock market output orientated decision making process supports the assumption of importance of stockholders in development strategy of stock listed firms with 43,39% of such factors. Pure uncertainty factors of 7,16% do still have significant strategic meaning, which has considerable influence in the decision-making process

3.5. Nonparametric efficiency models with ensemble machine learning

The main application of machine learning algorithm is hidden pattern recognition. Therefore, the uncertainty feature is integrated into the set. The goal of normalization is to eliminate redundancy in the datasets, because balanced data attempts to give all attributes an equal weight. Incorporating results of nonparametric analysis of inputs and outputs allows to assess efficiency as classification problem. The Random Forest has the most weight coefficient (0.9078), following by Support Vector Machines (0.822). The Artificial Neural Networks (0.024) and the mean algorithms do not comply with any practical meaning in this research.

There is no clear one algorithm, which can be applicable in any situation. The Random Forest approach give slightly better result, but Support Vector Machines can handle structured problems a better way.

The above analysis implies that Random Forest and Support Vector Machines are suitable for the classification task in the decision making systems with nonparametric efficiency assessment models. In particular, the integration of Support Vector Machines with the various kernel functions and Data Envelopment Analysis method achieved the best results. Proper method selection is necessary for the supplier evaluation which may guarantees firms evaluation optimum solutions when compared with other artificial intelligence approaches. Especially for Support Vector Machines, making an appropriate choice for kernel function is the key to construct a classification model which may enhance the prediction performance according to the above experimental results. Valid experiments using statistical test suggest that Data Envelopment Analysis score is a useful feature to improve the classification performance.

3.6. Empirical Research Results and Discussion

The objective of the empirical research part is to present generalized results on the efficiency assessment under uncertainty and to encourage scientific discussion on the research and the obtained results. The empirical research shows, that established on the theory of the structural risk minimization principle to estimate a function Support Vector Machines is shown to be very resistant to the overfitting problem, eventually achieving a high generalization performance. Due to the advantages of Support Vector Machines algorithm in solving nonlinear problems, it can be used to capture and provide explanatory power of underlying uncertainties.

It is confirmed that Machine Learning algorithms empowered with pre-defined kernel functions are good at coping with data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments over creating applicable yet easy interpretable knowledge about given issue.

CONCLUSIONS

Definition and practical implementation of an effective decision support system under uncertainty is not a trivial task. As the result, the study expresses the idea that in the modern environment there is always space for new investigations and researches.

The study proved, that it is important to involve theoretical and empirical aspects of uncertainty, nonlinearities, complexity and bounded rationality as the major assumption of the framework, but not equilibria based theories. The assessment of efficiency under uncertainty is defined by various sources of uncertainty, which cannot be quantified within other than a hybrid model. The study underpins that the quality of the data is very important factor. The ensemble methods in machine learning require accurate measurement of both the inputs and outputs, construction of datasets for uncertainty. The nonparametric modelling should be regarded with cautious due to its subjective nature. There is no common approach while handling missing data. But within the proposed model in this study the treatment of missing data is one of the important tasks.

This study is one of the first attempts to assess efficiency within both classification and regression model. The study among other researches investigate Random Forest, Artificial Neural Networks and Support Vector Machines classifiers to face uncertain knowledge in datasets. It is important to show how uncertainty data in knowledge datasets can be treated in ensemble methods of machine learning by employing robust optimization. The future of the machine learning is in combination of different approaches, because fully supervised algorithms are useful but not feasible from practical point of view.

The study has explicitly proposed to treat uncertainty not as a dummy variable, but phenomenon dissected within the proposed model on different layers: data-mining uncertainty, analytical framework uncertainty and uncertainty as a factor. Unlike the existing approaches, the combinations of machine learning techniques in this study do not require to think in terms of hypothetical assumption. Mathematically machine learning leads to the identification of implicit restrictions to weights, so there is a fundamental difference in these approaches, emerging from the way in which the data gathered. In each process the uncertainty is emerging in different qualities and it should be assessed with respective techniques.

Depending on a number of factors, it is crucial to elaborate a theoretical framework, which can embrace as many factors dynamically. Therefore, any research on efficiency assessment under uncertainty should have a broader scope and should not be limited on country-specific parameters but include configurations in clusters. Uncertainty have been proven to be unstable factor, with the variations being most vividly seen during the crises, requiring researchers' attention. Due to the reasons mentioned here, the assessing efficiency under uncertainty is relevant in both theoretical and empirical aspects. This study focuses on both aspects. The study confirms that uncertainty is persistent phenomena in economics and it must be faced continually by policymakers. The measuring of macroeconomic uncertainty and understanding its impact on economic activity is crucial for assessing risks in the current macroeconomic situation. Uncertainty should not be oversimplified. The phenomena affect individual sectors of the economy in totally different manner with different impact and different degrees of persistence.

PUBLICATIONS

1. S. Kornilov, T. Polajeva (2016). Economic Complexity: the future of the knowledge-based regional development. SGEM 2016. ISBN 978-619-7105-72-8 / ISSN 2367-5659. Book 1, Vol. 3, P. 239-246, DOI: 10.5593/SGEMSOCIAL2016/B13/S03.032.
2. S. Kornilov, S. Ridala, A. Aasma (2016). Dynamics of economic adjustment under uncertainty: MS-VAR model for Baltic Dry Index. SGEM 2016. ISBN 978-619-7105-76-6 / ISSN 2367-5659. Book 2, Vol. 5, P. 189-196, DOI: 10.5593/SGEMSOCIAL2016/B25/S07.025.
3. Sergei Kornilov, Tatjana Põlajeva (2012). Infrastructure development: Economic Growth Effects // The 7th international scientific conference «Business and Management 2012», May 10-11, 2012, Vilnius, Lithuania: selected papers. Vilnius: Technika, 2012. ISBN 9786094571169. ISSN 2029-4441. eISSN 2029-929X. P. 156–161.

CONFERENCES

1. EAEPE (European Association for Evolutionary Political Economy) 2013, Italy. Impact of innovation and infrastructure development on growth. Estimation of agent-based model approach.
2. VGTU Scientific Conference "Business and Management", 2012. Vilnius. Infrastructure development: the risks assessments and economic growth effects.
3. International PhD Summer School 2011, Vinistu, Estonia. Adopting Value Creation Policy in Economic Shocks.
4. ICEM 2011. Economics and management, Kaunas. Strategic approach in risk-adjusted tariff policy.

CV

SERGEI KORNILOV
Ph: +372 56 215 980
Email: sergei.kornilov@gmail.com

ACADEMIC BACKGROUND

2019–2020	**Mykolas Romeris University** Doctoral studies	*Lithuania*
2011–2017	**Tallinn University of Technology** Doctoral studies	*Estonia*
2008–2010	**Tallinn University of Technology** MBA	*Estonia*
1999–2003	**University for Applied Sciences of Furtwangen** Bachelor degree	*Germany*

WORK EXPERINECE

From 2019	**EVER SOLUTIONS OÜ** Data Scientist	*Estonia*
From 2018	**MTÜ KOOL 21. SAJANDIL** Member of the board, scientific supervisor	*Estonia*
2007–2018	**E.R.S. AS** Head of IT Department	*Estonia*
2003–2005	**GOMAB MÖÖBEL AS** IT Project manager	*Estonia*
2002–2003	**ROBERT BOSCH GMBH** Internship	*Germany*
2001–2002	**SIEMENS LTD.** Internship	*China*

LANGUAGES

Russian	Native speaker
English	Fluent
German	Fluent
Estonian	Fluent

Kornilov, Sergei

ASSESSING ORGANIZATIONAL EFFICIENCY UNDER MACROECONOMIC UNCERTAINTY IN DECISION SUPPORT SYSTEMS: ENSEMBLE METHODS IN MACHINE LEARNING WITH TWO-STAGE NONPARAMETRIC EFFICIENCY MODELS: daktaro disertacija. – Vilnius: Mykolo Romerio universitetas, 2020. 272 p.

Bibliogr. 155–176 p.

The modern environment, where we all live in, is the subject of constant changes over time. There are evidences both from mass-media and science, that the modern economic setting is characterized by increasing information flow gathered for decision making process, growing global competition on the macroeconomic level and limited physical resources. With developing of technologies, the opportunity cost is getting higher at the explosive scale. Thus, the assessing of effectivity plays enormous role in the decision-making process. The large number of studies shows that assessment of efficiency analysis has become an important topic in operational research, public policy, energy-environment management, and regional development. There is a clear shift to more intelligent decision support systems adopting a wide range of information sources from financial ratios, financial statements to mathematical modeling and evaluations. The research methods used in the study comprise analysis, synthesis and comparison of scientific literature to characterize uncertainty and efficiency. Due to the growing interest in the machine learning techniques and BigData, data-driven approaches are becoming very important in many scientific areas and real-world applications. The main focus of the study is to elaborate approaches to carry out a framework for nonparametric efficiency assessment, which is from one hand is reinforced by economic science and on another hand take advantage of the machine learning algorithms to create plausible estimation result.

Šiuolaikinė aplinka, kurioje mes visi gyvename, laikui bėgant nuolat keičiasi. Tiek žiniasklaidoje, tiek moksle yra duomenų, kad šiuolaikinei ekonominei aplinkai būdingas didėjantis informacijos srautas, renkamas sprendimų priėmimo procesui, auganti pasaulinė konkurencija makroekonominiame lygmenyje ir riboti fiziniai ištekliai. Tobulėjant technologijoms, alternatyviosios sąnaudos sparčiai didėja. Taigi efektyvumo vertinimas vaidina didžiulį vaidmenį priimant sprendimus. Daugybė tyrimų rodo, kad efektyvumo analizės vertinimas tapo svarbia operacijų tyrimų, viešosios politikos, energetikos ir aplinkos valdymo bei regionų plėtros tema. Akivaizdžiai pereinama prie intelektualesnių sprendimų palaikymo sistemų, naudojančių daugybę informacijos šaltinių, pradedant finansiniais rodikliais, finansinėmis ataskaitomis ir baigiant matematiniu modeliavimu ir vertinimais. Disertacijoje naudojami tyrimo metodai apima mokslinės literatūros analizę, apibendrinimą ir palyginimą neapibrėžtumui ir efektyvumui apibūdinti. Dėl augančio susidomėjimo mašininio mokymosi metodais ir „BigData" duomenimis pagrįsti metodai tampa labai svarbūs daugelyje mokslo sričių ir taikyme realiame pasaulyje. Pagrindinis disertacijos tyrimo tikslas yra sukurti neparametrinio efektyvumo vertinimo metodiką, panaudojant ekonomikos mokslo žinias ir mašininio mokymosi algoritmus, kad būtų sukurtas patikimas vertinimo rezultatas.

Sergei Kornilov

ASSESSING ORGANIZATIONAL EFFICIENCY UNDER
MACROECONOMIC UNCERTAINTY IN DECISION SUPPORT SYSTEMS:
ENSEMBLE METHODS IN MACHINE LEARNING
WITH TWO-STAGE NONPARAMETRIC EFFICIENCY MODELS

Daktaro disertacija
Socialiniai mokslai, ekonomika (S 004)

Mykolo Romerio universitetas
Ateities g. 20, Vilnius
Puslapis internete www.mruni.eu
El. paštas roffice@mruni.eu
Tiražas 20 egz.

Parengė spaudai Aurelija Sukackė

Spausdino BĮ UAB „Baltijos kopija“
Kareivių g. 13B, 09109 Vilnius
spauda@kopija.lt
http://kopija.lt