The University Of Sheffield.

Informal entrepreneurship and tax evasion in Bulgaria

By
Lyubomir Rumenov Mishkov

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Social Science
School of Management

Submission Date: 19.05.17

Contents

Executive Summary

Purpose
Tackling large informal economies is particularly challenging due to their self-perpetuating nature and complex participatory motivations. This thesis reveals the strongest determinants of business tax evasion, suggesting new evidence for more adequate policy approaches.

Literature
The institutional asymmetry paradigm (IAP), adopted in this study, poses specific empirical challenges. Indeed, how to measure empirically the intrinsic circularity between tangible and intangible phenomena, affecting tax evasive behaviour. The contribution to the literature debates is through appraisal of these issues, proposing an economic model of partly informal entrepreneurship.

Research aims
The main research aim is to develop a holistic account of the socio-economic influencers of the processes of tax evasion and (in)formalisation. A comprehensive review of tax policy management has been conducted, exposing the three major environmental pressures (institutional, economic and political) on Bulgaria's National Revenue Agency (NRA). It has been studied how institutional quality, procedural fairness and distributive justice transform into tangible economic effects on small businesses' evasive behaviour.

Methodology
Mixed methods in exploratory sequential design. The in-depth semi-structured interviews with key NRA officials have been supplementary analysed with NVivo and IBM SPSS Modeler for increased findings' objectivity. This informed a quantitative questionnaire to investigate entrepreneurs' informal tactics. Utilising proven multi-disciplinary methods (Decision trees, Neural and Bayesian networks), the study advances an innovative predictive socio-economic profiling of business tax evaders.

Findings and contributions
Operating in the most challenging EU economic environment, Bulgaria's NRA experiences political influence for asymmetrical tax gap closure, which leads to unjustifiable regulative formalism within an outdated deterrence orientated approach. 76.5% of the entrepreneurs evade taxes due to unfair competition ensuing survivalist reasoning. Evidently, the informal economy share is much higher than current macro-measurement methods report (31.2% of GDP). These tough policy challenges could be tackled only through inducing financially and socially efficient voluntary tax compliance.

Chapter One

1. Introduction.

The current socio-economic dynamics of the EU raise several welfare questions amongst its leaders, governments and citizens. With the most recent EU SME definition (European Commission, 2005, p. 8), updated in 2014 (European Commission, 2014), a special emphasis is put onto entrepreneurship development in the form of micro and small businesses. Economically SMEs are 99% of all businesses in the EU and for the last 5 years created around 85% of new jobs, providing two-thirds of the total private sector employment (European Commission, 2016). Given the widely accepted fact that small firms engage in turnover underreporting to much greater extent (Kamleitner, Korunka & Kirchler, 2012; Tedds, 2010, p. 2471) in comparison with large companies, there is a significant problem of lost revenues for national governments. Therefore, informal entrepreneurship and associated tax evasion becomes an important research field with great potential to improve economies.

One of the biggest obstacles in tackling the informal economy has often been the fact that it was being "conceptually framed as a "tame" problem (i.e. a problem that is complicated but easily solvable, often with a discrete response that can be replicated anywhere)" (Williams, 2014c, p. 3). Governments tend to underestimate how sophisticated this socio-economic phenomenon is and to appraise the full scale of micro and macro participatory motivations. This is so, due to the lack of regular deployment of appropriate empirical inquiries seeking to establish what a country's specific root causes of informality are. Although, contemporary academic literature accumulated sufficient evidence in recent years to support informed and effective policy response in most countries, it is usually unpopular or difficult to utilise by government officials at a policy making level. This study aims to contribute in this respect via a novel way of analytical tools' deployment in an accessible and easier to understand way.

Attempting to reduce the cost of formality many informal enterprises employ strategies, which position them in sophisticated interdependence links with other actors upward or downward in their supply chains. Some authors call this the "informality trap" (Institute for Market Economy, 2004, p. 61-70). This study suggests its own descriptive term – informal chain of interdependence (ICOI). Such phenomena create necessity of "tackling the barriers to formalisation" (Williams, Nadin, Barbour & Llanes, 2012a, 19) and becomes the most important side of the same loop – the first one could be described as tackling the enablers of informal entrepreneurship. Taking the route to formalisation asynchronously from any partnering enterprises within ICOI could ruin the well concealed tax evasive practice and make it very easy for tax authorities to recognise what is happening. So, the partly formal or informal firm willing to decrease the level of informality and thus their tax evasive pattern, may be incapable to afford that or will risk to lose a big part of their business if not the whole of it. Quite often micro-enterprises are driven out of the formal way of conducting their business or have their informality level increased by other informal business actors. The coping (cost-cutting) strategies of entrepreneurs with varying degree of informality are determined by various motivations, but the economic rationale plays significant role especially in transition economies. This study aims to model the Bulgarian informal entrepreneurship in its dynamics of economic reasoning. How entrepreneurs, who operate micro or small businesses, take decisions to reduce the costs of formality and thus face the costs of informality, which also need appropriate coping strategies.

The main research aim in this study was to develop a holistic account of the socio-economic influencers of the processes of business tax evasion, formalisation and informalisation and as a result to inform adequate policy responses. This task required emergence of two investigative levels – institutional and entrepreneurial. The first research level was attained with a qualitative method to achieve an in-depth understanding of the institutional context with a focus on tax policy management. In particular, revealing how the three major environmental pressures (institutional, economic and political) affect NRA's role in revenue raising policies and the associated institutional asymmetry. The main qualitative research question pursued in this thesis can be expressed as follows: How does a national revenue agency try to tackle tax evasion in an institutional context that makes operating within the informal economy a practically-viable alternative to operating in the formal economy? Full operationalisation of this research objective is available in chapter 4 (see Figure 5). To explore this central question, four lines of inquiry have been reported: discovery and exploration of agency strategies; agency operational approaches and the predominant type in the mix of policy measures; and the triggering procedures and methods of enforcing control and compliance – tax audits and inspections. The analysis and findings have been presented in chapter 5. The second investigative level was attained with a quantitative method (a business survey) deployed amongst entrepreneurs. The main quantitative research questions can be expressed as follows: What the strongest determining factors of business tax evasion are and whether the intrinsic circularity between tangible and intangible socio-economic phenomena is affecting tax evasive behaviour? Full operationalisation of this research objective is available in chapter 4 (see Figure 5).This research task required the development of an empirical instrument capable to appraise the root causes for partial informality amongst micro and small businesses. In doing so, the employed questionnaire and associated analysis contribute in two main aspects. Firstly, it advances the so called predictive economic profiling of business tax evaders exploiting an innovative application of proven multi-disciplinary analytical methods. Secondly, the quantitative empirical instrument enriches the literature debates about informal economy measurement's approaches – still arguing about the superiority of either macro or micro-economic methods.

The thesis is structured as follows:

Chapter Two: Literature review
The chapter appraises the literature on various definitions of the informal economy and entrepreneurship. As will become evident, there is lack of consensus and this has specific complications in regards to measurement, main drivers and policy responses to the informal entrepreneurship. A review of contemporary policy approaches and size estimation methods has been conducted to outline the current state of knowledge and establish areas of worthy contributions.

Chapter Three: The Bulgarian informal economy
The chapter presents the major Bulgarian socio-economic indicators characterising the research context. Bulgaria is the least developed EU state with the lowest OISC index. This is of particular importance to the understanding how entrepreneurs decide their informal strategies given the social and business setting in which they operate and what the policy response should be.

Chapter Four: Methodology
Chapter 4 starts with presentation of the research aims and objectives from within the philosophical perspective of the study. The study employs mixed methods in exploratory sequential design. The chapter explains the rationale behind the particular research design choices, methods' utilisation and the empirical strategy required to

investigate this sensitive topic. The presented empirical instruments' development and analytics are considered to be methodological contributions in their own right as this enables discovery of important findings in the latter chapters. The main research aim is to develop a holistic account of the socio-economic influencers of the processes of business tax evasion, formalisation and informalisation and as a result to inform adequate policy responses.

Chapter Five: Findings of semi-structured in-depth interview
Chapter 5 addresses the first out of two investigative stages to achieve this. The key qualitative contribution of this research is unveiling the three major environmental pressures upon NRA's strategic and operational approaches in tax policy management characterised by institutional incongruence status. While the first two – economic and institutional environments are amongst the worst in socio-economic terms within the EU, the political influence for asymmetrical tax gap closure as well as consequential misuse of collected revenue, appear to have the most significant impact on the agency's work. In particular, this creates two major types of difficulties – firstly, this impedes the work of senior officials who want to modernise tax policy in order to secure a greater level of operational independence and thus higher procedural fairness. Secondly, the findings indicate that current institutional practices disturb the social feeling for procedural fairness to such a degree that the institutional asymmetry and entrepreneurial informality are being progressively deepened.

Chapter Six: Findings of business survey
Based on analysis of a business survey, chapter 6 has two major parts: the summary statistics for individual variables and the so called modelling (predictive analytics). The second investigative level has been attained with a questionnaire, which was deployed amongst entrepreneurs to research their informal tactics. It was studied how institutional quality, procedural fairness and distributive justice transform into tangible economic effects on small businesses' tax evasive behaviour. Applying proven multi-disciplinary analytical methods (Decision trees, MLP Neural and TAN Bayesian networks), the quantitative analysis advances an innovative predictive socio-economic profiling of business tax evaders. The quantitative analysis utilised Multinomial Logistic Regression to verify the key findings. It has been demonstrated that unfair competition from informal businesses is the most frequent and strongest predictor of tax evasion. In addition, entrepreneurs exercise tax evasion in order to survive and as a corresponding response to the poor tax and administrative policies, which cause them to spend more valuable resources than deemed acceptable.

Chapter Seven: Discussion of findings and contributions
While the previous chapter concentrates on quantitative findings with the emphasis on the rationale and strengths of the methods, this final chapter appraises the significance of the results in relation to the extant literature. In doing so, it becomes evident that Bulgaria's NRA operates in the most challenging economic environment in the EU. Within its limited manoeuvrability, the agency experiences political influence for asymmetrical tax gap closure, which leads to unjustifiable regulative formalism within an outdated deterrence orientated approach. It was discovered that 76.5% of the entrepreneurs underreport their turnovers with the purpose to evade taxes (mainly to survive). This amounts to a significant proportion of the country's GDP and suggests that the proportion of GDP in the informal economy is much higher than current macro-measurement methods report for Bulgaria (31.2% of GDP). The logically clear and transparent decision trees and neural networks unveiled that unfair competition is the most frequent and strongest predictor of business tax evasion in conditions of very high real cost of doing business. The discovered high *frequency of societal involvedness in the informal economy* multiplied by *high Intensity of societal involvedness* constitutes a tough policy challenge. It could be tackled only through

inducing financially and socially efficient voluntary tax compliance. Indeed, to increase tax compliance and reduce the share of the informal economy, a qualitatively new policy approach is required to decrease the institutional asymmetry, based on two pillars. Firstly, to achieve a widely acceptable level of procedural fairness, namely equivalent treatment of everybody trough equal rules enforcement in the tax management domain. Secondly, to address the major deficiencies in the distributive justice mechanisms and in particular the perceived value of public goods and services as a function of the collected taxes. In order to achieve an appropriately sustainable social and economic system, the taxpayer should feel that they bear a reasonable tax burden within an acceptable real cost of existence in relation to others.

1.1 **Research context.**

Brief political-economy history of Bulgaria.

In 1945, after the Second World War, Bulgaria became a communist state with a Soviet-style planned economy. The communist regime continued for 45 years until 1990 when the government allowed multi-party elections. They were unsurprisingly won by the Bulgarian Socialist Party (BSP – a new name of the Communist Party). Twenty-seven years later the "Transition" period to open market economy still continues, although some politicians would argue that it has been successfully completed. The processes of initial creation of business capital were corrupted in their genesis, so that they resulted in tarnished capitals installation schemes, which were deeply concerning the Bulgarian society. Private economic activity was not allowed during the communist era, so there was no private seed and working capital. Instead, the Bulgarian Communist Party entitled selected individuals (from their upper echelon) with the privilege to use state sourced funds for private business. As a result, powerful oligarchic circles emerged with strong political embeddedness, which continues to impress even nowadays. Their political influence was enabling legislative development in support of their economic interests.

Some main socio-economic indicators.

In socio-economic perspective, Bulgaria was accepted as a full EU member state in January 2007. Unfortunately, the country remains one of the least developed economies in the EU27. Bulgaria has a 7.33 million population (Eurostat, 2013). The volume index of GDP per capita in Purchasing Power Standards (PPS) is the lowest while the risk of poverty is the highest in comparison with the other EU member states (see Appendices 2 and 3). Furthermore, the social cohesion index (OISC) is one of the lowest in the EU, while the shadow economy is the highest (Figure 4).

1.2 **Rationale: need for this research, its importance and significance.**

There is growing interest amongst academics about entrepreneurs' economic motives in choosing their level of informality. In particular, why and how business owners decide to offset the expenses associated with formality and consequently pursue informal strategies. This study aims to fill that gap and provide an empirically verified explanation of these phenomena. Pursuing such research aims requires specific methodological choices and it is proposed that Bulgaria satisfies them for the following reasons:

- Bulgaria has very high costs of formality and as such provides highly distinctive grounds for studying the main corresponding phenomenon – costs of

informality. The entrepreneurs' coping strategies are more explicit and subject to more idiosyncratic economic reasoning.

- Bulgaria is the least economically developed EU member state (see Appendix 2 and 3) with the highest shadow economy share (Buehn & Schneider, 2011). Thus the findings of this study would be highly relevant and beneficial – to inform adequate and necessary policy response.

The earliest trace to inventing the two central research concepts ("cost of formality" and "cost of informality"), leads to De Soto's seminal work (1989, p. 132-177). He argues that cost of formality is dependent on two corresponding concepts coined by Instituto Libertad Democracia (ILD): "the costs of access" to an economic activity and the "costs of remaining" in it. It is how much financial resource and efforts the entrepreneurs are required to spend to establish themselves formally in business and remain in compliance over time. De Soto's findings are that decisions to conduct informal activities are largely based on rational evaluation of the cost of formality. Furthermore, informality is discovered to "involve tremendous costs" (De Soto, p. 152), which people attempt to offset in inadequate ways. Based on De Soto's conceptual definitions, it becomes clear that they could not be looked at in isolation since they are strongly correlating phenomena.

In order to outline the significance of the research it needs to account for the following worthwhile implications:

- Cost of formality and cost of informality have been identified and studied conceptually and to certain degree empirically, but has not been operationalised in transition economies like the Bulgarian one. This is an important investigation considering that the economic challenges for the least developed EU state due to the largest informal sector might be the greatest. Most importantly, the main object of these concepts – the entrepreneur – has not been able to benefit from academic developments in the area. This needs changing through enabling the entrepreneurs to consume practicable applications of such academic work.

- In essence, there is concrete necessity to equip policy responses to the hidden economy with analytical tools, which would enable the inducement of financially and socially efficient voluntary tax compliance.

- The social importance of informality needs constant re-evaluation in relation to cost of formality and informality. Such an appraisal would be different for different economies. It could be argued that in less developed economies both cost of formality and informality are higher in comparison with more developed countries. That is why studying those phenomena have particular importance for Bulgaria and the potential for extrapolating of the findings to other transition economies is feasible.

There is always an implicit cost of transacting informally, which is difficult to predict by entrepreneurs who chose this mode of doing business. The main research aim of this study is to explore the relationship between the level of cost, time, and procedures, which push entrepreneurs to function partly/fully informally or impede formalisation and how these variables affect their tax evasion behaviour. We can observe an increased effort by governments and their policy units to influence and measure "shadow activities". The Italian tax authorities have recently created a self-diagnostic application (Redditest) to tackle the average tax evasion, which is widespread. The Australian Tax Office created an "Online Business Viability Assessment Tool" in their

effort to limit the risk of non-payment of tax debt. The Latvian Confederation of Employers launched a campaign “Against the shadow economy – for fair competition”. It included an online test for measuring the impact of an individual’s “shadow” activity. While this study’s rationale is parallel to those initiatives, it develops a new methodology in predictive diagnostics and profiling of business tax evaders. The thesis aims to build a holistic account of the main determinants of entrepreneurs’ tax evasive decisions as a function of the socio-economic environment in which the studied small firms operate. In doing so, it contributes by showing that intangible phenomena (such as institutional quality, procedural fairness and distributive justice) transform into tangible economic influencers of compensatory evasive behaviours.

1.3 **Theoretical framework and important research insights.**

The study appraises various theoretical frameworks, which explain the problems and root causes of informality (see section 2.3, 2.4). They all have different deficiencies in offering adequate theory to describe engagement with informal economic activities. The lack of contemporary explanatory framework until recently has been attributed to a couple of issues. With some exceptions (Williams, Horodnic & Windebank, 2015a; Williams & Horodnic, 2015d) the institutional theory on tax evasion have neglected studying empirically the interdependence of formal and informal institutions (Gërxhani, 2004b, p. 741). Helmke and Levitsky (2004, p. 727) description of formal and informal institutions has been selected as the most applicable to this study:
"informal institutions as *socially shared rules, usually unwritten, that are created, communicated and enforced outside of officially sanctioned channels.* By contrast, *formal* institutions are rules and procedures that are created, communicated, and enforced through channels widely accepted as official. This includes state institutions..." Therefore, institutional asymmetry meaning is to be understood as “the misalignment between formal and informal institutions” (Williams & Vorley, 2014d, p. 2). Adopting IAP, this study argues that large institutional asymmetry is guaranteed by default in transition economies (such as Bulgaria), which logically leads to significant tax evasion. Institutional asymmetry is now widely recognised as being positively correlated with informal economic activities – both at entrepreneurial and employment level (Williams et al., 2015a; Williams et al., 2015b; Williams et al., 2015c). In comparison with previous theorisations of informality, IAP is much more holistic in its appraisal capabilities. In particular, with its integrative understanding that tax evasive behaviours are emergent to a complex set of phenomena such as the socio-economic dynamics of the environment in which tax payers exist, as well as the dynamics of their perceptions for injustices and inequalities (procedural fairness and distributive justice).

However, precisely this holistic capability of IAP to explain the root causes of informality, poses specific empirical challenges from research perspective. Indeed, that is to discover and measure empirically intangible phenomena, which relate to social actors and enterprises involved in illicit economic activities. This thesis contributes to the developing literature debates in that regard, through operationalisation and investigation of the main concepts underpinning what appear to be a distinctively visible institutional asymmetry in a transition setting. To achieve this task, the study employed mixed methods in exploratory sequential design. Firstly, it investigated the formal institutional framework with specific focus on the National Revenue Agency and tax policy management. Secondly, it appraised how intangible phenomena such as institutional quality, procedural fairness and distributive justice transform into tangible economic effects on small businesses’ tax evasive behaviour.

Given the largest informal economy share, perceived institutional corruption and the

poorest socio-economic environment in the EU (see chapter 3, Figure 14, 15, 16, Appendices 2 and 3), it is not surprising that most small businesses must exercise huge effort to just maintain economic viability. It has been discovered that the most frequent motive for tax evasion is to simply survive economically (see sections 6.2.1 and 7.1). Evidently, the problems of self-perpetuating informality are affecting the fundamental constructs of the public contract in the Bulgarian society. One of the challenging circularity effects is that the individual cost of being honest in a corrupt society becomes significantly higher in comparison with societies where most individuals are honest (Andvig & Moene, 1990). This affects horizontal and vertical trust between economic actors and the state, fuels even greater tax evasion and the so called informal chains of interdependence (ICOI). An important contribution to knowledge is the appraisal of the *real cost of doing business* and how in particular this affects firms' economic behaviour, depending on their (in)formal economic status (see Figure 26, section 7.3). It has been discovered that (section 6.2.3, Table 22, Appendix 35) the strongest predictor for participation in ICOI is the disagreement with fair expenditure of tax money. 86% of all firms willing to formalise from within ICOI have full disagreement with how tax money is being spent and opinion that they are not treated in a fair and professional manner by the state institutions. Informality chains of interdependence are one of the significant self-perpetuating mechanisms of spreading informality and simultaneously obstructing formalisation via the power of their economic coercion. This turns them effectively into enablers of informality and barriers to formalisation. The study offers empirically informed policy solutions to two of the most challenging problems in regards with ICOI's economic coercion towards informalisation (see section 7.3).

In terms of cost of doing business, there are four substantial drivers of tax evasion, which develops as a responsive-compensatory strategy at entrepreneurial level. Firstly, unfair competition from fully informal or partly informal businesses is the most frequent and strongest predictor of business tax evasion. The current institutional capacity to tackle this issue and provide a levelled business environment is clearly inadequate. Secondly, the problem with institutional corruption at all levels, which causes unfair and expensive public service delivery – the so called significant costs of "getting things done" from entrepreneurial perspective. Considering the fact that small firms are highly likely to engage in corrupt activities within low institutional efficiencies (Tonoyan, Strohmeyer, Habib & Perlitz, 2010), this contributes to increasing informality. Thirdly, the high cost of doing business according to the rules due to bureaucratic inefficiencies as a result of poor tax and administrative policies. Lastly, there is a political influence, which affects formal institutions and the perceived procedural fairness and distributive justice amongst business tax payers. This study suggests new empirical insights, based on a novel methodology of data acquisition and analysis.

Chapter Two

2. Literature review.

This chapter unfolds with a discussion on the complexities and lack of consensus on defining what the informal economy and entrepreneurship is, including the emerging theoretical consequences. It gradually builds the argument that the best way of defining the informal economy and entrepreneurship is to employ a modular principle, depending on a widely accepted definition, which is fine-tuned to reflect a particular research context. This is important for two reasons. Firstly, to enable an accurate and exhaustive definition of the various informal economies, which tend to differ substantially across national borders by their nature of existence. Secondly, to enable appropriate design of tackling policy measures. Development of adequate policy approaches therefore is subservient on precise definition and comprehensive understanding of the root causes of informality at a national level. A working definition of informal entrepreneurship was developed for the purpose of this study.

The literature appraisal for the purpose of this study is constructed around the so called systematic literature review to ensure higher level of objectivity and precision in the positioning of the main arguments. To enhance research transparency and replicability a detailed description of how this has been achieved is subsequently offered including supporting appendices. Given the research aim to understand entrepreneurs' informal strategies and the basis of their choices to conceal economic activities, an account of the main competing literature theorisations is then presented. It will become evident, that neither of the current explanatory paradigms are sufficiently adequate. A group of scholars started to address the need of a new holistic theorisation, suggesting the institutional asymmetry paradigm (Williams et al., 2015a; Williams et al., 2015d; Williams et al., 2014d). Adopting IAP, this study aims to address its empirical challenges – indeed, how to measure empirically the intrinsic circularity between tangible and intangible phenomena, affecting tax evasive behaviour. The contribution to the literature debates is via appraisal of these issues, proposing an economic model of partly informal entrepreneurship through innovative socio-economic profiling of business tax evaders.

Estimating the real size of a shadow economy has implications about what policy responses should be utilised. In consideration of the self-perpetuating nature of a large informal economy (namely the Bulgarian one) a review of current macro and micro economic measurement methods was developed. As a result, the reader is enabled to appreciate the measurement capability of the specifically designed quantitative instrument in respect of its measurement potential. The final part of the literature review is dedicated to introducing the main socio-economic indicators characterising the research context. It aims to equip the reader with an understanding of how the studied informal entrepreneurship strategies and their institutional framework relate to their native setting.

There are currently three major literature streams studying and explaining informal entrepreneurs' motives to engage in shadow activities. These could be classified as: the literature on informal economy and entrepreneurship, literature on tax evasion and non-compliance, and economic psychology literature on taxation. Each of these streams inform corresponding policy measures based on its findings, but there are very few attempts to integrate their research approaches and findings into a holistic policy response to the problems of informality. Djankov, Lieberman, Mukherjee and Nenova (2003) were amongst the first authors who looked to develop cost-benefit analysis of informality. Their main rationale is that "*All informal activity has one common feature: the entrepreneurs who operate in the informal economy perceive the*

benefits of doing so to outweigh the costs of going formal" (p. 63). Simeon Djankov is the inventor of the World Bank's Doing Business report and the methodology being used in the report's series is similar. The authors calculate business registration time and attempt to establish full-cost measure of the entrepreneurs' time to comply however there is no explicit quantifiable model of this. Of course, we need to be aware that rational cost-benefit reasoning could be limiting in explaining entrepreneurial behaviour. This study aims to reveal the economic rationale behind informal strategies and choices, while recognising that that there are socio-psychological motives, which could be even predominant. However, it is argued that in a less developed entrepreneurial environment cost-benefit reasoning may have bigger weight in the decisions for involvement in informality.

According to a study of the Institute for Market Economy (2004, p. 18) "*Whenever the costs of operating in the formal sector exceed those of operating 'in the shadows', there are grounds for informal economy expansion*". Oviedo (2009) introduce stimulating argument about an increased number of theoretical papers, which motivates the existence of informal entrepreneurs as the result of rational economic choices. This study will build upon what has been discovered so far in costing formality and informality, but will aim to develop more practicable contribution in the entrepreneurial domain. In particular, revealing the root causes and motivations for entrepreneurial informality via innovative application of analytical methods.

There is an increasing interest towards the so called systematic literature review (SLR) and its significance for management research. In discussing the advantages of systematic literature review, Saunders, Lewis and Thornhill (2012, p. 113) consider that it will ensure that the report is clear, allowing others to assess the strength and weaknesses of the study. Easterby-Smith, Thorpe and Jackson (2008, p. 46) further clarify the point that the review criteria should be rather systematic than simply personal choice and "*transparent for others to see what you have done and so to judge the relevance and substantive nature of your review.*" Both groups of authors base their discussion on an earlier paper of Tranfield, Denyer and Smart (2003, p. 220) who believed that a *"systematic review can be argued to lie at the heart of a "pragmatic" management research, which aims to serve both academic and practitioner communities".* SLR main advantages are most easily visible when there is too much or too little literature in a specific research area (see Table 1, Appendix 4). The literature on informal economies increased substantially in size over recent years, hence the SLR approach is a sensible method to engage with such a field. Mishkov (2012) provides a pragmatic example of SLR applied within a small scale management research study. It is a particularly useful approach in providing evidence base for policy and practice in social sciences (ESRC, 2001).

Obviously, every SLR has certain limitations (Petticrew & Roberts, 2006), which emerge from the very nature of the approach. It is important to note, that this study appreciate that fact and aims to present an example application of SLR on the topics of interest. Each SLR is dependent to the research perspective of the author. Although SLR bears the potential of a much less biased interpretative approach, we should be aware of the risk that SLR could create a very well structured subjective analysis. If this happens, then it is easily visible to the detached reader. SLR is being used to an extent to discover the relevant areas of the study and position the research efforts with sufficient precision and justification. This means that the power of SLR is utilised at different stages selectively in an attempt to avoid the constraints, which SLR may impose on the research direction and creativity. Every possible attempt is made to use all advantages of a SLR approach, but giving priority rather to creativity, research flexibility and value than to the literature reporting frame.

2.1 **Literature review – systematic research protocol.**

The guidelines of Petticrew et al. (2006) presented in Appendix 1 have been used in constructing the SLR approach. The literature review of this study is developed around three scoping literature screenings (SLS) and four conceptual thematic areas, which as every other part of the protocol will be updated accordingly.

Scoping literature screenings:

1. Key expressions (see Table 2)
2. Peer-reviewed journals (see Table 3).
3. Key authors (subjectively/thematically determined).

Conceptual thematic areas:

1. Definitions of the informal economy.
2. Explanations – theorisations.
3. Methods estimating the size of the hidden economy.
4. The Bulgarian informal economy.

The initial SLS were executed in Primo Central and ScienceDirect. Full list of all the databases included in Primo central is provided by University of Sheffield (2013). The searches were implemented without date of publication restriction and with syntax to include all words from each search expression. This means that every result set contained every word from each expression. The approach led to discovery of 104 journals containing relevant publications. The generated search results (journal articles) were examined to ensure that each included journal contain at least one thematically relevant publication, which has a discussion in regards to a key expression. Most of the journals contained several relevant papers. Particular sets of interconnected articles have been selectively appraised to focus the scope of the emerging literature review. Some articles and journals accordingly, have been excluded, because the discussion of the topic was only marginal.

2.2 **Definitions of the informal economy and entrepreneurship.**

It is of central importance to this study to develop its arguments based on analytical appraisal of the literature on various definitions of the informal economy and entrepreneurship. As will become evident, there is lack of consensus and this has specific complications in regards to measurement, main drivers and policy responses to the informal entrepreneurship. The specifically designed systematic research protocol has been used to achieve maximum objectivity and rigour in evaluation of the various viewpoints.

Keith Hart is recognised to be the first author (Chen, WIEGO, 2012, p. 2; Gërxhani, 2004a, p. 269) defining the concept of the informal sector, which was done in a study of a Third World context. His understanding of the informal sector is that one is a participant based on whether their income is generated in a wage rewarding job or it is coming from an outside source like from self-employment (p. 68): *"The distinction between formal and informal income opportunities is based essentially on that between wage-earning and self-employment".* Hart (1973, p. 68-69), developed the earliest typology of the informal sector acknowledging the difference between legitimate and illegitimate activities.

Despite of the early contribution of Hart, it is widely considered (Batini, Levine, Kim & Lotti, 2011, p. 30; Bromley, 1978, p. 1036; Chen, WIEGO, 2012, p. 2; Gërxhani, 2004a, p. 269; Huang, 2009, p. 406; Mead & Morrisson, 2004, p. 1617; Peattie, 1987, p. 853;

Siggel, 2010, p. 94) that the founding research on the informal sector is done by the International Labour Office (1972). With the empirical evidence they present it becomes clearer that the informal sector activities could be *"economically efficient and profitmaking, though small in scale and limited by simple technologies, little capital and lack of links with the other ("formal") sector" (ILO, p. 5).* ILO has an intriguing view to encourage a positive attitude of governments towards the promotion of the informal sector. This view is grounded in *"wide range of low-cost, labour-intensive, competitive goods and services" (ILO, p. 21)*.

Gershuny and Pahl (1979, p. 127) develop an interesting argument about the genesis of the informal economy. It starts from the technological advancement in industrial production and hence consequential unemployment due to the lower demand for labour. So, the only way to keep the same level of employment is through restructuring to the service sector and increase in public expenditure. The authors' prediction is that we cannot escape the development of a dual economy, but we need to control it in an acceptable form.

It is inquisitive to trace the transformation of how Feige (1979, p. 6) chose to call the "irregular economy". He decided to employ this term from the sociologist Louis Ferman and thought that terms like "subterranean" or "underground" are too readily prone to journalistic sensationalism. According to Feige *"The term "irregular economy" is used to refer to those economic activities that go unreported or are unmeasured by the society's current techniques for monitoring economic activity."* Some 11 years later, Feige (1990, p. 991) acknowledged that there is no single underground economy, but there are many.

Gërxhani (2004a) developed a thorough study of the literature to provide a comparative account for the contributions on the informal sector in developed and less developed countries. The paper focuses on the criteria used in the literature by different authors to define the informal sector, the relationship between the formal and informal economy, tax evasion, and public choice analysis. The author stresses that the distinction between the two types of countries is of key importance in regards to how the informal sector is defined. Despite the strengths of Gërxhani's general overview of the contributions on the informal sector literature, it is surprising not to see evaluation in the analysis of two of the key authors on the informal economy – Colin Williams and Friedrich Schneider.

Wallace and Latcheva (2006) adopt more locally specialised approach when defining what the informal economy is in the transition countries of Central and Eastern Europe. They acknowledge that defining the informal economy should be done at a country level, since this could differentiate significantly. A particular trend in the countries of Central and Eastern Europe is the fact that various activities become legal and illegal (frequent legislation changes) in a pace that is impossible to build into the typology model the authors develop (p. 85).

There is an excellent account of informality definitions (ILO & WTO, 2009, p. 51-54) , which incorporates how actually the concept has become more operational. The supported viewpoint in this joint study of ILO and the WTO is that informality could be defined at the level of production units or firms, or at the level of workers. The main criteria being used to define informal enterprise are their size and their registrations status. Under the upgraded statistical definition of the informal sector (called later the informal economy) – "*informal employment" is understood to include all remunerative work – both self-employment and wage employment – that is not recognized, regulated or protected by existing legal or regulatory frameworks and non-remunerative work undertaken in an income-producing enterprise*" (p. 53). It is

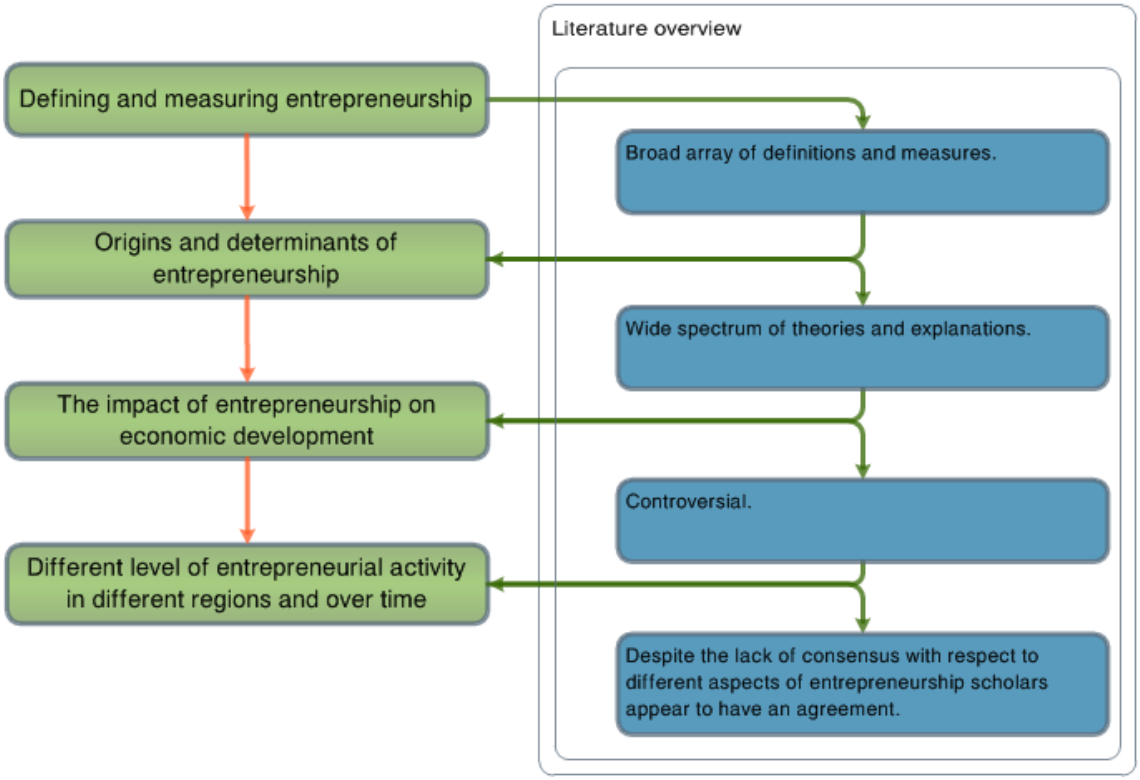
emphasized by ILO&WTO that due to the broadening of the definitions in time and associated consensus difficulties, a new integrated approach is currently arising based on the idea of multi-segmented labour markets. The main idea of this approach is that the informal economy is constituted of different segments that are occupied by different actors (p. 45). This appears to be highly potent literature trend, which is recognised and supported by ILO&WTO.

Andrews, Sánchez and Johansson (2011) review the informal economy through cross-country research perspective in order to expose the link between informality and corresponding public policy. They argue that "*There is no unique definition or measure of the informal economy. It typically refers to economic activities and transactions that are sufficiently hidden so that they are unmeasured or untaxed, and it is presumed that economic agents are at least passively aware that bringing these activities to the attention of authorities would imply tax or other legal consequences.*" Andrews et al. support a popular perspective (in accordance with Feld and Schneider, 2010; Easton, 2001) that the informal economy's definition that is most useful depends on the policy concern that motivates the analysis and data availability.

Two distinguished authors on the informal economy Schneider and Williams (2013, p. 23-27) provide excellent and concise discussion about definition of the shadow economy. The authors underline the important fact that correct measurement is dependent on a clear definition. In order to avoid any ambiguities and controversies they developed a table of "Sample activities and the shadow economy" (p. 26), which provides very useful example activities included or excluded from the shadow economy. For the purpose of the measurement methods in their monograph, Schneider and Williams accept that the shadow economy:
"*includes all market-based production of legal goods and services that are deliberately concealed from public authorities for the following reasons:*
- *to avoid payment of income, value added or other taxes;*
- *to avoid payment of social security contributions;*
- *to avoid having to meet certain legal labour market standards, such as minimum wages, maximum working hours, safety standards, etc.; and*
- *to avoid complying with certain administrative obligations"* (p. 25)

The authors exclude illegal underground activities and the informal household economy. This narrower definition fits and corresponds to earlier definitions, see for example (Williams & Round, 2007a, p. 426; Williams, Round & Rodgers, 2007b, p. 403; Williams & Windebank, 2002, p.232). Defining entrepreneurship and its informal modes is a difficult task, which has been addressed by various authors, but there is lack of consensus (Williams, 2007, p. 2; Williams, 2008, p. 1042). Grilo and Thurik (2004, p. 2) also confirm that fact and provide excellent literature review on entrepreneurship in Europe.

Figure 1: Literature review of entrepreneurship in Europe
Adapted from Grilo et al. (2004, p. 2).



It could be argued that the definitive approach to informal entrepreneurship should mirror the specific considerations we have when providing definition of the informal economy. Informal entrepreneurship should be considered and evaluated in correlation with the local socio-economic norms. In appraising the entrepreneurs' involvement in off-the-books activities Williams (2008, p. 1041) makes an important claim that studies "*have so far failed to evaluate the involvement of entrepreneurs in off-the-books work*". He finds prevalence of those practices amongst early-stage entrepreneurs and in later co-authorship work develops arguments about theories of informal entrepreneurship and evaluation of their motives. In their study of Russian informal entrepreneurs, Williams and Round (2009) discovered that both necessity and opportunity-drivers were co-existent in the decision to start up a business. Moreover, there is a shift towards opportunity-drivers with the better establishment of the business. Williams and Nadin (2010b, p. 363) suggest a definition of informal entrepreneurship:
"*Informal entrepreneurship, in sum, is here defined as involving somebody actively engaged in starting a business or is the owner/manager of a business that is less than 42 months old who participates in the paid production and sale of goods and services that are legitimate in all respects besides the fact that they are unregistered by, or hidden from the state for tax and/or benefit purposes.*"

An example definition of informal entrepreneurship in the Bulgarian context is suggested:

Informal entrepreneurship is dynamics of economic activities exercised by the entrepreneur or other people managed by them while acting on their behalf. These activities are legitimate in all respects, but their resulting economic profit is either fully or partly unregistered for the purpose of tax evasion.

Based on the literature debate about defining the informal economy and entrepreneurship, there are some important remarks to be made. There is no consensus on a single definition of the informal economy or informal entrepreneurship, but there are more widely accepted definitions, which are supported by the majority of authors. It is argued that the best way in defining the informal economy and entrepreneurship is to employ a modular principle. One could start from a general widely accepted definition, which is applicable to various country specific scenarios and then add set of descriptive characteristics to fine tune the definition to the research context. As a result, certain set of definitions of the informal economy would be more appropriate for developed countries in comparison with other definitions, which are more suitable for less developed countries. Therefore, defining the informal sector should be done with careful consideration of country specific factors. As informal entrepreneurship is phenomenon within the informal economies across the world, it is logical that the definitive approach should mirror the specific considerations we have when developing definition of the informal economy.

Another important definition related factor is the comparative cross-country analysis. In particular, scholars should depend on more widely accepted definitions, which is more likely to produce accurate results and satisfy the academic audience in terms of contributions' validity. Some authors present an argument for the nonexistence of systematic data available to research the informal sector phenomena and hence the difficulties to invent a single definition. However, there is substantial empirical evidence that there is need of country specific definitions to reflect on many regional factors of economic and socio-legal aspect. For instance, designing and undertaking economic activities for the purpose of tax avoidance could be legal in one country and illegal in another. The informal economy is not a separate phenomenon – it is not just another economy. It is rather part of the socio-economic country specific dynamics. This includes the state of the national economy, tax and business legislation, institutional and regulative quality, tax morale. In this perspective, the informal sector in any country will always respond to changes of the socio-economic dynamics.

2.3 **Theorisations of informal entrepreneurship.**

In order to understand entrepreneurs' informal strategies and the basis of their choices to conceal economic activities we need to be aware about the different explanations of informal entrepreneurship in the literature. These competing theorisations emerge from the diverse perception of the entrepreneurs' characteristics, types, motives and role in the economy. Williams et al. (2010b, p. 365) find out that different theories of entrepreneurship apply more in some populations than others and introduce key types of informal entrepreneurs – wholly or partly trading off the books. As soon as they become more established in business, we could distinguish between those entrepreneurs who become "*serial users of informal trading practices and those whose enterprises are in transition toward legitimacy*". According to Williams and Nadin there are 4 perspectives on the existence of informal entrepreneurship in economies, which will be reviewed below:

- Modernisation perspective.
- Structuralist perspective.
- Neo-liberal perspective.
- Post-structuralist perspective.

2.3.1 Modernisation perspective.

It was widely accepted that the formal economic configuration and organisation of

labour specialisation will take over the traditional production of goods and services. In essence, the informal entrepreneurship would be a diminishing mode of production, which would not fit into the modern economic system characterised by the enterprise culture. Consequently, the existence of informal entrepreneurship was considered to indicate under-development, since the demand for formal work by the modern enterprise should absorb the output of informal entrepreneurs and thus formalise them. Lewis (1955) developed a political economy perspective on the entrepreneurs, taking into account "The will to economise", "Economic institutions" and "Capital requirements". His modernisation view of the theory of economic growth makes him introduce the term "backwardness" (p.136, 153, 204) – where this condition of certain economies could enable spectacular growth based on the desire for countries not to fall behind others. Fallers (1963, p. 158-220) present a similar argument reviewing the entrepreneur from the prism of structural modernisation. Indeed, he makes the development distinction between the "owner-entrepreneur" and the "the modern corporation" accounting for the agent theory (i.e. ownership and management division). In Faller's view the modern occupational structures are those, which would bring welfare-state policies for the workers and help them out of their traditional informal work mode.

Gilbert (1994, p. 57-75) argued that the absence of the welfare state (in Latin American context) is forcing people to participate in the informal sector either at a low or higher skilled level. Due to rapid industrialisation and urbanisation too many people became involved in low productivity economic activities, a process identified as "tertiarisation" (p. 60) or remained unemployed. Their effort to earn a living had to inevitably include informal economic activities. The state did not manage to intervene adequately providing support and achieving a more balanced development process in order to catch up. This implies that the informal entrepreneurs are a phenomenon supposed to disappear in developed economies. It is well known that the level of informality is comparatively lower, but they still exist. The informal economy and entrepreneurship (Dickerson, 2010; Jütting & De Laiglesia, 2009; Minard, 2009) are expanding and developing diverse forms in many countries. Evidently, the modernisation perspective is inadequate to explain the growth of informality, so other theorisations need to be considered.

2.3.2 Structuralist perspective.

This is the first school of thought to recognise that the informal sector and entrepreneurship are fostering and expanding in the conditions of open unregulated economy. While the structuralists perceive the informal entrepreneurs as insecure, low paid and reliant self-employed, they appreciate that there is a realistic and dynamic interdependence of the formal and informal sector. The informal workers are viewed as "survival experts" and "survivalist entrepreneurs" (Gallin, 2001, p. 531-532; Valenzuela, 2001, p. 349). The accumulation of economic wealth in the capitalist organisation of production leads to a level of informality through the demand for labour. Consequentially, that creates survival and choices' necessities for certain workers forced to become part of the informal economy in order to sustain a living or to pursue their values.

2.3.3 Neo-liberal perspective.

The neo-liberal perspective interprets the growth of informality in the light of burdensome and expensive bureaucratic framework of existence for entrepreneurs and their logical wish to exit the over-regulated economy. In reasoning that, De Soto

(1989, p. 131-177) argued the concepts for costs of formality and informality. Through an experiment of actual business settlement, fulfilling all legal requirements, the author demonstrated that an individual's access to a small formal industry is prohibitively high. Further evidence is presented to support that informal entrepreneurs are choosing informality to reduce oppressive administrative regulations and the associated time and effort (Small Business Council, 2004, p. 35). Perry and Maloney (2007, p. 2) based on Hirschman, 1970, developed a cost-benefit analysis with bi-directional perspective to explain attitudes towards formalisation/informalisation. They come to the conclusion that firms and individuals undertake careful cost-benefit analysis to decide which mode is worthwhile and to what extent. The neo-liberal prism, which emphasises the economic rationale, is highly relevant to transition economies (like the Bulgarian one).

2.3.4 Post-structuralist perspective.

Until recently, the scholars interested in the growth of informal entrepreneurship were presented with the choice between the first three theories (Williams, Nadin & Rodgers, 2012b, p. 7). However, a better understanding of the entrepreneur's existence in the increasingly complex social environment has led to the development of the post-structuralist perspective. Some of the properties of this explanation are that informal entrepreneurship is exercised for: closer relationships, more for social reasons than for financial profit (Round, Williams & Rodgers, 2008; Williams, 2004), as a response and resistance to exploitation (Kudva, 2009) or new individual and identity motivations (Snyder, 2004). Some later contributions (Evans, Syrett & Williams, 2006; Williams & Nadin, 2010a) pursued a more advanced and integrative approach. It has been identified (Williams et al., 2010a, p. 361) that "*different theorisations of informal entrepreneurship apply more in some populations than others*". In their future development the post-structuralist theorisations will need to account not only for the more refined economic and social stratification, but for the micro motives of the informal entrepreneurs in that regard.

Some authors began to recognise that integration of entrepreneurship theorisations will reflect the reality better and that structuralist explanation might apply to relatively deprived entrepreneurs while enterprise leaning theorisations would describe more affluent groups (Evans et al., 2006; Williams et al., 2010a, p. 350-355). Williams et al. (2012b, p. 8) acknowledged that there is no explicit evaluation of the validity of these competing theorisations. The authors filled that gap proposing a typology of informal entrepreneurship that joins together the contrasting theorisations. Most importantly however one entrepreneurship's theorisation would not remain static – the entrepreneur's rationale changes over time (p. 16). It could be argued that the Bulgarian informal entrepreneurs in early stages possess survivalist characteristics from a structuralist perspective. As they become more established, the cost-benefit analysis (from neo-liberal perspective) dominates their level of informality.

2.4 **Institutional asymmetry as a contemporary explanatory theory.**

Despite the existence of extensive theoretical literature, until recently the institutional theory on tax evasion have neglected studying empirically the interdependence of formal and informal institutions (Gërxhani, 2004b, p. 741) in that regard. See (Williams et al., 2015a; Williams et al., 2015d) for some later notable exceptions. The economic and particularly economic psychology literature has recognised in the recent past the lack of sufficient research for the effects of socio-economic environments on tax evasive decisions (Cowell, 1992, p. 522): "In particular, it neglects the possibility that

people's participation in evasion may be driven by their perceptions of injustices or inequities in the tax system, or inequity in the system of economic rewards on which the tax system has been based." Some of the constructs of contemporary institutional asymmetry explanations of informal economies rely on concepts described by the theory of optimal tax evasion behaviour (Benjamini & Maital, Gaertner W and Wenig A, 1985). Where the two central aspects to individuals' decision to evade taxes are the subjective probability bias and perception of other people's behaviour.

Criticising the capability of optimal taxation theories to explain the decision to evade due to the lack of empirical support, Gërxhani (2004a, p. 288) draws attention to the explanatory importance of public choice theory where "all agents involved are assumed to be rational utility maximizers". The supporters of this school of thought come to the conclusion that individual evasion is higher the bigger the difference is between the expected public goods' provision and the actual one. Investigating tax evasion in a transition from socialism to capitalism Vihanto (2000) asserts the view of inherently disturbed social contract, which is consequential to the fundamental conflict "with the common opinion and interests of the people leaving extensive distrust of the government". Therefore, the status of institutional asymmetry is argued to be guaranteed by default in transition economies, which logically leads to significant evasion. This study builds upon previous findings in the area of how institutional environment influences the development of entrepreneurship in its formal and informal forms (Williams et al., 2014d) and advances a new methodological approach for measuring, predicting and tackling tax evasion.

Moving onto the definition of formal and informal institutions, Helmke et al. (2004, p. 727) description is selected as the most applicable to this study:
"informal institutions as *socially shared rules, usually unwritten, that are created, communicated and enforced outside of officially sanctioned channels.* By contrast, *formal* institutions are rules and procedures that are created, communicated, and enforced through channels widely accepted as official. This includes state institutions..." Therefore, institutional asymmetry meaning is to be understood as "the misalignment between formal and informal institutions" (Williams et al., 2014d, p. 2).

In consideration of the intrinsic circularity (Casson, Della Giusta & Kambhampati, 2010) between institutions' quality and economic development, it is obvious that in many instances the informal rules supersede the formal ones in terms of coping strategies. The less advanced an economy is the more likely informal norms would be superior to behavioural compliance with the formal institutional framework. Taking the case of Bulgaria as an example, since the move to market economy after the communist regime fell, a great effort was invested in advancing the formal institutions, but as argued by Williams et al. (2014d, p. 2) the reform of informal institutions has lagged behind. This process deepened the institutional asymmetry while major socio-economic indicators were deteriorating, thus making economic survival a challenge both in entrepreneurial and employment perspective. As a result, the choice of whether to comply with formal rules became non-existent for the majority of economic actors as their informal practices presented the only viable alternative. Adopting the institutional asymmetry explanatory theory, this study hypothesises that entrepreneurs in transition setting adopt fairly rational economic decision approach in their informal strategies. IAP's advantageous holistic appraisal capacity, in comparison with previous theories, poses specific empirical challenges from research perspective. Indeed, that is to discover and measure empirically intangible phenomena, which relate to social actors and enterprises involved in illicit economic activities. This thesis contributes to the developing literature debates in that regard, through operationalisation and investigation of the main concepts underpinning what appear to be a distinctively visible institutional asymmetry in a transition setting.

Having reviewed the key theorisations of informality, the next sections will critically appraise the socio-economic significance of informality and tackling policies. The direct and indirect measurement methods for estimating shadow economies' sizes will be discussed, evaluating their strength and weaknesses. In order to provide clarity about the business environment into which informal entrepreneurs operate, the most important socio-economic indicators, characterising the Bulgarian research context will be then discussed. This is necessary to put into perspective their choice of particular informal strategies and suggest appropriate tackling policy response.

2.5 **Socio-economic significance of informality and tackling policies.**

It is widely accepted that informal economy is not only a widespread phenomenon and growing relative to the formal sector in emerging markets, but also in Western societies (Thai & Turkina, 2013, p. 14). Almost two thirds (1.8 billion) from all 3 billion workers in the world are informal (Jütting & Laiglesia, 2009, p.17-26), while informal commercial activities account for an estimated 30% of economic activity around the world (Thai et al., 2013). From a more regional perspective, the cumulative informal economy share amongst EU countries in recent years is estimated to be approximately 20% from their GDP by MIMIC method (see Figure 3). This section offers a critical appraisal of the policy approaches used to deal with informality.

2.5.1 Wider economic and societal significance of informality.

In the light of the reported facts, the economic impact of shadow economies across the world is therefore clearly significant in terms of loss of tax and associated public goods and services. There is convincing empirical evidence that higher informality is recognised to increase income inequality, lower average GDP growth and reduce international trade (ILO et al., 2009, p. 88). Larger informal economies tend to have self-perpetuating nature due to their degenerative effects on social cohesion, procedural fairness and distributive justice. One of the problematic circularity effects is that the individual cost of being honest in a corrupt society becomes significantly higher in comparison with societies where most individuals are honest (Andvig et al., 1990). As it will be revealed in the case of Bulgaria, the relatively low OISC, horizontal and vertical trust between economic actors and the state, fuels even greater tax evasion and the so called informal chains of interdependence.

Evidently, the problems of informality are reaching out to the fundamental constructs of the public contract within societies and hence the necessity to appraise them accordingly in a holistic way. Unsurprisingly, economic psychology literature increasingly recognises the significance of social variables such as procedural fairness and distributive justice on tax compliance behaviour (Van Dijke & Verboon, 2010). Logically, the vast majority of tax compliance studies evaluating procedural fairness issues includes aspects of distributive justice (Hartner, Rechberger, Kirchler & Schabmann, 2008, p. 138), because the two concepts are closely related. While in earlier studies, procedural fairness has been appraised from authoritative and citizens' decision making perspective (Tyler, 1988), later studies included the group value model of procedural justice and the relational model of authority (Hartner et al., 2008, p. 139). In the context of this study, procedural fairness refers to the extent to which people believe that they are paying their fair share compared with others and receive fair treatment by institutions (Williams, 2014c, p. 26). Distributive justice refers to the feeling that everyone pays their fair share of taxes (not too much or too little), simultaneously exercising equal access to publicly funded resources (Verboon & van Dijke, 2007). Therefore, it is important to investigate whether institutional practices

disturb the social feeling for procedural fairness and distributive justice, henceforth increasing the institutional asymmetry, and informality accordingly.

Understandably, governments across the world have strong incentives to tackle their informal economies in an attempt to foster economic development and international competitiveness. It is however, important to recognise that informal economies are multi-faceted socio-economic structures, with a level of systemic dynamism and cultural embeddedness. There is an increasing body of scholars appreciating the intense relationship between formal and informal activities and the fact that "informal economic activity in general remains an important and growing element of contemporary society within advanced industrial economies" (Evans et al., 2006, p. 87).

2.5.2 Policy approaches to informality.

One of the biggest obstacles in tackling the informal economy has often been the fact that it was being "conceptually framed as a "tame" problem (i.e. a problem that is complicated but easily solvable, often with a discrete response that can be replicated anywhere)" (Williams, 2014c, p. 3). Governments tend to underestimate how sophisticated this socio-economic phenomenon is and to appraise the full scale of micro and macro participatory motivations. This is so, due to the lack of regular deployment of appropriate empirical inquiries seeking to establish what a country's specific root causes of informality are. Although, contemporary academic literature accumulated sufficient evidence in recent years to support informed and effective policy response in most countries, it is usually unpopular or difficult to utilise by government officials at a policy making level. This study aims to contribute in this respect via a novel way of analytical tools' deployment in an accessible and easier to understand way.

It has been already argued that the best way in defining the informal economy and entrepreneurship within it, is to employ a modular principle, depending on a widely accepted definition, which is then fine-tuned to reflect a particular country specific context. This is important for two reasons. Firstly, to enable an accurate and exhaustive definition of the various informal economies, which tend to differ substantially across national borders by their nature of existence. Secondly, to enable appropriate design of tackling policy measures. Development of adequate policy approaches therefore is subservient on precise definition and comprehensive understanding of the root causes of informality at a national level. One of the aims of this thesis is exactly that – to suggest effective tackling policy solutions against typical informal economy from a transition setting (namely Bulgaria). In order to cultivate the theoretical argument, a review of the various options available to policy makers should be firstly carried out.

Williams (2014a; 2014c) created a typology framework of policy measures arguing that a paradigmatic shift is required to achieve the desired effects in tackling the root causes of shadow economies. In accordance with the conceptual framework, policy approaches could be direct or indirect ones. The so called direct controls are of two types: deterrents ("sticks") and incentives ("carrots"). For instance, such policy measures are data matching and sharing, increasing sanctions, advertising penalties, simplify compliance, direct and indirect tax incentives, supply chain responsibility, support and advice to start-ups, supply-side incentives e.g., society-wide amnesties, voluntary disclosure, smoothing transition to formalization. The indirect controls are directed towards change of formal institutions (laws, regulations and codes) and informal ones (values, norms and beliefs) with the aim to reduce the institutional

incongruence. This way, decreasing the share of the informal economy. Typical policy initiatives from this type are aimed at tax education, normative appeals, awareness raising of benefits of declared work, procedural fairness and justice, redistributive justice, wider economic and social developments. From the perspective of the policy maker there are four hypothetical policy choices (Williams, 2014a, p. 65) – doing nothing, deregulating the declared economy, eradicating the shadow economy and moving the shadow economy into the declared economy.

Doing nothing

Policy makers could potentially choose to do nothing against the informal economy. One of the possible rationales for adopting this approach is that a shadow economy is simply relatively small and cost-benefit analysis indicates unfavourable return to expensive institutional and policy interventions. However, there is no clear evidence what share of informal economy is relatively small in order to offset the financial rationale of seeking to increase otherwise lost tax revenue as a result of tackling policy efforts. To a certain extent informality could assist entrepreneurship in its early start-up phase (Andrews et al., 2011) and thus nurture enterprise development. However, at the same time leaving the informal sector to flourish creates unfair competition to legitimate businesses while making employment conditions worse and leaving consumers' rights exposed unprotected. Adopting the "do nothing" approach is therefore socially and economically unacceptable. Such policy attitude could be justifiably engaged towards micro enterprises and the small scale paid favours' sector which is "arguably a source of active citizenship in contemporary monetized market societies" (Williams, 2014a, p. 66).

De-regulating the formal economy

De-regulating the formal economy as a policy approach option rests on the neo-liberal theorisations of informality. Indeed that the growth of informality is due to burdensome and expensive bureaucratic framework, which increases the cost of formal economic activities. As argued by De Soto (1989, p. 131-177) individual entrepreneurial access to a small formal industry could be prohibitively high and therefore push economic actors to informality. There is evidence that informal entrepreneurs are choosing informality to reduce oppressive administrative regulations and the associated time and effort (Small Business Council, 2004, p. 35) and that easing the processes will enable formalisation. Another aspect of informalisation processes due to perceived high cost of formal economic activities is the pressure towards the use of the so called envelope wages. This is quite popular practice in transition economies where some employers offset the high costs of doing business according to the rules via underreporting the real remuneration in their firms and therefore materialise savings. An employee may receive an official salary and an additional sum ("paid in an envelope"). This leaves room for the employer to pay only partial social contributions on the basis of the reported official salary.

De-regulating the formal economy could have positive economic impact in terms of reducing the cost of doing business in multiple aspects and as such might reduce informalisation while encouraging formalisation. However, it has to be accounted that the significance of regulations for informalisation decisions is different in different economies. For instance the importance of taxes, regulations and financial constraints to informalisation decisions in developing countries is reduced in comparison to the crucially important quality of legal systems (Dabla-Norris, Gradstein & Inchauste, 2008). Therefore, the wider socio-economic gains from deregulating the formal economy could be questionable when taking into consideration the negative

effects on employment relations in terms of working conditions and fair remuneration. If reducing the tax rates is considered as one possible deregulatory measure, there is evidence suggesting no significant association between tax rates and the extent of unofficial activity in Eastern Europe (Johnson, Kaufmann, McMillan & Woodruff, 2000). It could be argued that deregulating as a policy option would be more beneficial in tackling shadow activities in economies with higher costs of doing business. Such policy choice should be always carefully appraised in terms of its impact on levelling down the working conditions and utilised with complimentary measures to reduce such negative side effects.

Eradicating the shadow economy

This option looks attractive from a government perspective particularly when taking into consideration potential benefits in tackling the well-known negative effects of informality. However, one practical problem exists (Williams, 2014a, pp.69-70) relating to the fact that beyond a certain point the cost of eradication may far outweigh the benefits of doing so. In addition, as already discussed, it must be considered that small share of informal economy acts as an entrepreneurial incubator environment and thus has positive economic effects. What share of informal economy would be economically useful to a given society is currently unclear. Arguably, that would be correlated to the development level – the more advanced an economy is the smaller share of informality would be beneficial. Pursuing an eradication strategy from formal institutional perspective is traditionally perceived via the utilisation of deterrence policy measures. However, there is convincing literature evidence (Hofmann, Hoelzl & Kirchler, 2008; Van Dijke et al., 2010) suggesting that direct tax compliance enforcement through increased penalties and audits is less effective than cultivating dynamics of mutual trust between the state and the taxpayers.

To evaluate what is the predominant mix of policy measures being used in Europe, a review of EUROFOUND (knowledge bank of policy measures against informal practices) has been conducted. The general policy imperative becomes clearly visible – indeed that EU governments use primarily direct control policies (deterrents and incentives) and rarely rely on indirect controls (changing formal and informal institutions). It is obvious, that the preferred tackling approach against non-compliant economic behaviour is still firmly deterrence and punitively grounded. At the same time, several EU economies are still facing larger shares of informality than economically viable (see Figure 2 and Figure 4). EUROFOUND was created by Professor Colin Williams (with Regioplan) and consists of 186 policy initiatives from 33 countries including all EU member states. This is the only source of 'good practice' policy measures available to governments throughout the world and as such is also a very good source for analysis of policy makers' attitudes towards tackling informality. Appendix 37 presents detailed classification of all policies by type and target with a summary of conclusions.

Given the discussion of facts above, an eradication policy approach should not be employed alone as the ultimate solution to informal economy phenomena. On the other hand, it could be successfully utilised as part of a more holistic answer – such as one, which combines direct and indirect policy measures in a responsive way.

Moving the shadow economy into the declared economy

While moving the shadow economy into the declared economy is theoretically possible, it should be noted that practically (after a certain point) it could be only

partially achieved via securing a great level of voluntary compliance. There are several justifications to invest institutional efforts into such direction and their dualistic nature could be reviewed from the perspective of the benefit or detriment caused.

From a business perspective, an immediate benefit to fully compliant enterprises would be the reduction of unfair competition. At the same time, ex semi of fully informal firms would benefit from their formalisation in terms of not having to pay the real costs of informality. Such as lack of access to bank credit, proper marketing channels, commercial enforcement of contracts. Obviously, previously evaded taxes would be payable in order to access the privileges of fully compliant business. Quite often smaller, fully or partly informal firms absorb higher costs of informality without knowing it – in fact greater than the costs of full compliance. Formalisation would benefit those formerly providing they labour in the undeclared economy. Employees who have been receiving supplementary "envelope wages" could see them converted to fully paid social security contributions in addition to an increased industrial protection and working rights. At the same time, full conversion of the shadow economy into the declared economy will eliminate the entrepreneurial testing environment for start-ups and could lead to a reduced entrepreneurial activity.

Another perspective is this of the customers of products and services of informal economies. Inevitably, they would be deprived from the ability to consume lower priced goods and services as the full cost of legitimate production must be paid. However, consumer rights would be significantly improved in terms of quality, guarantees, standards of service and adherence to commercial contracts with potential for appropriate enforcement.

Governments' rationales for converting the informal economy into the declared one are multifarious. Most obviously, it would improve revenue (Evans et al., 2006) and thus enable social integration and mobility to be pursued (Williams & Windebank, 1998). There would be an increase in provision of jobs, which are easier to control in terms of work conditions. However, governments would need to invest more into entrepreneurship start-up schemes to compensate for the lack of a test bed, which would be otherwise available should informality is represented in a greater share of the economy.

Towards a mixed approach

As argued by Williams (2014c) the current debate is not so much over whether to use direct or indirect controls to tackle informal economies, but rather which specific policy measures are most effective and in what combination. Shifting shadow economy into the declared economy is indeed the most viable approach and it should be achieved via overarching strategy combining "responsive regulation approach" with the "slippery slope framework" (Williams, 2014a). The policy focus should be firstly on the use of indirect policy controls to induce compliance in a voluntarily and engaging manner. Only when this does not work, resorting to incentives, followed lastly by deterrents should be employed.
Synergies of potentially effective policy mixes would need to be customised in regards with any target country context. A key prerequisite to achieve success is the ability to investigate and diagnose the various motivations of the socio-economic actors in any given society. This thesis advances a qualitatively new methodology to acquire and appraise specifically entrepreneurial rationales for operating in the informal economy. The suggested analytical techniques offer the so called reverse predictive probability to diagnose informality with the associated motivations and could be utilised by formal institutions, policy makers and scholars interested in the field.

2.6 **Methods estimating the size of the shadow economy.**

Different economies have diverse structures of their shadow economies and require appropriate measurement methods accordingly. Estimating the real size of a shadow economy has implications about what policy responses should be utilised. Moreover, measurements could provide answer to the question about cost of informality, but this time from the perspective of the society, not the entrepreneur. Large informal economies pose comparatively much more complex challenges to tackle beyond their existence and impact on the relevant economic domains – such as deep institutional asymmetries and low levels of social cohesion. The self-perpetuating nature of such big informal economies also changes the motivations at a wider societal level for participation including the processes of informalisation and formalisation. For that reason, studying the various motivations of partly informal entrepreneurship should be undertaken with close attention to the magnitude of a given informal economy size. While this study does not aim to measure the size of the Bulgarian informal economy directly, its quantitative empirical instrument has an embedded capability to do so amongst the small and medium sized businesses. It demonstrates a significant difference between what contemporary macro-measurement approaches report and the latest findings derived here. Therefore, it is important to outline the main macro-economic estimation methods and later on appraise the value of the suggested micro-economic method.

There are inherent difficulties, which emerge from the fact that it is problematic to measure what is being hidden by the participants. Schneider (2005) made a significant contribution to the discussion of various measurement methodologies. He generalised that there are three main groups of methods: direct, indirect approaches and the model approach. The purpose of this section of the thesis is to provide details about the different estimation methods, their advantages, disadvantages and relevant critique. It is important to note that the different measurement methods and in particular their results should be always examined in the light of the implicit definition they employ.

2.6.1 Direct approaches.

In general, the direct approaches are micro-economic approaches, which "*employ either well designed surveys and samples based on voluntary replies or tax auditing and other compliance methods*" (Schneider, 2005, p. 619). Surveys with different level of representativeness are used in various countries to estimate the size and involvement in the shadow economies. Isachsen and Strom (1985) have extensively used the survey method and found that is extremely difficult to arrive at a precise and reliable estimate of the shadow activities not only in Norway, although they had 70% participation rate. Direct survey methods should always employ a particular questionnaire strategy, which accounts for the regional and cultural characteristics of the respondents. For instance, the concealment level in Eastern Europe is quite higher and thus the questions should be more indirect.

Another way of producing estimates of the hidden economy is via comparison of the declared income for tax purposes and that captured through selective tax audits. The discrepancy will be indicative for the size of the hidden economy. However, there are specific weaknesses, which need to be accounted for. As noted by (Schneider, 2005, p. 620) this method could employ possibly biased sample from the population.

A mutual weakness of the described direct approaches is that they lead to estimation of only part of the shadow activities, which then should be extrapolated to the entire

population. But their narrowness also provides capability for detailed stratification of the different hidden activities. In summary, all micro measurement methods are methodologically pre-disposed to exhibit substantial errors, but if designed and deployed correctly could also achieve good veracity.

2.6.2 Indirect approaches

The indirect approaches are based on macro-economic and other indicators, which reflect the development of the hidden economy. Currently there are 5 widely recognised approaches which are also called indicators (Schneider, 2005; Schneider & Enste, 2000). Over the next paragraphs these macro-economic approaches will be discussed with their advantages and disadvantages.

2.6.2.1 The discrepancy between national expenditure and income statistics.

In principle, there should be parity between the national income (GNP) and the national expenditure part of GNP, but this is not so in practice. This gap could be used as a measure of the informal sector. However there is tendency that the national accounts will be statistically adjusted as suggested by Feige (1989, p. 30), so the discrepancy would appear smaller than it is in reality. The national accounts' managers will also inevitably embed a standard error in the calculations due to their general complexity and inability to account for everything in a balanced way. This would cause an additional deviation. Therefore, it is suggested that the first estimate is taken as the most accurate representation of the informal economy size (Schneider, 2005, p. 620). It could be argued that more developed economies have systems for more precise calculations of the both sides of the GNP equation – the expenditure and the income. That provides advantages for the application of this approach in developed compared to developing economies.

2.6.2.2 The discrepancy between the officially registered and actual labour force.

This is quite self-explanatory macro approach in terms of measuring the spread between officially registered work force in the formal economy and its potential outflow. However, any economy could have decline in the formal work opportunities due to other economic dynamics, but not only limited to an increasing informal sector. For instance, the officially registered workers are those reported by companies to be part time or full time workers. The actual labour force is consisted of all social actors free of health related work impairments and those who are temporarily disabled to work with legitimate health reasons. Both numbers (officially registered workers and actual labour force) could be usually obtained through the offices for national statistics or commercially licensed companies, which sell such information. In cases where an outflow from the officially registered work force is observed, without relevant demographic reasons, it could be argued that this is in favour of informal workers' increase.

Using this method, Contini (1981) proposed a definitional equation, which employs some simplifying assumptions about the Italian labour market in order to compensate for the inherited methodological weaknesses. A good example of the method in American context is presented by O'Neill (1983) who attempted to measure the unregistered underground employment and concluded that the US underground unemployment is actually less than what other researchers implied. Nastav and Bojnec (2007) discovered that the shadow economy that relates only to unemployment discrepancies in Slovenia amounts on average to around 6 % of the

official economy. Simultaneously, a more detailed activity-level unemployment analysis gives much higher values of up to 20%. Therefore, the authors conclude that this approach should not be used for policy development. Another disadvantage of this method (Schneider, 2005, p. 621) is the ability of workers to have a formal job and be involved in informal activities as a second complementary part time job. Therefore, the labour discrepancy approach should be considered with particular care in the context of its flaws. In countries similar to Bulgaria, with particularly well developed envelope wages culture and only partial registration of employment contracts, the applicability of this method would be very limited.

This is so, because we can actually talk about the share of the formal economy and not the other way round (see the findings and discussion of findings). This rare phenomenon amongst EU member states in combination with the methodological weaknesses of the method renders it as particularly inaccurate approach for the Bulgarian context.

2.6.2.3 The transactions approach.

This method was best developed by Feige (1996). Intriguingly Schneider (2005, p. 621) argued that the approach is based on the well-known Fisherian equation: M*V=p*T (with M money, V velocity, p prices, and T total transactions). However, upon closer examination of Feige's work, I have failed to find any use of that equation. Feige has built his hypothesis from empirical evidence looking at the US underground economy. In the centre of the approach is a puzzling macroeconomic anomaly called "*currency enigma*", which consists of a stock and flow component. In general, there is missing currency, which in the case of US is 300 billion dollars. Feige's core argument to that approach is that bigger proportion of that missing (not in banks) currency is held abroad for transaction purposes, but the other part is kept domestically and is being used to pay for underground economic activities. Estimating the domestic currency component will result in estimation of the shadow economy size. There are two sub-methods suggested to calculate the currency proportions held domestically and abroad: the monetary demographic model and note ratio model. Without going into greater detail suffice it to say that the transaction approach needs to employ significant number of assumptions in order deploy its reverse logic and arrive at an estimate. That results in extremely high empirical requirement should we want to reduce the number of assumptions or make them less significant. The use of the approach in practice should account for those characteristics.

Thomas (1999, p. 382) provided detailed criticism based on the fact that the approach rests on three "heroic assumptions": "*First, a year (or quarter, depending on the available data!) must be identified in which the black economy did not exist. Secondly, transactions in the black economy are carried out exclusively using cash. Finally, the velocity of circulation of cash is the same in both the nonblack and black economies*." Rejecting these assumptions Thomas quoted Windebank and Williams (1997) who provided evidence against the assumption that only cash is used in the black economy.

2.6.2.4 The currency demand approach.

The currency demand approach was firstly used by Cagan (1958) to evaluate the shadow economy. His concept emerges from the well-known public demand for currency (cash) as a relative function to the total money supply. This results in a currency ratio into whose behaviour and determinants Cagan was interested. A key

element of the concept (from informal perspective) is the variable rate of tax and its effects on transactions and in particular the proportional correlation with cash/currency demand – the higher the income tax is the higher demand for currency would be, which consequentially results in lower currency ratio. The financial institutions like banks will be left with smaller supply of currency in relation to the total supply of money. Providing that the tax rate is only one variable affecting the currency ratio, it becomes questionable what part of the increased demand for cash is caused by unregistered businesses activities. As Cagan reflected, such an estimate could be subject to a considerable error.
Nineteen years later Gutmann (1977) used the same approach, but without statistical verification. His logic is based on the composition of the stock of money. As one economy is more developed, more transactions are executed intangibly without the exchange of real paper money, which means that the "demand deposits" outgrows the necessity for currency. Again, the discrepancy could illustrate the underground economy size. However, this leads to questions about applicability of this measurement method to traditionally more cash orientated economies. For instance, the Bulgarian SMEs are known to use predominantly cash payments where velocity of exchange is important for the limited cash flow potential of those micro and small enterprises. It could be argued that the bigger the cash aspect of an economy is the more influenced the measurement results would be.
Tanzi (1983) made significant improvements of Cogan's approach further into a multi-variable equation, but it does not seem that he has fully acknowledged Cogan's contribution. Tanzi claimed that he has provided estimates for the US during 1930-80 using a method developed by himself (p. 283): "*Briefly, as the level of taxation rises, individuals are encouraged to engage in tax-evading activities that are facilitated by the use of currency, as this practice leaves no traces; therefore, the use of currency rises*" (p. 290). Intriguingly, there is a fierce critique of Tanzi's approach by (Feige, 1986), which could be traced to Tanzi's reply to Feige, Thomas and Zilberfarb (Tanzi, 1986). In principle, Feige states that Tanzi failed to "*state clearly precisely what he sets out to measure, and finds fault with several of the assumptions that underlie his procedure*" (p. 768).

One of the main assumptions of the currency demand approach typical for all authors is that the unrecorded transactions are executed exclusively with cash. As we have already seen above, Windebank and Williams (1997) presented evidence against that. However that seems to be one of the most commonly used approaches, which has been applied to many OECD countries (Schneider, 2005, p. 622).

2.6.2.5 The electricity consumption method.

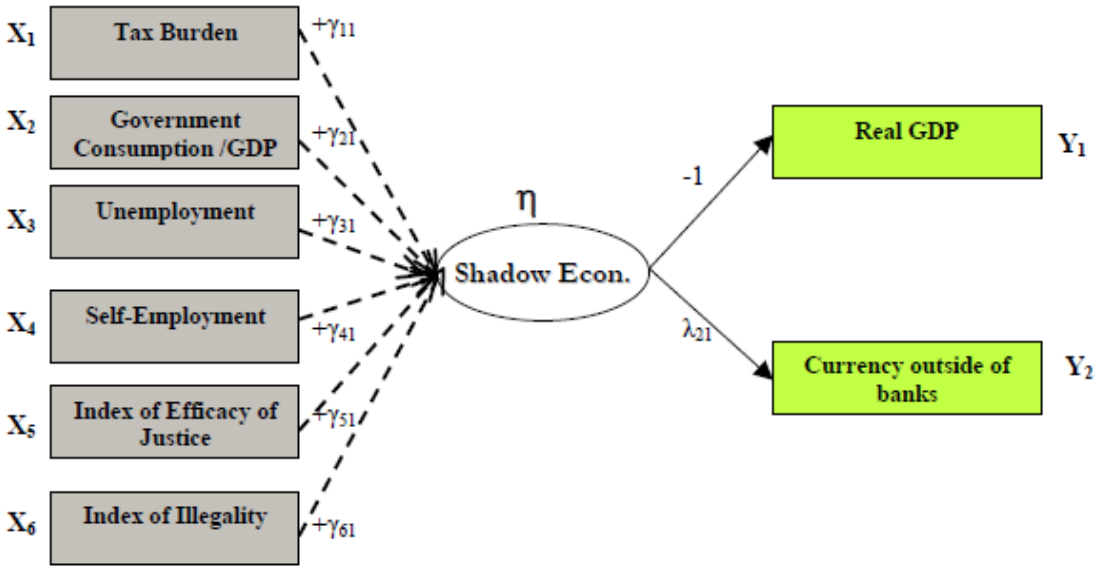
Kaufmann and Kaliberda (1996) developed and tested the macro-electric approach in post socialist economies and in particular Ukraine. This method rests on the assumption that official GDP is good measure of the formal economy output since informal economy has remained unaccounted for by official statistics. Empirical evidence from across the world is available that electricity consumption is in straight correlation with the increase of GDP with elasticity close to one. Electrical consumption growth represents economy growth (formal and informal). However, some informal activities require less electrical power and thus even their significant increase will have only marginal effect. In summary, this is an attractive method providing it could be estimated what bias the unit elasticity would impose on the informal economy measurement.

2.6.3 The model approach and discussion.

As we have seen all measurement methods concentrate on a single indicator or correlation to capture the scale of the hidden economy. But the shadow economies affect labour recourses, production potential and money/cash flows simultaneously. The amount of taxation as a problem is taken as a predominant cause with bigger weight than other typical causes. This creates the necessity for completely different measurement viewpoint, which in turn could generate different and more adequate approaches.

To address some of those issues Frey, Weck and Pommerehne (1982) developed the so called "soft modelling" approach based on Kofler and Menges' (1976) mathematical theory (Limited Probability Information). The strength of the approach is in the ability to give appropriate weights (rankings) to the different contributing factors and hence adjust the model to a particular economy. With no earlier traces from the literature, it appears that in one of their later papers, Frey and Weck-Hanneman (1984) have been the first to analyse the size of hidden economies as an "unobservable variable". They have proposed radically different approach emerging from the "*statistical theory of unobserved variables which considers multiple causes and multiple indicators of the phenomenon to be measured*" (p. 35). Following Frey and Weck-Hanneman's seminal work, Dell'Anno (2003) applied the Multiple Indicators and Multiple Causes (MIMIC) on the Italian economy. The model is able to estimate the size of the shadow economy (as a latent variable) based on a number of observable indicators from one side and a set of observed causal variables on the other.

Graph[26] 1: MIMIC 6-1-2



The figure is reproduced exactly as it is from Dell'Anno (2003, p. 8).

To summarise, it is argued that the micro-economic approaches are more locally sensitive in terms of questionnaire design and sampling. Direct questionnaire design should be used with care since it might greatly affect the quality and quantity of captured information. It should be accounted for cultural characteristic of the general population where the micro approaches are deployed. It becomes evident that using a carefully designed business survey (i.e. micro-economic approach) the potential to derive a more realistic evaluation of the shadow economy is greatly enhanced. This

thesis presents evidence (see discussion of findings) that the actual share of the Bulgarian informal economy is significantly bigger than what current macro-economic measurements report. The macro-economic approaches tend to rest on multiple assumptions, which are either very broad or very difficult to be verified empirically. But they provide better basis for estimation of the social cost of informality.

Chapter Three

3. The Bulgarian informal economy.

This chapter is opening the discussion with some highly disturbing figures and comparisons from Bulgarian perspective. It aims to provide the reader with the major socio-economic indicators characterising the research context. Highlighting the following important aspects of the economic environment is necessary to establish grounds for the discussion of findings. In particular, how entrepreneurs chose their informal strategies given the social and business setting in which they operate and what the policy response should be. As we can see from Figure 2 and 3 (on the next page) the following conclusion could be made:

- In comparison with 27 European countries, Bulgaria has the highest share of informal economy as a percentage from its GDP. The estimation is done by the MIMIC method and shows that the Bulgarian shadow economy is 31.20% from GDP in 2013 (Figure 2).

- The Bulgarian shadow economy is well above the 27 European countries' averages for all years in 2003-2013 term (Figure 3). It is 1.6 times higher in 2002 and 1.7 higher in 2013.

- For a 10 years period (2003-2013) the share of the Bulgarian hidden economy dropped only with 4.7 points (35.90%-31.2%, Figure 2).

- A decrease of the informal economy is observed from 2003 to 2013 for all countries, but for most of them, it is negligible (Figure 2).

- As Schneider (2012, p. 2) report "*we have an increase of the size of the shadow economy from west to east*" (Figure 2).

3.1 **Socio-economic facts, figures and trends.**

Figure 2: Size of the shadow economies of 27 European Countries over 2002 – 2013 in % of GDP by MIMIC.

Note: This figure is produced by the author based on data from (Schneider, 2013)

Figure 3: The Bulgarian informal economy in comparison with 27 EU countries' averages – % of GDP by MIMIC method.

Note: This figure is produced by the author based on data from Schneider (2012); (Schneider, 2013).

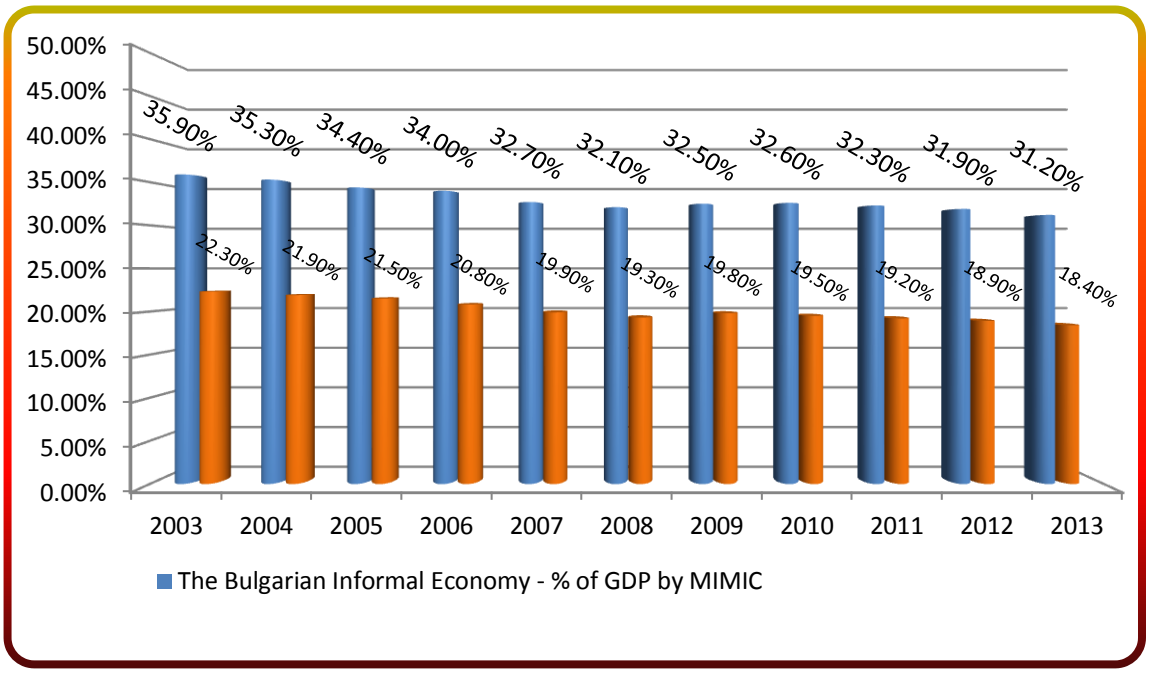
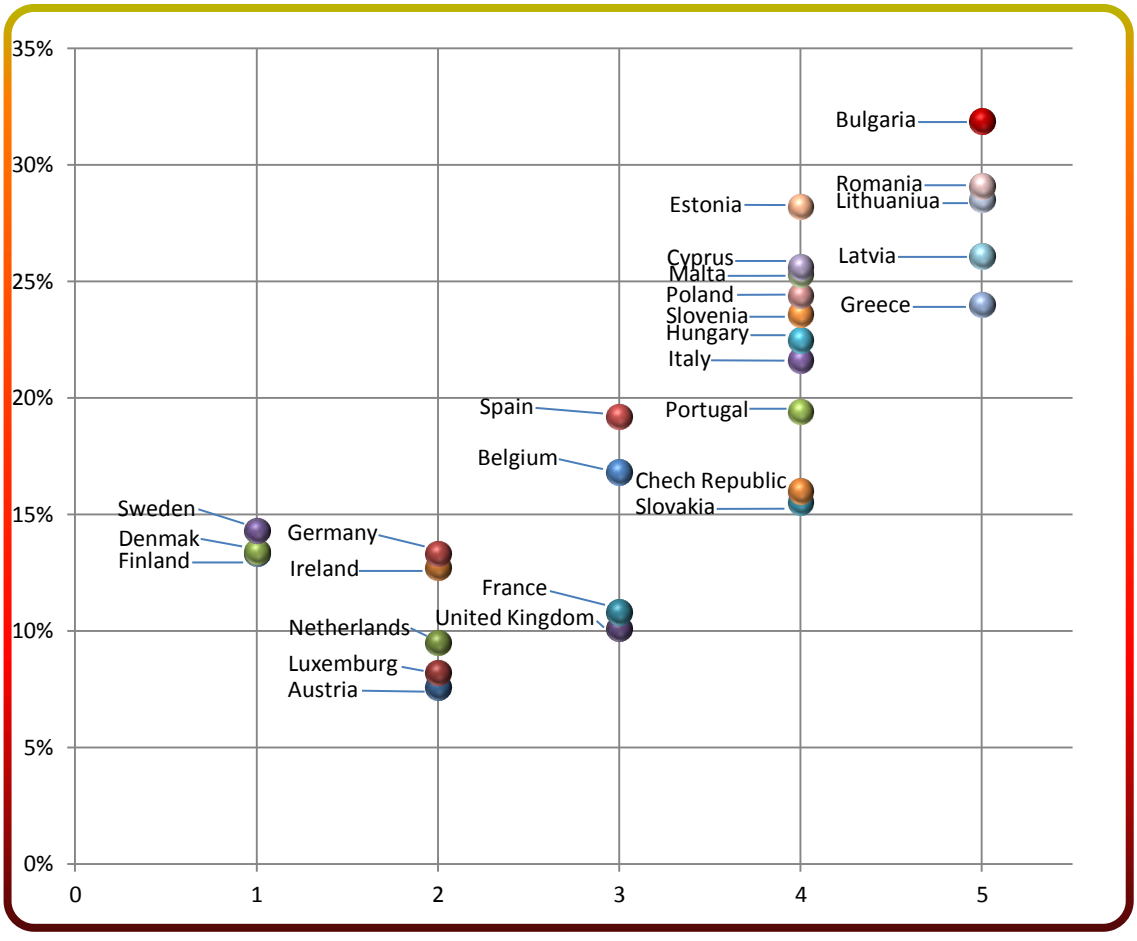


Figure 4: Simple correlation between shadow economy size and overall index of social cohesion.

Note: This figure is produced by the author based on data from Schneider (2013) for the shadow economies' sizes and data from Dragolov, Ignácz, Lorenz, Delhey and Boehnke (2013) for the overall index of social cohesion.



X-axis: overall social cohesion index, where 1 represents the highest value and 5 the lowest social cohesion.

Some very intriguing results are presented in Figure 4. It embodies a simple correlation between shadow economy size and overall index of social cohesion (OISC). The correlation is developed based on data from Schneider (2013) and a very topical study by foundation Bertelsmann Stiftung (Dragolov et al., 2013). The authors developed a definition of social cohesion to measure a quality of society comprised of nine dimensions: "*A focus on the common good manifests itself in the level of solidarity and helpfulness, people's willingness to abide by social rules and the extent to which they participate in society*" (p. 13). The nine dimensions have been measured by 58 indicators and are sub-entities of 3 top level spheres: Social relations, Focus on the common good and Connectedness. Dragolov et al. (2013) conducted a comprehensive international study, which lead to development of overall social cohesion indices for the EU and Western OECD countries.

For the analytical purpose of Figure 4 only 27 EU countries with their social cohesion indices are selected to correlate with the data on their shadow economies. Dragolov et al. (2013) used coloured ranking scheme in their work for each country, but for the purpose of building the functional analysis and graphical representation this needed numerical codification. The X-axis shows the values of the overall social cohesion index, where 1 represents the highest value and 5 the lowest social cohesion. The values on the X-axis are in inverted order – the higher the number the lower the value. The 0 and 5.5 values are outside of the used measurement (which is 1 to 5) and have only illustrative role. The following comments and conclusions could be made:

- In general, there is significant correlation between the social cohesion index and the size of the shadow economy. As a general trend, a higher social cohesion index correlates positively with lower share of the informal economy. However, there are countries with higher social cohesion index, which have bigger share of the shadow economy than others with lower social cohesion index (for instance Sweden and United Kingdom).

- Bulgaria exhibits the highest share of informal economy and the lowest OISC. Bulgaria, Romania and Lithuania (Figure 2) are the 3 countries with biggest informal economy share in descending order. It is obvious from Figure 4 that they have the lowest social cohesion index.

These results outline some contemporary research questions about the connection between the sizes of the shadow economies and the quality of those societies measured by their OISC. Two major areas of future research could be drafted:

- Why and how a lower OISC of a society leads to a bigger shadow economy?
- Why and how certain societies still have smaller shadow economies, although their OISC indices are relatively smaller than others? This is contrary to the observed general correlation.

Before continuing with the discussion, a short historical overview will be provided to enlighten potential causes of why Bulgaria might have the lowest social cohesion index. Around 1388 after several smaller invasions, Bulgaria was conquered by the Ottoman Empire (modern Turkey nowadays). Bulgaria entered very dark era of suppression, planned elimination of the intellectual class and misgovernment, which continued for 500 years. During that period, many Bulgarians were forced to surrender their Christianity and accept the Muslim religion or otherwise face death penalty. During that catastrophic period, the Turks were using various usurpation methods to rule Bulgaria in every possible aspect, but most importantly culturally. They have favoured and comprehensively rewarded certain Bulgarians who were prepared to work as spies helping the Turkish cultural assimilation. Most of them were given a simplified choice to accept the offer to serve or be exterminated. This process

created gradually strong community mistrust and solidarity between the members of the Bulgarian society in their effort to survive. The Russian-Turkish war brought liberation to the country in 1878. In the period between 1878 and 1945, Bulgaria has participated in the Balkans wars and in the 2 World Wars, but managed to progress significantly in economic and cultural aspect. With the communist regime (1945-1990) and the methods of the State Security agency (DS/ДС) the grounds of social mistrust were resurrected with great success. The Bulgarian people were greatly stratified in their well-being between the communists and the rest of the people. Special DS agents were spying on people's lives, their "contributions" to communism and majority of the intellectual class were killed in special secret camps similar to those in Germany during the World Wars. All these historic processes made the cultivation of social cohesion very difficult and contributed to its current very low level. It could be argued that the Bulgarian shadow economy, in addition to the well-known causes of this phenomenon, has very strong cultural foundation. In the given context, the cultural aspects of the hidden economy require specific tackling approaches to account for the lowest OISC in Europe. An overview of some specific measures and initiatives against the Bulgarian shadow economy will be provided over the next paragraphs. This review is based on the knowledge bank against the undeclared work developed by an EU body called Eurofound – the European Foundation for the Improvement of Living and Working Conditions[1].

Joint effective control on undeclared work

The Bulgarian labour law has been recently changed with the updated Labour Act and the Labour Inspection Act (from 15.02.13 and 01.01.2011 accordingly). The purpose of the legislation changes was to equip labour inspectors and government bodies with stronger sanctioning tools and authority. This affected their rights to inspect how the legislation is applied in practice and to enforce penalising procedures. Article 414 from the Labour Act defines severe financial penalties for all failures to adhere to the law, which includes the lack of employment contract. The fines (if heavier penalties do not apply) are from 1500BGN to 15 000BGN for the employer and 1000BGN to 10 000BGN for the employee. These are substantial fines in relation to the Bulgarian life standard – 1GBP = 2.2BGN. This legislative counter-measure is aimed at all undeclared work practices, but is particularly addressing a popular labour legislation violation. Employees were often working full time, while being registered on part-time contracts. This is being done for social security and tax evasive purposes benefiting predominantly the employer. The labour inspectors have also the authority to stop temporary the business of the offending company.

The new legislative changes have led to better integration and coordination of special government control agencies and trade unions. The following government bodies and labour unions are able to be more effective in their mutual effort against undeclared work:

- General Labour Inspectorate (GLI)
- National Revenue Agency (NRA)
- Confederation of Independent Trade Unions in Bulgaria (CITUB)
- Confederation of Labour Podkrepa (CLP)

[1] Eurofound asked Professor Williams to expand the knowledge bank to 31 European countries and he was awarded the European Commission contract to evaluate the feasibility of establishing a European platform for tackling undeclared work (University of Sheffield, 2012). Based on Eurofound (2013c) the initiatives and measures against undeclared work and the informal economy in general are in 6 spheres:

- Joint effective control on undeclared work
- Connecting fiscal devices to the National Revenue Agency
- Mandatory registration of individual employment contracts
- 'Come into the light' initiative
- Rules for Business Centre

As reported by (Eurofound, 2013b): "*The coordinated approach (based on GLI's cooperation agreement with the trade unions and joint inspections with the NRA) to detect and prevent undeclared work proved to be successful. As a result of the joint inspections of the NRA and GLI, over 30 000 people have passed from a part-time (four-hour) contract to a full-time labour contract, and 10 361 workers detected to be working without a contract have signed written contracts. Social security contributions made as a result of the joint inspections of the NRA and GLI are over BGN 20 million.*"

Connecting fiscal devices to the National Revenue Agency

Bulgaria has been severely affected by the economic crisis of 2008 and this resulted in high revenue losses and higher level of tax evasion. This was partly due to decline in income from existing businesses, limited start-ups and increased effort by enterprises to underreport turnover. In order to address the situation the government imposed new legislative rules with Ordinance 18/2006 from Minister of Finance (lastly updated on 12.10.12). All businesses should connect their fiscal devices with NRA. There was gradual approach in implementing the requirement. The reporting of all fiscal transactions is done in real time via a mobile connection – this practically eliminates discrepancies between real amounts of cash transactions and accounted for amounts. However, the new approach still cannot detect cases where the trader or enterprise decides to conceal a business transaction without issuing the compulsory fiscal receipt from their fiscal device. This measure proved to be more effective for bigger enterprises where there is greater difficulty to do business without the compulsory issue of fiscal receipts. As reported by (Eurofound, 2013a) "*The NRA registered growth in the taxable VAT base, with BGN 4.3 billion for the second quarter of 2012 (since all companies were connected) compared to the same period of 2011. According to the tax administration, the main reason for this increase is linking fiscal appliances with the NRA*." However, the remote fiscal connection as a policy response, applied reverse pressure on smaller businesses that are biased not to issue sales receipts to trusted clients. It could be argued that part of the increased tax base amongst bigger companies is offset by decreased tax base for bigger number of smaller companies.

Mandatory registration of individual employment contracts

With the changes to the Labour Act (article 62) and Ordinance 5/29.12.02 the Bulgarian government introduced mandatory registration of the employment contracts with NRA. The purpose of this measure was to reduce the number of workers without employment contract and those with not-corresponding contracts – the envelope wages phenomenon, see (Williams, 2010a). This practice also increased the social insurance base and tax revenues.
"*As a result of the negotiated new levels of contributions, in 2003 the revenues from the social security contributions in the state social security insurance funds – as a net effect of negotiations – increased by about BGN 405 million (€206 million) compared with 2002. In 2004, these revenues increased by BGN 305 million (€155 million) compared with 2003, and in 2005 by BGN 183 million (€93 million). It is expected that the revenues will amount to over BGN 500 million (€255 million) in 2009*" (Eurofund, 2009).
While this proves to be an effective measure, the administrative burden in registering the contracts increases the employers' expenses in terms of time and effort. It also leads to over-utilisation of the existing workforce to minimise the necessity of new employment.

'Come into the light' initiative

This initiative was established in 2007 by two of the largest representative employer organisations in Bulgaria (CITUB and CLP), supported by eight national media and government agencies. A dedicated web site was developed (www.nasvetlo.net) where special publications are made available for the benefit of the employers, employees and the state. The main aim of this initiative is to bring to the attention of the wider public the substantial grey economy problem, which starts to cause alarming damage to the national economy.

Restriction and Prevention of Informal Economy – Rules for Business Centre

This special project (14.09.2009 – 31.10.2014) is implemented by the Bulgarian Industrial Capital Association (BICA) with the financial support the Operational Program Human Resources Development, co-funded by the European Social Fund under contract BG051PO001-2.1.05 (see Rules for Business Centre). The project research and results have various dissemination channels, but it is still early to judge upon their impact. There are extensive publications over 1000 pages (see for instance the main report by the Bulgarian Industrial Capital Association (2013)) on the various aspects of the informal economy, methods and policies for tackling it.

3.2 **Some concluding remarks.**

Bulgaria has one of the lowest GDP in the EU (Dragolov et al., 2013, p. 41), which correlates to low social cohesion index as well. Due to the reported socio-economic factors and trends, it could be argued that the economic rationale is the predominant one in determining informal practices. This is assumed to be so in informal employment and for tax evasive purposes in the case of informal entrepreneurship. In support of this argument, although concerning the first part of it, Williams (2013) discovered that informal employment is smaller in countries with higher levels of GDP per capita. Therefore, studying the economic motives of informal entrepreneurs who engage in shadow activities and in particular how they take such decisions is important. That would help in developing policy responses to influence all those decisions leading to informality, which are having the greatest weight in entrepreneurs' decisions. Based on the literature review, the research agenda will be to review how informal entrepreneurs decide their level of informality. In that process, they are assessing the connected costs of formality and the consequential costs of informality quite logically. It is expected that the knowledge to be derived would be highly relevant to other transition economies with similar characteristics to the Bulgarian one.

Evidently from previous studies (Bitzenis & Nito, 2005; Valdez & Richardson, 2013) macro-level institutional environment is significantly associated with entrepreneurial patterns in a given country's context. At the same time the likelihood of small firms to engage in corrupt activities is strongly influenced by low institutional efficiencies (Tonoyan et al., 2010) particularly in Central-Eastern Europe. Xheneti and Bartlett (2012, p. 618) argued that entrepreneurs who successfully learn to weave their way through a corrupt institutional environment, acquire a competitive advantage over others and experience business growth as a result of corruption. Obviously, this illustrates precisely one of the self-perpetuating aspects of large informal sectors, such as the Bulgarian one, in respect of poor formal institutional framework. Furthermore, it is important to note that development of enterprise policy in Bulgaria experienced the so called policy implementation gap – where consecutive governments failed to transfer and align working EU programs to the local context. Similarly to another transition economy, various programmes have been adopted for

the wrong rationales (Xheneti, 2017; Xheneti & Kitching, 2011) doing little in solving the problems of the business community. The described socio-economic environment is expectedly predominantly oppressive to micro and small businesses. This study suggests insights, which could serve improved policy efforts aiming to address deficiencies from the recent past.

To reinstate the key points from the literature review, it is important to note that the current lack of consensus on a single definition of the informal economy or informal entrepreneurship has specific implications. It has been already argued that the best way in defining the informal economy and entrepreneurship within it, is to employ a modular principle, depending on a widely accepted definition, which is then fine-tuned to reflect a particular country specific context. This is important for two reasons. Firstly, to enable an accurate and exhaustive definition of the various informal economies, which tend to differ substantially across national borders by their nature of existence. Secondly, to enable appropriate design of tackling policy measures. Development of adequate policy approaches therefore is subservient on precise definition and comprehensive understanding of the root causes of informality at a national level. A working definition of informal entrepreneurship was developed for the purpose of this study.
The main competing theorisations of informal economies and entrepreneurship have been discussed, emerging from the diverse perception of the entrepreneurs' characteristics, types, motives and role in the economy. It became evident, that neither of the current explanatory paradigms are sufficiently adequate. A group of scholars started to address the need of a new holistic theorisation, suggesting the institutional asymmetry paradigm (Williams et al., 2015a; Williams et al., 2015d; Williams et al., 2014d). Adopting IAP, this study aims to address its empirical challenges – indeed, how to measure empirically the intrinsic circularity between tangible and intangible phenomena, affecting tax evasive behaviour. The contribution to the literature debates is via appraisal of these issues, proposing an economic model of partly informal entrepreneurship through innovative socio-economic profiling of business tax evaders.

Chapter Four

4. Methodology.

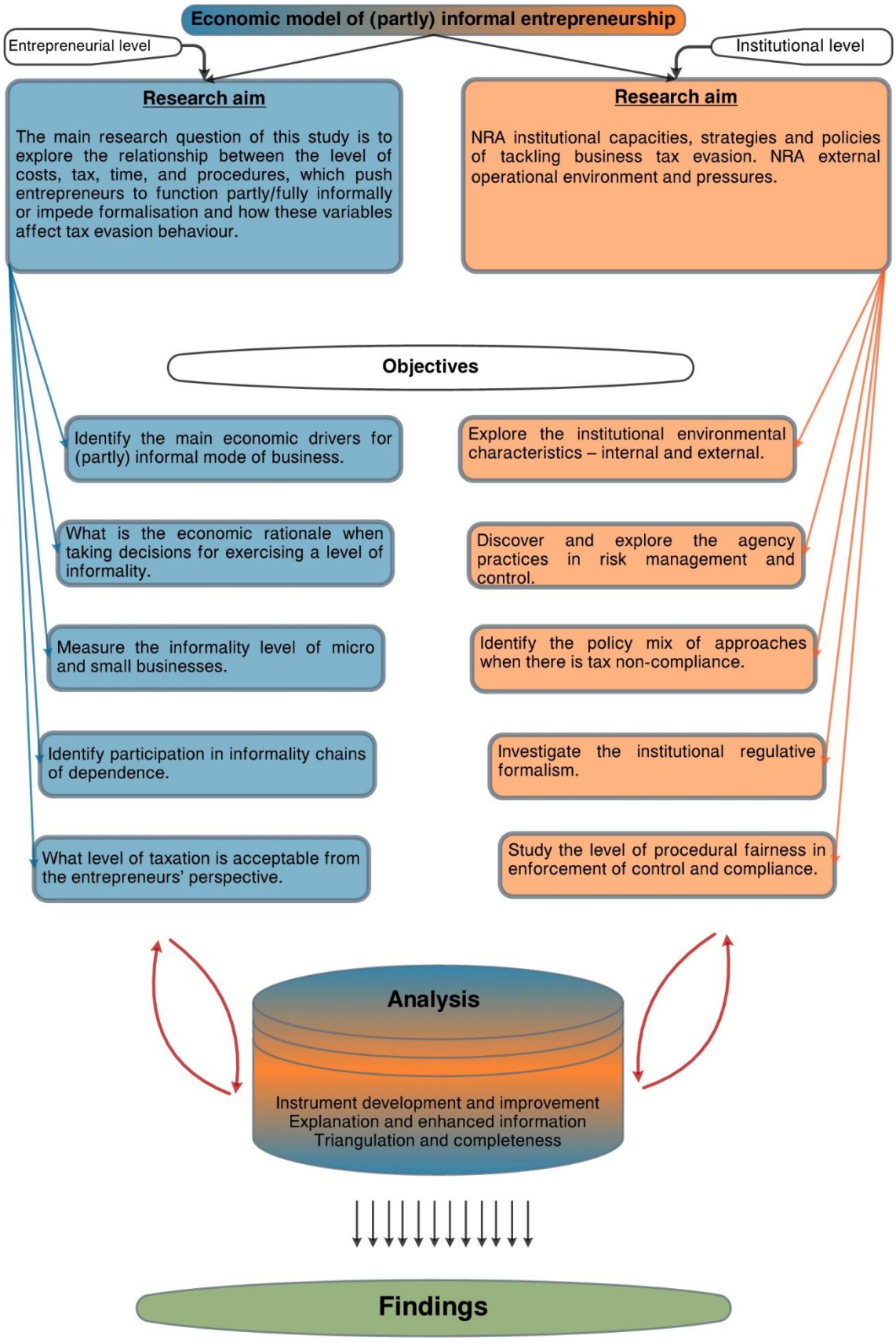
This research project aimed to build an integrative account of the main determinants of entrepreneurs' tax evasive decisions as a function of the environment in which the studied small firms operate. Chapter 4 sets the research aims and objectives underpinning the study along with the philosophical perspective of the study. It explains the rationale behind the particular design choices and methods' utilisation. The chapter offers a comprehensive justification of the empirical strategy at qualitative and quantitative level in order to secure the desired research access to carefully guarded institutional information. A detailed discussion of the employed analytical software solutions then unfolds to justify how the use of the best industry tools enabled an enhanced analytical rigour and findings accordingly.

In order to fulfil the research aims of the study, it is required to integrate quantitative and qualitative methods. The research objectives are at two levels – institutional and entrepreneurial level. Therefore, there is need for three layers of investigation – exploration, explanation and integration of findings. Figure 5 below visualises the research aims and objectives in their dynamics, followed by a detailed explanation of the methodology.

The text continues on the next page.

4.1 Research aim and objectives, philosophical perspective of the study.

Figure 5: Research aims and objectives.



The main research aim of this study is to explore the relationship between the level of costs, tax, time, and procedures, which push entrepreneurs to function partly/fully informally or impede formalisation and how these variables affect tax evasion behaviour. In that regard, the research aim is rather based on evaluating the economic rationale than on other types of motivations while investigating the entrepreneurs' decisions for their level of informality. Building a model of the entrepreneurial economic rationale to operate in the shadows would serve for future research to incorporate other types of motivation and hence enable reflection of the reality with higher precision. It is also a theoretical aim of the study to consider the applicability of such an informality model at a policy level, so that formalisation efforts will become more effective. In some occasions enablers of informality and barriers to formalisation could overlap as we can see from the study of Johnson et al. (2000) who discovered that bureaucratic corruption is significantly associated with hiding output. This is a distinctive phenomenon for countries like Bulgaria and these characteristics would need parallel exposition and analysis.

At the first level of investigation, it is required to explore the rationale of the institutional response (National Revenue Agency – NRA) to factual and suspected informality. To critically appraise the risk management strategy of the NRA and in particular their attitude towards controlling informal entrepreneurship activities. What are the triggering mechanisms and methods of enforcing control and compliance – tax audits, license and regulations' compliance audits. NRA's has institutional advantageous position in terms of actionable tax intelligence, which makes them the leading coordinating body in investigating and controlling informal economic activities in Bulgaria. Therefore, the policies, which they develop, constitute powerful stimuli for formalisation-informalisation practices and thus affect economically productive entrepreneurship. This research aims to contribute as one of the most comprehensive studies of NRA's tackling policy strategy. Therefore, the empirical focus asserts a qualitative approach as appropriate and necessary. Building upon in-depth interviews with the key NRA officials, involved in risk management policy creation, it has been aimed to reveal how policy attitude connects with the local entrepreneurship environment. In order to foster entrepreneurship and increase formalisation levels it is required to create more responsive and objective risk assessment system. Therefore, a relevant policy effort should be informed by an exhaustive exploration of the institutional response to factual and suspected informality.

At the second level (entrepreneurial), it is required to discover the economic rationale employed by entrepreneurs to exercise a preferred or imposed level of informality. In doing so, it is necessary to expose the causal relationship of their strategies to reduce costs of formality and consequentially informality. The third set of objectives is to discover and incorporate the causal relationships between the first two sets of objectives into an economic model of informality. This would enable understanding of the entrepreneurs' economic rationale to operate in the shadows and how their behaviour may trigger institutional response. A study offering such an integrative inquiry, deployed simultaneously at institutional and entrepreneurial level in a transitional economic setting, provides a degree of exhaustiveness, which is required to fully appraise these complex phenomena. The utilised empirical strategy allowed gaining research access to carefully guarded institutional context, which informed the quantitative stage and enabled the generation of valuable findings. The next paragraphs will provide an account for the philosophical stance of the study before justifying the implemented methodological approach.

It is important that this study makes its philosophical assumptions and perspectives clear to the reader. This will allow the author to expose his thinking and findings for others to interpret through the relevant theoretical and methodological commitments.

The research design and methods should emerge from:

- The philosophical stance of the researcher.
- The research questions and objectives.
- The necessity to acquire maximum reliability and validity of the empirical data in support of the anticipated findings.

The challenges in front of social science research grow proportionally with the increasing sophistication of social phenomena. Hence, the importance for researchers to justify their philosophical stance – if not always explicitly, then at least quite clearly from the argumentative body of their work. This would constitute the required transparency and rigour of their work. The ontological and epistemic affiliation of the researcher brings certain methodological commitments and when one is not adhering to them, they should provide a good reason for that.

The clear distinction between quantitative and qualitative methods loses grounds and the complementary use gain greater popularity. Echambadi, Campbell and Agarwal (2006, p.1816) argue that qualitative methods allow discovery of new variables, which were not accessible in previous quantitative studies. But this leads us to another contemporary issue in management research to which we need to pay attention if we would like to be engaging social researchers – the different forms of "*reflexivity – each with distinctive implications for the role of the management researcher in terms of aims, processes, and outcomes*" (Johnson & Duberley, 2003a, p. 1279). The awareness and openness towards the issues discussed above would allow for the creation of more relevant and impactful research. The rationale for positioning this study in the perspective of critical realism is the natural perceptive association with its main philosophical beliefs and in particular that:

- "*the defining feature of realism is the belief that there is a world existing independently of our knowledge of it*" (Sayer, 2000, p. 2)

- "*Critical realism distinguishes not only between the world and our experience of it, but between the real, the actual and the empirical, defining these in a special way*" (Bhaskar, 1978).

- "*Furthermore, critical realism argues that the world is characterized by emergence, that is situations in which the conjunction of two of more features or aspects gives rise to new phenomena*" (Sayer, 2000, p. 12).

These philosophical assumptions appear highly relevant and capable to enlighten a research perspective, which is the most suitable to study the domain of the informal economy. We have already drawn the parallel between use of methods and their emergence from our philosophical stance. Some additional important remarks to be made are that "*no methodology can be constructed as epistemically superior to any other – all are partial and fallible modes of engagement*". So that, even if one equip themselves with "*epistemic reflexivity*" in their effort to avoid the trap of "*inevitable circularity of epistemological issues*", the best that could be done is to develop a better constructed knowledge claim without the right to declare it superiorly valid (Johnson & Duberley, 2003b, p. 168, p. 187, p. 177).

To justify why critical realism perspective is ideally positioned to advance this study theoretically and empirically, we should consider one of the major characteristics of large informal economies. Indeed, this is the fact that they tend to be self-emergent and as such are progressively fuelled by the deepening conflict between the three social world stratums, which they further incite – in the sense of Bhaskar (1978) understanding of the real, the actual and the empirical. "The real" is the existing social

structures and causal powers, which can create events or how a given society is organised. “The actual” is the real world events, which are happening and “the empirical” is the social agents’ experiences. Obviously, due to incongruities (or institutional asymmetries) between the “the real” and “the actual” layers of the world, there are emerging individual empirical experiences, which could be socially polarising. The Bulgarian economic environment makes a distinct case to demonstrate and study exactly that. As evident from the findings (see 7.2), social actors (entrepreneurs in this study) appear to be strongly dissatisfied by their “empirical” experience of “the real” institutional framework including tax and administrative policies. Key state institutions (such as NRA) fail to actually exercise their powers in a fair and equalising manner and thus affect procedural fairness and distributive justice. This leads to raising the costs of formal entrepreneurship to levels deemed unacceptable and logically small firms engage in informal compensatory strategies. As a result, the development of informal chains of interdependencies and unfair competition act as strong influencers of even larger informal economy. Therefore, the individual stratification of reality due to the emergent social system entities in the sense of O'Mahoney, Vincent and Edwards (2014, p. 9), provides logical environment for the construction and deployment of mixed methods design, which is the intended approach in this project. This serves adequately the usual complexity of management research into which mono method research approaches are not always sufficiently rigorous. Mingers (2006, p. 202) argues “*that multimethodology (the combination of a plurality of intervention of research methods) is an appropriate and complementary methodology*” in critical realism. Given the fact that critical realism is regarded almost as a methodological tradition within economics (Fleetwood, 1999, p. 128), it is appropriate to appraise the socio-economic influencers of tax evasive behaviours through this philosophical stance.

4.2 **Research design and methods.**

In appraisal of the research design choices for this study, it is required to review certain integrative issues, which may arise. Incorporating qualitative and quantitative research requires careful consideration of the rationale to do so and quite often mixing methods does not reflect the methods themselves (Bryman, 2006). This leads to a loss of participants’ time, research resources and interpretation difficulties. In order to justify the utilised approach in this study, two important queries should be taken into consideration – what is the role of the research question in terms of methods’ superiority and what the quality criteria for mixed methods are. There are two major discourses discovered amongst researchers and their attitude towards the research design. Particularistic discourses (Bryman, 2007) are part of the traditional view where mixed methods are considered suitable only when this is determined by the research question. Whereas the universalistic discourse accepts the superiority of the mixed methods and that they will “*provide better outcomes more or less regardless of the aims of the research*” (p. 8).
The research questions of this study demands the use of mixed methods in terms of discovery of causalities and information, which the researched participants are expected to wish to withhold. This formulates the rationale for use of mixed methods, which implicitly employ features of the universalistic and particularistic discourses in this specific research scenario.

4.2.1 Design.

Before explaining the planned research design and methods into greater detail, we need to account for some important general characteristics. Miller and Tsang (2011,

p. 151) argue that *"Mixed designs combine both extensive and intensive designs, thereby utilizing their complementary strengths and weaknesses.*" From one perspective, this study would like to expose certain generalities through quantitative method (extensive approach) and from other perspective, it needs to uncover explanatory mechanisms of certain entrepreneurship's behavioural causalities (through intensive qualitative approach).

There are particular challenges in collecting primary data from respondents involved in shadow activities and usually that involves specific inference techniques. These difficulties spread over the design of the methods and how each data collection procedure needs to occur. In recognising those issues, this study will build upon two tested approaches. The International Labour Organization (2012) suggests specific order of questions deployment and gradual approach in the data acquisition steps. This is to minimise the respondents' stress from sensitive data disclosure and increase the information quality. Another study (Institute for Market Economy, 2004) done in Bulgarian context employed similar strategy at firms' level, i.e. to detect informality with series of indirect questions. Specific combinations of answers indicated informal activities.

In the research context of the study, it could be argued that exploratory sequential design will be best suited for its purpose. A qualitative method (in depth semi-structured interviews) has been used to collect primary data from key revenue officials. The collected intelligence has been analysed to inform the creation of quantitative method – a survey was used to collect data from entrepreneurs and small business owners. The rationale for using this particular research design is evident from the overarching structure of the two main research questions (Figure 5), which is one of the two common approaches (Teddlie & Tashakkori, 2009, p. 133). Obviously, the qualitative method is required to provide enhanced information leading to ability for constructing more comprehensive account of the studied phenomena. Furthermore, that would enable triangulation of the findings – to compare quantitative and qualitative findings for corroboration. Looking at the institutional level of the empirical collection phase, the practicable rationales are:

- There is a specific research question at institutional level, which requires qualitative approach – semi-structured interview with institutional officials in order to gather information about their practices and attitudes. This part of the study aims to explore NRA's risk management and control policies.

- Instrument development and enhancement – the results from the in-depth interviews will assist in fine-tuning the survey for entrepreneurs.

These mixed-methods have been designed right from the outset to be mutually informative in order to enable integration of the findings. It is projected that this data collection strategy will allow development of an economic model of informality – incorporating entrepreneurial and institutional strategies to deal with the costs each party impose on the other.

4.2.2 Methods.

The following data collection methods have been used:

1. Collection and analysis of secondary data from:
    - Government web sites (National Revenue Agency, National Statistical Institute, Ministry of Finance), reports and publications.
    - Existing legislation, regulations.

- Media accounts, newspapers.
- European Union publications – including survey based secondary data like Eurobarometer.
- Findings of other similar studies.

2. Collection and analysis of primary data from:
   - Semi-structured in-depth interviews with institutional officials.
   - A survey to entrepreneurs.

The questionnaire (survey) – entrepreneurial level

Contemporary data collection tools need to be flexible and capable to acquire maximum information with minimum efforts from the respondents. Participants should see only questions, which apply to their specific circumstances. In order to select a suitable provider, 23 leading data collection software developers were compared. They were also screened to find out if their software solution is capable to meet the following criteria (Mishkov, Maguire DS, 2012, p. 25-30):

1. Display logic.
2. Skip logic
3. Branching logic
4. Custom web pages visualisation based on responses.

In the presence of the above criteria, it appeared that Qualtrics possess all these features in their academic research suite and they seem to be most easily accessible and customizable. Therefore, this was the choice of survey software.

4.3 **Semi-structured in-depth interviews with key NRA officials.**

The section provides explanation for the development and design of the first component of the mixed methods methodology – the in-depth interviews with NRA officials. A brief review of the institutional context and the revenue agency is offered in order to navigate the reader into the method's choice and design. Justification of the empirical strategy, the development of the semi-structured interview questions as well as the choice and use of text analysis software solutions have been deployed in that order.

The semi-structured interviews with institutional officials were conducted in order to acquire in depth information about institutional practices in tackling tax evasion as a result of informality practices (including risk management and control strategy). All senior NRA officials who are involved at decision management level in policy development have been interviewed in order to achieve maximum representativeness. The relevant NRA directorates are: Risk Management and Control.

The three main institutions responsible for developing and implementing control functions in regards with the informal economy in Bulgaria are the National Revenue Agency (NRA), the General Labour Inspectorate (GLI) and National Social Security Institute (NSSI). The strategic cooperation between the three institutions is established in the so called Agreement 85 (General Labour Inspectorate, 2010). Dzhekova and Williams (2014, p. 64-67) provided detailed discussion about their role, functions and main performance indicators. What becomes evident is the fact that NRA has implemented a huge amount of operational checks during 2011, 2012 for which data is already publicly available. Even though the high amount of imposed sanctions, the overall efficiency remains questionable – if it is accounted for the collected income

and expense per check for that period. Over the past few years, the Bulgarian Revenue Agency (NRA) has become more sensitive to enhancing its efficiency and operational capacity. Its current inspection and auditing decisions are based upon a 'selection process' (risk assessment and profiling) incorporating a number of risk indicators as well as systematic selection procedures. Furthermore, NRA has started implementing a contemporary software solution based on the same platform used by HMRC, UK – Oracle Business Intelligence Enterprise edition Plus. The software system should have been implemented by September 2014, but it is expected that the configuration process will spread over at least a couple of years' period. Amongst its diverse features, it would enable:

- Actionable tax intelligence and real time responses.
- Tax non-compliance diagnostics – calculation of cost.
- Tax compliance enforcement – calculation of cost.
- Ad hoc Analysis and Interactive Reporting.
- Proactive Detection and Alerts.
- Lesser subjectivity in enforcement of checks and audits.
- Scorecard and Strategy Management.

At this moment of time, even before the implementation of the new IT system, NRA is the leading coordinating body in investigating and controlling informal economic activities in Bulgaria. This is due to its advantageous position in terms of its actionable business intelligence. However, it should be also noted that despite these innovations, NRA experiences shortages of qualified and well-trained analysts, which may compromise the potential effectiveness of the system. Particularly financial investigations of complex tax evasion cases require significant improvement in terms of operational and analytical capacity, while political intervention with high-profile investigations remains a serious problem. At the same time, the NRA has become much more service-oriented towards taxpayers. Recently it has released four new e-services for companies, including access of companies to documentation with regards to ongoing control and audit procedures, as well as electronic submission of relevant documents within audit procedures (Dnevnik, 2014). This enables better communication between taxpayers and the NRA and greater transparency and accountability in relation to its control function. These ongoing efforts should be reinforced through learning from successful similar experiences of other countries, in order to improve objectivity of control activities, greater use of e-services by taxpayers in the communication with the tax office, as well as more targeted investigations and risk assessment.

Procedural fairness (Williams, 2014c, p. 26) refers to the extent to which people believe that they are paying their fair share compared with others and receive fair treatment by institutions. It is worthwhile noting, that the appreciation of procedural fairness amongst Bulgarian individual and small business tax payers is below the acceptable social minimum (see section 7.2). For instance, big tax payers often receive much more favourable treatment by the tax authorities and that they could even avoid situations (not paying the lawful amount of tax) while the smaller tax payer would be treated to the letter of the law and exposed to much more institutional scrutiny. The reasons behind these phenomena lie in the domains of high level of corruption (see Figure 16, Appendix 17) and doubtful political protection for specific tax payers (Noutcheva & Bechev, 2008). While procedural fairness is an intangible social phenomenon, it has tangible footprint on the business tax payers and consequently affect the economic aspect of their rationale to engage in tax evasive practices. That is why it is highly desirable and applicable for Bulgaria to achieve greater level of automation and objective detection in its institutional control and enforcement measures (via NRA), starting from more responsive risk assessment system.

Figure 6 summarises how the main research aim at NRA level is materialised via three investigative lines of inquiry embedded in the qualitative instrument.

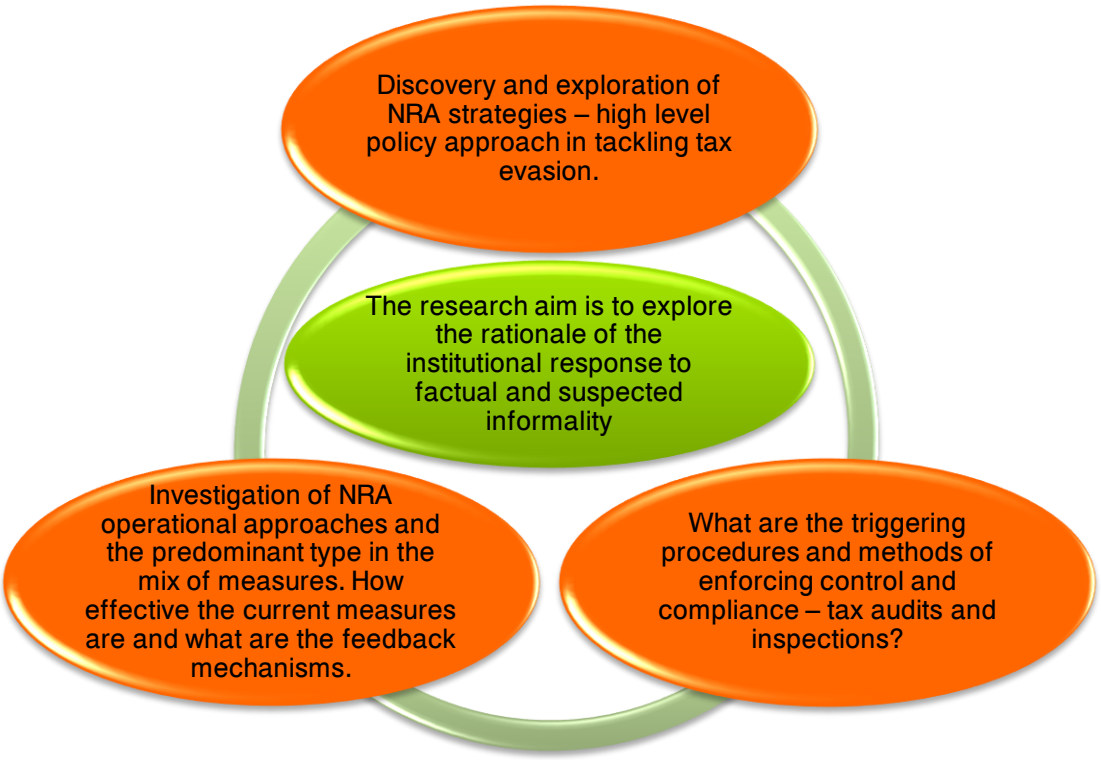


Figure 6 – research aims – NRA.

4.3.1 Empirical strategy.

This section provides rationale and explanation for the chosen empirical strategy, which is based on secondary and primary data and in particular, how the access to this data was acquired. The secondary data has been used as evidence base to inform the development of a semi-structured interview in terms of questionnaire formulation, sequence and precise wording.

According to the Bulgarian law, access to institutional information could be provided based on a specific request submitted usually in writing (Bulgarian National Assembly, 2011). Public information is all information created, collected and managed, by institutions except in the cases where this information is protected by another law – for instance personal data, classified information. While all requests for access to public information should be treated in alignment with what the law dictates, sometimes institutions may process different information request in their own view. In such cases where the organisational response is not satisfactory, the enquirer has the right to escalate their information request through a legal procedure.

The main task of the Bulgarian National Audit Office (BNAO) (2014) is to "control the reliability and truthfulness of the financial statements of budget-funded organisations, the legality, efficiency, effectiveness and economy of the management of public funds and activities, as well as to provide the National Assembly with reliable information thereof." In order to implement its main functions BNAO has been granted auditing authority over any government institution. BNAO could acquire access to internal documents and information, which then serve as evidence base for the creation of various extensive audit reports published on their web page. Therefore, it is an

effective empirical strategy to commence a study in a particular area of research with reading the evidence base utilised by these reports. They usually contain specific information, which is difficult to be accumulated or acquired otherwise.

The most relevant to this study and up to date BNAO audit report is called "A performance audit of the control and enforcement actions undertaken by the National Revenue Agency to enforce compliance with the tax and social security legislation" (Bulgarian National Audit Office, 2012). This report is produced using extensive evidence base of documents and empirical information acquired from NRA. The best possible way to a rigorous development of an interview questionnaire intended for key decision makers is to conduct an independent research analysis of the extensive evidence base of internal documents. Following this empirical strategy there were two possible lines of action:

1. To submit access to information request to NRA.
2. To submit access to information request to BNAO for the same documents.

Considering the sensitive nature of the information in question, it was decided that it is rather unlikely that NRA should process favourably such an information request. They might have refused to provide evidence based documents on the grounds that some of them are classified as a professional secret. In principle, each information request submitted in accordance with the Law for access to public information (Bulgarian National Assembly, 2011), is processed by the legal department of the relevant institution to ensure it meets the requirements of the law. The legal department also guarantees that their decision is legally grounded. BNAO is the ultimate public control body and this implies higher level of accountability in comparison with other institutions in front of the society.

It was decided to submit an information request (IR) to BNAO instead to NRA assuming greater chance of acceptance. Both written information requests specified explicitly that the information is required in regards with PhD research, although there is no such a legal requirement – to explain the reasons for the demand of public information. The purpose of this statement was to openly provide the legal reviewers with the genuine purpose of the enquirer, which is to use all the information for conducting academic research and eventually disseminate the findings. The first information request asked for an electronic copy or a paper copy of the information. BNAO granted only visual access to the required information, although most of the data was available in electronic form. This means that the enquirer has the right to review the authorised documents in an electronic or paper form (as they are available) in a dedicated room. After an initial review of the documents, it became evident that:

- There are about 440 files – predominantly excel and word files. The massive amount of evidence made the visual access at BNAO unfeasible.

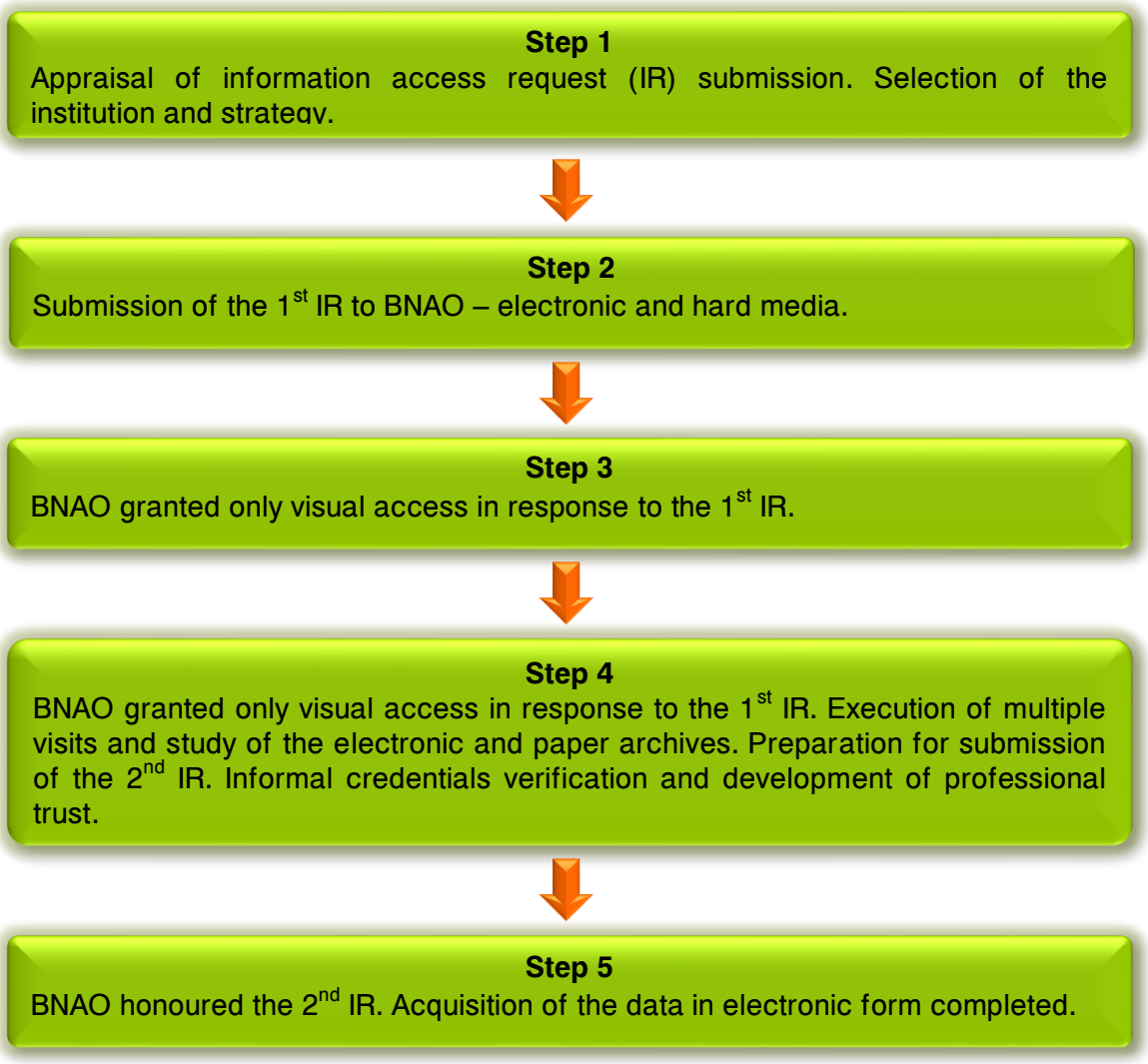
- It has been discovered that there are some more documents of interest not specified in the initial information request. They were included in the second one.

During the visual access following the first IR, the two BNAO institutional officials responsible for providing the access to information raised several inquiries about the role and intentions of the researcher. A good working rapport was established during a couple of a few hours' sessions (visits). Consequentially, the researcher was advised to prepare and submit a new IR. The second IR had to specify the newly discovered relevant documents and an electronic form of access to all sources where possible

with paper versions in case that there are no electronic versions. BNAO has approved the second IR within the lawful time limit. The originals of the first and second IR are presented in Appendix 10 and 11. This is a partial description of what evidence base the IR has been authorised to receive access to – each data stream (folder) has multiple documents in it created by NRA:

- Control
- Collection
- Tax and social insurance methodology
- Risk management
- Investigation of special cases
- Modelling and design of business processes
- Annual program for compliance and minimize risk in the NRA for 2010, 2011 and 2012
- Analysis of the assessment and management of risk in the NRA
- Analysis of the process of establishing violations, development of administrative proceedings and issuance of sanctioning orders in during inspections by the revenue department of the «Operational checks»

The following flowchart summarises the empirical strategy in regards with acquisition of data:



4.3.2 Development of semi-structured interview questions.

Given the project's overarching aim to develop an in-depth understanding of tax policy management within institutional incongruence setting, nine in-depth semi-structured interviews with key NRA officials have been conducted. The average length of the interviews was 55 minutes each. The transcribed text in English is approximately 41 150 words. NRA headquarters are organised in directorates and the two ones responsible for developing and implementing policy approaches are Risk Management and Control Directorate (NRA organisational structure is presented in Appendix 9). Each of these departments has a director, deputy director and two senior officials. These high level executives are responsible for collaborative development of the risk and control management strategies of NRA, which represents the core mission of the organisation. Thus, the interview schedule includes all key executives in order to collect information from everybody involved in risk and control management policy – 9 people in total. In pursuit of such specific research findings, it could be argued that the only feasible methodology is a semi-structured interview deployed amongst all key decision-makers in the area of study.

The conducted interviews were semi-structured and were deployed using the questionnaire outlined in Appendix 7. The semi-structured element allowed for some implicit and relevant issues, which were not on the interview schedule, to be explored further using additional clarification questions. Some of these issues were raised by the respondents and some of them provoked by the interviewer. As it is evident from the research objectives and Appendix 7, the interview questions are focused to collect business intelligence on specific aspects of NRA attitudinal approach – in terms of developing and implementing risk and control management strategies. Such information is not publicly available and it is very carefully institutionally guarded. Specific aspects of the sought business intelligence, for instance the concrete criteria for attracting audits and inspections of suspected non-compliant firms, is protected by law (Bulgarian National Assembly, 2013, article 14, paragraph 14). For this reason the wording and the sequence of the questions are developed with particular care using graduated approach, so to:

- Collect as much information as lawfully possible.
- Expose the anticipated findings from multiple angles (the same respondent answering nuanced questions across the interview). There is some intended question overlapping to achieve this purpose (up to 10%).

The graduated approach is a special well-known approach in investigating sensitive topics (International Labour Organization, 2012; International Labour Organization, 2013; Kazemier, 2014; Kazemier & van Eck, 1992). The interview questions were informed by thorough analysis of the information gathered with the authorised IRs by BNAO. All these files are electronically available as part of this thesis, however access to them needs to be properly documented and authorised, because of obvious legal reasons. Appendix 8 provides explanation for each question of the semi-structured interview: rationale/justification of the question, what type of information it is designed to collect and its interconnectivity with other questions. All questions are developed emerging from the study of the extensive evidence base – accessed via the authorised information requests.
The initial section of the interview consists of four introductory and general questions, so to put the respondent in comfort mode and establish rapport with the interviewer. The aim of this section is to gather information about the main tasks, responsibilities and methods being used to perform them by the two NRA directorates. An additional implicit purpose of the first interview section is to explore the cohesion level in terms of intra-organisational understanding of these tasks, responsibilities and methods

accordingly. It is aimed to discover the actual level of alignment amongst the key decision makers' opinions and thus to build a picture of departmental role cohesion and potential related issues.
The next seven questions are part of the so called "Special block of questions" aiming to collect the most sensitive intelligence. This section's purpose is to reveal NRA response upon discovery of non-compliance, but most importantly how systematic it is in reality as well as objective and adequate to contemporary evasive practices.
The last part of the questionnaire presents an opportunity for cross verification of previous and impending questions with the formerly provided answers. This last section aims to reveal additional aspects of NRA tackling tactics with an accent on the type of action from policy perspective. Furthermore, it also asks some questions to be used for later correlation analysis when additional primary data is collected from a business survey amongst entrepreneurs.

4.3.3 Choice and use of text analytics software.

The use of Computer Assisted Qualitative Data Analysis Software (CAQDAS) is acquiring significant credibility in qualitative data research and this could be correlated to the amount of publishing activity over the last 15 years (Rademaker, Grace & Curda, 2012, p. 3). The purpose of this section is to briefly review some of the advantages of CAQDAS software in qualitative analysis and provide rationale of its use on this occasion. It only considers two representatives of this type of software and in particular NVivo and IBM SPSS Modeler. The reasons for this are that NVivo is perhaps the most popular and widely used solution in social science research utilising statistical algorithms in text analytics. In doing so, we could talk about partial quantification of qualitative research as a contemporary trend. At the same time IBM SPSS Modeler is the pioneer of advanced linguistic technologies and Natural Language Processing (NLP), which is being used to rapidly process unstructured text data. IBM describes in the help documentation of their product how this modern technology works:

"IBM® SPSS® Modeler Text Analytics relies on linguistics-based text analysis. This approach offers the speed and cost effectiveness of statistics-based systems. But it offers a far higher degree of accuracy, while requiring far less human intervention. Linguistics-based text analysis is founded on the field of study known as natural language processing, also known as computational linguistics."

Both software products in question demand intervention from the analyst in terms of building and applying the analytical means, but at significantly different levels. NVivo uses statistics-based algorithms while IBM SPSS Modeler employs natural language processing capability. With both software solutions the researcher needs to bring his own understanding during the analysis but this may happen at different times. The two main differences could be summarised:

- NVivo requires involvement of the researcher from the very beginning and all justifications are statistically based on frequencies of speech parts. There is no artificial intelligence involved.

- IBM SPSS Modeler employs artificial intelligence and NLP and thus allows for the analyst to advance the study much quicker. The product allows involvement of the researcher from the very beginning or at a later analytical stage. IBM SPSS Modeler tends to understand more like a human will do and machine learning and training could improve this process for a particular area of analysis.

- Once the initial key concepts (categories/nodes) are identified the visualisation capabilities of NVivo appear to be more developed than IBM SPSS Modeler ones. This provides better grounds to present the findings to the readers. Said differently, the justifiable explanations could be presented in a more visually appealing manner using NVivo.

CAQDAS software offers a range of benefits, which increase the rigour of analysis. Some of the advantages of using CAQDAS software relates to the creation of an "auditable 'footprint' of the progressive dialogue between the researcher and their data" (Sinkovics & Alfoldi, 2012, p. 5). Employing CAQDAS solution allows building and conceptualising of a qualitative study in a structured way, which is not otherwise possible – for instance when a manual approach is adopted. The analytical approach, which characterises CAQDAS solutions, could be discussed in the following aspects:

- Provision of standardised coding schemas with high level of legitimacy.
- Development of justifiable explanations – quantification of qualitative analysis. Escalation from descriptive to analytic stages – from concepts to categories and text link analysis.

Provision of standardised coding schemas with high level of legitimacy

Majority of CAQDAS qualitative analyses, independently from the source types, start with the creation of an initial coding scheme. This is the stage where the main concepts are being discovered using various methods (automated and non-automated). These main concepts are then grouped in categories, sets, hierarchies. It is important to note that the different software solutions have different working terminology and approaches accordingly, but their principle of generalisation is quite similar – this is the escalation from low to advanced level of study through concepts, categories, themes and models. Utilising an industry standard text analytics solution secures higher level of investigative legitimacy. The use of CAQDAS software provides the researcher with a better understanding of the analysis process undertaken (Wickham & Woods, 2005, p. 687), which arguably increase the quality and quantity of meaningful insights. Any relevant line of inquiry could be traced and verified by an external reader to the lowest level of descriptive nugget and then could be traced forward to the emergence of advanced level explanations. Inevitably, this increases the validity of research and analysis.

Development of justifiable explanations.

The development of justifiable explanations is possible through the so called partial quantification of qualitative analysis (ability to apply statistical and/or NLP algorithms in qualitative analysis) – offered by CAQDAS. Otherwise, analysis of qualitative data would remain exposed to the subjectivist biased questioning from the critics. As Ryan (2009, p. 158) argues such software solutions "enable researchers to make visible their methodological processes for a more 'trustworthy' study". Furthermore, partial quantification of qualitative analysis is the only method to date which contributes to lesser subjectivity in qualitative analysis through the use of software. The below figure 7 illustrates what is a typical workflow of text analytics in IBM SPSS Modeler. Following an industry approved data mining workflow creates explanations with widely recognised justifiable potential.

Figure 7 – IBM SPSS Modeler workflow adapted from (IBM, 2014, p. 7)

The ultimate purpose of employing the so called dual solution software rationale (click here for more details) is to achieve maximum validity of findings, transparency and objectivity. To achieve this task, both industry benchmark products (Nvivo and IBM SPSS Modeler Premium) have been utilised in dual mode. At the same time, it is important to note that they have been engaged from the perspective of being analytical research enhancer tools. All findings and the software inquiries to derive them have been double and triple verified via software to software logic comparison and software to researcher's logic and knowledge simultaneous appraisal. It is the author's view that this dual solution approach is in itself a methodological contribution to the way qualitative analysis might be conducted. It allows quicker, deeper and higher level analysis to be implemented in a very visible and trustworthy way. Chapter 5 and the associated appendices contain the fine detail of this process.

4.3.4 Integrative qualitative analysis with IBM SPSS Modeler and NVivo.

In recognition of the substantial amount of interview data (in excess of 41 000 words) and for the purpose of achieving maximum validity of findings, transparency and objectivity, a dual software solution approach has been employed with two of the industry benchmark solutions. In particular, IBM SPSS Modeler and NVivo as major representatives of CAQDAS, have been utilised from the perspective of being analytical research enhancer tools. Headings 4.3.3 and 4.3.4 provide extended rationale for the respective choices. This section consists of two parts: first analytical phase using IBM SPSS Modeler and second analytical phase using NVivo. The discussion of findings takes place in chapter 5. This section presents the rationales and technicalities of the qualitative analysis, while the next chapter discusses the findings in regards with the research context and summarises the relevant conclusions. The dualistic, integrative application of the two software suites enabled:

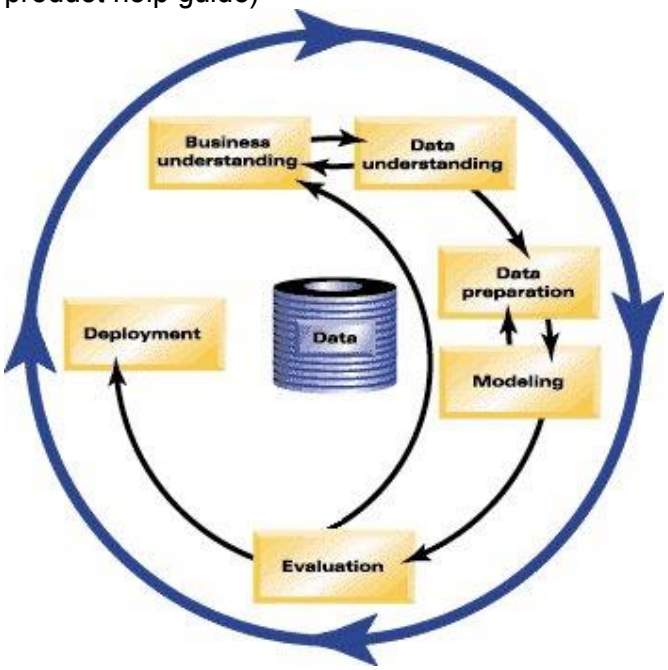
1. Utilisation of specific advantages and functionalities of these products at each stage of analysis – NLP and visualisations for instance. More details are provided later in the section (here).
2. Facilitation of cross-verification and objectivity of the coding and thus increased validity of the findings.

First analytical phase using IBM SPSS Modeler.

The qualitative analysis of the in-depth interviews started with the so called concept extraction from within IBM SPSS Modeler Premium (v. 16). It is based on a combination of linguistic and statistical methods. The linguistic algorithms of the product deal with ambiguities of the language and help the processing stage in terms of providing artificial intelligence during extraction and interpretation. Simultaneously, the statistical algorithms are responsible for classifying/ordering the extracted key concepts. Such an algorithmic combination guarantees high level of objectivity in terms of key concept extraction in comparison with manual coding. The key concepts represent the essence of information the texts contain and are generally the simplest analytical construct.
Text analytics with IBM SPSS Modeler is based on several stages – the next section will briefly describe what steps have been implemented for this particular qualitative analysis. The level of explanatory detail is quite low in comparison with the product user manual of approximately 336 pages – it provides concrete information for every capability, so that the researcher could take an informed decision about which ones are suitable and how to use them. Every qualitative analysis with IBM SPSS Modeler fits into the embedded workflow – CRISP-DM, which stands for Cross-Industry Standard Process for Data Mining (see Figure 8 below). That is the industry-proven

way to guide data mining efforts and it is spanned over several semi-structured stages. There is a certain level of flexibility in terms of approaching and conducting this process, but it is limited to ensure that the analyst follows a benchmark industry standard.

Figure 8 – The data mining life cycle (based on IBM SPSS Modeler Premium v. 16 product help guide)



Stages of the text mining approach with IBM SPSS Modeler:

1. Extraction of key concepts.
2. Creating and assigning types to the extracted key concepts. Several passes of extraction and prioritisation of key concepts.
3. Building and refining of categories, based on key concepts.

Stage 1: Extraction of key concepts.

In order to extract the key concepts from the in-depth interviews multiple IBM SPSS techniques have been employed (text tokenization, text normalization, candidate term extraction, part of speech tagging). Consequentially about 387 key concepts were extracted from within approximately 41 150 words of interview texts.

Stage 2: Creating and assigning types to key concepts.

During the second stage majority of the concepts have been assigned a specific type (see Figure 9). A certain amount of concepts has been left without an assigned type due to the decision that they are insignificant for the analysis. This decision has been taken using a cumulative assessment based on three criteria:

1. What is the percentage of appearance of a particular key concept in comparison with the others.
2. In how many transcribed interview documents a given key concept presents itself.

3. A subjective expert assessment by the researcher about the significance of a given key concept.

In principle, the necessity to assign a "type" to key concepts relates to some later stages of the text analysis and in particular discovering patterns and relevant text links (text link analysis – TAP). A "type" in IBM SPSS Modeler terminology is a higher-level concept that contains one or more terms – several key concepts may belong to one "type". There are some default types (like *organisation, product, person, and location*) however, the researcher must create custom types for every particular text analysis. In doing so, the relevant IBM SPSS Modeler language libraries are being updated and accustomed to the specific analysis scenario. When a new extraction phase of key concepts is implemented again, the freshly created types are assigned as per the linguistic interpretation by the software engine (statistical algorithms might be used too). This means that the researcher trains the artificial intelligence of the software to assign certain types to certain key concepts. This leads to the so called interactive working mode of several passes for key concept extraction, which makes every later stage more accurate and intelligently educated by the software algorithms. For illustrative purposes, the following picture is an example of what types were assigned to some of the top key concepts. There is a colour correspondence between key concepts and their relevant types. This is what have been followed for this particular analysis.

Figure 9: Text analytics with IBM SPSS Modeler – key concepts representation.

Concept	In	Global	D(	Type
inspections		166 (7%)	...	<Control>
audit		151 (6%)	...	<Control>
individuals		128 (5%)	...	<Persons>
selection		125 (5%)	...	<Risk>
measures		113 (4%)	...	<Process>
right		110 (4%)	...	<Legislation>
taxes		93 (4%)	...	<Tax>
risk		87 (3%)	...	<Risk>
system		80 (3%)	...	<Product>
people		76 (3%)	...	<Persons>
nra		71 (3%)	...	<Organization>
information		70 (3%)	...	<Information>

An important phase (as a sub-stage of Stage 2) is to assign all important concepts to a category. This has been done manually exercising the expert knowledge of the texts. Most of the concepts have been already automatically assigned by IBM SPSS Modeler using categorisation rules. This could be seen in the column "In" where there is an icon available (Figure 9 above). This automated category assignment has been verified by the researcher and this has been done on several passes.

Stage 3: Building categories

At this stage, the process of text mining included building advanced level themes – categories. These represent the information, which is considered important from the responses. A category is "a group of closely related concepts, opinions, or attitudes" (IBM, 2010, p. 60) and it could also include specific conditional rules to recruit key concepts. The so called key concepts, types and patterns from the previous stages are the building blocks of the categories. It is very important to select and apply the most adequate categorisation approach, considering that the software product offers several ones to suit different analytical scenarios. The correct categorisation

technique(s) should be based on the following selective methodology, which was adhered to in this qualitative analysis:

1. The researchers familiarise with what each categorisation method offers in terms of functionalities.

2. Different sets of method combinations are then applied in several independent passes according to their suitability. The categorisation results are inspected by the researcher against their expert knowledge of the targeted texts for analysis.

3. The researchers elect these categorisation methods, which are likely to produce categories reflecting the expected analytical outcome. There is an expert subjectivism involved in this selection process and the analysis could have different bias grounded in the different choices made. However, independently of the categorisation algorithms mix employed, there will be a high level of linguistic and statistical justification of the various categories created.

Even the most relevant categorisation approach will generate categories, which will be imperfect with different combinations of inaccuracies in their names and hierarchical structure. The researcher will then need to change names, create rules of inclusion/exclusion, merge categories, delete or create new ones. This is lengthy process and it predetermines the quality of findings. This is the process of building and improving a categorisation model and it was implemented for the purpose of this study. Implicitly, during that process, several categorisation passes have been implemented and every later one generated better results. In general, there are two major categorisation approaches, which could be utilised – linguistic techniques or frequency techniques. The latter employ statistical algorithms to group key concepts into categories and as such are better suited for structured data like survey responses. The four linguistic methods (concept root derivation, semantic network, concept inclusion) provide a natural language-based approach to categorisation. A fourth technique (co-occurrence) uses co-occurrence rules to discover and group concepts that are strongly related (IBM, 2010, p. 60). It is important to note that utilising only the four linguistic methods one could generate approximately 16 different sets of categorisations in accordance with combining the different techniques with each other. A rigorous analytical approach would demand that the researcher evaluate and select from between all these. The suitability of “which technique when” is discussed in great detail in one of the product guides (IBM, 2010).

It has been found that in the given text analysis scenario the simultaneous use of the three linguistic methods generates the most adequate categorisation, so this approach has been used. There are three tasks which have been completed in the categorisation refinement:

1. The first one is to visually inspect and edit categories – their names, hierarchies and what concepts are being included. This is a process of improving the categorisation adequacy. During the implementation of this task some categories should be deleted and others created including processes of merging relevant ones. It is important to note that this involves changing the descriptors of each category and in doing so their significance and priority for later stages of the analysis.

2. The second task is to review again all remaining key concepts, which have not been included into categories to this moment of time. They need to be

manually assigned to the relevant categories. Return to task 1 and re-implementation.

3. The final task is to run the categorisation process once again, since this is about to generate slightly different results. A final adequacy review needs to be applied by the researcher, so that it could be verified that the newly generated categories are appropriate and precise.

After several categorisations passes the following 17 top categories have been created with their relevant number of descriptors:

Figure 10: Top categories from interview analysis.

Category	Descriptors
All Documents	-
Uncategorized	-
No concepts extracted	-
Risk Management	121
Control	48
Tax	46
NRA	44
Business	36
Behavior	22
finance	21
Institutions	17
Tax fraud	9
Legislation	7
bulgaria	6
police	5
Lawful prosecution	4
Environment	4
penalties	3
economics	2
features	2

The top 17 categories contain 89 sub-categories down to heading 3. Full representation of all categories is available in Appendix 13. The next important step is to create a custom Text Link Analysis package (TAP) from the results, so far. Text Link Analysis is a pattern-matching technology that can find relationships between the concepts in the text data based on user-defined patterns (IBM, 2010, p. 166). The initial categories and key concept have been created with IBM SPSS Modeler. Several passes have been implemented using various automated techniques (extraction algorithms) to refine the categories. Furthermore, the process was supported by exploration of Text Link analysis (TLA), which provides a way to identify and extract patterns from the texts and then use the pattern results. Subsequently, each key concept, category and pattern was analysed for appropriateness by the researcher. This was achieved using the expert knowledge of field - appraising the accuracy of the artificial intelligence analysis through the perspective of the author and making adjustments where required.

Second analytical phase using NVivo.

This phase involves NVivo 10, which have been used for the rest of the analysis. The rationale behind this dual solution approach ensures a couple of important advantages for the analysis. Explained differently that could be described as the best

of both worlds strategy. The two accents, which need consideration, are:

- The initial stages of key concept recruitment and categorisation have been implemented with IBM SPSS Modeler employing the most advanced NLP and extraction techniques available. Certainly, this is advantageous in comparison with the statistical algorithms, which NVivo uses exclusively for coding purposes. The differences have been already commented earlier in the thesis.

- The higher level study and representation of analysis is implemented using NVivo's superior project visualisation and query capabilities.

The technical process of exporting the project from IBM SPSS Modeler 16 to NVivo 10 is possible through some advanced features available in the latest versions of the two products. In particular, this includes the following major steps and functionality utilisation, which minimises the efforts' cost of this approach:

1. Key concepts and categories have been exported in csv and excel format from within IBM SPSS Modeler. They have been consequently imported into NVivo as nodes with the corresponding number of sub-nodes.

2. Depending on the researcher's knowledge of the topic and the analysed interviews, a few subsequent passes of coding and verification for adequateness have been implemented from within NVivo. The hierarchical structure of categories (nodes using NVivo terminology) has been further refined through merging into updated hierarchies, which led to a reorganised version – available in Appendix 14. It includes 9 top categories (nodes) and 67 sub-categories.

3. The coding process from this moment forward has been performed using an advanced innovative feature available in NVivo 10. It is called "Automatic coding using existing coding patterns". In general, the analyst must do some manual coding to "train" NVivo in which nodes to which parts of the text should refer – that is creating patterns. This manual coding level should be done for all nodes and their relevant sub-levels according to the technical requirements of the product. Once the process is completed, the automatic coding feature could be employed to examine all target texts and to deploy the required coding depending on pattern recognition and replication. Upon accomplishment, the researcher must perform random verification with sufficient level of detail to find out whether the auto coding feature performed with the right precision and relevancy. This has been done through exploring the newly created model of nodes with their relevant sources and references. Having implemented all these steps, it was discovered that the auto coding feature executed its functionality with high precision.

4.4 **Survey development and design.**

The section provides justification for the development and design of the second component of the mixed methods methodology – the business survey. In doing so, some key aspects of available survey methods will be reviewed here, which have development relevance to the quantitative empirical instrument. These could be generally divided into two types as presented in the table below.

Evaluation of data collection methodologies	1. Choice of data collection strategy. 2. Sampling method, sample size and survey distribution.
Questionnaire development and rationale	1. Gradual versus direct approach. 2. Relevance of the survey to the research aims and questionnaire's rationale 3. Pilot cognitive testing.

4.4.1 Evaluation of data collection methodologies.

Choice of data collection strategy

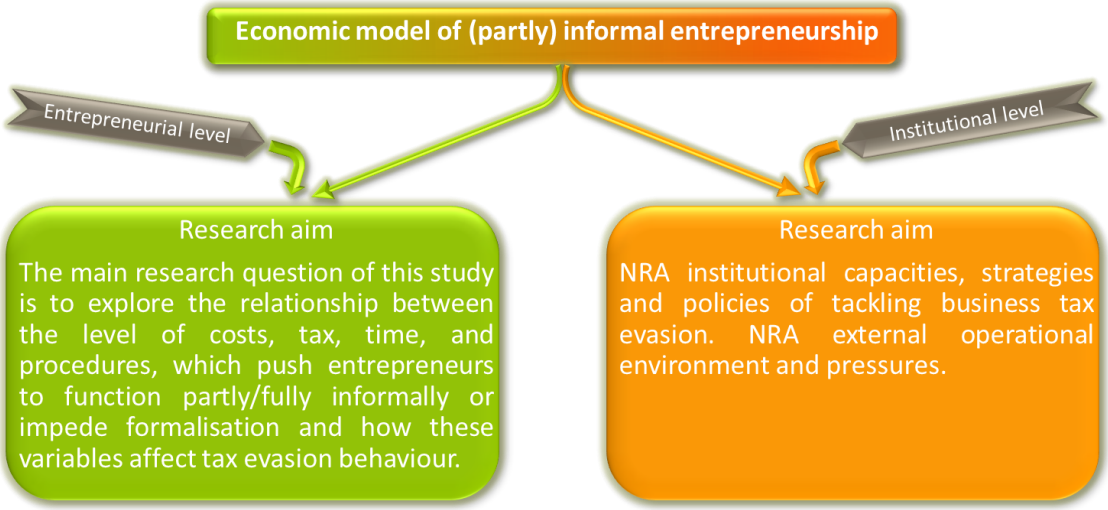
Survey methods are being increasingly employed to evaluate various aspect of the undeclared economy. Some notable cross-national direct surveys of participation in the undeclared economy emerged (Eurobarometer, 2007; Eurobarometer, 2014). They share the same definitions and statistical methodology in all EU member states, which provides important cross-country comparability and standardisation. While earlier studies had their accent on measurement of shadow activities, later ones became better equipped to discover the nature of the informal economy, the characteristics and motives of the participants – both on the supply and demand side of such illicit activities. Such is the aim of the employed business survey design in this study. In order to avoid variable errors, a particular attention is paid to "translating complex concepts into operative, communicable questions placed in the best context and sequence" (International Labour Organization, 2013, p. 80).

The data acquisition probability in direct surveys on the informal economy is greatly influenced by the unit of analysis and how they are approached. The substantial differences between surveys on the informal sector with the household as a unit of analysis and those with the enterprise instead, demand specific strategies designed to cope with such complexities. Direct approaches interested in non-business respondents have been available for several decades and have been extensively used. Reasonably, it could be argued that such methods are in advanced stage of development (scientific maturity) and generate reliable data, which is in synergy with the findings of macro-economic methods. However, direct surveys intended for the informal enterprise should be considered to function in their own comparatively underdeveloped domain. A significant consideration when studying the informal economy, and activities, is the sensitive and expensive nature of approaching business entities in comparison with households. This is a weighty consideration, particularly when surveying sensitive topics such as tax evasion. Informal economic activities are considered more socially acceptable when exercised by individuals and the other way round when undertaken by businesses. Domestic interviewees are therefore much more likely to be less secretive and hence anticipated to disclose potential involvement in the undeclared economy (Jones, Ram & Edwards, 2004; Williams, 2001; Williams, 2004). Simultaneously, partly or fully informal enterprises are logically much more likely to attract institutional attention if they chose to reveal evasion behaviour, which could easily result in negative consequences for the business. The questionnaire design for the purpose of this study recognises and accounts for these issues' involvedness – it is presented exactly how later in the chapter.

Williams (2006) analysed the first direct survey (in the advanced economies) of

business perceptions regarding the magnitude of the shadow economy. An important methodological implication is the way of asking about involvement in shadow activities. Instead of asking the question directly, the respondents are instead enquired about involvement in cash-in-hand transactions by other businesses from their sector. Since then, this indirect evaluation approach becomes the standard method of inquiring about illicit activities where the enterprise is the unit of analysis. It is also successfully implemented in this study, signifying notably bigger share of informal economy in comparison with the one established by macroeconomic methods. In the comparative evaluation of data collection techniques Williams (2015a, p. 6), comes to the conclusion that there is no "clear-cut conclusion that can be reached from the experiments with different data collection techniques so far conducted, about which is most effective to use." It could be argued that the choice and design of empirical instruments to study the informal economy should take into consideration the socio-economic and cultural characteristics of the research context and the actual sample of respondents. For instance, in societies with low perceived institutional quality, procedural fairness, justice and OISC index, it is more likely to engage prospective participants in fairer disclosure of sensitive information (such as tax evasion) if unattended forms of interviewing are being used. This was confirmed during the pilot cognitive testing and later again through the actual data collection process.

A major line of methodological criticism regarding surveys is the reduced confidence that they would fulfil their primary purpose – collect data about tax evasive practises, which are supposed to be kept secretive by the respondents. And then if they actually collect such data what would be its reliability and validity. In that respect, it could be argued that in order to determine the adequacy of direct versus indirect methods one has to perform a weighted findings comparison – for instance measuring the informal economy size with a direct nationally representative survey against the findings of macroeconomic methods. Whether both methods will generate similar results and if not, which method is more likely to better represent a given reality. However, it could be argued that direct methods enable exploration of motives and behavioural patterns in tax evasion, which remain totally inaccessible for indirect methods. Because of the discussed characteristics and in respect of the research aims of the thesis (see the entrepreneurial level in the figure below), the direct business survey appeared the best data acquisition approach.



For that reason, a business survey was used to detect and evaluate informal practices amongst formally registered micro and small businesses, meeting specific economic profile criteria. It is important to underline that diagnosing a level of informality, together with the background motives, amongst formally registered enterprises is

actually the most methodologically challenging study task possible (in exploration of the informal economy). The businesses, which are actually reporting their economic activities to the authorities and only concealing part of their transactions, are exposed to higher risks than fully informal businesses. This is so, because fully informal enterprises do not bear the cost of doing business according to the rules, which is being paid by the registered business although not fully tax compliant. In potential detection scenario, a partly informal business, which benefited only marginally from their tax evasive behaviour could be penalised similarly to fully informal enterprise, which benefited to maximum effect due to the complete concealment of turnover. Therefore, the incentive for partly informal businesses to remain undetected is arguably much stronger.

The best known institutionally implemented direct survey on international scale with modules on informality, taxes and regulations simultaneously with the enterprise as a unit of analysis is the one developed by the International Finance Corporation (2013) as part of the World Bank group. The available micro data (survey series for Bulgaria) includes 2002, 2005 and 2012-2013. The International Labour Organization also undertakes the so called "mixed household and enterprise surveys" (International Labour Organization, 2012, p. 60), which are "designed to obtain detailed information on the characteristics of the owners of informal sector units, of their enterprises and of the persons working with them, if any". Both international organisations employ gradual approach in their questionnaire where there are a certain number of non-contentious profiling questions gradually escalating into sets of questions intended to detect and evaluate informality. As a result of the pilot cognitive testing of the survey in Bulgaria, it was discovered that the gradual questionnaire approach delivered the anticipated level of informal activities disclosure. This provided rationale for this particular design choice in the final survey deployment.

The Business Customer and Strategy unit as part of HM Revenue and Customs UK, conducted an extensive evaluation of quantitative methods for analysis of the hidden economy. They came to the conclusion that direct methods are the most favoured ones by experts. Based on these findings HMRC (2012, p. 3) considered possible and feasible to design their own business survey method in order to capture meaningful insights on:

- "Hypotheses as to why businesses enter, and continue to operate in, the hidden economy
- The motivations and wider social and societal drivers which drive some into the Hidden Economy
- The business strategies and processes they adopt to remain hidden
- What levers might be successful in encouraging businesses back into the formal economy"

The exhaustive appraisal of quantitative methods in other relevant studies is important in two aspects. Firstly to review and benefit from their investigative experience of such sensitive topic. Secondly, to adapt and incorporate lines of inquiry, which are relevant to the research context and the research aims. Tables 4 and 5 are partial expression of this effort.

Another highly relevant survey (Fries, Lysenko & Polanec, EBRD, 2003), using the enterprise as a unit of analysis in transition economies, was conducted by the European Bank for Reconstruction and Development (EBRD) and the World Bank to investigate the extent to which government policies and practices facilitate or impede business activity. The first round of Business Environment and Enterprise Performance Survey (BEEPS) took place during 1999 in Central and Eastern Europe and the Commonwealth of the Independent States – it covered 4,104 firms (excluding farms) in 25 countries. The second survey round was performed in 2002 amongst

6,153 firms in 26 countries of the region. The analysis of quantitative measures of the business environment indicates strong correlation with business obstacles, added costs and constraints due to inefficient state governance. Based on these regionally relevant findings it could be argued that the high cost of doing business due to the inadequate business environment is a significant driver for informality amongst enterprises. This hypothesis is explored in greater detail and confirmed valid for the Bulgarian micro and small business environment.

The Rockwool Foundation Research Unit has been using surveys on the extent of the black economy in Denmark since the end of the 1980s. During the nineties, the Rockwool Foundation also took interest in their neighbouring countries Sweden and Norway. Later on it was decided to extend this work for comparative purposes to Germany and Great Britain (Pedersen, 2003) addressing one of the main methodological difficulties in comparative measurement – the definitions of the observed phenomena and the way the questions have to be formulated. It was discovered that although the same definitions and questions' formulation were used, this actually led to some variations of the respondents' understanding and answers. This was due to the country specific contexts, nevertheless their similar West European economic profiles. However, this study is one of the few to contribute to cross-border comparison through adjusting the questions to fit various countries' tax legislation and techniques of adding questions about activities which are taxable in one country but not in another.

The next section will describe the sampling method, sample size and survey distribution. A detailed survey justification is presented afterwards – the implications of the gradual versus direct approach are also discussed. The relevance of the survey's sections to the research aims of the thesis are illustratively presented in a table. The data collection rationale behind each question and its interconnectivity with other questions are in the business survey's analytical matrix, which follows. Finally, the implications from the pilot cognitive testing are then discussed.

Sampling method, sample size and survey distribution.

The companies' contact data was acquired through one of the two most reputable commercial data providers in Bulgaria: Ciela. Their software product Ciela Info combines information in real time from the various state registers – the commercial register, BULSTAT register, data from the registers of the tax administration, data from State Gazette, as well as information provided by Ciela research teams. The data integration and cross verification from various official and unofficial sources makes for the most exhaustive source on business data available.

The following table represents the characteristics of the targeted population of businesses. As of 30.04.15 the entire population consisted of 76 298 entities. This includes all types of "for profit" registered businesses.

Enterprise characteristics	**Criteria**
Legal form	all types of "for profit" registered businesses
Place of business registration	Sofia
Date of registration	from 30.04.2005 to 30.04.2015
Socially insured persons	1 – 50
Health insured persons	0 – 50
Business status	Active
Turn over	0 – 500 000
Capital	0 – 100 000

The business registration period is from 30.04.2005 to 30.04.2015 (start of the survey), which means that the longest operating business from the population functioned for up to 10 years maximum. The longer a business operates formally registered the less likely it would be involved in undeclared activities. This selection criterion was introduced to control for businesses, which are not established for longer than 10 years and therefore could be characterised by significantly reduced possibility to be involved in off-the books transactions. The study aims to collect insights from within the partly informal enterprises and hence should attempt to collect data from as many as possible. The business status criterion excludes businesses, which for some reasons declared cease of economic activity – this is to ensure that all population members are currently operationally active enterprises. The place of business registration is selected to be Sofia (reasoning is available on the page below).

The European Commission (2014) introduced an upgraded definition for micro, small and medium-sized enterprises (SMEs), which represents 99% of all businesses in the EU. For the purpose of the study the survey was distributed to micro and small businesses only where the level of informality could be expected to be higher (Williams & Martinez, 2014b; Williams, Nadin, Newton, Rodgers & Windebank, 2011). However, the turnover criterion was intentionally violated (explanation after the table). According to the EC definition (2014) the following ceilings apply:

Company category	**Employees**	**Turnover**	or	**Balance sheet total**
Medium-sized	< 250	≤ € 50 m		≤ € 43 m
Small	< 50	≤ € 10 m		≤ € 10 m
Micro	< 10	≤ € 2 m		≤ € 2 m

As it is evident from the economic characteristics of the targeted population in Bulgaria, the maximum turnover allowed was 500 000 BGN or approximately 250 000 EUR. The number of staff remains as specified by the EC. This selective change was intentionally implemented to account for the fact that Bulgaria is significantly less developed member of the EU and its micro, small and medium-sized businesses have proportionally smaller turnovers. Furthermore, due to relatively cheap labour costs, Bulgarian SMEs tend to utilise a bigger workforce for a given profitability index.

In order to estimate the adjusted sample size (confidence level of 95%, margin of error 5% and 50% response distribution), the following steps were performed:

1. It was assumed that the response rate would be approximately 15% (85% – refusal to participate, unreachable, break-off, partial response or ineligible to participate). Response rates for web based surveys vary between 7 and 44% (LaRose & Tsai, 2014, p. 20; Schonlau, Fricker & Elliott, 2002). In consideration of the sensitive study topic, it could be argued that 15% is realistic expectancy.

2. A population of 76 298 needs a minimum sample size of 383 valid responses to fulfil the above requirements. Having more than 300 responses ensures that “test parameters tend to be stable regardless of the participant to variable ratio” according to Kass and Tinsley (1979) in (Field, 2009, p. 647).

3. Assuming conservative 15% response rate, the adjusted sample size has to be 2555. $\mathbf{n}^{\mathbf{a}} = \mathbf{n} / \mathbf{re}\%$ (Saunders et al., 2012, p. 266-269)

where:

n^a	is the actual sample size required
n	is the adjusted minimum sample size to achieve 5% margin error, 95% confidence level
re%	is the estimated response rate expressed as a percentage, in this case = 15%.

Qualtrics' systematic random sampling procedure was used over the entire population in order to generate the adjusted sample for the survey deployment. This decision was made based on the following decision tree (Saunders et al., 2012, p. 271):

Question	**Answer**
Must statistical inferences be made from the sample?	Yes
Must the sample represent the population?	Yes
Does the research require face-to-face contact?	Yes and No
Is the population geographically concentrated?	Yes
Does the sampling frame contain relevant strata?	No
Does the sampling frame contain periodic patterns?	No

All prospective participants were presented with the opportunity to be interviewed face to face if they so wish or if they want to meet up the principal investigator for any reason – such as additional information acquisition, clarification and/or identity verification. The principal investigator was physically present in Sofia for the entire data collection period to satisfy contact requests and to be able to deal with the survey effectively. In consideration of the scale of the required responses, options for face to face contact with the respondents and the limitations of the empirical collection capabilities within this PhD study, it was not possible to make it a nation-wide survey. In the given circumstances, a choice of one city was required. The capital of Bulgaria was selected for the following reasons:

- It was assumed that the informal strategies would be more sophisticated (therefore more valuable to study) in comparison with similar ones employed by SMEs outside of the capital. This assumption rests on the fact that institutional compliance pressure is higher on businesses operating from within the capital. In addition, general competition is stronger. Therefore, SMEs have to be more creative and secretive if they employ informal strategies – enhance their competitiveness and remain undetected.
- It was assumed that SMEs in the capital have generally better conditions for doing business and potentially are expected to sustain a lower level of partial informality in comparison with other parts of the country. As a result the economic motives behind evasive practices were expected to be more meaningful for analysis and policy recommendations. According to the macro-economic methods, the Bulgarian informal economy is the biggest in the EU. It was hypothesized that the actual share of the informal economy is much higher than what is currently known to be the case and it was attempted to measure it (through the survey) where it should be least pronounced – SMEs in the capital. That would have provided good comparative grounds on the basis of extrapolation – if SMEs in the capital exercise X% of turnover underreporting on average, then SMEs from the rest of the country are likely to exhibit Y% of turnover underreporting, where Y>X.

The survey was deployed to the business owners (entrepreneurs) and they have been requested to fill the online questionnaire themselves or by a business manager designated by the owner to do so, on their behalf. The survey invitation contained the so called information sheet and information consent as per the ethics policy of University of Sheffield. The information sheet provides answers to several

standardised questions in regards with the research involved including extensive reasoning on anonymity and purpose of data collection.

4.4.2 Questionnaire development and rationale.

This section presents the logic behind the survey development as an empirical instrument and in particular its role within the thesis. The connection between the questionnaire's building blocks and the research aims is mapped in a table format as well as the rationale behind each individual question. The advantages of the gradual information acquisition approach are discussed and the technicalities of pilot cognitive testing. The provided level of detail is intended to equip the reader with the background knowledge for interpretation of the quantitative analysis and discussion of findings accordingly.

Having adopted the institutional asymmetry paradigm, this study aimed to expose and evaluate the asymmetries between the formal and informal institutional frameworks. The first investigative level appraised the formal institutions with a specific focus on tax management policy, revealing the major environmental pressures (institutional, economic and political) on NRA. The second investigative level has been designed to research the strongest determinants of business tax evasion, emerging from the dynamics between one of the most challenging socio-economic environments in the EU and the small firms operating from within. Given the recent development of IAP as a contemporary theorisation of informality, the majority of research (Williams et al., 2015a; Williams & Horodnic, 2015b; Williams & Horodnic, 2016; Williams et al., 2015d) utilises existing data sets such as Eurobarometer. This study contributes with specifically developed empirical instruments (in-depth interviews and a business survey) to collect and analyse primary data at both investigative levels.

IAP poses a specific empirical challenge – indeed, how to measure empirically the intrinsic circularity between tangible and intangible phenomena, affecting tax evasive behaviours. To clarify this aspect, it should be noted that formal institutional quality could be decomposed to several intangible properties: institutional reputation amongst tax payers, lack of corruption, procedural fairness, distributive justice, vertical trust. Tables 4 and 5 provide detailed operationalisation of the quantitative research objectives, explaining how each survey question relates to the extant literature and what its purpose is to deliver in regards with new IAP insights. These two tables also explain how particular groups of questions function as meaningful investigative sets. As evident (see sections 6.2.2, 7.2), this thesis shows that intangible formal institutional characteristics play strong role in determination of the diversely motivated business tax evasion and therefore have tangible and circular socio-economic impact.

The questionnaire development and design is based on four main pillars:

1. Utilisation of NRA's extensive experience on detecting and diagnosing informal business practices. As every revenue agency, NRA developed a special methodology to detect with high probability when an enterprise could be involved in evasive practices. It is based on combination of regionally specific profiling indicators. They are developed and updated in Risk Management directorate with contributions from other directorates too (mainly Control). The information access requests (see Appendix 10 and 11 described under Empirical Strategy heading) granted partial access to the methodologies employed in the development of these indicators. Consequentially, they have been embedded in the questionnaire for the same purpose – to detect informality amongst the recruited enterprises.

2. Review of the established survey methods in evaluation of the hidden economy with a focus on these ones, which study the enterprise as a unit of analysis. Particular questions have been extrapolated and adjusted where necessary to the specific regional context.

3. Development and use of own questions in combination with the above to supplement the data acquisition and verification process.

4. Use of the so called branching, skip and display logic in order to profile the participants correctly and expose them only to circumstantially relevant questions, based on the answers provided.

Gradual versus direct approach.

The business survey utilised gradual, but not direct approach in terms of how the questions have been ordered and conveyed, based on the benefits revealed by previous studies on similar sensitive topics. The main advantage of the gradual approach is in building rapport with the respondent and gaining their confidence in terms of asking less sensitive questions first. The participants emerge into the practise of providing answers to the introductory questions and are then gradually taken into more sensitive topics. It could be argued that respondents develop a level of psychological engagement, trust and advancement in the questionnaire, so it becomes more difficult for them to leave a later question unanswered or simply drop-off. In a study of the structure and magnitude of the hidden labour market, implemented by the Netherland Central Bureau of Statistics (Kazemier et al., 1992), eight survey variants were thoroughly appraised in terms of suitability to obtain valid information. The gradual approach in the questions presentation was found to acquire the highest participation rates. Later studies on the same topic (Kazemier, 2014, p. 39) were successfully utilising the gradual approach and also signified that "The outcomes do not differ between face-to-face and online surveys". Comparing the Eurobarometer survey's results with other relevant studies, Williams (2015a, p. 7) presents evidence in support of the gradual approach's capability to elicit higher response rates due to lengthier priming questions instead of a shorter lead-in.

Utilising the gradual approach in the current business survey, it consisted of 8 blocks, containing 31 questions in total. The first survey block is the introduction and informed consent, while block 8 is the end of the survey with the "Thank you" page. Effectively, the actual data collection is carried out in blocks 2 to 7 inclusive. Appendix 18 represents the entire questionnaire. It is important to note that neither the block names nor the branching logic was available to the participants – they have been exclusively created for the purpose of the study and its audience. Survey block 2 contains general profiling questions and therefore serves the gradual approach purpose – to lead-in participants to more particular questions later in the questionnaire. The questions from the block sound very insensitive to participants, however most of them are actually NRA's informality indicators – they are designed to detect high probability of informality involvement amongst SMEs.

Table 4: Relevance of the survey to the research aims and questionnaire's rationale.
The table below briefly describes the connection of each survey section to the research aims of the thesis.

Block 1 – Introduction and informed consent.	
Block 2 – General business profile and initial informality indicators.	Main purpose is to profile the business and diagnose probability for informal economic activity with special lead-in questions. Use of NRA's informality detection questions in combination with some World Bank Enterprise survey's questions.
Block 3 – Institutional quality and fair use of public money. Procedural fairness and distributive justice.	Provides opportunity to voice dissatisfaction with the institutional quality and NRA specifically, their service to businesses and the way public (tax) money are being spent. Proactive, engagement of the respondent with the topic and good purpose of the survey – data collection towards a positive change.
Block 4 – Diagnostics of informality, motivations and perceptions.	Investigates the economic motives for a given level of informality. Detection and evaluation of turnover underreporting. Cost-benefit analysis for the reasons for informality. Studies the tolerance towards tax evasion and bribery and their interdependence.
Block 5 – Interdependence of informal businesses (informality chain of interdependence).	Detection of informality chains of interdependence – economic barriers to formalisation.
Block 6 – External business environment and cost of doing business. Taxation, regulations and licensing frameworks.	What level of taxation/licensing is acceptable by the entrepreneurs. Attempt to cost the entrepreneur's time – institutional cost of doing business. Perceived level for customer service and policy approach by NRA – from the perspective of the entrepreneur.
Block 7 – Unfair competition from informal businesses.	Evaluation of unfair competition from informal businesses and cross-detection of informal economic activities.
Block 8 – End of survey – free text box.	

Table 5: Analytical survey matrix.
The following table describes the main purpose of each question and serves as an analytical survey matrix. Secondly, it shows the interconnectivity with other questions from the survey and thus reveals the data collection strategy in the light of the gradual approach. Thirdly, any relevance to other similar studies is indicated.

Question number	Main purpose	Interconnectivity with other questions	Relevance to other studies
1.	Acquire an informed consent.	Not applicable	Not applicable
2.	Business subject profiling function. In addition, could contribute to detection of probability for informality together with other questions. For instance a business trading with goods and not using an office/warehouse. At the same time having an outstanding debt and not undertaking collection activities.	Questions 3, 5, 9 and 10.	(IFC, 2013)
3.	Activity profiling question – part of the 6 NRA's indicators to detect high risk of informality.	Questions 2, 5, 9 and 10.	NRA methodology P.5.1
4.	Business profiling question and informality indicator. Detects use of undeclared labour, envelope wages and consequential tax evasion – envelope wages could not be claimed as business expenses. The reported number of full and part time staff could be compared with the officially reported number for a particular business available from government registers.	Not applicable	Similar questions are available in most subject related business surveys.
5.	Business profiling question – part of the 6 NRA's indicators to detect high risk of informality.	Questions 2, 3, 9 and 10.	NRA methodology P.5.1
6.	Business profiling question – supportive to NRA's indicators to detect high risk of informality.	Questions 2, 3, 5, 9 and 10.	NRA methodology P.5.1

Question number	Main purpose	Interconnectivity with other questions	Relevance to other studies
7.	Start-up informality indicator – evaluation of whether small business start-ups test-trade in the informal economy. Supplementary to informality risk detection in conjunction with NRA's indicators.	Questions 2, 3, 5, 9 and 10.	(Williams et al., 2014b) (International Finance Corporation, 2014, p. 52)
8.	Conditional business profiling question – reported year of business formalisation could be compared to official state registers and potential discrepancies could be appraised together with finding from question 7.	Question 7	(IFC, 2013, p. 4)
9.	Economic activity profiling question – part of the 6 NRA's indicators to detect high risk of informality. Finding out the level of outstanding debt and what collection actions (question 10) are being performed. Lack of adequate debt collection in relation with debt's size while trading predominantly with goods, could indicate informal chain of dependence.	Questions 2, 3, 5, and 10.	NRA methodology P.5.1
10.	Economic activity profiling question – part of the 6 NRA's indicators to detect high risk of informality. Works together with the previous question.	Questions 2, 3, 5 and 9.	NRA methodology P.5.1
11.	Economic activity profiling question – part of the 6 NRA's indicators to detect high risk of informality or tax fraud. There are two main points: fully formal businesses could rarely function without external forms of finance and secondly, using external funding with significantly different interest rates (to market ones) is an indication for tax fraud (question 12 below).	Questions 2, 3, 5, 9 and 10.	NRA methodology P.5

Question number	Main purpose	Interconnectivity with other questions	Relevance to other studies
12.	Economic activity profiling question – conditional of question 11.	Questions 10 and 11.	NRA methodology P.5
13.	This is a key question evaluating perceptions for procedural fairness and distributive justice – fairness of tax administration and public expenditures. Have to be discussed in terms of tax morale and tax compliance in similar cultural contexts. Together with question 21 and 24 it provides insights on institutional asymmetry.	Questions 14, 21 and 24.	(Cummings, Martinez-Vazquez, McKee & Torgler, 2006) (Williams et al., 2015a; Williams, Franic & Dzhekova, 2015c; Williams et al., 2015d; Williams & Shahid, 2014c; Williams et al., 2014d)
14.	Key informality profiling question evaluating the economic motives for informality, but also institutional asymmetry determinants.	Questions 14, 21 and 24.	(Putnins & Sauka, 2014a, p. 35) (Putnins & Sauka, 2014b)
15.	Key informality profiling question – engaging in an unofficial payments to “get things done” almost inevitably signifies partly informal practices. It could be argued that the turnover underreporting is at least equal to bribes paid.	Questions 15, 16, 17, 18, 19 and 20.	(Putnins et al., 2014a, p. 32)
16.	Key informality profiling question – to diagnose turnover underreporting (tax evasion) including the motives and attitudes behind.	Questions 15, 17, 18, 19 and 20.	(Putnins et al., 2014a, p. 35)
17.	Key informality profiling question – measurement of turnover underreporting (tax evasion). Works in collaboration with question 16.	Questions 15, 16, 18, 19 and 20.	(Putnins et al., 2014a, p. 31)

Question number	Main purpose	Interconnectivity with other questions	Relevance to other studies
18.	Informality profiling question – part of the 6 NRA's indicators. Detect informality chain of interdependence and attitude to formalisation together with question 19 and 20. Businesses who are acquiring part of the products/services they need without invoices are actually forced to either accept such net loss (to the reported value) or achieve similar level of informality to offset the costs.	Questions 19 and 20.	NRA methodology P.5
19.	Works together with question 18 and 20 – see the description above.	Questions 18 and 20.	In support of NRA methodology P.5.
20.	Works together with question 18 and 19 – see the description of question 18.	Questions 18 and 19.	In support of NRA methodology P.5.
21.	Evaluates perception of NRA's institutional quality and its effect on the ease of doing business. Together with the interconnected questions appraise attitudes towards corruption.	Questions 14, 24, 25, 26, 27, 28.	(International Finance Corporation, 2014, p. 16) (Putnins et al., 2014a, p. 30)
22.	Evaluates perceptions of the tax and social security burden on the cost of doing business.	Questions 23 and 24.	(International Finance Corporation, 2014, p. 15)
23.	Evaluates acceptable/desired level of taxation. Therefore, assumptions could be made about potential risk of tax evasion/informality in comparison with the actual tax burden.	Question 22.	
24.	Evaluation of external business environment constraints – indirect institutional costs of doing business.	Questions 22, 23, 25, 27, 28 and 29.	(International Finance Corporation, 2014, p. 147)

Question number	Main purpose	Interconnectivity with other questions	Relevance to other studies
25.	In combination with question 26, it attempts to cost enterprises' time dealing with taxation, regulations and licensing frameworks.	Question 26.	(IFC, 2013)
26.	In combination with question 25, it attempts to cost enterprises' time dealing with taxation, regulations and licensing frameworks.	Question 25.	(IFC, 2013)
27.	Examines relationship with NRA and potential institutional pressure for bribery. Works together with question 28 and 29.	Questions 28 and 29.	
28.	Evaluates cost of doing business due to NRA interventions.	Question 26, 27 and 29.	
29.	Examines relationship with NRA and potential institutional pressure for bribery. Works together with question 28 and 29.	Question 27 and 28.	
30.	It detects unfair competition from informal businesses. This question works together with question 31 and provides insights for economic barriers to formalisation.	Question 31.	(IFC, 2013) (Williams, 2006)
31.	Evaluation of unfair competition from informal businesses. This question supplements question 30 and provides insights for economic barriers to formalisation.	Question 30.	In support of (IFC, 2013)

Pilot cognitive testing

One of the major challenges in unattended data collection methods (where there is no interviewer present) is the cognitive understanding of the actual questions' meaning. It is sufficiently challenging to translate the research aims into meaningful questions with avoidance of overly-complex terminologies. However, it is even more difficult to ensure that the average cognitive understanding of the questionnaire is similar to majority of the respondents. The questions' clarity, difficulty and assumptions being raised have to be designed in a way that the targeted audience see them in very uniform way. Even so, the most important factor in sensitive topics' research (such as the informal economy) is the cultural alignment of the questionnaire – what could be asked and in what way for a given cultural context, so that the non-answer and drop-out rate is kept at minimum. Phrasing a question about underreporting of business turnover in Eastern Europe has to be without doubt different in comparison with Western Europe in order to secure appropriate and relevant engagement. Very few studies report any pre-testing process and results accordingly and until recently there was little methodological research on pretesting available (Presser, 2004). To improve on this aspect of data acquisition, a cognitive testing approach was employed.

Because of the specific role of this phase of instrument development, the interviewer had to be present face-to-face at all times of conducting the pilot cognitive testing. All pilot participants were in compliance with the pre-determined economic profile. There were the following reasons for pilot testing the questionnaire:

- Preliminary evaluation of the questions to avoid any misunderstanding or misinterpretation by the targeted audience.
- Appraisal of the gradual approach in terms of ability to acquire insights about evasive activities.
- Assessment of the preferred data collection (telephone, postal, face to face or internet survey).

The survey was pilot tested amongst 10 entrepreneurs recruited via a snowball sampling technique. All 10 of them were known to be engaged in partly informal mode of doing their business – about 30-50% of their turnover was concealed from the authorities. The snowball sample started with 3 entrepreneurs, who then were able to assist with information access to some of their business partners with whom they have formed the so called informal chain of dependence – mutual strategic double accounting techniques for the purpose of tax evasion with limited probability of detection. Having been through the entire questionnaire, all entrepreneurs were asked whether they would have preferred an alternative interview technique such as the same questionnaire deployed via post, telephone, Internet or face-to-face. The main concern, which the respondents have raised, was the strict confidentiality they all required in order to agree to participate. The interviewed entrepreneurs were presented with the following choices and asked to informally discuss the best option from their perspective.

- On-site face-to-face interview.
- Telephone interview.
- Postal questionnaire.
- Internet based survey on University of Sheffield branded web address of the type (computer and mobile device compatible): https://sheffieldmanagement.eu.qualtrics.com

The following points and insights were acquired:

- The interviewees wanted extensive and verifiable guarantee that the findings of this research will be exclusively used for academic purposes. They somehow distinguished the academic research from other types of research and did not feel threatened if this point could be proved adequately.

- All of them expressed preference to be able to fill the questionnaire in the best possible time and most of the time it would be out of business hours and out of the business premises. Later results from the survey indicated exactly that – majority of the responses were collected after outside usual business hours, not insignificant part of which during unsocial hours and over more than one session per completion (the survey not finished at once, but during an additional session at a later time).

- Nobody considered the telephone option due the high amount of prank and marketing calls and inability to verify the identity of the interviewer. The respondents said it is highly likely they would deter such attempts to be interviewed and also most probably would ignore the postal questionnaire.

- The internet based survey gathered the biggest support due to satisfying majority of the expressed requirements, but most importantly proving the pure academic character of this research and related strict ethical procedure adherence (explained in the information sheet). The respondents felt most comfortable and confident in the internet based version due to the verifiable internet address and explicit association with University of Sheffield.

Qualtrics is one of the world-wide leaders in the provision of contemporary tools for marketing and academic research and was selected for the purpose of the business survey of this study. In conjunction with the agreement between Qualtrics and University of Sheffield, internet based surveys could be deployed using verifiable and trackable institutionally possessed web domains. This was perceived by the respondents as an assurance that the research in question is associated with a renowned academic institution. The choice of survey deployment was therefore considered most likely to generate the highest response rate as a result of the additional credibility in front of prospective participants and its associated functional convenience.

4.5 **Predictive analytics.**

This heading (Predictive analytics) is designed to communicate the advantages of the utilised analytical tools and therefore justify their use for the purpose of this study. It starts with brief methods discussion and technical explanation of the information acquisition functionality of the survey. The next sub-sections include contextual introduction and applicability of decision trees and neural networks as appropriate analytical tools. Introduction to predictive modelling, its main terms, definitions and the modelling workflow framework are then presented, so that the reader is driven to sufficiently similar investigative stance as the author. The essence of the methods' application and the emerging findings are presented in chapter 6 (section 6.2).

Purely statistical methods would not be sufficient to reveal all multi-layered relationships within the acquired data. Fitting statistical models to the data and evaluating the statistical significance as an outcome depends on the knowledge of the researcher of the expected effects and/or findings. The researcher has to form a theory about a relationship, convert it to a hypothesis and then test it with a statistical

model. The purely statistical mode of analysis means that the investigator tests hypotheses of some prediction – the alternative hypothesis or the null hypothesis. The alternative hypothesis cannot be proved using statistics, but the rejection of the null hypothesis provides support for it. Based on the lack or presence of an effect, we are led into potential rejection of the null hypothesis, which consequentially provides support for the alternative hypothesis. It could be argued that one can fit a large number of statistical models to the data even there, where they do not expect any meaningful insights and thus actually discover some, which deserve attention. Evidently, this is a manual deductive approach. However, this imposes the risk of non-discovery of statistically significant effects, which obviously translates into loss of meaningful research explanations. The more complicated the research data is the higher the risk of missing on significant relationships and findings becomes. If the research data contains some unusual or unexpected findings, which contradict or at least do not conform to current scientific knowledge in the area, then classic statistical methods may not be enough to expose them. Or at least the researcher is open to failure in their relevant application while they are trying to discover them. Obviously, testing research theories through decomposition to statistically testable hypotheses is to some degree dependent on the researcher's scientific bias, correlated to their particular knowledge of the studied phenomena.

A more advanced analytical approach is required to put to work in order to surmount such deficiencies. A dual mode analytical inquiry into the empirical data combines predictive analytics (data mining) techniques with classic statistical models. In other words, this is a specialised interrogation of the research data in order to discover relationships from which theories could be built and then tested with classic statistical methods where necessary. It could be argued that this is an abductive analytical approach – an amalgamation between the inductive capabilities of data mining (predictive analytics) with the deductive nature of classic statistical models.

In general, advanced data mining techniques are based on mathematical algorithms (linear and non-linear) which are packaged in artificial intelligence software solutions. The main strength comes from the fact that the software solution, but not the researcher, identifies any important features and relations of the data. Any exposed insights are then scrutinised by the knowledge of the investigator and adjustments are made where necessary into the data mining strategies and techniques. The advantages of this approach are self-explanatory – maximum robustness and objectivity where the expert's knowledge is combined with active analysis techniques supported by artificial (computer based) intelligence. Independently from the software solution used, there are two major stages in data mining with several product specific sub-stages. The two major steps are:

1. Machine learning or the so called modelling. This is where various data mining models are deployed into the data and their settings are being tuned until the models achieve maximum precision. This is the stage of training the model(s) using the available data. The tuning of the data mining models requires the expert knowledge of the researcher in order to facilitate the strength of the machine learning capabilities.
2. Deployment of the trained models into the real world (or in new data sets) in order to predict outcomes from similar or completely different variable inputs. This is the so called predictive analytics aspect of data mining.

IBM SPSS Modeler Premium was employed as one of the two software solutions in the qualitative analysis phase of this thesis. At that stage of investigation, its natural language processing capabilities were used in combination with text analytics streams and models to analyse semi-structured data (the in-depth interviews). IBM SPSS Modeler is currently the industry benchmark, which establishes various important

standards for data analysis, including the Cross-Industry Standard Process for Data Mining (see Figure 8 for CRISP-DM). As explained earlier in the relevant chapter, that is the industry-proven way to guide data mining efforts and it is spanned over several semi-structured stages. There is a certain level of flexibility in terms of approaching and conducting this process, but it is limited to ensure that the analyst follows a benchmark industry standard. The quantitative analysis stage relies on IBM SPSS Modeler 18 Premium and IBM SPSS Statistics 23 Premium with the advanced modelling and analytical options available in these versions accordingly. Both products have been employed in dual and/or sequential mode so that the validity and robustness of findings are at their best.

One of the viable and easy to demonstrate descriptors of the complexity of the acquired research data is the economic profiling function of the business survey questionnaire. In the context of the impending quantitative analysis, it is worthwhile to remind the reader about the capabilities of the questionnaire to adapt dynamically to a given respondent's profile. This is achieved with the so called display logic questions, which are circumstantially relevant questions, based on previous answers provided. Therefore, the survey instrument allows for creating substantial amount of different economic profiles for respondents with significant precision. This is actually achieved via real time automated construction of the questionnaire in front of the respondent based on their answers as they advance through it. There are 2048 functionally possible variants of the questionnaire. The explanation of how this is technically achieved follows below.

The survey questionnaire consists of 19 constant questions and 11 display logic ones. In addition, there is an introduction question responsible for the acquisition of informed consent. The 19 constant questions were designed to appear in front of all respondents independently from their previous replies. The 11 display logic questions were developed to appear only in specific circumstances, based on conditional logic of the type: if a respondent provides a particular answer to a previous question, then they must see a relevant display logic question. To serve their purpose, the display logic questions have been staged at 11 specific points throughout the questionnaire and they have two functional modes of existence – display or do not display condition (see Appendix 18, which shows the embedded display logic in detail). Having in mind the binary mode of appearance of every display logic question, the possible variants are represented by the following mathematical expression: $2^{11} = 2048$. This is the symbolic expression of the binary nature of the display logic questions, with the intention of demonstrating how the number of all possible combinations is derived.

Display question 1	Display question 2	Display question 3	And so up to 11
Display = 0	Display = 0	Display = 0	Display = 0
Do not display = 1	Do not display = 1	Do not display = 1	Do not display = 1

Number 2 in the expression ($2^{11} = 2048$) accounts for the binary nature of display logic questions and the grade, which on this occasion is 11, represents all possible combinations between the display logic questions. Evidently, this design of the survey allows for great theoretical flexibility in profiling respondents via 2048 different versions of the questionnaire. In other words, the survey may create 2048 different tools for economic profiling of entrepreneurs. Having in mind the 413 responses in total, this means that a maximum of 413 variants has been exhausted out of 2048, but only if all utilised paths through the questionnaire are truly unique. It is unlikely to be so and most probably, the actual number is smaller through some similar patterns. The shortest path through the survey is for a respondent to face only 19 constant questions and neither one from the 11 possible display logic ones. The longest path through the survey is if a respondent faces all 19 constant question and based on the

provided replies trigger all of the 11 display logic questions. This is equal to answering all 30 possible questions. These two described combinations are obviously the two extremes from all possible versions of the questionnaire. This aspect of the business survey demonstrates the actual complexity of the empirical information acquired, the possible interconnectivity and potential for significant correlations between variables.

4.5.1 Decision trees – a brief contextual introduction and applicability.

The described level of data complexity requires advanced analytical techniques in order to generate reliable models, which could be used for predictive deployment into the real world. The major purpose of the business survey is to provide structured data to serve the modelling of economic behavioural patterns of entrepreneurs, which could be consequentially used by policy makers for precision interventions. One way of achieving this purpose is through the so called decision trees. They are based on different algorithms ranging from popular statistical tests such as chi-square and regressions to combinations of linear and non-linear mathematical modelling via decision forests (forests of decision trees). A fundamental benefit to decision and policy makers is the fact that all types of decision tree algorithms provide some form of significance of their conditional/branching structure. Their approach, which is similar to rule induction, is always visible in terms of reasoning – it is available from visually browsing the tree. In addition, the decision tree algorithms include only predictor variables, which contribute to the accuracy of the model and all others are excluded. As a result, they are invaluable instruments in appraising how a particular change of a certain condition will affect the dependant ones. This functionality will become evident via the demonstrated applications in the coming sections.

The main objective of a decision tree is to expose a distinct pattern of characteristics and/or behaviour with the conditional support of a rule structure from the type "If/Then". This is achieved via the logical branches of the tree and the associated rule set, which could be derived. The main building components of a decision tree are a number of antecedents (predictor variables) and a consequent (the target variable). In cases where a statistical algorithm supports the growing or in other words the branching of a decision tree, it is its significance, which ensures adequacy of the model in relation to the available data. In general, decision trees may have a hierarchal visual structure and/or purely conditional expression as a rule set. Depending on the type of analysis, both expressions could be utilised to present a problem-solution configuration.

As a powerful instrument of analysis, decision trees are popular in many disciplines including structuring policy problems (Dunn, 2012, p. 25). It could be argued that their particular usefulness in that respect is due to their, as described by Rokach and Maimon (2008, p. 3), properties of discovery-orientated approaches in contrast with verification-oriented methods such as classic statistical methods. In comparison with the extensively used logistic regression, evidence suggests that decision trees retain superior accuracy even when research data becomes less accurate (Coussement, Van den Bossche & De Bock, 2014, p. 2752). However, decision trees have some well-known drawbacks such as "over-sensitivity to the training set, to irrelevant attributes and to noise" (Rokach, 2016, p. 114). There are circumvention methods to compensate for such weaknesses (like boosting, cross-validation, bagging), but the most powerful and perhaps advanced solution is through creation of decision forests.
After extensive evaluation of various tree algorithms, it has been discovered that Exhaustive Chi-squared Automatic Interaction Detector (E-CHAID) trees, Classification and Regression Trees (C&RT) and C5 trees are best suited to the type of empirical information available. The evaluation process was conducted according to IBM CRISP-DM standards and established the following major considerations:

- The characteristics of the data satisfy the assumptions of the described models.

- The above mentioned tree algorithms tend to produce models with the highest meaningfulness and deployability. The predictive analytics power is the highest in comparison with other similar algorithms.

Two important properties of decision trees are further supporting the rationale for their use in the current analysis. Firstly, their automated interaction identification capabilities – this allows detections of significant relationships between variables, which is a powerful dimension reduction technique. The so identified relationships could be further explored (if necessary) in formal parametric models. It could be argued that in highly sophisticated data sets such as the one used here, the knowledge of the researcher greatly benefits from such analytical enhancing techniques. Secondly and most importantly, the prediction capabilities of decision trees which allow forecasting of future behavioural patterns. The following relevant examples are offered as target deployment scenarios for decision trees:

- If a segment of micro and small businesses tend to evade taxes for purely economic survival reasons how would changing aspects of their business environment do affect evasive patterns. Policy makers will need to know how impactful their interventions would be in order to even consider potential implementation.

- If a segment of micro and small firms operates in partly informal mode of doing business due to a mixture of unfair competition, poor institutional quality and being part of informal chain of dependence, then what type of policy efforts would have the greatest positive impact towards a higher level of formalisation.

The following paragraphs provide a brief discussion about the most appropriate tree algorithms. The CHAID algorithm was originally developed by Kass (1980). It is a very powerful statistical model of tree growing based on statistical tests (chi-square or F-test) as a criterion. The algorithm depends on a built-in significance testing with the effect of using the most significant predictor (rather than the most explanatory). This is particularly useful for identification of unexpected knowledge patterns and hence principally useful to the current research aims. Multi-way splits (in contrast to binary) or more than two branches for some splits are also possible. This allows for exhaustive descriptive logic of explanation, which is another important feature required. As a higher algorithm level, the so called exhaustive CHAID was developed later by Biggs, De Ville and Suen (1991, p. 49) to eliminate some of the weaknesses of CHAID. “Unlike Kass's algorithm, the algorithm does not favour simple partitions (low values of k) nor does it discriminate against free-type (no restriction on order of values) predictor variables with many categories.” E-CHAID can find the best split for each predictor variable through examination of all possible splits, and then choose which predictor to split on by comparing the adjusted p-values (IBM, 2015a, p. 79). In the context of all possible versions of the questionnaire and the complex predominant categorical structure of the data, this becomes highly relevant functionality.
Classification and Regression Trees (C&RT) were originally developed in depth in the monography of Breiman, Friedman, Stone and Olshen (1984). These algorithms are quite robust in the presence of missing data in combination with large amount of variable fields. This is exactly the case with the so called system missing information in sets of variables (questions) not displayed to respondents due to the display logic employed in the questionnaire. Another advantage and suitability in the given circumstances is the fact that C&RT trees have very straightforward structure for interpretation. The modelling assumption accommodates for continuous as well as categorical variables. In those terms, the utilisation of C&RT in the analysis is

dependent on their strengths in regards with the research question.
C5 decision trees are state of the art algorithms. They have been designed to analyse substantial data sets with immense amount of numeric, time, date, or nominal fields. Ross Quinlan was developing tree based models such as the ID3 in the 1970s which later evolved into C4.5 as described in his seminal work (Quinlan, 1986; Quinlan, 1993). The main strength of the algorithm comes from the inductive inference capabilities, which are particularly useful in unusual knowledge discovery. It could be argued that due to enhanced handling of missing data and large number of variable inputs in combination with reduced errors in tree pruning (Pandya & Pandya, 2015), C5 are currently extensively used in critical automated computer-aided diagnosis (Azar & El-Metwally, 2012). The described advantages of C5 algorithms obviously secure the required level of precision for critical applications. Moreover, they appear to be the most popular tree algorithms in the machine learning communities (Salzberg, 1994).

4.5.2 Neural Networks.

In addition to decision trees, neural networks have been used to take advantage of their nonlinear approach to classification and pattern recognition. It is argued that traditional linear regression is a special case of neural networks, but it is analytically disadvantaged due to the rigid model structure and set of assumptions that are imposed before learning from the data (IBM, 2015c, p. 1). Given, the need to unveil the complexity of motivations in informal business practices, the best analytical technique is to enable unlimited flexible function fitting (Cherkassky, Friedman & Wechsler, 1994, p. 107) of various models into the data. This is achieved through ensembles of multilayer perceptron (MLP) neural networks using more than two layers, which enables generation of arbitrarily complex decision boundaries (Bishop, 1995). The number of hidden units could create issues with network complexity (Ripley, 1996, p. 168) and robustness and this issue is controlled through automating the selection via IBM SPSS Modeler. The use of neural networks provides an additional analytical advantage in terms of greatly limiting any research bias and prior knowledge assumptions, so that the discovery of unusual or unexpected causal relationships is greatly enhanced.

The property of neural networks to approximate a wide range of linear and nonlinear statistical models depending on variables' relationships is invaluable for the current tax evasion model building. Modelling of economic (tax elusive) behaviour is generally nonlinear dynamical system, which aims to predictively simulate social actors' behaviour. Therefore, neural networks are employed here in their natural "neurodynamical context" (Haykin, 2009, p. 589) to solve such problems. In order to answer some of the research questions of this study (see tables 4 and 5) the appraisal of the predictors' importance through neural networks models served dual purpose:

- Firstly, it reveals what are the strongest influencers for entrepreneurs to engage in tax evasive activities. This is achieved with neural computation of relationships' significance between predictor and target variables.
- Secondly, the current data set was used to train and make available in electronic form several neural networks, which could be used for future predictive data scoring (in later longitudinal data sets).
  All models with their exhaustive configuration and training details are available as part of this thesis' electronic data store.

4.5.3 Predictive modelling – terms and definitions.

The purpose of this section is to orientate the reader in the analytical modelling flow and therefore it provides explanations of the terms and definitions to be used. As described earlier, the two software solutions utilised for this purpose are IBM SPSS Modeler Premium version 18 and IBM SPSS Statistics Premium 23.

1. IBM SPSS Statistics is used in the so called mode of a "Helper Application" to IBM SPSS Modeler. This could be called a master-slave relationship, where IBM SPSS Modeler is the primary analytical solution. This means that for particular analytical tasks where further statistical tests would prove beneficial to the research claims, IBM SPSS Statistics is being revoked and used from within IBM SPSS Modeler.

2. Both software solutions are considered to represent the industry standard and therefore only brief descriptions for the used terms is provided in table 6 below.

3. Table 7 provides images and brief explanation of the most commonly used modelling nodes in the analysis.

Table 6: Brief definition of common modelling terms.

Software	Software term	Explanation
IBM SPSS Modeler	Modeling node	"A model is a set of rules, formulas, or equations that can be used to predict an outcome based on a set of input fields or variables." (IBM, 2015b, p. 5) There are several nodes, which employ simultaneous use of multiple other nodes.
IBM SPSS Modeler	Fields	In most cases, this is to be understood as equal to variables.
IBM SPSS Modeler	Target	When discussed in modelling context a "Target" is usually the variable being predicted as a function of one or more "Inputs" (predictor variables).
IBM SPSS Modeler	Target field	Usually a field in a model, which is designed to accommodate the target variable.
IBM SPSS Modeler	Input	When discussed in modelling context a "Target" is usually the variable being predicted as a function of one or more "Inputs" (predictor variables).
IBM SPSS Modeler	Modeling stream	"To build a stream that will create a model, we need at least three elements: • A source node that reads in data from some external source, in this case an IBM SPSS Statistics data file. • A source or Type node that specifies field properties, such as measurement level (the type of data that the field contains), and the role of each field as a target or input in modeling. • A modeling node that generates a model nugget when the stream is run." (IBM, 2015b, p. 6)

Software	Software term	Explanation
IBM SPSS Modeler	Overall accuracy	“The percentage of records that is correctly predicted by the model relative to the total number of records.” (IBM, 2015b, p. 60)

Table 7: Commonly used modelling nodes in IBM SPSS Modeler.

Modelling nodes	Brief explanation in accordance with (IBM, 2015b)
	“The Statistics File node reads data from the .sav or .zsav file format used by IBM SPSS Statistics, as well as cache files saved in IBM SPSS Modeler, which also use the same format.”
	“The Type node specifies field metadata and properties. For example, you can specify a measurement level (continuous, nominal, ordinal, or flag) for each field, set options for handling missing values and system nulls, set the role of a field for modeling purposes, specify field and value labels, and specify values for a field.”
	“The Association rules node associates a particular conclusion (the purchase of a particular product, for example) with a set of conditions (the purchase of several other products, for example).”
	“The Auto Classifier node estimates and compares models for either nominal (set) or binary (yes/no) targets, using a number of different methods, enabling you to try out a variety of approaches in a single modeling run. You can select the algorithms to use, and experiment with multiple combinations of options. For example, rather than choose between the quick, dynamic, or prune methods for a Neural Net, you can try them all. The node explores every possible combination of options, ranks each candidate model based on the measure you specify, and saves the best models for use in scoring or further analysis.”
	“The CHAID node generates decision trees using chi-square statistics to identify optimal splits. Unlike the C&R Tree and QUEST nodes, CHAID can generate nonbinary trees, meaning that some splits have more than two branches. Target and input fields can be numeric range (continuous) or categorical. Exhaustive CHAID is a modification of CHAID that does a more thorough job of examining all possible splits but takes longer to compute.”
	“The C5.0 node builds either a decision tree or a rule set. The model works by splitting the sample based on the field that provides the maximum information gain at each level. The target field must be categorical. Multiple splits into more than two subgroups are allowed.”

Modelling nodes	Brief explanation in accordance with (IBM, 2015b)
	"A neural network can approximate a wide range of predictive models with minimal demands on model structure and assumption. The form of the relationships is determined during the learning process. If a linear relationship between the target and predictors is appropriate, the results of the neural network should closely approximate those of a traditional linear model. If a nonlinear relationship is more appropriate, the neural network will automatically approximate the "correct" model structure."
	"The Ensemble node combines two or more model nuggets to obtain more accurate predictions than can be gained from any of the individual models. By combining predictions from multiple models, limitations in individual models may be avoided, resulting in a higher overall accuracy. Models combined in this manner typically perform at least as well as the best of the individual models and often better."
	"The Table node creates a table that lists the values in your data. All fields and all values in the stream are included, making this an easy way to inspect your data values or export them in an easily readable form. Optionally, you can highlight records that meet a certain condition."
	"The Analysis node enables you to evaluate the ability of a model to generate accurate predictions. Analysis nodes perform various comparisons between predicted values and actual values (your target field) for one or more model nuggets. Analysis nodes can also be used to compare predictive models to other predictive models."

4.5.4 Modelling workflow framework.

This section contains the so called *Modelling Workflow Framework* (Table 8) used for the analysis. It was especially developed to serve the current management research needs. It is presented here in order to provide full transparency to the reader for the logic behind the selection and use of any predictive modelling techniques. As a result, it also delivers the logical means for reverse examination and scrutiny of the modelling employed. Simultaneously, it could be argued that the suggested MWF is a robust methodology, which could be used beyond the management research focus here. The MWF is represented with 3 major stages, with sub-stages.

Table 8: Modelling workflow framework (MWF).

Modelling Workflow Framework (MWF)
1. Discovery of the most appropriate and meaningful predictive models to be used. 1.1. Using association algorithms (APRIORI, CARMA, and Association Rules) discover patterns and relationships in the data, including any unknown to the researcher (if such exist). Based on researcher's knowledge verify which variables' relationships need analysis. If there are any known important relationships missed by the automated association algorithms, explore those either with manual modelling or include them in the next step. This is the deductive interrogation of the data. In doing so, satisfy any particular modelling assumptions – every model has specific requirements for the data in terms of targets and inputs' characteristics. The criterion for model selection: if there is more than one model with satisfied assumptions, then select the one, which generates the highest accuracy-deployability ratio. 1.2. Using the Auto Classifier and Auto Numeric algorithms perform a structured automated model evaluation to explore the relationships identified in 1.1. This is the inductive interrogation of the data for the purpose of new knowledge discovery. 1.3. Fine-tune the top performing models from 1.2 based on accuracy-deployability ratio. 2. Examination of whether models created in stage 1 could be used in the so called model ensembles in order to derive a higher level models and research findings accordingly. This is the abductive interrogation of the data, as it requires the specific knowledge of the researcher in building the model ensemble. It also presents a degree of inductive knowledge discovery thanks to the automated application of the model ensemble to the empirical data. Model ensembles utilise the predictive power of at least two models to achieve more accurate predictions than can be gained from any of the individual models if engaged on their own. 3. All models from stage 1 and 2 need to be evaluated in several aspects, so the investigator could decide to keep or dismiss them. The main question is whether these models present any meaningful and significant research findings, which are worth reporting. 3.1. If yes, then evaluate the model predictive accuracy and deployability. Ensure satisfactory comparability with the initial evaluation by the auto classification algorithms. Keep or dismiss the model based on this evaluation. Depending on the model and the associated research findings, it is a viable compromise between accuracy and deployability. Very high model accuracy may result in poor deployability and vice versa (generalisable predictive power). 3.2. If the model does not represent any meaningful and significant research findings, then it must be dismissed and the search for an adequate one continued.

4.6 **Ethical and other issues raised by the methods or the approach (e.g. language, cultural factors etc.)**

The study's topic requires disclosure of information, which could affect the respondents if not handled ethically. In that regard, there is a demand for a greater level of trust in the communication between the researcher and the participants in comparison with what would be sufficient in other more conventional research scenarios. The principal researcher is a native Bulgarian and although he has been living in the UK, but not in Bulgaria for the last 5 years prior to the study, he would be recognised as a native Bulgarian by the respondents. This would have removed any additional cultural and language barriers from the research relationship and will have a beneficial advantage.

Ethical principles:

The main ethical principles have been that the researcher:

- While working within the University of Sheffield ethical approval rules, applied whichever national ethical standards are the highest between UK (University of Sheffield) and Bulgaria.
- Provided information sheet about the study and the relevant research method (survey/interview). The information sheet/cover letter will contain confirmation that full anonymity and confidentiality will be secured for the informants.
- Report and disseminate results only in aggregate form.
- Particular care has been exercised to inform the respondents about their rights (to stop the interview at any time, to share only information they want to share).
- Acquire informed consent from the respondents.

General compliance, scientific integrity and ethical conduct:

This project abided by the University's 'Good Research Practice Standards': and by the University's 'Ethics Policy Governing Research Involving Human Participants, Personal Data and Human Tissue

Chapter Five

5. Findings of semi-structured in-depth interviews.

This thesis employed mixed methods in exploratory sequential design and the current chapter addresses the first out of two investigative levels. Indeed, to analyse the in-depth interviews with key NRA officials and therefore contribute to the comprehensive review of the institutional context with a focus on tax policy management. To achieve this task, four lines of inquiry have been reported: discovery and exploration of agency strategies; agency operational approaches and the predominant type in the mix of policy measures; and the triggering procedures and methods of enforcing control and compliance – tax audits and inspections. Particularly important is the illuminating discovery of the three major environmental pressures (institutional, economic and political) which affect NRA's role in revenue raising policies and the associated institutional asymmetry. This line of qualitative investigative analysis (chapter 5) complements the quantitative one (chapter 6) and both inform a holistic account of the socio-economic influencers of the processes of business tax evasion, formalisation and informalisation.

5.1 **Qualitative analysis – visualisation and discussion.**

National revenue agencies face the problem of balancing revenue-raising at sufficient rates to enable states to function and doing so at levels regarded as socially acceptable among economic actors. If revenue raising policies are deemed unacceptable, it can encourage avoidance of the formal economy in favour of remaining within, or entering, the informal economy. This balancing act becomes particularly acute in economies with weak formal institutions and significant informal institutions because operating within the informal economy is a practically-viable alternative to operating in the formal economy. Previous research (Williams et al., 2014d) exposed the so called institutional asymmetry, which affects negatively entrepreneurship in Bulgaria. Many transition economies like Bulgaria face just this situation, where formal institutions vie for control over informal institutions, but little is known about how revenue agency officials work to manage this balancing act.

Institutional asymmetry is now widely recognised as being positively correlated with informal economic activities – both at entrepreneurial and employment level (Williams et al., 2015a; Williams et al., 2015c; Williams et al., 2015d). The formal institutions of a country consist of the political ruling model with its supportive economic and legal structures-agencies. The informal institutions are embodied in the fundamental social norms in a given society, which translate into various beliefs, value systems and social behavioural norms. There is an inherent interconnection between the formal and informal institutions, which affects entrepreneurship and economic development in general. According to North's (North, 1990) theory of institutional change, formal institutions are representation of informal ones and they both co-evolve. Casson et al. (2010) conducted an intriguing critical review of the literature, revealing the intrinsic circularity between institutions and economic development. In their view "the quality of institutions in a country will influence its development process, while at the same time, the process of development will influence the quality of these institutions". Further evidence (Valdez et al., 2013) in support of the relationship between the quality of macro-level institutional environment and the sought alignment with people's norms, perceptions, acceptance and experiences, signifies that actually the latter have stronger economic benefit for a given country context. This part of the analysis is with a particular focus on providing insights on how officials seek to balance their duty to enforce the formal tax system while managing the risk that entrepreneurs seek to

avoid the formal institutions and operate in the informal sector. Until now, there is no such an in-depth study of the National Revenue Agency's strategic methodology, attitudes and policies to counteract tax evasive activities. This qualitative analysis aims to present summarised evidence from which some potential lessons will be highlighted for other transition economies.

Appendix 15 is a graphical representation of all nodes from within NVivo. All top level nodes (9 in total) are represented in rounded rectangular shapes coloured in green.
It becomes evident, which top nodes possess the biggest number of sub-nodes. It is top to bottom order (Appendix 15) where the higher the root level node is located the most sub-levels it possesses. However, from the following figure 11 it is visible that, although the top node "Risk management" possess the most sub-nodes it is on third place in terms of text references – parts of text coded to it.

Figure 11: Descending classification of root level nodes according to number of coding references.

Name	References
Tax	235
Control	213
Risk management	177
NRA	97
Legislation	72
Business	35
Institutions	33
Behaviour	20
Political situation	2

When working with nodes in a hierarchy, one can gather all the materials from child nodes and roll it up to the parent node. This is how the total number of references is accumulated in the figure above. For the purpose of the analysis, we can say that although "Risk Management" node possesses the biggest number of sub-nodes it is on third place in terms of number of references. To identify the significance of a theme, the analyst needs to account for both – number of references, and number of sub-themes. As is evident, all nodes are actually representing a particular theme (topic) of certain importance.
Cluster analysis is an exploratory technique in NVivo, which could be used to visualize patterns in any given project by grouping sources or nodes that share similar words, similar attribute values, or are coded similarly by nodes. Cluster analysis diagrams are available in different shapes to provide a graphical representation of sources or nodes. For the needs of the analysis, the following figure has been developed using Pearson correlation coefficient (cluster analysis).

Figure 12: Cluster analysis of top nodes (themes).

Nodes in the cluster analysis figure that appear closer together are more similar (by words similarity) than those that are far apart. In addition, from the node's location in relation with other nodes we could observe the relation between particular themes. In the given example, Risk Management is of central position to the other root level nodes, which indicates its central importance to what is being done in the NRA.

The approach called frequency technique is a useful method to visualise the most frequently coded words (concepts). This is being done by estimating the intensity of coding to a particular concept. As is evident from figure 13 below, some of the most frequently coded concepts are: selection, tax, inspections, and audits. If the font being used is bigger, then this represents a more frequently coded concept.

Figure 13: Coding by concepts' frequency.

Table 9 represents the key themes (relating to particular questions), which have emerged from the interviews. The in-depth interview (Appendix 7) contained 26 predetermined (structured) questions and table 9 below presents 13 of them. Thanks to the semi-structured nature of the interviews, the researcher had the chance to ask relevant clarifying questions and thus capture enriched information about the most significant topics. The first column of the table presents selected questions from the interview template. The second column provides additional information about the research purpose of the question. The third column summarises some of the emerging key accents and the final column contain illustrative quotations.

Table 9: Key interview themes and summary of responses.

Question	Research purpose	Key accents	Illuminating quotes
What challenges do you encounter in the implementation of your tasks and responsibilities?	This question aims to reveal what are the difficulties, which are currently encountered by NRA officials. The question is asked in such a way to provoke the respondents to speak about their real professional challenges and thus share their impressions about things that needs addressing.	Political environment and pressure, difficult organisational management, bank secrecy as problem, need for new data management systems, need for specific training	"Quite a large issue. Too slow response to some external stimulus, too difficult change... slow system... Or restriction in relation with other organizations that could assist and support us. Others are related to... how to say... not just criticism of the political environment but rather the political situation. I mean the administration is sometimes used for certain purposes other than what it should do. But let's say that this is normal for the political situation in Bulgaria." "Another example is bank secrecy... very restrictive regime, it is difficult to take bank secrecy which in turn hampers the detection of tax fraud." "it was difficult to find the right people for the job"
One of NRA's strategic aims is to encourage voluntary tax compliance. What initiatives and/or activities do you undertake in that regard?	The purpose is to discover how NRA satisfies one of its strategic aims. Is "encouraging voluntary tax compliance" just an advertised strategic aim of NRA or it is genuinely pursued. Based on the collected information it could be determined to what degree indirect policy measures are being favoured by NRA.	slow development of voluntary compliance, cooperation with business organisations, need for increased provision of electronic services, letter and telephone campaigns, use of treatment mix	"Here is an example of a very slow things happening in the administration." "In addition, our cooperation with business organisations, I think, is getting better every year. It's easier to know and to feel that the NRA is your partner and not an opponent..." "actually in this sense are the efforts of the NRA for encouraging voluntary compliance by trying to create more clear laws, more access to electronic services, more access to explanations and methodological guidelines and so..."

Question	Research purpose	Key accents	Illuminating quotes
Does NRA have specific methods or systems in place with which you reveal the ways and means used by the tax payers to evade their tax duties or lie to you – for instance concealment of business turnover?	This question is designed to reveal facts about the current approach in tackling informal practices exercised by businesses. It is generally questioning whether there is some level of automation (systems) employed to tackle tax evasion.	lack of systematic information, systems underdeveloped and underutilised, insufficient system synchronisation, Oracle Business Intelligence Suit	"we have a remote connection to the cash registers, which helps a lot to locate places where such income is concealed...This can be done automatically and the colleagues do it." "In general we have a lot of information, but not always process it successfully enough. This is one of the problems facing the NRA. Now there are works on a new control system, hopefully thanks to it things will be happening more successfully." "But generally we use some purely statistical methods such as regression models, for example, which can be used as risk profiling of the individuals."
Are the assignments and management of tax audits and inspections automated to some extent or not at this moment of time?	This is a very topical question in regards to the high corruption level in Bulgaria – affecting institutions and administrations of all types. It actually aims to expose the level of subjectivism and opportunities to influence impending or on-going audits and inspections.	opportunities to circumvent the automated distribution, manual intervention option, new system impending	"Currently only the distribution is automated, that is, who will conduct an audit of a certain taxpayer. But by the middle of August the whole process will be automated in this information system." "In fact, the audit is distributed automatically to an employee at random. It is not subjective. Nobody decides, but the system decides." "Although there are internal rules that are actually able to affect this process. It is not fully and completely.... right" "if I'm honest, there were some possibilities to circumvent this automated distribution. But now with the new system as though it will be avoided."

Question	Research purpose	Key accents	Illuminating quotes
In your opinion, is there any subjectivism in the selection, delegation and management process of tax audits and revisions? Or is this avoided through a level of automation of those processes?	This last question from the special section is designed to collect and confirm personal opinion about any subjectivism in regards with the implementation of tax audits and inspections.	available level of subjectivism	"I think that, in theory, subjectivism can be avoided. We live in the 21st century. Talking about artificial intelligence. But the question is whether we want it to be avoided and whether it should be avoided. I think the extremes are always harmful and it should not be avoided. It should be limited, the subjectivism. That is, subjectivism should be exercised on an objective basis." "Specifically for the selection I cannot tell you. I would wish there not to be." "So there is, of course there is. Subjectivism cannot be absent."
Is there a process for analysis of the possible measures, the implementation expenses for them and the expected effect?	It is an established fact that contemporary tax offices are deciding upon their engagements (such as audits and inspections) based on embedded cost benefit ratio procedures. So, this question is looking to find out how it is at NRA – operational efficiency and real world feedback.	newly developed cost-benefit methodology, backward cost-benefit capability, few officials aware about the new methodology	"Yes. This is part of risk management...We try here to evaluate how much an inspection will cost us, how much an audit will cost us, how much a media campaign will cost us and what effect it will have." "there is no such risk which is so small in terms of damage to the budget, so that to fail to cover the cost of treatment...We can measure what we have achieved, but we still cannot predict what we will achieve." "There is no procedure. The only thing I know that has been developed is an average price how much an audit costs the administration, namely that it costs about ten thousand levs." "what costs this control action is worth to the NRA and the expected collection. Well at least I am not aware such a thing to exist."

Question	Research purpose	Key accents	Illuminating quotes
How is it decided into which areas of tax non-compliance to invest efforts? Said differently, which types of tax infringements attract most attention?	This question relates to two other questions (13 and 16). It is part of the so called gradual approach in extracting related information. Again, it seeks information about a relationship between applied compliance measures and expected outcome (cost benefit relationship).	lack of agreement about risk management	"I do not know how it turns out that some risks are prioritized... I'm sorry but one of the risks was "a risk of tax evasion by individuals conducting driving lessons". That type of risks, which simply I do not know how they manage to find such risks! How it got such a high priority, that risk!" "To be honest I do not know how it is decided. So we have a strategic plan, some strategic objectives are laid in it. One of these objectives is for example to combat tax fraud." "well it is logical to assume that where there are the biggest tax losses, we should be directed there and I think this is leading." "Well, generally avoiding to report turnovers, to display revenue. This is the main we observe particularly in Risk Management. Others already are... non-compliance with labour legislation, this is seriously practiced."
What are the most frequent reasons/motives for turnover underreporting by the Bulgarian businessmen (entrepreneurs)?	The question is after NRA opinion about the most frequent reasons/motives for turnover underreporting by the Bulgarian entrepreneurs – the formal institutional perspective.	lack of study on the topic, unfair competition problem, most officials were interested to know the real motives for evasion	"Unfortunately, we do not yet have made such a slender study on this topic. So far we have only read external studies..." "I think it is a herd behaviour. Just all do it, why not do it myself." "Some people do not want to pay taxes because they think that they are not spent properly from the budget after that." "for me a big role is played by the competition, sometimes unfair...People do not like to pay taxes, it is clear."

Question	Research purpose	Key accents	Illuminating quotes
If you have noticed change in the behaviour of firms and/or individuals as a result of counter evasive activities, is there any mechanism of changing any future counter activities and if yes how is this happening?	This question is connected with 13, 16, 17, 20, 21 and 22. It aims to reveal any institutional culture of responsiveness.	procedures for risk treatment feedback are still not fully developed, lack of real world feedback mechanism	"Yes, because when we see that a risk does not respond or there is extremely small effect of our actions, rather we have to look to change our impact or to be directed to other risks." "it is very fragmented. I can hardly say that such a process exists, systematic." "There are indicators and benchmarks for evaluating whether it is influenced significantly." “If we consider the particular risk and we have evaluated it is greatly influenced, it may not be treated in the next one or two years.”
At this moment of time what methods or campaigns are being used to encourage voluntary tax compliance of firms and traders?	It aims to reveal to what extent NRA focuses their action on indirect methods.	less emphasis on information campaigns, lack of developed campaigning methods	"There are some campaigns being carried out. Although I would not say that they are exactly for the promotion of a correct tax behaviour. But there are some campaigns on compliance with tax legislation." "Informational campaigns are being used. Every year there is also a campaign in the resorts - sea or mountain resorts."

Question	Research purpose	Key accents	Illuminating quotes
From your perspective, what are the characteristics of the firms and traders who allow themselves to declare lower tax base than the real one? Is there something particularly common between them or something which differentiate them from the other tax payers who are compliant?	The main purpose is to find out whether NRA observed specific features and patterns characterising the informal firm or these firms, which regularly exercise partially informal practices.	specific cultural attitude to evade taxes, evasion for profit maximisation, smaller businesses are likely to evade more taxes	"we will not be able to change the stereotypical behaviour of Bulgarian folk-psychology." "The first thing that comes to my mind is that they are Bulgarians. I have already said, the attitude, mentality of the people..." "These are normal companies. They just do it because others are doing it." "Yes, there is. For example the size of the certain trader is of essential significance. The structure of the trader also, the structure of management in sole traders and in the so called family firms, where there is little staff and they are related." "But for those who actually do business, the only reason I can think of to conceal turnovers is to pay less taxes." "The business world is very colourful. It is hard to make such a generalization. Some common feature, in my opinion, can hardly be displayed. They have complex features and they are interconnected. In general, those who conceal income are usually small and medium-sized companies in terms of volume of turnover."

Question	Research purpose	Key accents	Illuminating quotes
In your view, what kinds of difficulties are being faced by the Bulgarian businessmen (entrepreneurs) in their efforts to be tax compliant?	It is aimed to discover the institutional perspective about administrative and other difficulties experienced by Bulgarian businessmen, which drive them into informal practices.	frequently changing legal framework, heavy bureaucratic procedures, lower level of customer service	"My point is that there are still many documents we require from tax payers...And sometimes they have to provide us with quite a lot of information and it hampers them." "Maybe there are heavy bureaucratic procedures, the way of service" "I guess one of the reasons is the complexity of the legislation." "if the rules cannot change so often." "it may be demanded from the administration to be friendlier to business. We are not accustomed to this thing." "What I've heard from the business is that it is easier to conceal everything than to conceal one part. It is much easier for you to be entirely grey than trying to be compliant and someone came to audit you and to still find something."
Is there any information, which is of particular interest to you about the firms and traders evading taxes, which you would like to know?	This last question has couple of implicit purposes: ➢ Firstly, it attempts to discover what kind of information NRA is after in regards to informal entrepreneurs (businesses). Thus, some conclusions about their approaches could be made. ➢ Secondly, the question aims to collect insights and inform the creation of question(s) for a business survey.	lack of information about motives for evasion, need for behavioural models of evaders, need for feedback about NRA	"So what we don't know here is about the behavioural models of the different groups. That is, firstly we do not have a clue i.e. do not have stratification of our taxpayers...From a behavioural point of view, right." "Yes. I wonder for example what they are not satisfied with, to receive feedback on what they are not satisfied and what they believe needs to be improved in the NRA." "I would go even further - what would that thing be, which will change them, what would that thing be that will make them not do it. Whether the fear of an audit or something else." "Now I think that if there is an opportunity, to ask under what circumstances they would declare absolutely everything."

In order to enable holistic appraisal of NRA methods, policy approaches and attitudes, it is required to consider that they are embedded into a specific socio-economic context. The next 3 figures (14, 15 and 16) objectively describe some of the key indicators characterising the institutional environment in which NRA operates.

Figure 14 below presents the ease of doing business and selected related activities. It is based on data from World Bank and IFC (2015b). Instead of using the regular ranking indicators to produce the figure, the so called distance to frontier (DTF) score was used. DTF is a much more robust indicator due to its capability to demonstrate the absolute level of regulatory performance and how it improves over time (World Bank & IFC, 2015a) for a particular country. This measure shows the distance of each economy to the "frontier," which represents the best performance observed on each of the indicators across all economies in the Doing Business sample since 2005 (World Bank et al., 2015a). Zero represents the lowest performance and 100 represents the frontier (best performance) – the higher the DTF score is the closer to the frontier a country is.
We can observe that there is a decrease in majority of the DTF scores (some of them preserve the same value) for Bulgaria in the period 2010 to 2015 – the decrease is more distinct in the period 2013-2015. The cumulative ranks of 41 for 2014 and 49 in 2015 (World Bank et al., 2015a) are even more affected (relatively) with a negative decrease – there is a decrease of 8 positions.

The methodology used by "Doing Business" report includes small business according to the EU definition (European Commission, 2014). Businesses, which qualify as small, are actually big in comparison with the average business environment in Bulgaria. It could be argued that the ease of doing business has a financial expression for a given business – the more bureaucratic the institutional environment is the more expensive running a business would be. Majority of the institutional procedures are fixed in terms of efforts and costs and the bigger the business is they would be a smaller proportion of its earning potential. Therefore, the DTF scores (if calculated) for the average firm in Bulgaria are expected to be poorer.

Figure 14: The ease of doing business and related activities in Bulgaria.
Figure 14 is based on data from the World Bank et al. (2015b). A customised report about the ease of doing business, paying taxes and enforcing contracts is available in Appendix 16.

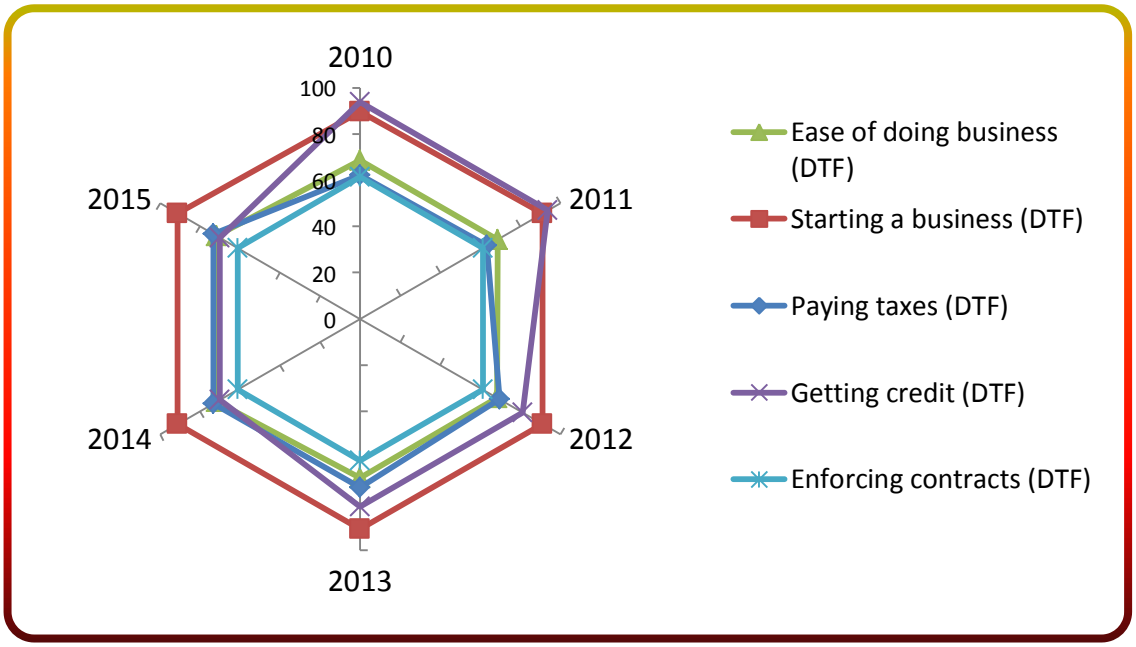


Figure 15 presents the global competitiveness index (GCI) rankings for Bulgaria in the period 2010-2015. The Y-axis represent the GCI index – the lower the value the better competitiveness a country would have. It could be observed that there is a small improvement in the GCI index in that period, but still there are countries in 2015 like Puerto Rico and Chillie, which have approximately 70% better GCI – Puerto Rico GCI 33, Chillie 32, Bulgaria 54 (World Economic Forum, Schwab K, Forum WE, 2015, p. 13). Taking into consideration that Bulgaria is an EU member since 2007, the value of this indicator signifies some serious macroeconomic inefficiencies.

Figure 15: The Global Competitiveness Index (GCI) rankings – Bulgaria for 2011-2015. This figure is produced using data from World Economic Forum (2015) – the data set used is available in Appendix 17.

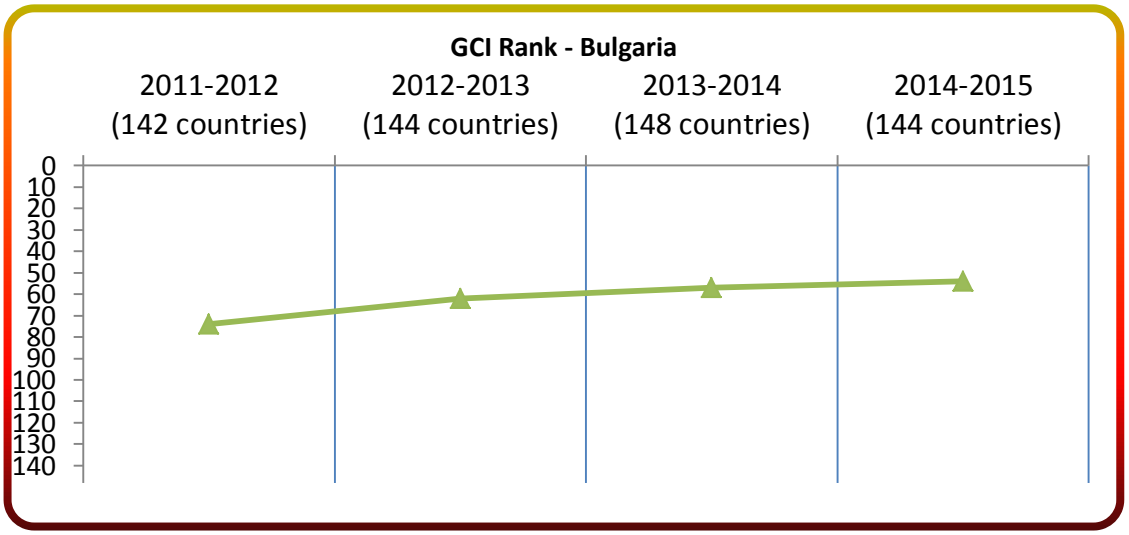
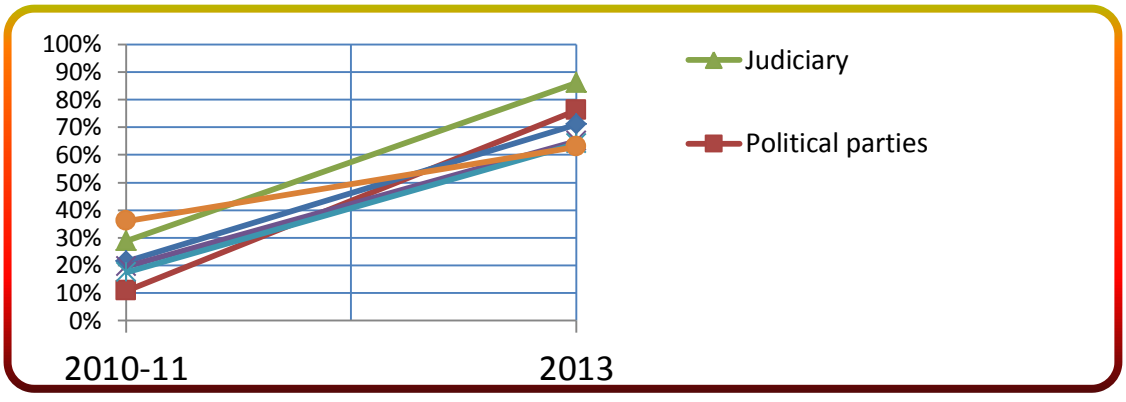


Figure 16 is a visual representation of the latest data from the Global Corruption Barometer (Transparency International, 2013) for the years 2010-11-2013. The Y-axis represents what proportion of the surveyed people thought that a particular institution is corrupt/extremely corrupt, while the X-axis provides the values for two measurement points – 2010-11 and 2013. It could be observed that disturbingly increasing proportion of the surveyed individuals between 2010-11-2013 believe that some of the key state institutions are *corrupt or extremely corrupt.* In appraising the link between corruption and entrepreneurial informality, Yalamov (2012) suggests that bribery is used primarily when companies are engaged in hidden economic activities and circumvention of the law. This means that these phenomena could not be tackled separately, but rather they need a holistic approach.

Figure 16: Institutional corruption levels in Bulgaria.
This figure is produced using data from Transparency International (2013) – the data set used is available in Appendix 17.



To summarise the socio-economic indicators, it should be emphasized that Bulgaria continue to be the poorest EU member state with the highest share of informal economy and corruption. Unsurprisingly, at the same time it is one of the least competitive countries.

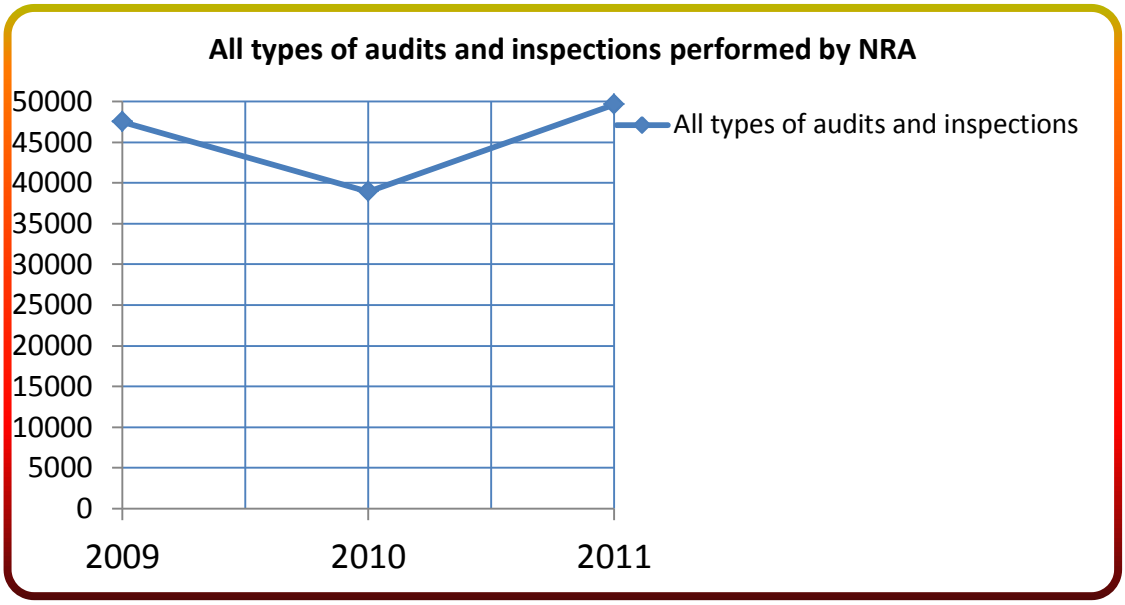
Williams et al. (2014d) performed a critical analysis of how institutional asymmetry affects entrepreneurship in a transition setting. The authors defined institutional asymmetry "as the misalignment between formal and informal institutions, with the formal being generally supportive of entrepreneurship and the informal, unsupportive. The asymmetry develops over time as formal institutions are reformed to support entrepreneurship" (p. 2). Further research into institutional asymmetry (Williams et al., 2015c) finds out that the greater the asymmetry between formal and informal institutions is, the greater is the likelihood of participation in the undeclared economy, and vice versa. One of the contributing factors for the increasing gap between the development of formal and informal institutions in Bulgaria, is the fact of the economically early acceptance as an EU member state due to geo-politic reasons and consequential inadequate policy transfers.

Several Bulgarian governments attempted to adopt the best EU institutional practices at policy and operational bureaucratic level. In these transfer scenarios, little attention was paid to the fact that they were borrowed directly from advanced old EU members states, where the alignment between formal and informal norms is very high. Inevitably, the regulative pressure on the immature micro and small business sectors in Bulgaria increased disproportionally to their potential to adapt. This in combination with the low levels of social cohesion (see Figure 4), very high levels of corruption and mistrust, contributes to the significant and widespread informal economy and related phenomena. If a greater focus is granted to a further improvement of the formal institutions and their efficiency (without simultaneous development of the informal ones), this will inevitably lead to even greater share of the informal practices. As Williams et al. (2014d, p. 2) argued informal institutions matters "at least as much as formal institutions for fostering entrepreneurial activity" and therefore proportionate policy effort must be invested into their advancement. However, it should be underlined that policy influence over historically embedded cultural factors is limited even if designed appropriately, but this fact does not remove the necessity to create and maintain the right policy framework. Unfortunately, to this moment of time, neither the right policy framework existed, nor sufficient number of adequate policy transfers occurred in Bulgaria.

The socio-economic context of NRA was enlighten, so that the discussion of its policy approaches, methods, attitudes and issues could be developed in an interconnected line. Figure 17 represents the total number of all types of audits and inspections (for businesses and citizens) performed by NRA. A very large number of controlling activities by NRA is evident having in mind that population of Bulgaria is only 7.2 million (approximately). The data available is for the years 2009 – 2011, but there is no significant shift in tax policy since then, so it is highly likely that the figures will be similar for the years after. This is an indirect expression of the fact that the efforts to tackle informal economic activities have been focused on direct policy approaches, aiming to increase the risk of detection and cost of informality using disincentives. From an institutional perspective, it is intriguing to consider what the cost-benefit ratio is for this policy approach. The number of NRA audits has recently decreased to nearly 10 000 per year and the cost per audit is almost 10 000 BGN (based on data from the interviews). Noting that Bulgaria has the highest share of informal economy in the EU (Figure 2), this demonstrates that the aggressive and expensive deterrence policy approach is proving inefficient to tackle this problem. There is evidence (Williams et al., 2014c) that reducing the asymmetry between the formal and informal

institutions will improve the level of formalisation of entrepreneurs. As it is apparent from the data presented so far, firstly putting the emphasis onto direct measures is highly inefficient and secondly it deepens the institutional asymmetry even further.

Figure 17: All types of audits and inspections performed by NRA. This figure is based on data acquired through an information access request as per Appendices 10 and 11. The data acquired, contained information for the years 2009, 2010 and 2011.



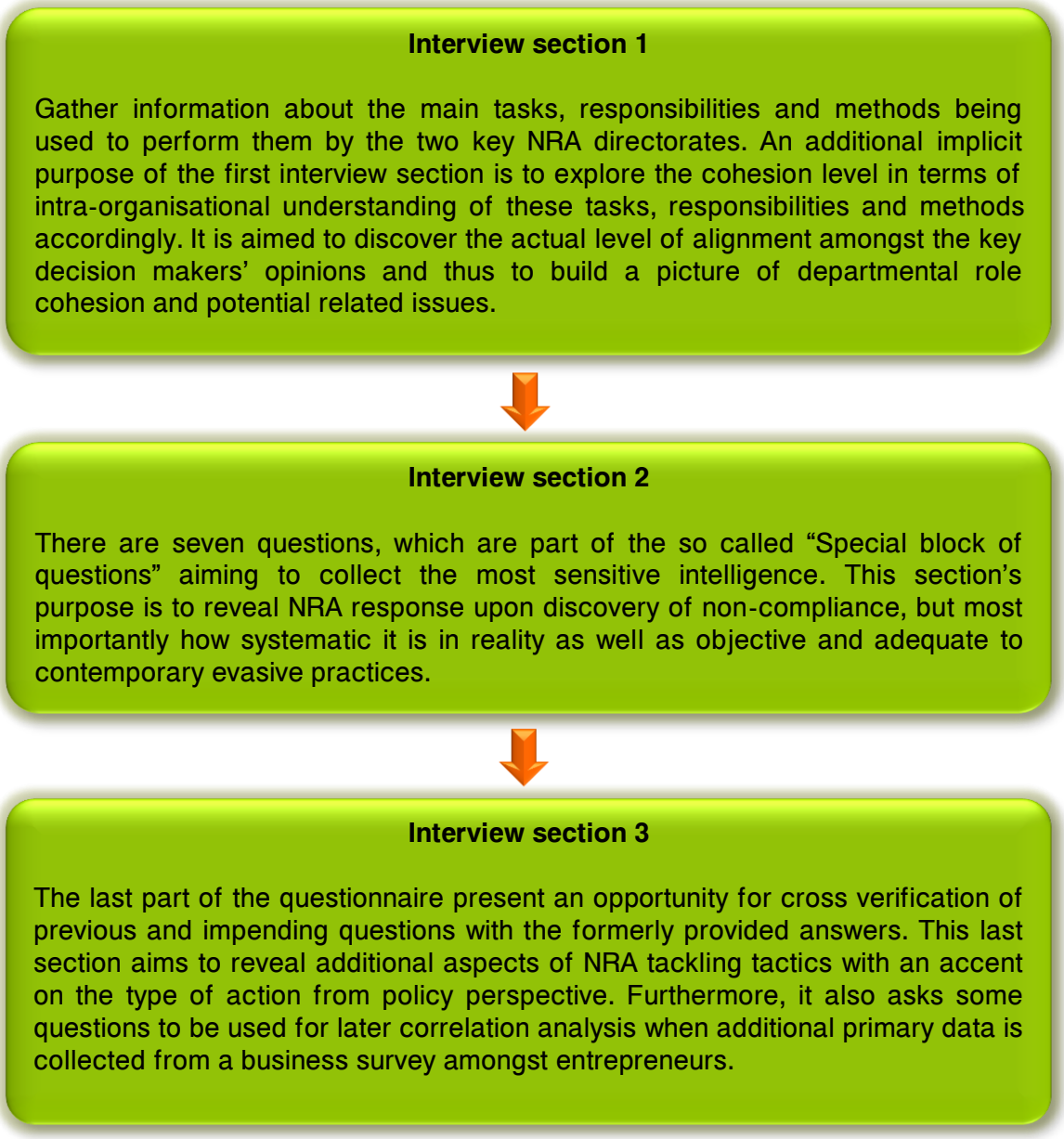
Koettl and Weber (2012) utilised an intriguing standpoint about the role of labour taxation and social benefit design in regards to formal/informal work. It is important to note that the authors use a legalistic definition which includes the self-employed and employers amongst the informally employed if they conceal all their income. Although the authors have not conveyed explicitly, their approach belongs to a relatively new trend in the literature, which is concentrating on the rationale decision based model – social actors taking decisions about participating and to what extent into any form of informality. Koettl et al. (2012, p. 2) introduced an innovative measurement called formalisation tax rate (FTR) to analyse the disincentives for formal work: "The novelty of the latter is that it measures disincentives stemming not only from labour taxation, but also from benefit withdrawal due to formalization." The analysis is implemented using data about six Eastern European countries (including Bulgaria). The authors argue that the higher the FTR that an individual is facing the higher the chance is that they would work informally. This could be considered in accordance with three possible policy approaches for NRA (sticks or carrots or a mixed approach) – if the government is using deterrence measures predominantly or otherwise in combination with incentive measures. It is argued that NRA can positively affect formalisation rates much more effectively through adjusting the FTR in Bulgaria. This is possible with an increase of the benefits for formalisation in comparison to the current trend in exercising intensive deterrence/direct measures approach.

The semi-structured interview questions (to senior NRA executives) utilised a graduated approach over three interrelated sections. Considering, the fact that the topic of the in-depth interviews connected with information, which is not publicly available and is very carefully institutionally guarded, the main aims were:

- Acquire as much intelligence as lawfully possible (see heading 4.3.2 and Appendix 10, 11 for additional details).
- Achieve the above without exposing NRA beyond its duty for transparency and openness to policy scrutiny.

Therefore, the main purpose of the specific research design was to increase the quality and validity of the acquired information. For clarity, a brief summary for each interview section is presented below in Figure 18.

Figure 18: Summary of the in-depth semi-structured interview's sections.



The key interview findings are described in Table 10. The first column from the left presents some major categories into which insights have been acquired. The other column to the right contains the most important accents, which emerged from the analysis of the interviews. The next sections will critically appraise the key findings and connect them with the relevant literature.

Table 10: NRA – categorisation of main accents.

Topic category	Key accents
Operational environment	➢ Political pressure
Policy approach	➢ Accent on direct measures (deterrence approach) ➢ Less emphasis on information campaigns ➢ Lack of developed awareness campaigning methods ➢ Slow development of voluntary compliance
Information management software	➢ Systems underdeveloped and underutilised ➢ Opportunities to circumvent the automated case distribution ➢ Available level of subjectivism in case distribution ➢ Insufficient system synchronisation ➢ Need for new data management systems ➢ Lack of systematic information ➢ Need for specific training ➢ Need for increased provision of electronic services
Organisational management	➢ Difficult organisational management ➢ Lack of agreement about risk management ➢ Low level of customer service
Legislative framework	➢ Bank secrecy as problem ➢ Frequently changing legal framework ➢ Heavy bureaucratic procedures
Real world bidirectional feedback	➢ Newly developed cost-benefit methodology ➢ Retrospective cost-benefit capability only ➢ Few officials aware about the new methodology ➢ Procedures for risk treatment feedback are still underdeveloped ➢ Most officials are interested to know the real motives for evasion, but they do not have information acquisition practice ➢ Need for behavioural models of evaders ➢ Need for feedback about NRA ➢ Specific cultural attitude to evade taxes ➢ Evasion for profit maximisation ➢ Smaller businesses are likely to evade more taxes ➢ Unfair competition problem

As Calciolari, Cristofoli and Maccio (2013) suggest public sector organisations are simultaneously subject to three types of environmental pressure: institutional, economic and political. Having enlightened the economic setting in which NRA operates, it is required to discuss one of the most pressing issues, which emerged

through the qualitative study – the fact of strong political pressure over NRA in terms of collecting income for the government. In particular, the case when NRA is strategically utilised as a revenue generating instrument to fill in political needs – such as financing election campaigns (and other improper income expenditures) or scrutinising the tax matters of business circles around the potential government in shadow. Such practices have been so frequently employed that for some of the high level NRA members of staff, they are perceived as "normal for the political situation in Bulgaria". Logically, there are several multifaceted effects from these political pressures:

- NRA strategic behaviour becomes dependent on the needs of any current government. Consequently, their policy of favoured direct measurement approach is even more emphasized, but in *a disproportionate and highly subjective manner*.

- Controlling and punitive policy methods are deployed amongst two types of businesses in order to minimise the tax gap, but also to obstruct business support for the political opposition. The first categories of affected businesses are micro and small ones, which are not politically associated by default. The other category is the medium and large businesses, which are connected with the non-ruling political parties. This creates the issue of intentional *asymmetrical closure of the tax gap*.

- Furthermore, such institutional practices disturb the social feeling for procedural fairness and henceforth increase the institutional asymmetry and informality accordingly. As Williams (2014c, p. 26) discuss procedural fairness as the extent to which people believe that they are paying their fair share compared with others and receive fair treatment by institutions. It is worthwhile noting (Williams, Dzhekova, Baric, Franic & Mishkov, 2014a, p. 49) that the appreciation of *procedural fairness amongst Bulgarian individual and small business tax payers is below the acceptable social minimum*.

It is an established fact that the state administration in Bulgaria is very subservient to the governing political class (Noutcheva et al., 2008, p. 115) and high level public officials are often forced to act in greater alignment to political interests than in alignment with the legislative framework. Refusal to comply, usually results in position dismissal and in some cases disciplinary proceedings against such public servants. This is an issue that is even more acute for officials, who already accepted (at least once) to act purely as per the political will. In the context of the described phenomena, it would be unrealistic to expect efficiency orientated reforms supported by objective policy implementation at NRA. That would be only possible if there is genuine and appropriate political will to allow it, no matter how good the professional qualities of NRA high level executives are. In relation with the need for strategic management reforms to the link between NRA and government, it is important to emphasize that the current situation undermines what Cerase and Farinella (2009) determine as "public service motivation". Thanks to the qualitative approach employed, it has been discovered that this is another significant problem, which is arguably widespread in the entire state administration. Obviously, its effects are more distinctive in revenue agencies (NRA and Customs).

Amongst the most distinct information technology issues, which hamper NRA's work is fact that the current systems are underdeveloped and underutilised in addition to the state of insufficient system synchronisation. The current systems are developed or acquired on a modular (ad-hoc and add-on) principle, which makes their synchronisation strongly limited. As apparent from the study findings, there are

opportunities to subjectively circumvent the automated case management, which arguably harms the perceived procedural fairness amongst the targeted tax payers (business and individual). This is especially valid for micro and small businesses, which bear the biggest proportion of compliance and asymmetrical gap closure pressure. As Hartner et al. (2008) discussed in case of perceived unfair treatment and processes of unfair decision making, taxpayers resist paying the whole amount of their taxes due. Additional issue between the two key NRA departments (Risk Management and Control) is the fact that risk management strategy becomes increasingly complicated due to the difficult economic situation and intra-organisation management issues. Therefore, risk management demands high level of standardisation in the development as well as in its enforcement process. As part of the difficult organisational management, both departments recognise that there are some mutual disagreements in both of these processes. This inevitably translates into disproportionate enforcement to tax payers and is also interpreted by them as an unfair treatment, which leads to the inter-connected questions of distributive justice and customer attitude. That is why it is highly desirable and applicable for NRA to achieve greater level of automation and objectivity in its intuitional control and enforcement measures, starting from more efficient risk assessment system.

The discussion about political pressure over NRA and how it could be partly addressed, lead us to one of the logical ways forward. NRA should equip modern business intelligence software like other tax administrations such as HMRC UK and ATO in Australia). Such information technology solution would enable:

- Significantly higher level of automation, which results in greatly reduced subjectivity in distribution and management of audits and inspections.

- External auditing of NRA activities (for instance by BNAO) to be much more robust and informative. NRA work will be based on a transparent foundation.

- Obstruct political pressure for asymmetrical closure of the tax gap – any tax management activities, which are not according to the established rules, would leave a visible trace and could be scrutinised at any time.

- Reduction of the non-effective tax management actions – where the cost exceeds the potential income.

At the time of writing the implementation of Oracle Business Intelligence Enterprise Edition is still ongoing (click here for more details) and the outcome of this innovation will become evident in the years to come. The efforts to improve the regulatory performance, cost-efficiency and transparency of NRA are based on bottom-up approach where the initiative is coming from within the agency. From one perspective, NRA high level executives need to address the strategic link between the agency and the government – this would make their work more independent through capacity for wider public accountability. This would secure much higher price political parties will pay if they need to exercise inappropriate institutional pressure. From another perspective, NRA feels the need to simplify and modernise their work internally (in intra-departmental way) and externally (e-customer service perspective).

In consideration of the financial aspects of similar systems in Italy, Romania and Slovenia, Dečman, Stare and Klun (2010) argued that although the cost-benefit ratio over short periods (up to 3 years) is not favourable, there are some very important non-financial benefits. The reduction of formal control activities (shift in policy), higher taxpayer satisfaction and immediate availability of tax return data to mention some, could prove very profitable in over 3 years terms. Nevertheless, it is prudent to note that software solutions in support of tax administrations are a subject of an intense

debate in the literature. It is required to acknowledge that adoption of modern information technology solutions to enhance tax compliance is “no substitute for political will…computerization can easily serve as a way to absorb technical assistance, and camouflage the fact that little or no fundamental progress is being made on the reform” (Jenkins, 1996, p. 3). This danger is particularly relevant to the Bulgarian institutional environment and specific effort must be made to avoid its realisation.

Another important aspect of having a fully integrated business intelligence software solution (at NRA) must be appraised here. In particular, the purpose of having prospective and retrospective cost-benefit functionality and how this relates to responsive regulation and enforcement. It could be suggested that the cost-benefit regulative policy approach has distinctive strengths to countries with low distributive justice and procedural fairness (like Bulgaria). To put this question into context, it should be said that modern tax administrations have two major types of controlling approaches:

1. Enforcement activities applied at risks and in sectors where the expected benefit will outweigh the cost of the action(s).

2. Enforcement activities applied at risks and in sectors where the moral effect will have wider effect in the society – to support a culture of voluntary compliance and tax commitment. These types of actions are undertaken predominantly on non-profitable basis – i.e. the expected monetary return is smaller than the cost of the action(s).

The latter controlling approach is not usually a subject of cost-benefit analysis, because the intangible gains are considered to be worthy to the intended investments by revenue agencies. However, the first approach is based on detailed cost-benefit evaluation, for a number of reasons. From one perspective, to ensure cost-efficiency for tax administration – operational expenses to generated income ratio. From another perspective, to ensure that the most harmful tax evasive practices receive appropriate enforcement attention. The cost-benefit analysis is done prospectively and retrospectively, the results are then compared and changes to evaluation methodology implemented. The prospective part is based on the cost of known administrative and other resources in addition to prognosis for the expected revenue. The retrospective part is based on after treatment data through internal (revenue agency information) and external macro-economic information in addition to internal cost of the compliance actions. Both parts of the cost-benefit analysis are possible if a given tax administration possess real time feedback about its own costing procedures and a centralised way of extracting intelligence about tax compliance effects. This is only possible through an integrated business intelligence suit, which accounts for the default non-linearity of these processes.

From a policy perspective, the cost-benefit approach to enforcement activities provides multi-faceted justifiable rationale. Firstly, public money is only spent on revenue collection activities, which are deemed to be sufficiently rewarding. Secondly, enforcement efforts are fairly allocated to risks and sectors, which deserve deterrence approach initially while commitment building resources are preserved for other segments of the taxpayers. This is encouraging and rewarding to already compliant business and individual taxpayers, strengthening their institutional trust. On the contrary, lack of solid cost-benefit reasoning could expose a revenue agency to political pressures, external and internal misuse of regulative recourses and compromise its reputation amongst the taxpayers. In an intriguing theorisation about using the taxpayer choice to target tax enforcement Raskolnikov (2009) propose a reform to IRS where taxpayers are forced to reveal their motivations for paying taxes.

Those who chose to cheat the system are assigned to deterrence measures and the rest to significantly cheaper cooperative enforcement measures. Practically this process has cost-benefit stratification characteristics. Precise targeting of sophisticated and non-sophisticated evaders versus mass deterrence has much lower social price too. It could be suggested that Bulgarian entrepreneurs are quite refined evaders when they engage in playing the system and therefore it is even more important to use a responsive regulation enforcement approach with them. There is evidence (Job, Stout & Smith, 2007, p. 98) in support of the fact that cultural change to responsive regulation of tax policy results in greater income for revenue offices. This is possible due to the readiness of tax offices to strengthen the taxpayers' motivational posture of "thinking morally" (Braithwaite, 2007; Braithwaite, Murphy & Reinhart, 2007).
In a study of ATO who are the innovator of the responsive regulation approach, Waller (2007) discovers that there is strong connection between the perceived institutional integrity and the success of a responsive regulation. The author confirms that in order for such responsive attitude to be efficient, institutional integrity should be present at the level of measures design and their everyday implementation. Unfortunately, NRA field officers' reputation is very questionable amongst Bulgarian tax payers(Pashev, 2005). However, based on the data from satisfied information access requests (Appendix 10 and 11) and the qualitative findings, it could be suggested that the level of integrity is high at department and strategy management level.

Due to the lack of holistic information management solution (at the time of study) NRA has only recently developed cost-benefit methodology in Risk Management department. In general, their methodology is based only on retrospective indicator tracking, which diagnose how particular risks reacted to the treatment mix applied and how this affected the tax gap. Inevitably, the development and application of such evaluation approach is very complicated and resourceful and therefore it signifies the strong will of the relevant department to align with modern risk management practices. As a result of the methodology's complexity only few officials are able to implement and benefit from it. As it became obvious, there was an agreement that actual risk treatment feedback is still underdeveloped. It should be emphasized that prospective cost-benefit evaluation function is not currently possible, because there is no technical and functional capacity to support it. This organisational shortcoming allows for a greater level of subjectivism in risk prioritisation and consequentially creates increased chances for enforcement pressure over already sufficiently compliant tax payers (asymmetrical tax gap closure). Verboon et al. (2007) conducted an intriguing self-interest analysis of distributive justice and tax compliance. The authors acknowledge that self-interest considerations are strong determinant of why people care about distributive justice. In doing so, they confirm that economic and psychological approaches are not necessarily as divergent as is often thought in explaining the phenomenon. If we extrapolate their findings to the situation of the Bulgarian entrepreneur, it could be suggested that favourable tax treatment outcomes would have a positive effect on compliance, while unfavourable treatment will have much stronger opposite effects. This would be so, due to the very low social cohesion index (Figure 4), which could suggest stronger correlation between self-interest and distributive justice in tax compliance.

The new business intelligence suit (more details here), which is expected to be deployed at NRA will also prepare the revenue administration for full integration into the so prolonged pursued e-government (specifically through the provision of e-taxation). It is questionable though, to what level such an advanced solution would be utilised. The experience from a geographically close regional scenario in Turkey (Stafford & Turan, 2011) signifies two important aspects. Firstly, the fact that the e-revenue function is a paramount to every e-government and secondly the fact that "e-

government lies in the likelihood of close public scrutiny of revenue-producing and revenue-linked e-government systems". Bulgaria is known with the repeated expensive failures to introduce e-government services. It could be argued that this is not due to the lack of technical capacity, best practices examples, or legislative basis, but it is due to the lack of political will. Government after government demonstrates unwillingness to perform transformational paradigm shift and support this process. Such significant institutional reform would leave greatly reduced capabilities for corruption according to the evident experience in countries, which adopted this governance approach. Clearly, NRA will be endangered to politically ground regulative and normative pressures during its own way of bottom-up modernisation.

NRA self-critiquing perspective is that their level of customer service should change in its attitude and way of delivery – in particular to adopt the customer orientated approach and to offer more e-services. The attitude of treating the taxpayers as customers is popular amongst contemporary tax administration and is part of the so called soft policy approach. In principle, it is a complicated organisational change for a regulative agency to address the "users" of its services in practice as customers in reality. Via an impactful policy study of this contemporary question Tuck, Lamb and Hoskin (2011) outline the main difficulties, which public institutions may encounter, specifically the regulative ones such as HMRC in the UK. The main obstacle to such attitudinal paradigm shift comes out of the fact that tax administrations must regulate taxpayers/customers in accordance with the law and they cannot choose the way they are regulated or choose which rules and legislation apply to them. But, still consumer orientation change in tax administration and the balance between control and service continue to be developed over the past decades. It would be much easier to treat the taxpayers as clients in countries where the share of informal economy and tax evasion in general is relatively low (<15%). This would be so, due to the greatly decreased necessity for enforcement activities in comparison with countries with share of informal economy in excess of 15%. In a comparative study of American and Norwegian tax systems Aberbach and Christensen (2007) finds out that administrations such as IRS, which adopt the doctrine of consumer friendliness aggressively, are environmentally forced to balance their approach due to the fact that they still need to collect revenue. But if IRS is one the positive side of extreme consumer orientation while Norwegian tax administration occupies a position in the middle, then it could be suggested that NRA has a significant capacity towards customer friendliness before any diminishing returns start to appear.

5.2 **Conclusions.**

To sum up, it should be emphasized that NRA functions under great environmental pressures. The most significant one is the political influence to asymmetrical gap closure and the consequential misuse of collected revenue. NRA is forced to eradicate non-compliance from sectors and businesses left without government protections while, others who enjoy protectionism are left to sustain much lower compliance levels. Consequentially, the tax administration is pushed to a state of unjustifiable regulative formalism, which is both lacking adequate information technology support and is based to debatable cost-benefit rationale. This has significant negative effects to procedural fairness and distributive justice, which is suggested to be below the acceptable social minimum for entrepreneurs.

As evident from the qualitative findings from the interviews, the current NRA policy approach is distinctively deterrence orientated, although there are some commitment campaigns, which start to gain internal and external popularity. Although, NRA officials feel that the level of customer service and friendliness should improve, the

organisational culture has still much to evolve in order to prepare for a responsive regulation approach. NRA officials are keen to acquire knowledge about the real motives and behavioural models of tax evaders, but they still do not have an established practice to do so (except some very limited attempts). NRA's perspective on evasion is that this is culturally embedded phenomenon, it is being done (at business level) for profit maximisation and that, smaller businesses are likely to evade more taxes. Generally, NRA suffers from information deficiency in regards to real evasive motives and the reasons for that are in the limited institutional resources, but not due to the lack of research interest. Therefore, NRA's compliance measures are predominantly being informed by very inflexible regulative and legislative framework, but not by evidence based approach. In order to acquire enriched picture of the situation, it will be explored (at entrepreneurial level) what Bulgarian businessmen think about the level of customer service and policy approach they are facing.

Based on the qualitative findings it became evident that another big obstacle to NRA tax compliance methodology is the so called bank secrecy. In order to match declared income with financial and non-financial assets, access to bank accounts is required. In general, such access could be acquired in the presence of good reasons and granted accordingly by a court of law. This makes NRA's work very difficult, slow and on certain occasions prohibitively expensive (from institutional perspective). In order to resolve this issue legislative changes are required. That would allow much more efficient expenditure of institutional efforts and will make NRA's work significantly less intrusive.

An important correlation between the shadow economy size and the overall index of social cohesion (OISC) was demonstrated where the lower OISC of a society is a higher level of informality exists. This link could provide rationale for a policy shift towards indirect measures of influencing two important socio-economic aspects – the OISC and the institutional asymmetry to decrease the size of the informal economy. The economic and institutional environments are amongst the worst in the EU in terms of socio-economic indicators. Furthermore, the state administration in Bulgaria is traditionally subjugated by the governing political class through illegitimate methods – public officials are often forced to act in greater alignment to political interests than in alignment with the legislative framework. Although, this has harmful effects on public service motivation, it was found out that NRA has some very promising management potential. Some key officials exercise "bottom to top" initiative of internal reformation in order to modernise NRA's practices and secure a greater level of operational independence. However, NRA's reforms should overcome strong political resistance in order to align the strategic power relationship with ruling governments. It is only time, which will prove whether genuine institutional reforms could be successful in Bulgaria if not started from top to bottom.

Chapter Six

6. Findings of business survey.

The previous chapter addressed the first out of two investigative levels (institutional and entrepreneurial), describing the specifics of the utilised CAQDAS qualitative analysis and the findings. It revealed the three major environmental pressures (institutional, economic and political) affecting NRA's role in revenue raising policies and the associated institutional asymmetry. Chapter 6 advances the study in the development of a holistic account of the socio-economic influencers of the processes of business tax evasion, formalisation and informalisation. The second investigative level (entrepreneurial) has been attained through the analysis of the business survey. Applying proven multi-disciplinary analytical methods (Decision trees, MLP Neural and TAN Bayesian networks, Multinomial Logistic Regression), the quantitative analysis advances an innovative predictive socio-economic profiling of business tax evaders. This chapter is entirely concentrated on the technicalities of the analysis itself and does not offer a discussion of findings in relation to the extant literature, which happens in the last chapter 7.
This chapter is consisted of two main sub-headings as follows:

1. Summary statistics for individual variables.
2. Modelling.

Part 1 contains traditional summary statistics for every individual variable from the business survey. It follows the structure of the questionnaire and reveals the main meaningful characteristics of the sampled population. In doing so, important inter-variables' relationships become evident and consequently underpin the elevated modelling inquiries, which emerge in Part 2. It provides advanced form of analysis and modelling looking into the complex relationships between interconnected variables and the associated findings. As findings materialise naturally from the analysis they all exist in the relevant sections of part 1 and 2 accordingly rather than being separated from the analysis. It is intentionally designed to be so and as a result enabling the readers to track how any given analytical method generated a particular finding. The findings presented in part 1 and 2 satisfy the quantitative research aims of this thesis (see tables 4 and 5).

6.1 **Summary statistics for individual variables.**

The first step in the statistical data analysis is to perform descriptive statistics for the individual variables and check for any unusual or unexpected values. This initial task reveals the main meaningful characteristics of the sampled population. It creates the basis for inferential statistics regarding the strength of the observed findings in relation with the studied population. The statistical frequencies are presented below following the order from the questionnaire for all variables accordingly. In terms of variables' naming convention, it should be noted that variable Q2 corresponds to question 2 from the business survey, variable Q3 to question 3 and so on. Every variable has been described in a table format for clarity and they have been grouped together by type. All descriptive statistic summaries precede the relevant cumulative tables for every variable. The key points are described within one paragraph per variable and at the end of it is indicated the corresponding table and variable number accordingly. There are two terms used to indicate whether a particular variable has missing values – user missing values and system missing values. The first term indicates that a respondent was presented with the question, but for some reason they have not provided a response. The second term (system missing) indicates that according to the display logic of the questionnaire, a particular question was not

asked accordingly to a group of respondents. Therefore, there is a number of missing answers.

As is evident (Table 11, variable Q2), the predominant type of businesses amongst the surveyed entrepreneurs are those trading with services at 50.6 percent. Trading with goods and services is 23.2 percent, while trading with goods only is 13.3 percent.

13.3 percent of the entrepreneurs stated that their main business subject is trading with goods only. 29.1 percent from them (3.9% out of total) require substantial stock and 56.4 percent (7.5% out of total) require non-substantial stock in order to operate (Table 11, variable Q3).

93.7 percent of the businesses employ up to 10 full time workers and 97.3 percent employ up to 10 part time workers. Companies employ on average 4 full time and 1 part time workers (Table 11, variable Q4).

In total 62.7 percent of the businesses are operated from within an office, while 31.2 are run from home (Table 11, variable Q5).

47.9 percent of the entrepreneurs run their business from a rented office and 13.1 percent own it. 37.3 percent operate their business from home. The system missing values indicate the cases where the question was not applicable based on respondents' earlier reply (Table 11, variable Q6).

89.3 percent of the businesses reported that they have been formally registered from the very beginning while 9.4 said that this has happened at a later stage (Table 11, variable Q7).

9.2 percent of all entrepreneurs reported that their business has been formally registered at a later stage (not simultaneously with commencement of trade). (Table 11, variable Q8).

28.1 percent of the entrepreneurs report that the debt to their business is in medium/acceptable amounts, while for 25.2 percent it is only small debt. For 8.2 percent the debt to their business is in substantial amounts (Table 11, variable Q9).

36.3 percent (28.1+ 8.2) of the entrepreneurs reported that there was debt to their businesses in medium or substantial amounts (variable Q9). From within this group, 26.7 percent (9.7% out of total) do not undertake debt recovery activities in that regard while 68 percent (24.7% out of total) do undertake (Table 11, variable Q10).

51.8 percent of the entrepreneurs do not use any forms of external business funding. 26.9 percent use business loans from banks, 21.5 percent use friends and family financial support, 15.3 percent use funding from non-bank organisations, 14 percent are utilising loans from business partners (suppliers and/or clients) and only 4.1 percent use credit cards. This is a multiple answer question and therefore the % out of total sum is greater than 100% (Table 11, variable Q11).

30.5 percent of the entrepreneurs were using business loans from banks and managed to achieve on average 7.39 percent interest rate. 13.4 percent of them reported interest rates of up to 6%. 23.7 percent of the entrepreneurs were using credit cards as external business funding and managed to achieve on average 7.21 percent interest rate. 39.8 percent of them reported interest rates not higher than 17% for this type of funding. 5.8 percent of the entrepreneurs were using business loans from non-bank organisations and managed to achieve on average 1.59 percent

interest rate. 7.7 percent of them reported interest rates not higher than 10%. 10.2 percent of the entrepreneurs were using business loans from their business partners and managed to achieve on average 1.41 percent interest rate. 16.8 percent of them reported interest rates not higher than 6%. 9.7 percent of the entrepreneurs were using business loans from friends and family and managed to achieve on average 1.88 percent interest rate (Table 12, variable Q12).

60 percent of the entrepreneurs either disagree or strongly disagree that the tax money is fairly spent. 19.6 percent either agree or strongly agree that the tax money is fairly spent. 20.1 percent of the entrepreneurs neither agree, nor disagree with the fairness in spending the tax money. 41.4 percent of the entrepreneurs either disagree or strongly disagree that NRA has fair professional attitude and treatment approach. At the same time, 30.8 percent either agree or strongly agree that NRA has fair professional attitude and treatment approach. 27.4 percent of the entrepreneurs neither agree, nor disagree. 54.3 percent of the entrepreneurs either disagree or strongly disagree that the state institutions have fair professional attitude and treatment approach. 21.1 percent either agree or strongly agree that the state institutions have fair professional attitude and treatment approach. 24.2 percent of the entrepreneurs neither agree, nor disagree (Table 13, variable Q13).

60.6 percent of the entrepreneurs either agree or strongly agree that doing business according to the rules is too expensive. 19.6 percent either disagree or strongly disagree that doing business according to the rules is too expensive. 19.4 percent of the entrepreneurs neither agree, nor disagree that doing business according to the rules is too expensive. 50.6 percent of the entrepreneurs either agree or strongly agree that tax evasion is the answer from the business to the poor tax and administrative policies. 28.1 percent either disagree or strongly disagree that tax evasion is the answer from the business to the poor tax and administrative policies. 20.6 percent of the entrepreneurs neither agree, nor disagree. 32.2 percent of the entrepreneurs either agree or strongly agree that tax evasion is an acceptable phenomenon in the current business environment. 44.5 percent either disagree or strongly disagree that tax evasion is an acceptable phenomenon in the current business environment. 22.5 percent of the entrepreneurs neither agree, nor disagree. 24.2 percent of the entrepreneurs either agree or strongly agree that using bribes is an acceptable phenomenon in the current business environment. 61.7 percent either disagree or strongly disagree that using bribes is an acceptable phenomenon in the current business environment. 13.3 percent of the entrepreneurs neither agree, nor disagree (Table 13, variable Q14).

Only 10.2 percent of the entrepreneurs reported that businesses from their sector are not pushed to pay informally “to get things done” – such as administrative services, tenders, licensing, regulative permissions and alike. According to 57.9 percent, the activity of paying "to get things done" costs between 1 to 35 percent of the business turnover (Table 14, variable Q15).

82.2 percent (29.3 + 19 + 17.4 + 15.6 + 0.9) of all responses are distributed amongst the first 5 possible answers (Q16_1, Q16_2, Q16_3, Q16_4 and Q16_5), indicating that for one reason or another there is turnover underreporting in order to evade taxes. 316 cases out of 413 have selected at least one of the first 5 possible answers (Q16_1, Q16_2, Q16_3, Q16_4 and Q16_5). This means that according to 76.5 percent of the entrepreneurs other businesses from their sector underreport turnovers in order pay less tax with varying motives. The three most frequent reasons for underreporting turnovers are paying less tax in order to survive (29.3% of responses, 45.7% of all cases), reducing costs to increase profits (19% of responses, 29.7% of all cases) and to pay less tax (17.4% of responses, 27.3 of all cases). In 65.7 percent

(29.3 + 19 + 17.4) of all turnover underreporting cases, this is caused by one of the previously described three top reasons or a random combination between them. Question 16 from the business survey is a multiple choice one with up to three possible answers. There are 59.3 percent of the respondents who selected one answer, 24.5 percent who selected 2 answers and 15.7 percent who selected 3 as the maximum allowed number of answers (Table 15, variable Q16).

Those who stated that there is underreporting of turnovers in their sector (N = 304) assessed that on average the concealing is 37.8 percent of the actual turnovers (Mean = 37.82, SD = 19.64). Variable Q17 represent question 17. There are no user missing values, but only system ones, due to the fact that this question was only asked to some of the respondents based on display logic (Table 14, variable Q17).

Variable Q18 represent question 18. There are only 2 user missing values. The question has been asked to all respondents. 58.6 percent of the entrepreneurs have bought goods or services for their businesses without the lawfully compulsory receipt or had purchased from suppliers who were not formally registered. 38.7 percent did not have such experience (Table 11, variable Q18).

Variable Q19 represent question 19. This question has been only asked to some of the respondents, based on their previous answers and associated display logic (N = 251). 33.5 percent of the entrepreneurs who experience purchasing of goods and services without a receipt manage to acquire invoices for up to 80% of their consumption as businesses. The entrepreneurs who experience purchasing of goods and services without a receipt, do usually receive fiscal document on average for 81.51 percent of what they buy as a business (M = 81.51, SD = 19.46). (Table 14, variable Q19).

Variable Q20 represent question 20. This question has been only asked to some of the respondents, based on their previous answers (N = 218). 87.2 percent of the respondents to this question would prefer all of their suppliers to provide them with invoices. Those who answered the question this way constitute 46 percent of all cases (Table 11, variable Q20).

Variable Q21 and Q21_Text represent question 21, which is a multiple choice single answer type and has been asked to all respondents. 28.4 percent of the entrepreneurs consider that NRA's procedures and administrative capacity has medium hampering effect on their businesses. 21.7 percent (12.7 + 9) consider the hampering effect to be from significant to enormous. 19.1 percent of the respondents reported only small hampering effect, while for 19.1 percent there was no such effect (Table 11, variable Q21).

Variable Q22 and Q22_Text represent question 22, which is a multiple choice single answer type and has been asked to all respondents. 79.8 percent of all entrepreneurs consider the current tax rates and social security contributions to be impediment to their business to varying degree. 12.2 percent of the respondents do not consider these to have negative impact on their business, while for 4.2 percent the tax rates have a stimulating effect (Table 11, variable Q22).

Variable Q23 represent question 23, which is a slider scale one from 0% to 100% – it has been asked to all respondents. According to the entrepreneurs the tax rate on business profit should be on average 9.92 percent (M = 9.92, SD = 6.64). (Table 14, variable Q23).

Variables Q24_1 to Q24_10 including Group and Rank variable modes represent

question 24, which is a complex pick, group and rank type of question. It has been asked to all respondents. The top three business impediments creating the most difficulties for entrepreneurs are as follows:

- For 53.9 percent of all cases it is the unfair competition from the informal sector.
- The corruption in the state administration – 49.9 percent of all cases.
- Low efficiency of the state administration – 42.1 percent of all cases.

Respondents were able to select and rank up to three business impediments on a scale of 1 to 3 where 1 signified the most obstructing factor and 3 the least obstructing factor.

- Where unfair competition from the informal sector was one of the three possible factors, it was most frequently ranked 1 as the most obstructing one.
- Where corruption in the state administration was one of the three possible factors, it was most frequently ranked 1 as the most obstructing one.

Where low efficiency of the state administration was one of the three possible factors, it was most frequently ranked 2 as the second most obstructing one (Table 16, variable Q24).

Variable Q25 represent question 25, which is a slider scale one from 0 to 20 days – it has been asked to all respondents. 74.7 percent of all entrepreneurs spend on average from 1 to 5 days per month to deal with administrative requirements and regulations (N = 407, Mode = 5, Mean = 4.93, SD = 4.02). For 25.1 percent of the entrepreneurs it takes 5 days per month to deal with administrative requirements and regulations (N = 407, Mode = 5, Mean = 4.93, SD = 4.02). For 16.7 percent it is 2 days and for 14.3 it is 3 days per month. These are the three top highest frequency responses (Table 14, variable Q25).

Variable Q26 represent question 26, which is a slider scale one from 0 to 3000 BGN per day. The question has been asked to all respondents. The average cost of doing business per day is 466.5 while the most frequent value is 100 (N = 402, Mode = 100, Mean = 466.48, Median = 253, SD = 613.9). (Table 14, variable Q26).

Variable Q27 represent question 27, which is a single answer one. The question has been asked to all respondents. 62.5 percent of the entrepreneurs have been inspected by NRA tax office officials while 32.9 percent have not (Table 11, variable Q27).

Variable Q28 represent question 28, which is a slider scale one (from 0 to 30 days). The question has been asked to specific respondents only, based on their previous answers. Entrepreneurs spent on average 8 days to deal with an audit by NRA officials, 68.7 percent reported from 0 to 8 days expenditure in that regard. (N = 252, Mode = 2, Mean = 8.11, Median = 5, SD = 8.47). (Table 14, variable Q28).

Variable Q29 represent question 29, which is a multiple choice single answer one. The question has been asked to specific respondents only, based on their previous answers (N = 375). 59.5 percent of the entrepreneurs who answered this question (54% out of total) have not been expected or encouraged to provide informal benefits (bribes) to NRA officials regarding the conducted inspections. 10.4 percent said that they have been expected to behave that way, while 20.3 percent preferred not to answer this question. 9.9 percent of the entrepreneurs opted to use the free text box. 4 respondents out of the 37 (9.9 percent) confirmed that they have been expected to provide bribes – this is 1 percent from all respondent who answered the question (Table 11, variable Q29).

Variable Q30 represent question 30, which is a multiple choice single answer one.

The question has been asked to all respondents. 68 percent of the entrepreneurs reported that they are exposed to unfair competition from informal businesses while 26.6 percent were not (Table 11, variable Q30).

Variable Q31 represent question 31, which is a multiple choice single answer one. The question has been asked to specific respondents only, based on their previous answers (N = 281). 32.4 percent (from N = 281, or 22 percent from N = 413) of the entrepreneurs stated that they experience significant difficulties in regards with exposure to unfair competition from informal businesses. 61.2 percent experience difficulties, but they are able to cope with them. It could be argued that strictly formal businesses are supposed to suffer most from unfair competition from informal businesses. This is so, because they lack capabilities to deal with informal competitors due to the fact that they function according to rules and therefore pay the full cost of doing business formally. Their grey competitors do not pay the full price though. Enterprises, which function in partly informal mode with predominant adherence to the rules of business and tax, are also experiencing difficulties due to unfair competition. However, they are likely to cope with them much better in comparison with fully formal businesses. It could be argued that businesses, which experience significant difficulties due to unfair competition, do not engage in illicit compensatory strategies and are most likely operating in a fully formal mode (22% of all businesses). Following this line of thought, it could be suggested that 78 percent (100%-22%) of all entrepreneurs are functioning in partly informal mode (being in an informal chain of dependence). The value of 78 percent corresponds to a finding from variable Q16 that 76.5 percent of the entrepreneurs underreport turnovers in order pay less tax with varying motives (Table 11, variable Q31).

The first column from the left (Table 11 below) indicates the variable's number in accordance with the business questionnaire (see Appendix 18).

Table 11: Summary statistics.

No	Variable	Number (%)
2	What is your main business subject?	
	Trading with goods.	55 (13.3)
	Trading with services.*	209 (50.6)
	Trading with goods and services.	96 (23.2)
	Something else – free text.	53 (12.8)
	Total	413 (100)
	*Mode = 2	
3	Do you need to keep stock and in what quantities?	
	Yes, the nature of our business requires maintenance of substantial stock availability.	16 (3.9)
	Yes, but not substantial stock availability.*	31 (7.5)
	No, the nature of our business does not require maintenance of stock supplies.	8 (1.9)
	Total	55 (13.3)
	*Mode = 2	
5	Does your business have an office or you run it from home?	
	Yes, there is an office.*	259 (62.7)
	No, there is no office – it is run from home.	129 (31.2)
	Something else – free text box.	25 (6.1)
	Total	413 (100)
	*Mode = 1	
6	Is the office owned by the business or is it rented?	
	The office is owned by the business/the company.	54 (13.1)
	The office is rented.*	198 (47.9)
	Something else – free text box.	7 (1.7)
	System missing	154 (37.3)
	Total	413 (100)
	*Mode = 2	
7	Have you formally registered your business since it began operating or this has happened at a later stage?	
	Yes, from the very beginning of business operations.*	369 (89.3)
	No, this has happened at a later stage.	39 (9.4)
	Something else – free text box below.	5 (1.2)
	Total	413 (100)
	*Mode = 1	

No	Variable	Number (%)
9	Are there any outstanding debts to your business and approximately what their size is?	
	Yes, there are, but in small/insignificant amounts.	104 (25.2)
	Yes, there are in medium/acceptable amounts.	116 (28.1)
	Yes, there are in substantial amounts.	34 (8.2)
	No.*	157 (38)
	Something else – free text box.	2 (0.5)
	Total	413 (100)
	*Mode = 4	
10	Do you undertake debt recovery activities to collect the outstanding amounts?	
	Yes.*	102 (24.7)
	No.	40 (9.7)
	Something else – free text box.	8 (1.9)
	System missing.	263 (63.7)
	Total	413 (100)
	*Mode = 1	
11	Do you use any forms of business finance/credit?	
	Business loans from banks.	111 (26.9)
	Business finance/credit such as credit cards.	17 (4.1)
	Business loans from non-bank institutions/organisations.	63 (15.3)
	Loans from your business partners (suppliers or clients).	58 (14)
	Loans from friends and family.	89 (21.5)
	Do not use any forms of external funding.*	214 (51.8)
	Total	552 (133.7)
	Mode = 6	
18	Have you ever bought goods or services for your business without an invoice or from suppliers not formally registered?	
	Yes.*	242 (58.6)
	No.	160 (38.7)
	Something else.	9 (2.2)
	User missing	2 (0.5)
	Total	413 (100)
	*Mode = 1	
20	Would you prefer all of your suppliers to provide you with invoices?	
	Yes.*	190 (46)
	It does not matter.	16 (3.9)
	No.	7 (1.7)
	Something else.	5 (1.2)
	System missing	195 (47.2)
	Total	413 (100)
	*Mode = 1	

No	Variable	Number (%)
21	How would you assess NRA's work (procedures) and in particular its administrative capacity in regards with your business?	
	There is stimulating effect on the business.	2 (0.5)
	It does not hamper the business.	111 (27.1)
	There is small hampering effect.	78 (19.1)
	There is medium hampering effect.*	116 (28.4)
	There is significant hampering effect.	52 (12.7)
	There is enormous hampering effect.	37 (9)
	Something else – free text box.	13 (3.2)
	User missing	4 (47.2)
	Total	413 (100)
	*Mode = 4	
22	What is your perception about the way the current tax rates and social security contributions affect your business?	
	They constitute small impediment to the business.	55 (13.4)
	They constitute medium impediment to the business.*	155 (37.9)
	They constitute significant impediment to the business.	88 (21.5)
	They constitute enormous impediment to the business.	32 (7.8)
	They do not constitute impediment to the business.	50 (12.2)
	The tax rates have stimulating effect on the business.	17 (4.2)
	Something else – free text box.	12 (2.9)
	User missing	4 (1)
	Total	413 (100)
	*Mode = 2	
27	Have you ever been inspected by NRA tax officials?	
	Yes.*	258 (62.5)
	No.	136 (32.9)
	Something else.	14 (3.4)
	User missing values	5 (1.2)
	Total	413 (100)
	*Mode = 1	
29	During these inspection(s) visits have you been expected or encouraged to provide any form of informal benefit to the official(s)?	
	Yes.	39 (9.4)
	No.*	223 (54)
	I prefer to leave this question unanswered.	76 (18.4)
	Something else.	37 (9)
	System missing values	38 (9.2)
	Total	413 (100)
	*Mode = 2	

No	Variable	Number (%)
30	Do you experience unfair competition from unregistered or informal businesses (grey economy businesses)?	
	Yes.*	281 (68)
	No.	110 (26.6)
	Something else.	17 (4.1)
	User missing values	5 (1.2)
	Total	413 (100)
	*Mode = 1	
31	What level of difficulties does your business experience due to unfair competition from the informal sector?	
	There are no any difficulties being experienced.	15 (3.6)
	We experience only difficulties, which we manage to cope with.*	172 (41.6)
	We experience significant difficulties.	91 (22)
	Something else – free text box.	3 (0.7)
	System missing values	132 (32)
	Total	413 (100)
	*Mode = 2	

The first column from the left (Table 12 below) indicates the variable's number in accordance with the business questionnaire (see Appendix 18).

Table 12: Descriptive statistics of continuous variables.

No	Variable	Number (%)	Std. deviation	Mean	Min-Max
4	How many people are working for your business?				
	Full-time workers.	not applicable	8.16	4.27	0-50
	Part-time workers.	not applicable	3.91	1.58	0-50
8	When was the business registered?				
	Registered after start of trade.	38 (9.2)	3.5	2008	1997-2013
	System missing	375 (90.8)			
	Total	413 (100)			
12	What is approximately the lowest interest rate, which you have managed to achieve for the finance type you have				
	Business loans from banks.	126 (30.5)		7.39%	0%-100%
	Credit cards.	98 (23.7)		7.21%	0%-24%
	Loans from non-bank institutions/organisations.	24 (5.8)		1.59%	0%-46%
	Loans from your business partners (suppliers or clients).	42 (10.2)		1.41%	0%-45%
	Loans from friends and family.	40 (9.7)		1.88%	0%-100%
	Total	330 (79.9)			

Table 13: Descriptive statistics of variables 13 and 14.

	What is your opinion about the correct and fair expenditure of tax money?	What is your opinion about fair, professional attitude and treatment by NRA?	What is your opinion about fair, professional attitude and treatment by the state institutions?	
	Number (%)	Number (%)	Number (%)	
Strongly agree.	49 (11.9)	49 (11.9)	38 (9.2)	
Agree.	32 (7.7)	78 (18.9)	49 (11.9)	
Neither agree, nor disagree.	83 (20.1)	113 (27.4)	100 (24.2)	
Disagree.	110 (26.6)	89 (21.5)	113 (27.4)	
Strongly disagree.	136 (32.9)	82 (19.9)	111 (26.9)	
Missing	3 (0.7)	2 (0.5)	2 (0.5)	
Total	413 (100)	413 (100)	413 (100)	
Mean	3.61	3.19	3.51	
Median	4	3	4	
Mode	5	3	4	
Minimum	1	1	1	
Maximum	5	5	5	
Strongly agree = 1 Strongly disagree = 5				
	Doing business entirely according to the rules is too expensive?	Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria?	In the current business environment, tax evasion is an acceptable phenomenon?	In the current business environment, using bribes is an acceptable phenomenon?
	Number (%)	Number (%)	Number (%)	Number (%)
Strongly agree.	177 (42.9)	119 (28.8)	69 (16.7)	61 (14.8)
Agree.	73 (17.7)	90 (21.8)	64 (15.5)	39 (9.4)
Neither agree, nor disagree.	80 (19.4)	85 (20.6)	93 (22.5)	55 (13.3)
Disagree.	39 (9.4)	54 (13.1)	72 (17.4)	57 (13.8)
Strongly disagree.	42 (10.2)	62 (15)	112 (27.1)	198 (47.9)
System missing	2 (0.5)	3 (0.7)	3 (0.7)	3 (0.7)
Total	413 (100)	413 (100)	413 (100)	413 (100)
Mean	2.26	2.63	3.23	3.71
Median	2.00	2.00	3.00	4.00
Mode	1	1	5	5
Minimum	1	1	1	1
Maximum	5	5	5	5
Strongly agree = 1 Strongly disagree = 5				

Table 14: Descriptive statistics.

No	Variable				Number			Mean	Mode	Std. deviation	Min-Max
					Valid	System and User missing	Total				
15	In your view, what percentage of the turnover do other businesses from your sector are pushed to pay informally "to get things done" – such as administrative services, tenders, licensing, regulative permissions and alike?				404	8	413	27.6	10	26.6	0%-100%
	0% from turnover	10.2%				1					
	Between 1% and 35% of turnover	57.9%									
17	According to your judgement, with what percentage usually the declared turnovers are lesser in comparison with the actual ones?				304	109	413	37.8	30	19.6	0%-100%
						0					
19	For what part of the goods and services, which you need as a business, do you usually receive an invoice?				251	162	413	81.5	90	19.5	9%-100%
	Up to 80%	33.5%				0					
23	What would be the acceptable level of tax on business profit?				408	0	413	9.9	10	6.6	0%-70%
						5					
25	Approximately how many days per month do you spend to cope with administrative requirements imposed by various state regulations (such as business permissions, tax return, communication with NRA, NSSI and others)?				407	0	413	4.9	5	4	0-20
						6					
	1 to 5 days	2 days	3 days	5 days							
	74.7%	16.7%	14.3%	25.1%							

No	Variable		Number			Mean	Mode	Std. deviation	Min-Max
			Valid	System and User missing	Total				
26	Approximately what is your cost of doing business per day (all business expenses including cost of salaries and others)?		402	9 2	413	466.5	100	613.9	0-3000
28	How much time approximately did you spend in this connection?		252	161	413	8.1	2	8.5	0-30
	From 0 to 8 days	68.7%		0					

Table 15: Descriptive statistics of variable Q16.

Tax evasive motivations.				
		Responses		% out of total
		Number	Percent	
Q16_2	Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to survive?*	188	29.3%	45.7%
Q16_3	Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to reduce their costs and increase their profits?	122	19%	29.7%
Q16_1	Businesses from your sector underreport their actual turnovers with the purpose to pay less tax?	112	17.4%	27.3%
Q16_5	Businesses from your sector underreport their actual turnovers with the purpose to pay less tax, because they are forced to do so and they do not have another choice?	100	15.6%	24.3%
Q16_7	You do not know whether businesses from your sector underreport their actual turnovers with the purpose to pay less tax?	70	10.9%	17%
Q16_6	Businesses from your sector do not underreport their actual turnovers with the purpose to pay less tax?	30	4.7%	7.3%
Q16_8	There is something else about businesses from your sector, which underreport their actual turnovers with the purpose to pay less tax?	14	2.2%	3.4%
Q16_4	You do not see anything wrong in the fact that businesses from your sector underreport their actual turnovers with the purpose to pay less tax?	6	0.9%	1.5%
Total		642	100%	156.2%
*Mode There were 2 user missing values out of 413.				

Table 16: Descriptive statistics of variable Q24.

Which three factors create the most difficulties for your business?							
	Answers	Responses		% out of total	Mean	Mode	Min-Max
		Number	Percent				
Q24_3	Unfair competition from the informal sector.*	215	18.6%	53.9%	1.59	1	1-3
Q24_4	Corruption in the state administration.	199	17.3%	49.9%	1.81	1	1-3
Q24_5	Low efficiency of the state administration.	168	14.6%	42.1%	2.04	2	1-3
Q24_9	Lack of adequately educated/skilled labour force.	145	12.6%	36.3%	2.3	3	1-3
Q24_8	Difficult and expensive access to financing.	143	12.4%	35.8%	2.15	3	1-3
Q24_6	The laws and regulations.	101	8.8%	25.3%	2.2	3	1-3
Q24_7	Licenses and regulative regimes.	73	6.3%	18.3%	2.16	2	1-3
Q24_2	The tax rate.	60	5.2%	15%	1.82	1	1-3
Q24_1	The National Revenue Agency (NRA).	39	3.4%	9.8%	1.9	1	1-3
Q24_10	Something else.	10	0.9%	2.5%	2.1	3	1-3
Total		1153	100%	289%			
*Mode = 3							

6.2 **Modelling.**

The previous section (6.1 Summary statistics for individual variables) provided initial insights into the collected empirical information while summarising every individual variable with the relevant descriptive statistics. This heading provides an advanced level of analysis looking into the complex relationships between interconnected variables (survey questions) and the associated findings. In doing so, it satisfies the research aims of the second empirical method used in the thesis – the business survey. In particular, the analysis emerges from the correspondence between the structure of the survey with its gradual information extraction approach and the research aims (see tables 4 and 5). The analytical approach being used here employs a variety of modelling methods taken from machine learning, artificial intelligence and statistics. In order to fulfil the research aims, this part of the quantitative analysis provides answers to the main objectives of the business survey as specified in Questionnaire development and rationale section. The two tables in heading 4.4.2 offer detailed explanation of each survey block and in particular, what the questions sought to discover, including inter-variable dependencies. This part of the quantitative analysis does not follow the structure of the questionnaire, but rather the research logic behind it. As it was already described, the questionnaire used gradual information extraction technique, where many questions were leading in and profiling

the respondents. This approach has been used in order to escalate progressively to the major questions seeking to discover the most important information. Therefore, an appropriate analytical method of data inquiry is to present the findings starting from these fundamental themes of interest (headings 6.2.1, 6.2.2, 6.2.3 and 6.2.4)

6.2.1 Tax evasion diagnostics and motivations.

One of the key survey questions is number 16 (Appendix 18). It is from the questionnaire's "Block 4 – Diagnostics of informality, motivations and perceptions". This question was developed to diagnose turnover underreporting, as an indicator for partial mode of informality, including capture of motives and attitudes behind such economic behavioural patterns. Question 16 contains 8 possible answers in total, from which the first 5 are confirmatory of turnover underreporting simultaneously offering associations with different motivations. The essence of the findings is presented at the end of section 6.2.1.

Conceptual description of the predictive modelling approach:

1. Based on MWF creation of 5 predictive models as decision trees (for each of the 5 confirmatory answers) with the capability to predict or classify future real world observations based on sets of decision rules. Effectively, this is economic profile modelling of partial mode of informality.

2. Evaluation if a higher level model (from all 5) is possible using ensembles to increase predictive accuracy at a viable model deployability level.

Implications for policy development and scientific value:

A decision tree model in such context provides invaluable information to policy makers in two aspects:

1. Firstly, it shows what the characteristics of entrepreneurs involved in tax evasion are. This is achieved via very clear reasoning process based on either relevant statistical tests' significance or alternative algorithms such as C5 trees.

2. Secondly, it shows (through the predictive power of the model) what policy changes could potentially influence evasive practices and to what extent. Consequently, prospective policy analysis and evaluation are fully enabled this way.

Variable Q16_1

1. Step 1:

Using the Auto Classifier node, 432 models have been automatically tested and classified based on the best overall accuracy achievable. The 432 models exhausted several settings' combinations in each of the following decision tree models: CHAID, C5, C&R tree and QUEST. More details are available in "Auto Classifier – Training summary for Q16_1" in Appendix 19.

The best overall accuracy of 80.3% has been achieved by a model named E-CHAID Model Q16_1. The model naming convention is the modelling method first, followed by the variable it concerns. The model is a decision tree based on the use of 10 fields (variables).

2. Step 2:

The model from step 1 has been evaluated using the analysis node after some adjustment in the settings, which resulted in a small overall accuracy gain. The predictive power of the model secures correct prediction in 80.6 percent of all cases. Appendix 19 provides the performance evaluation, coincidence matrix and the visual representation.

3. Step 3:

Targeting value 1 (Yes) from variable 16_1, a rule set has been generated with 84.6% confidence (see the symbolic expression in Appendix 19). According to this rule set, entrepreneurs who answered with "Yes" (1st of 8 options on question 16) have the following profile characteristics:

1. They underreport their turnovers with 21% to 65%.
2. They spend up to 10 days per month to cope with administrative requirements imposed by various state regulations (such as business permissions, tax return, communication with NRA, NSSI and others).
3. Have been inspected by NRA officials and during these inspection(s) have been expected or encouraged to provide forms of informal benefit to the official(s).
4. Do not agree that doing business entirely according to the rules is too expensive.

The above rule set is only one of the many possible, which could be derived from this E-CHAID decision tree. Every rule set, as a decision tree's derivative, is only a partial expression of the predictive power of a tree. Given the specific purpose of the analysis, this rule set is considered relevant and worthy to discuss. The same modelling steps have been implemented for variables Q16_2, Q16_3, Q16_4 and Q16_5 below.

Variable Q16_2

1. Step 1:

Using the Auto Classifier node, 570 models have been automatically tested and classified based on the best overall accuracy achievable. The 570 models exhausted several settings' combinations in each of the following algorithms: CHAID, C5, C&R tree, QUEST, Decision List, Support Vector Machine (SVM) and Nearest Neighbour Analysis (KNN). More details are available in "Auto Classifier – Training summary for Q16_2" in Appendix 20.

The best overall accuracy of 81.1% has been achieved by a CHAID decision tree, which used 13 fields for the model development. Another model, which was E-CHAID decision tree (E-CHAID Model Q16_2) achieved 80.6%, however it used 14 fields. It was selected due to the higher explanatory power against negligible loss of overall accuracy.

2. Step 2:

The model from step 1 has been evaluated using the analysis node. The predictive power of the model secures correct prediction in 80.6 percent of all cases. Appendix 20 provides the performance evaluation, coincidence matrix and the visual representation.

3. **Step 3:**

Targeting value 1 (Yes) from variable 16_2, a rule set has been generated with 92.6% confidence (see the symbolic expression in Appendix 20). According to this rule set, entrepreneurs who selected option 2 "Yes, in order to survive" (2nd out of 8 options on question 16) have the following profile characteristics:

1. According to their judgement, the declared turnovers are lesser with up to 30% in comparison with the actual ones.
2. They do not use loans from friends and family as forms of business funding.
3. Have experienced unfair competition from the informal sector, but they have been managing to cope with it.
4. The low efficiency of the state administration is either the first or the second (out of three) most significant impediments to their business.
5. Their opinion about the fact that doing business entirely according to the rules is too expensive varies between strong agreements to neutral.
6. The cost of doing their business per day is more than 103BGN.
7. They disagree or are neutral that tax money is being spent in a correct and fair way.

Variable Q16_3

4. **Step 1:**

Using the Auto Classifier node, 570 models have been automatically tested and classified based on the best overall accuracy achievable. The 570 models exhausted several settings' combinations in each of the following algorithms: CHAID, C5, C&R tree, QUEST, Decision List, Support Vector Machine (SVM) and Nearest Neighbour Analysis (KNN). More details are available in "Auto Classifier – Training summary for Q16_3" in Appendix 21.

The best overall accuracy of 82.6% has been achieved by a CHAID decision tree, which used 13 fields for the model development. Another model, which named E-CHAID decision tree (E-CHAID Model Q16_3), achieved 79.7%. The predictive power of that model for "Yes" response type was better, so it was selected.

5. **Step 2:**

The model from step 1 has been evaluated using the analysis node. The predictive power of the model secures correct prediction in 79.7 percent of all cases. Appendix 21 provides the performance evaluation, coincidence matrix and the visual representation.

6. **Step 3:**

Targeting value 1 (Yes) from variable 16_3, a rule set has been generated with 79.2% confidence (see the symbolic expression in Appendix 21). According to this rule set, entrepreneurs who selected option 3 "Yes, they do it in order to reduce their costs and increase their profits" (3rd out of 8 options on question 16) have the following profile characteristics:

1. According to their opinion, the declared turnovers are lesser in comparison with the actual ones with 29% to 40%.
2. Some of these entrepreneurs have been inspected by NRA. Some of the inspected

were encouraged to provide bribe(s) to the officials.
3. Have been pushed to pay informally "to get things done" with up to 39% of their turnovers or refused to answer this question.
4. The low efficiency of the state administration is either one of the three biggest business impediments ranked 3rd or is not selected amongst the possible options.
5. They experience forms of unfair informal competition.

Variable Q16_4

7. Step 1:

Using the Auto Classifier node, 570 models have been automatically tested and classified based on the best overall accuracy achievable. The 570 models exhausted several settings' combinations in each of the following algorithms: CHAID, C5, C&R tree, QUEST, Decision List, Support Vector Machine (SVM) and Nearest Neighbour Analysis (KNN). More details are available in "Auto Classifier – Training summary for Q16_4" in Appendix 22. The best overall accuracy of 99.5% has been achieved by a C5 decision tree, which used 14 fields for the model development.

8. Step 2:

The model from step 1 has been evaluated using the analysis node. The predictive power of the model secures correct prediction in 99 percent of all cases. Appendix 22 provides the performance evaluation, coincidence matrix and the visual representation.

9. Step 3:

Targeting value 1 (Yes) from variable 16_4, a rule set has been generated with 91.3% confidence (see the symbolic expression in Appendix 22). According to this rule set, entrepreneurs who selected option 4 "Yes and I do not see anything wrong in this" (4th out of 8 options on question 16) have the following profile characteristics:

1. There are between 2 and 4 people working full time for the business.
2. Loans from business partners (suppliers or clients) are not used as a form of business finance.
3. Using bribes in the current business environment is considered normal or at least have not been rejected as unacceptable practice.
4. These firms consume goods and services in order to operate, but receive invoices for only up to 81% of them.

Variable Q16_5

1. Step 1:

Using the Auto Classifier node, 570 models have been automatically tested and classified based on the best overall accuracy achievable. The 570 models exhausted several settings' combinations in each of the following algorithms: CHAID, C5, C&R tree, QUEST, Decision List, Support Vector Machine (SVM) and Nearest Neighbour Analysis (KNN). More details are available in "Auto Classifier – Training summary for Q16_5" in Appendix 23. Similarly to Q16_4, the best overall accuracy here has been achieved by a C5 decision tree with 95.6%, which used 12 fields for the model development.

2. Step 2:

The model from step 1 has been evaluated using the analysis node. The predictive power of the model secures correct prediction in 95.6% percent of all cases. Appendix 23 provides the performance evaluation and coincidence matrix.

3. Step 3:

Targeting value 1 (Yes) from variable 16_5, a rule set has been generated with 86.4% confidence (see the symbolic expression in Appendix 23). According to this rule set, entrepreneurs who selected option 5 "Yes, because they are forced to do so and they do not have another chance" (5th out of 8 options on question 16) have the following profile characteristics:

1. They disagree that NRA has professional attitude and treatment.
2. These entrepreneurs fully disagree that tax money are being spent in a fair way.
3. Credit cards have not been used as a form of business finance.
4. Have been inspected by NRA.
5. There is at least 1 person working full time for the business.
6. Their opinion about NRA's procedures and in particular its administrative capacity in regards with their business, varies from neutral effect to medium hampering effect.

In order to provide clarity and comparability Table 17 summarises the results from the decision tree modelling of entrepreneurs' tax evasive motivations.

Table 17: Decision tree modelling of entrepreneurs' tax evasive motivations (variable Q16).

Variable	Tested model algorithms	Best model accuracy	Rule set
Businesses from your sector underreport their actual turnovers with the purpose to pay less tax.	432 models' configurations from: CHAID, C5, C&R tree and QUEST	E-CHAID at 80.3%	At 84.6% accuracy: 1. They underreport their turnovers with 21% to 65%. 2. They spend up to 10 days per month to cope with administrative requirements imposed by various state regulations (such as business permissions, tax return, communication with NRA, NSSI and others). 3. Have been inspected by NRA officials and during these inspection(s) have been expected or encouraged to provide forms of informal benefit to the officials. 4. Do not agree that doing business entirely according to the rules is too expensive.
Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to survive.	570 models' configurations from: CHAID, C5, C&R tree, QUEST, Decision List, Support Vector Machine (SVM) and Nearest Neighbour Analysis (KNN)	E-CHAID at 80.6%	At 92.6% accuracy: 1. According to their judgement, the declared turnovers are lesser with up to 30% in comparison with the actual ones. 2. They do not use loans from friends and family as forms of business funding. 3. Have been experiencing unfair competition from the informal sector, but they have been managing to cope with it. 4. The low efficiency of the state administration is either the first or the second (out of three) most significant impediments to their business. 5. Their opinion about the fact that doing business entirely according to the rules is too expensive varies between strong agreements to neutral. 6. The cost of doing their business per day is more than 103BGN. 7. They disagree or are neutral that tax money is being spent in a correct and fair way.

Variable	Tested model algorithms	Best model accuracy	Rule set
Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to reduce their costs and increase their profits.	570 models' configurations from: CHAID, C5, C&R tree, QUEST, Decision List, Support Vector Machine (SVM) and Nearest Neighbour Analysis (KNN)	E-CHAID at 79.7%	At 79.2% accuracy: 1. According to their opinion, the declared turnovers are lesser in comparison with the actual ones with 29% to 40%. 2. Some of these entrepreneurs have been inspected by NRA. Some of the inspected were encouraged to provide bribes to the officials. 3. Have been pushed to pay informally "to get things done" with up to 39% of their turnovers or refused to answer this question. 4. The low efficiency of the state administration is either one of the three biggest business impediments ranked 3rd or is not selected amongst the possible options. 5. They experience forms of unfair informal competition.
You do not see anything wrong in the fact that businesses from your sector underreport their actual turnovers with the purpose to pay less tax.	570 models' configurations from: CHAID, C5, C&R tree, QUEST, Decision List, Support Vector Machine (SVM) and Nearest Neighbour Analysis (KNN)	C5 at 99%	At 91.3% accuracy: 1. There are between 2 and 4 people working full time for the business. 2. Loans from business partners (suppliers or clients) are not used as a form of business finance. 3. Using bribes in the current business environment is considered normal or at least have not been rejected as unacceptable practice. 4. These firms consume goods and services in order to operate, but receive invoices for only up to 81% of them.
Businesses from your sector underreport their actual turnovers with the purpose to pay less tax, because they are forced to do so and they do not have another choice.	570 models' configurations from: CHAID, C5, C&R tree, QUEST, Decision List, Support Vector Machine (SVM) and Nearest Neighbour Analysis (KNN)	C5 at 95.6%	At 86.4% accuracy: 1. They disagree that NRA has professional attitude and treatment. 2. These entrepreneurs fully disagree that tax money are being spent in a fair way. 3. Credit cards have not been used as a form of business finance. 4. Have been inspected by NRA. 5. There is at least 1 person working full time for the business. 6. Their opinion about NRA's procedures and in particular its administrative capacity in regards with their business, varies from neutral effect to medium hampering effect.

The next modelling steps involve development and testing of 5 MLP Neural Networks for each variable accordingly: Q16_1, Q16_2, Q16_3, Q16_4 and Q16_5. As described earlier in the text, this analytical step has dual purpose – to suggest neural net models for description of the current data set and make them available in electronic form for analysing future longitudinal data sets. In doing so, to identify and rank each input field based on the strength of its relationship to the specified target variable. The results are presented in figures 59, 60, 61, 62 and 63 accordingly. Each figure shows the top 5 strongest influencers (predictor variables), which contributed to a particular answer by respondents. Table 18 unites the results from all figures in a tabular format, however only the top three predictors are summarised in descending order for clarity.

In order to build the 5 MLP Neural Networks all input variables from the 5 decision trees were taken cumulatively into the modelling for each of the target variables (from Q16_1 to Q16_5). Therefore, the Neural Networks have been pruned and purified by initial design, concentrating on their pattern recognition capability on the most important input variables. Because all input variables from the 5 decision trees are taken into the modelling of each of the 5 target variables simultaneously, any research bias and preliminary knowledge have been eliminated even before the training of the networks began. The neural network model for Q16_1 has been provided with the input variables from all decision trees (Appendix 24). The same method was employed for Q16_2 to Q16_5 inclusive. As a result, each neural network had the chance to capture patterns and relationships while being trained, which might not have been exposed by the decision tree modelling stage. Coincidence matrices, model accuracy summaries and stream details for each of the 5 neural networks are provided in Appendix 25, 26, 27, 28 and 29 accordingly. To increase the predictive accuracy and models' robustness, the following parameters were applied:

Neural network model: Multilayer Perceptron (MLP)
Hidden Layers: Automatically compute number of units

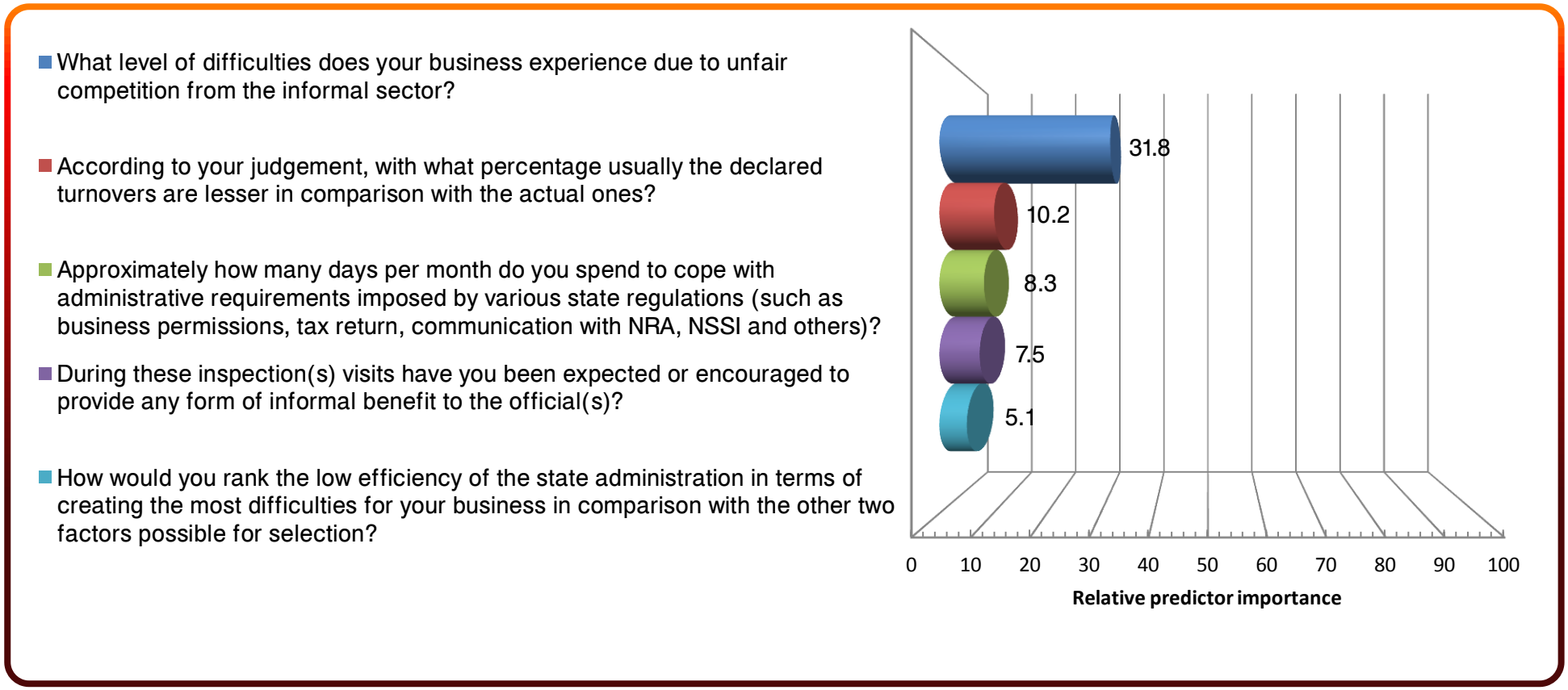
Overfit prevention set(%): 30.0
Replicate Results: true
Random seed: 229176228
Missing values in predictors: Impute missing values

Stopping Rules
Use maximum training time (per component model): false
Customize number of maximum training cycles: false
Use minimum accuracy: false

Ensemble
Number of component models for boosting and/or bagging: 170

Figure 19: Neural Network – relative predictors' importance for target variable Q16_1.

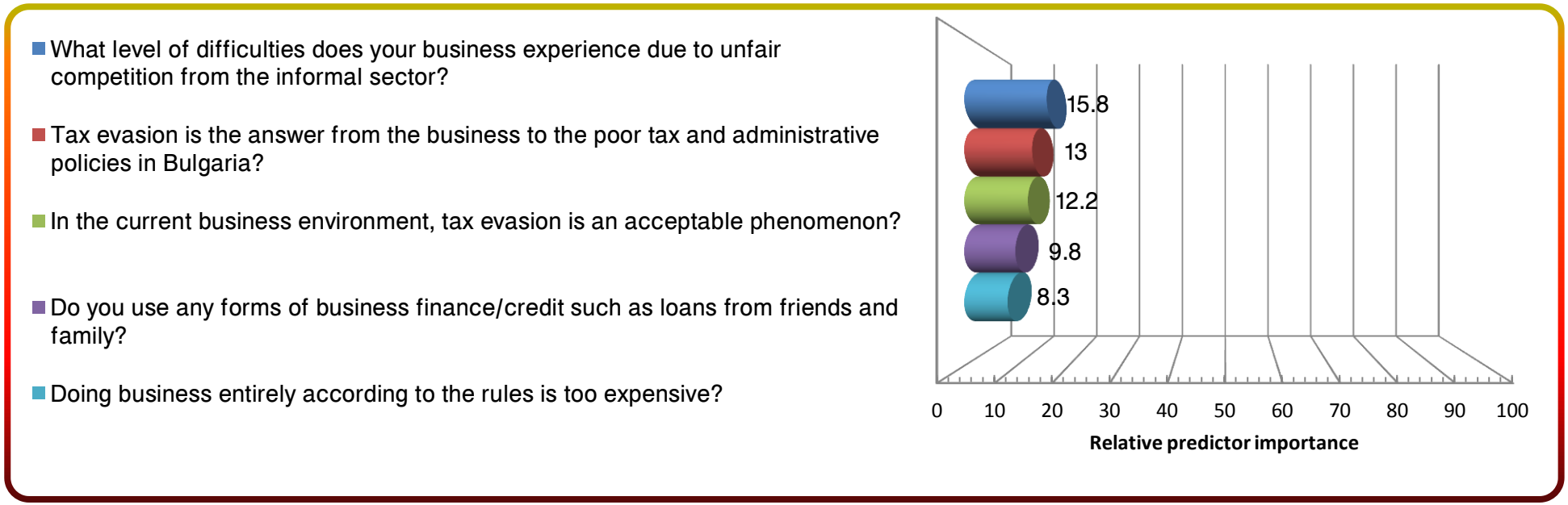
Target: Businesses underreport their actual turnovers with the purpose to pay less tax.



Note: Neural network modelling details are available in Appendix 25.

Figure 20: Neural Network – relative predictors' importance for target variable Q16_2.

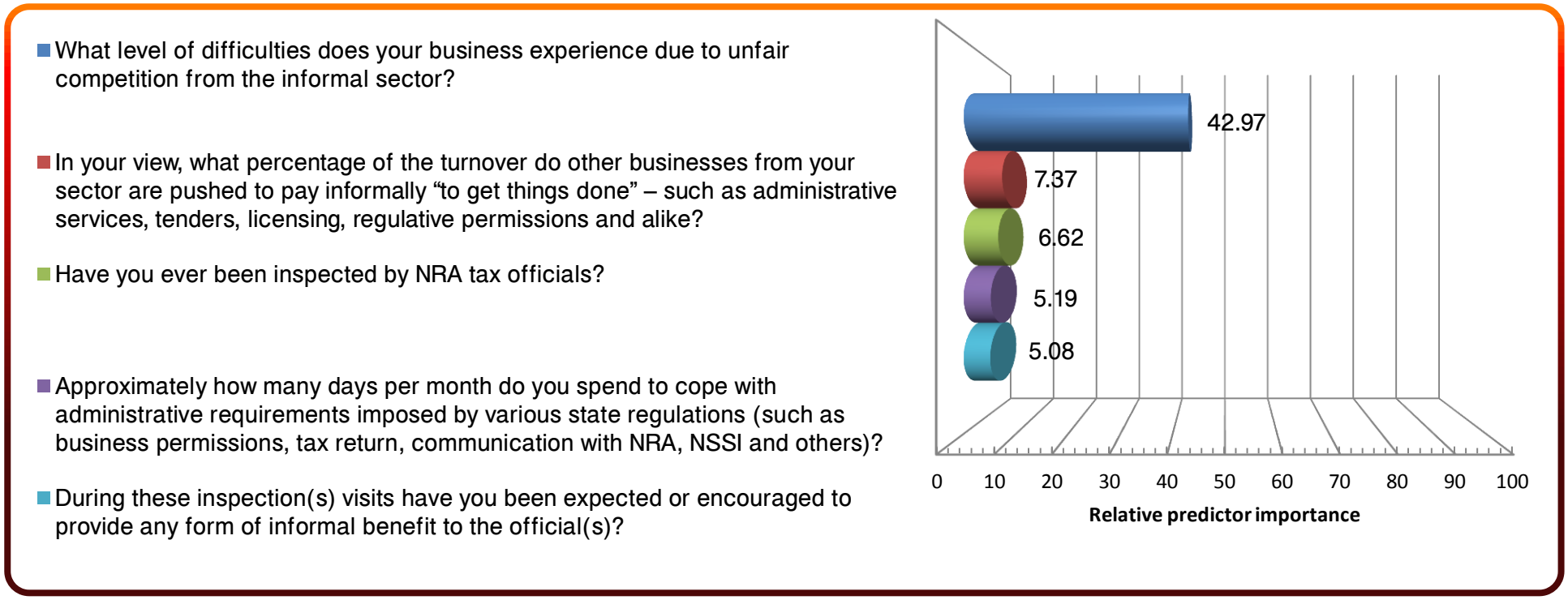
Target: Businesses underreport their actual turnovers with the purpose to pay less tax in order to survive.



Note: Neural network modelling details are available in Appendix 26.

Figure 21: Neural Network – relative predictors' importance for target variable Q16_3.

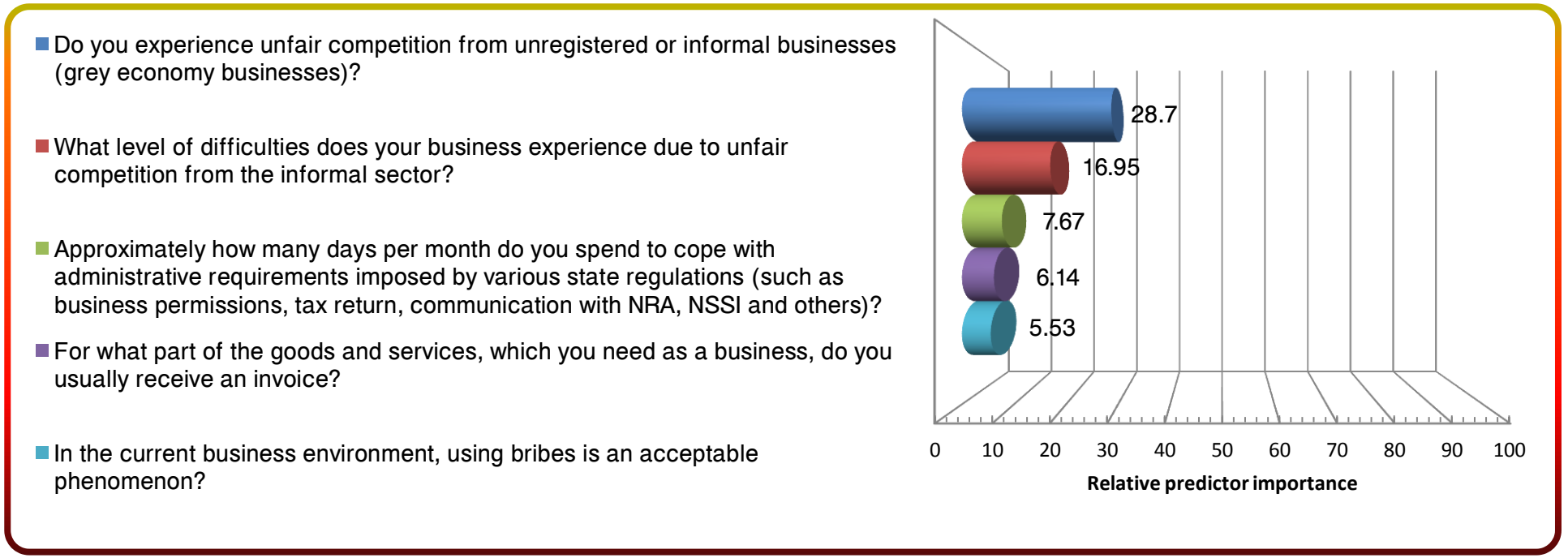
Target: Businesses underreport their actual turnovers with the purpose to pay less tax in order to reduce their costs and increase their profits.



Note: Neural network modelling details are available in Appendix 27.

Figure 22: Neural Network – relative predictors' importance for target variable Q16_4.

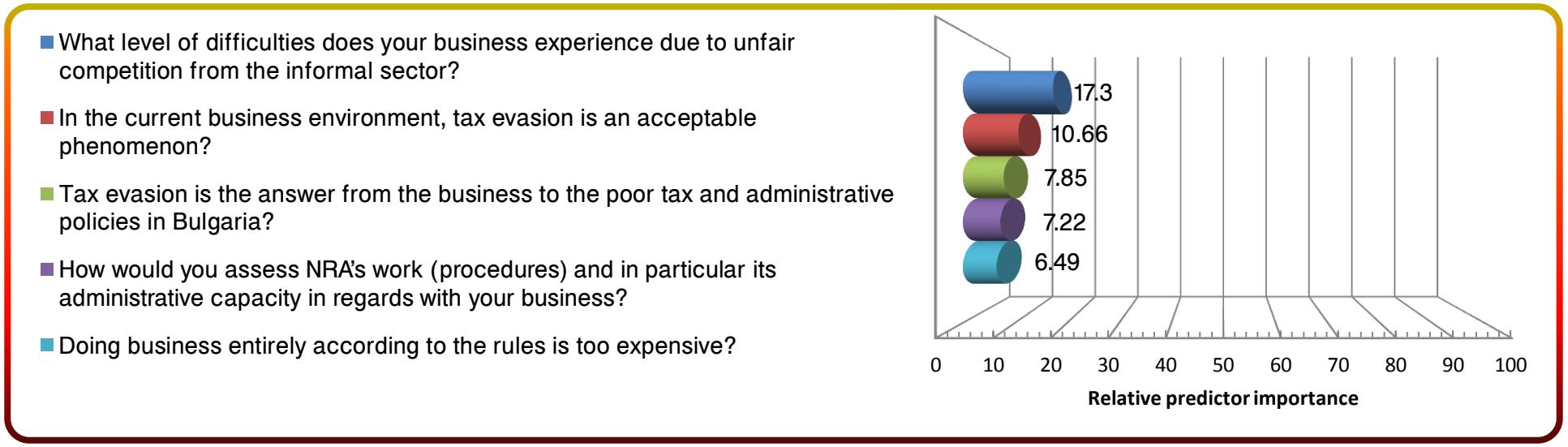
Target: There is nothing wrong in the fact that businesses underreport their actual turnovers with the purpose to pay less tax.



Note: Neural network modelling details are available in Appendix 28.

Figure 23: Neural Network – relative predictors' importance for target variable Q16_5.

Target: Businesses underreport their actual turnovers with the purpose to pay less tax, because they are forced to do so and they do not have another choice.



Note: Neural network modelling details are available in Appendix 29.

Table 18: Results from MLP Neural Network modelling of tax evasive motivations.

Variable	Selected answer	Predictors' relative importance in descending order
Q16_1	Yes.	What level of difficulties does your business experience due to unfair competition from the informal sector?
		According to your judgement, with what percentage usually the declared turnovers are lesser in comparison with the actual ones?
		Approximately how many days per month do you spend to cope with administrative requirements imposed by various state regulations (such as business permissions, tax return, communication with NRA, NSSI and others)?
Q16_2	Yes, in order to survive.	What level of difficulties does your business experience due to unfair competition from the informal sector?
		Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria?
		In the current business environment, tax evasion is an acceptable phenomenon?
Q16_3	Yes, they do it in order to reduce their costs and increase their profits.	What level of difficulties does your business experience due to unfair competition from the informal sector?
		In your view, what percentage of the turnover do other businesses from your sector are pushed to pay informally "to get things done" – such as administrative services, tenders, licensing, regulative permissions and alike?
		Have you ever been inspected by NRA tax officials?
Q16_4	Yes, and I do not see anything wrong in this.	Do you experience unfair competition from unregistered or informal businesses (grey economy businesses)?
		What level of difficulties does your business experience due to unfair competition from the informal sector?
		Approximately how many days per month do you spend to cope with administrative requirements imposed by various state regulations (such as business permissions, tax return, communication with NRA, NSSI and others)?
Q16_5	Yes, because they are forced to do so and they do not have another choice.	What level of difficulties does your business experience due to unfair competition from the informal sector?
		In the current business environment, tax evasion is an acceptable phenomenon?
		Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria?

Given the research aims of the study, it is important to discover the most powerful predictors (characteristics) of those respondents who think that tax evasion is an acceptable phenomenon. To answer this question a Multinomial Logistic Regression (MLR) has been employed for variable Q14_3. The reference category is the last one – "Strongly disagree". Of particular interest for the research aims are the respondents who answered with "Strongly agree" or "Agree" accordingly that tax evasion is an acceptable phenomenon. As evident from the case processing summary on the next page, there were 69 ("Strongly agree") and 64 ("Agree") observations. The distribution of observations amongst the factor (predictor) variables is also visible. The full distribution of observations for variable Q14_3 is available in table 13.

The chi-square test indicates a significant decrease (p = .000) of unexplained variance (850.91 – 644.95 = 205.95) in comparison with the baseline model (850.91). This means that the final model explains a significant amount of original variability, so it is a better fit than the original model. The Pearson and Deviance statistics are simultaneously strongly insignificant, which indicates that the predicted values are not significantly different from the observed ones – therefore, the model is a good fit to the data. The Cox and Snell's measure (.40) and Nagelkerke's adjusted value (.41) are similar and both indicate strong-sized effects.

Model Fitting Information						
	Model Fitting Criteria			Likelihood Ratio Tests		
Model	AIC	BIC	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	858.906	874.961	850.906			
Final	756.953	981.721	644.953	205.954	52	.000

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	485.214	460	.201
Deviance	403.875	460	.972

Pseudo R-Square	
Cox and Snell	.396
Nagelkerke	.413
McFadden	.159

The three most significant predictors (Table 19) of whether entrepreneurs strongly agree that tax evasion is an acceptable phenomenon are:

1. They strongly support that tax evasion is the answer from the business to the poor tax and administrative policies (p = .000).
2. Support that tax evasion is the answer from the business to the poor tax and administrative policies (p = .002).
3. They believe that underreporting of business turnovers is with the purpose to survive economically (p = .003).

The next subpopulation is defined by those who agree that tax evasion is an acceptable phenomenon (Table 19). As evident, there is little explanatory difference with the first group above. Their most significant predictors are:

1. The first three most significant predictors include the entrepreneurs who think that tax evasion is the answer from the business to the poor tax and administrative policies. Strongly agree, agree and neither agree, nor disagree are all significant at p = .000, p = .000 and p = .007 accordingly.
2. They believe that underreporting of business turnovers is with the purpose to survive economically (p = .016).

Case Processing Summary Multinomial logistic regression of variable Q14_3				
Variables		**Answers**	**Number**	**Marginal Percentage**
Dependent	In the current business environment, tax evasion is an acceptable phenomenon?	Strongly agree.	69	16.9%
		Agree.	64	15.6%
		Neither agree, nor disagree.	93	22.7%
		Disagree.	72	17.6%
		Strongly disagree.	111	27.1%
Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to survive?		Yes.	188	46.0%
		User not selected	221	54.0%
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria?		Strongly agree.	119	29.1%
		Agree.	90	22.0%
		Neither agree, nor disagree.	84	20.5%
		Disagree.	54	13.2%
		Strongly disagree.	62	15.2%
What is your opinion about the correct and fair expenditure of tax money?		Strongly agree.	49	12.0%
		Agree.	32	7.8%
		Neither agree, nor disagree.	82	20.0%
		Disagree.	110	26.9%
		Strongly disagree.	136	33.3%
What is your opinion about fair, professional attitude and treatment by the state institutions?		Strongly agree.	38	9.3%
		Agree.	49	12.0%
		Neither agree, nor disagree.	99	24.2%
		Disagree.	112	27.4%
		Strongly disagree.	111	27.1%
Valid			409	100.0%
Missing			4	
Total			413	

Table 19: Multinomial Logistic Regression for variable Q14_3.

Parameters estimates – Multinomial Logistic Regression for variable Q14_3								
The dependent variable is Q14_3. Strongly agree that in the current business environment, tax evasion is an acceptable phenomenon.	B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
							Lower Bound	Upper Bound
Intercept	-2.412	.600	16.155	1	.000			
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Strongly agree.	3.258	.606	28.856	1	.000	25.990	7.918	85.311
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Agree.	2.027	.669	9.169	1	.002	7.591	2.044	28.191
Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to survive? = Yes.	1.113	.372	8.967	1	.003	3.045	1.469	6.310
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Neither agree, nor disagree.	1.109	.659	2.836	1	.092	3.033	.834	11.031
What is your opinion about fair, professional attitude and treatment by the state institutions? = Neither agree, nor disagree.	-.690	.552	1.560	1	.212	.502	.170	1.481
What is your opinion about the correct and fair expenditure of tax money? = Disagree.	-.583	.516	1.275	1	.259	.558	.203	1.536
What is your opinion about fair, professional attitude and treatment by the state institutions? = Strongly agree.	.880	.968	.827	1	.363	2.411	.362	16.064
What is your opinion about the correct and fair expenditure of tax money? = Agree.	-.776	1.020	.579	1	.447	.460	.062	3.398
What is your opinion about fair, professional attitude and treatment by the state institutions? = Disagree.	-.366	.518	.498	1	.480	.694	.251	1.916
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Disagree.	.360	.801	.202	1	.653	1.433	.298	6.893
What is your opinion about the correct and fair expenditure of tax money? = Neither agree, nor disagree.	.132	.563	.055	1	.814	1.141	.379	3.439
What is your opinion about fair, professional attitude and treatment by the state institutions? = Agree.	-.049	.725	.005	1	.946	.952	.230	3.940
What is your opinion about the correct and fair expenditure of tax money? = Strongly agree.	-.052	.864	.004	1	.952	.949	.175	5.157

Parameters estimates – Multinomial Logistic Regression for variable Q14_3								
Agree that in the current business environment, tax evasion is an acceptable phenomenon.	B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
							Lower Bound	Upper Bound
Intercept	-2.933	.702	17.455	1	.000			
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Strongly agree.	3.192	.719	19.716	1	.000	24.329	5.947	99.535
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Agree.	3.360	.725	21.478	1	.000	28.784	6.951	119.195
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Neither agree, nor disagree.	1.946	.726	7.181	1	.007	7.001	1.687	29.060
Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to survive? = Yes.	.874	.363	5.803	1	.016	2.397	1.177	4.882
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Disagree.	1.348	.803	2.816	1	.093	3.848	.797	18.572
What is your opinion about fair, professional attitude and treatment by the state institutions? = Neither agree, nor disagree.	-.735	.549	1.792	1	.181	.480	.164	1.406
What is your opinion about the correct and fair expenditure of tax money? = Neither agree, nor disagree.	.254	.541	.220	1	.639	1.289	.446	3.721
What is your opinion about the correct and fair expenditure of tax money? = Agree.	.311	.833	.139	1	.709	1.365	.267	6.988
What is your opinion about fair, professional attitude and treatment by the state institutions? = Strongly agree.	-.270	1.145	.056	1	.813	.763	.081	7.195
What is your opinion about fair, professional attitude and treatment by the state institutions? = Disagree.	-.064	.495	.017	1	.897	.938	.355	2.476
What is your opinion about the correct and fair expenditure of tax money? = Disagree.	.036	.477	.006	1	.939	1.037	.407	2.644
What is your opinion about fair, professional attitude and treatment by the state institutions? = Agree.	-.020	.691	.001	1	.977	.980	.253	3.798
What is your opinion about the correct and fair expenditure of tax money? = Strongly agree.	-.021	.944	.000	1	.982	.979	.154	6.232

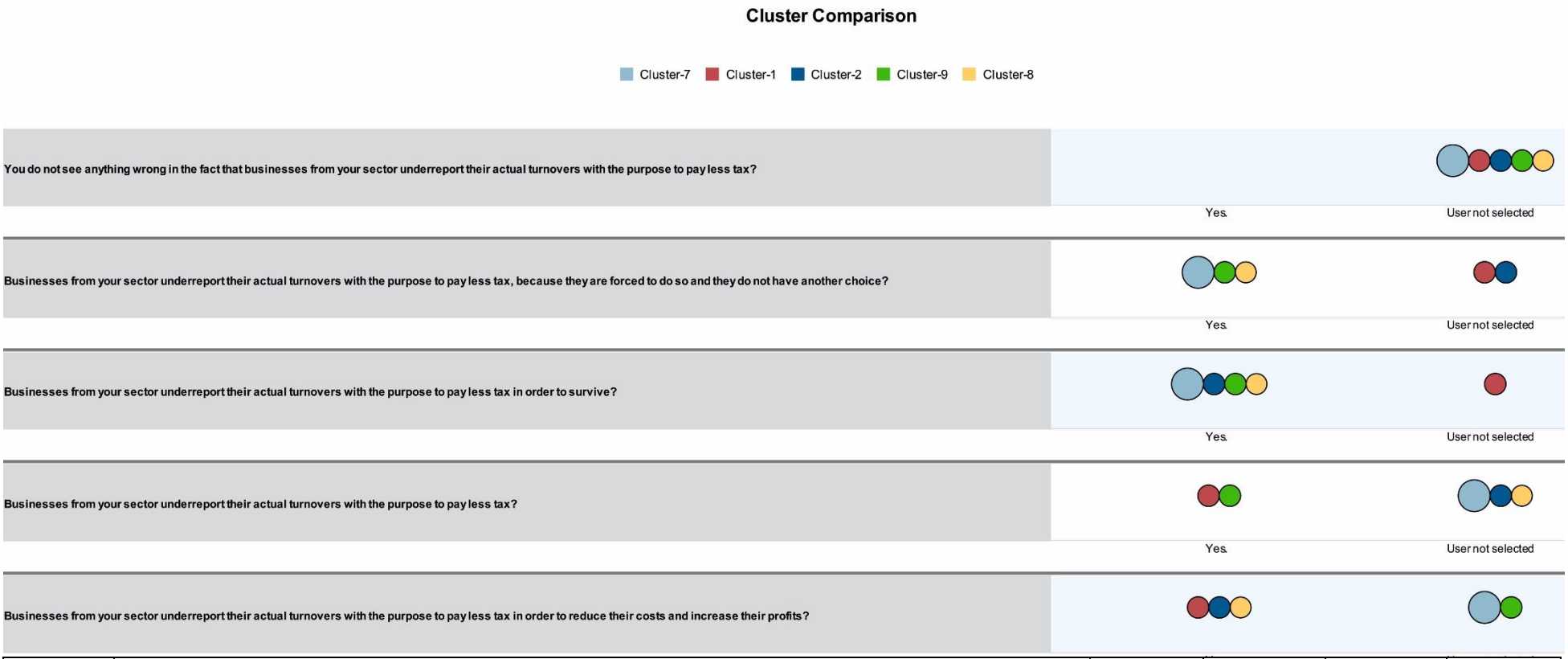
Parameters estimates – Multinomial Logistic Regression for variable Q14_3								
Neither agree, nor disagree that in the current business environment, tax evasion is an acceptable phenomenon.	B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
							Lower Bound	Upper Bound
Intercept	-2.634	.626	17.722	1	.000			
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Strongly agree.	3.221	.650	24.561	1	.000	25.060	7.010	89.589
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Agree.	3.217	.656	24.027	1	.000	24.951	6.894	90.304
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Neither agree, nor disagree.	2.618	.616	18.061	1	.000	13.707	4.098	45.844
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Disagree.	1.576	.690	5.223	1	.022	4.836	1.252	18.687
What is your opinion about fair, professional attitude and treatment by the state institutions? = Agree.	-1.403	.717	3.828	1	.050	.246	.060	1.003
What is your opinion about the correct and fair expenditure of tax money? = Agree.	1.441	.743	3.764	1	.052	4.223	.985	18.099
Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to survive? = Yes.	.530	.330	2.589	1	.108	1.700	.891	3.244
What is your opinion about fair, professional attitude and treatment by the state institutions? = Neither agree, nor disagree.	-.610	.481	1.611	1	.204	.543	.212	1.394
What is your opinion about the correct and fair expenditure of tax money? = Disagree.	.533	.431	1.532	1	.216	1.704	.733	3.965
What is your opinion about the correct and fair expenditure of tax money? = Neither agree, nor disagree.	.520	.500	1.079	1	.299	1.681	.631	4.482
What is your opinion about fair, professional attitude and treatment by the state institutions? = Disagree.	-.277	.455	.370	1	.543	.758	.311	1.849
What is your opinion about fair, professional attitude and treatment by the state institutions? = Strongly agree.	.510	1.084	.221	1	.638	1.665	.199	13.933
What is your opinion about the correct and fair expenditure of tax money? = Strongly agree.	-.415	.982	.178	1	.673	.660	.096	4.530

Parameters estimates – Multinomial Logistic Regression for variable Q14_3								
Disagree that in the current business environment, tax evasion is an acceptable phenomenon.	B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
							Lower Bound	Upper Bound
Intercept	-3.426	.745	21.117	1	.000			
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Agree.	3.669	.715	26.319	1	.000	39.220	9.654	159.329
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Neither agree, nor disagree.	2.729	.687	15.792	1	.000	15.324	3.988	58.887
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Disagree.	2.645	.695	14.490	1	.000	14.079	3.607	54.952
Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? = Strongly agree.	2.195	.804	7.456	1	.006	8.976	1.858	43.371
What is your opinion about the correct and fair expenditure of tax money? = Disagree.	.675	.483	1.953	1	.162	1.963	.762	5.056
What is your opinion about the correct and fair expenditure of tax money? = Agree.	.940	.786	1.431	1	.232	2.561	.549	11.950
What is your opinion about the correct and fair expenditure of tax money? = Neither agree, nor disagree.	.616	.533	1.335	1	.248	1.852	.651	5.267
What is your opinion about the correct and fair expenditure of tax money? = Strongly agree.	.493	.955	.267	1	.605	1.638	.252	10.644
Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to survive? = Yes.	.008	.359	.001	1	.982	1.008	.499	2.037
What is your opinion about fair, professional attitude and treatment by the state institutions? = Strongly agree.	.468	1.102	.181	1	.671	1.597	.184	13.840
What is your opinion about fair, professional attitude and treatment by the state institutions? = Agree.	-.066	.704	.009	1	.925	.936	.235	3.722
What is your opinion about fair, professional attitude and treatment by the state institutions? = Neither agree, nor disagree.	-.039	.558	.005	1	.945	.962	.322	2.873
What is your opinion about fair, professional attitude and treatment by the state institutions? = Disagree.	.347	.530	.428	1	.513	1.415	.501	3.997

As is evident from the modelling with decision trees and neural networks (summarised in Table 17 and 19), unfair competition from informal businesses is the most frequent and strongest predictor of tax evasion. It is highly likely that entrepreneurs who experience such competition would compensate their losses via turnover underreporting. In addition, entrepreneurs exercise tax evasion as a corresponding response to the poor tax and administrative policies, which cause them to spend more valuable resources than deemed acceptable. This finding was supported by the multinomial logistic regression according to which surviving economically is the second most significant predictor of tax evasion. On some occasion there were bribes demanding inspections from NRA, which appeared as a determinant of tax evasive behaviour.

The modelling workflows in this section (6.2.1), including the decision trees and neural networks, have been aimed at inter-variables' relationships between one target variable and several input variables. In order to fully exhaust this line of analytical enquiry (revealing tax evasive motivations), one more modelling technique has been utilised – a two-step cluster evaluation analysis based on log-likelihood as distance measure and Akaike Information Criterion (AIC). It has the advantage to test several cluster solutions selecting the best ones and to automatically estimate the optimal number of clusters for the training data (IBM, 2015b, p. 25). This method of clustering is necessary to unveil what is the associative pattern of respondents who provided more than one answer to question 16 (up to three out of 8 options). In particular, what is the pattern of answers if their selection is within the first 5 options, which are all indicating tax evasion with various possible motivations. The analysis is pointed only at those respondents who exercised their right to select 2 or 3 answers and all of them were within the first 5 options of question 16. Appendix 30 provides full details of the two step clustering. Nine clusters in total have been discovered and the model achieved average silhouette coefficient of 0.9. Considering that the silhouette coefficient is a "dimensionless quantity which is at most equal to 1" (Kaufman, Rousseeuw PJ, 1990, p. 87-88), evidently this cluster model is very robust in exposing the real patterns in the data. Figure 24 is an illustration of the cluster comparison. The upper section of the figure shows the patterns' distribution between the variables in a graphical form. For instance, it is visible that cluster 7 contains members (respondents) who provided two answers out of the maximum 3 allowed – two light blue circles. Members of cluster 9 provided 3 answers respectively – three green circles. The lower section of figure 24 is a table quantification of the very same data. Emerging from the cluster modelling, it is argued that the entrepreneurs who have provided more than one motive for tax evasion simultaneously have some important coincidence of their answer patterns. Firstly, turnover underreporting is exercised, so that entrepreneurs can survive economically and secondly, this illicit activity is practiced because they do not have another choice. Thirdly, tax evasion is performed for the reduction of operation costs and increasing profits.

Figure 24: Cluster comparison for variables Q16_1 to Q16_5 inclusive.



Variable		Clusters	Members	Clusters	Members
Q16_4	You do not see anything wrong in the fact that businesses from your sector underreport their actual turnovers with the purpose to pay less tax?	0	0	7, 1, 2, 9, 8	117
Q16_5	Businesses from your sector underreport their actual turnovers with the purpose to pay less tax, because they are forced to do so and they do not have another choice?	7, 9, 8	79	1, 2	38
Q16_2	Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to survive?	7, 2, 9, 8	98	1	19
Q16_1	Businesses from your sector underreport their actual turnovers with the purpose to pay less tax?	1, 9	38	7, 2, 8	79
Q16_3	Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to reduce their costs and increase their profits?	1, 2, 8	56	7, 9	61

6.2.2 Institutional quality, procedural fairness and distributive justice.

Question 13 is from the business survey's "Block 3 – Institutional quality and fair use of public money (see Appendix 18). Procedural fairness and distributive justice." It is represented with three sub-variables – Q13_1, Q13_2 and Q13_3. As is evident from the summary statistics for variable Q13, the bigger proportion of entrepreneurs disagree with the way tax money are being spent. The state institutions' services, including NRA, are deemed unfair and unprofessional from their perspective, which harms the perceived procedural fairness, and obstruct distributive justice accordingly. Most entrepreneurs spend on average more than 4 days per month to cope with administrative requirements imposed by various state regulations. There is a dual pattern of dissatisfaction amongst the entrepreneurs – simultaneously with the state institutions' quality, including NRA and the way tax money is spent. Evidently from the decision tree modelling below, entrepreneurs are likely to be pushed into evasive economic behaviour by the apparent poor quality of state institutions and NRA. The experience of unfair competition from informal businesses is an additional influencer. However, the same entrepreneurs still disapprove the use of bribes in the current business environment as a coping strategy and have registered their businesses formally from the very beginning of trading. Individual modelling for each of the three sub-variables is presented below. For clarity and comparability, Table 20 presents the cumulative findings.

Variable Q13_1

1. **Step 1:**

Using the Auto Classifier node, 3104 models have been automatically tested and classified based on the best overall accuracy achievable. The automated classification exhausted several settings' combinations in each of the following models: C5, Neural Networks, C&R Tree and CHAID tree. More details are available in "Auto Classifier – Training summary for Q13_1" in Appendix 31. The best overall accuracy of 81.5% has achieved by a C5 decision tree model. It is based on the use of 31 variables and is 17 levels in depth.

2. **Step 2:**

The model from step 1 has been evaluated using the analysis node. The predictive power of the model secures correct prediction in 80.9 percent of all cases. Appendix 31 provides the performance evaluation, coincidence matrix and the symbolic representation of the model.

3. **Step 3:**

Following the study's analytical line of enquiry, it is necessary to profile respondents who exhibited disagreement with the way tax money is being spent. Therefore, the target values of variable Q13_1 are "Disagree" and "Fully disagree" – mode 4 and mode 5 accordingly. Based on a rule set with 81.5% accuracy, the following table contains the most representative profile characteristics of those entrepreneurs (who think that public money is not fairly used).

Profile characteristics	Q13_1 value
Have registered their businesses formally from the very beginning.	"Disagree."
Doing business entirely according to the rules is too expensive.	
Experience unfair competition from informal businesses.	
Disagree that in the current business environment, using bribes is an acceptable phenomenon.	
Have been inspected by NRA tax officials.	
Disagree that NRA has fair and professional attitude.	
They strongly disagree that state institutions, including NRA have fair and professional attitude.	"Fully disagree."
They spend on average more than 7 days per month to cope with administrative requirements imposed by various state regulations (such as business permissions, tax return, communication with NRA, NSSI and others.	
In the current business environment, are supportive to tax evasion behaviour as an appropriate response, but disagree with the use of bribes.	
Doing business entirely according to the rules is not perceived as being too expensive.	

Variable Q13_2

4. Step 1:

Using the Auto Classifier node, 3104 models have been automatically tested and classified based on the best overall accuracy achievable. The automated classification exhausted several settings' combinations in each of the following models: C5, Neural Networks, C&R Tree and CHAID tree. More details are available in "Auto Classifier – Training summary for Q13_2" in Appendix 32. Similarly to variable Q13_1, the best estimated accuracy of 83.7% has been achieved by a C5 decision tree model. It is based on the use of 34 variables and is 14 levels in depth.

5. Step 2:

The model from step 1 has been evaluated using the analysis node. The predictive power of the model secures correct prediction in 94.2 percent of all cases. Appendix 32 provides the performance evaluation, coincidence matrix and the symbolic representation of the model.

6. Step 3:

In order to profile the respondents who think that NRA does not treat them in professional and fair way, mode 4 and 5 were the targeted ones – corresponding values "Disagree" and "Fully disagree" accordingly. Based on a rule set with 74.2% accuracy, the following table contains the most representative profile characteristics of those entrepreneurs. The logic of the question is reversed, so an answer "Disagree", means that a respondent disagrees that NRA has fair and professional attitude, and approach.

Profile characteristics	Q13_2 value
Disagree that the state institutions have fair and professional attitude.	"Disagree."
Spend more than 7 days per month on average to cope with administrative requirements imposed by various state regulations (such as business permissions, tax return, communication with NRA, NSSI and others).	
Have registered their businesses from the very beginning of trading.	
Strongly disagree that in the current business environment, using bribes is an acceptable phenomenon.	
Experience unfair competition from informal businesses.	
Disagree or strongly disagree that the state institutions behave in fair and professional way.	"Fully disagree."
Think that NRA's work and procedures have medium to enormous hampering effect on their business.	
Strongly disagree that tax money are being spent in a correct and fair way.	

Variable Q13_3

7. Step 1:

As with the previous two variables, the same types and number of models have been automatically tested and classified for Q13_3. The tested models were: C5, Neural Networks, C&R Tree and CHAID tree. More details are available in "Auto Classifier – Training summary for Q13_3" in Appendix 33. The best estimated accuracy has been achieved again by a C5 decision tree model at 94.6%. It is based on the use of 55 variables and is 17 levels in depth.

8. Step 2:

The model from step 1 has been evaluated using the analysis node. The predictive power of the model secures correct prediction in 82.8 percent of all cases. Appendix 33 provides the performance evaluation, coincidence matrix and the symbolic representation of the model.

9. Step 3:

In order to profile the respondents who think that the state institutions do not treat them in professional and fair way, mode 4 and 5 were the targeted ones – corresponding values "Disagree" and "Fully disagree" accordingly. Based on a rule set with 73.4% accuracy, the following table contains the most representative profile characteristics of those entrepreneurs. The logic of the question is reversed, so an answer "Disagree", means that a respondent disagrees that the state institutions have fair and professional attitude, and approach.

Profile characteristics	Q13_3 value
The laws and regulations are the first out of three factors creating the most difficulties to their businesses.	"Disagree."
They disagree that tax money are being spent in a correct and fair way.	
Have been inspected by NRA and on average spend more than 4 days monthly to cope with administrative requirements imposed by various state regulations (such as business permissions, tax return, communication with NRA, NSSI and others).	
Disagree that NRA has fair and professional attitude.	
Experience unfair competition from informal businesses.	
Have been inspected by NRA and have been encouraged to provide forms of informal benefit to the officials.	"Fully disagree."
Disagree that NRA has fair and professional attitude.	
In the current business environment, are supportive to tax evasion behaviour as an appropriate response.	
They disagree that tax money are being spent in a correct and fair way.	
Have formally registered their business since it began trading.	

Table 20: Decision tree modelling of entrepreneurs' perceptions of procedural fairness and institutional quality.

Variable	Tested model algorithms	Best model accuracy	Rule set	
Correct and fair expenditure of tax money.	3104 models' configurations from: Neural Networks, C5, C&R Tree and CHAID tree	C5 at 80.9%	At 81.5% accuracy:	
			Disagree	• Have registered their businesses formally from the very beginning. • Doing business entirely according to the rules is too expensive. • Experience unfair competition from informal businesses. • Disagree that in the current business environment, using bribes is an acceptable phenomenon. • Have been inspected by NRA tax officials. • Disagree that NRA has fair and professional attitude.
			Fully disagree	• They strongly disagree that state institutions, including NRA have fair and professional attitude. • They spend on average more than 7 days per month to cope with administrative requirements imposed by various state regulations (such as business permissions, tax return, communication with NRA, NSSI and others. • In the current business environment, are supportive to tax evasion behaviour as an appropriate response, but disagree with the use of bribes. • Doing business entirely according to the rules is not perceived as being too expensive.
Fair, professional attitude and treatment by NRA.	3104 models' configurations from: Neural Networks, C5, C&R Tree and CHAID tree	C5 at 82.8%	At 74.2% accuracy:	
			Disagree	• Disagree that the state institutions have fair and professional attitude. • Spend more than 7 days per month on average to cope with administrative requirements imposed by various state regulations (such as business permissions, tax return, communication with NRA, NSSI and others). • Have registered their businesses from the very beginning of trading. • Strongly disagree that in the current business environment, using bribes is an acceptable phenomenon. • Experience unfair competition from informal businesses.
			Fully disagree	• Disagree or strongly disagree that the state institutions behave in fair and professional way. • Think that NRA's work and procedures have medium to enormous hampering effect on their business. • Strongly disagree that tax money are being spent in a correct and fair way.

Variable	Tested model algorithms	Best model accuracy	Rule set	
Fair, professional attitude and treatment by the state institutions.	3104 models' configurations from: Neural Networks, C5, C&R Tree and CHAID tree	C5 at 94.2%	At 73.4% accuracy:	
			Disagree	• The laws and regulations are the first out of three factors creating the most difficulties to their businesses. • They disagree that tax money are being spent in a correct and fair way. • Have been inspected by NRA and on average spend more than 4 days monthly to cope with administrative requirements imposed by various state regulations (such as business permissions, tax return, communication with NRA, NSSI and others). • Disagree that NRA has fair and professional attitude. • Experience unfair competition from informal businesses.
			Fully disagree	• Have been inspected by NRA and have been encouraged to provide forms of informal benefit to the officials. • Disagree that NRA has fair and professional attitude. • In the current business environment, are supportive to tax evasion behaviour as an appropriate response. • They disagree that tax money are being spent in a correct and fair way. • Have formally registered their business since it began trading.

6.2.3 Informality chains of interdependence – economic barriers to formalisation.

Block 5 from the survey contains three questions and is designed to detect the so called informality chains of interdependence a business might be part of. This is a complex economic situation where an enterprise satisfies a proportion of their trade needs for goods and/or services without receiving invoices accordingly. In doing so, the participating firm cannot claim eligible expenses and develops strong motivation to offset the loss with a reciprocal action – not providing invoices to conceal taxable income. The informality chain of interdependence could be caused by pressure from other enterprises on the supply or the demand side of the business, or by both. To illustrate how it works in practice the following example is provided:

1. Firm A manufactures cardboard packaging. Firm A sells various cardboard packaging sets to firms B and C, which are French croissant manufacturers). They need the packaging in order to sell their products to their end customers. Firm A buys raw material to produce the cardboard packaging from firm D.

2. After a period of mutually beneficial business relationship, firm D suggest to firm A that they need to increase the price of the raw material with X%. However if firm A agrees to accept 30% of the monthly supply of raw material without invoice from firm D, then the prices would remain unchanged.

3. Firm A would be out of their competitive market niche if they have to transfer the raw material price increase with X% into their product (cardboard packaging). Firm A's clients (firm B and C) will not accommodate such an increase and will look to replace their supplier. Firm A, also considers if they are able to introduce any production optimisation to achieve lower cost of production, but they do not have the capital to do that. Therefore, firm A agrees to accept 30% of their monthly supply from firm D without invoice. It is either accepting that or losing market position and the financial feasibility of the entire business operation.

4. Firm A successfully negotiates a similar deal with firms B and C in order to offset the loss of eligible business expense claim. In the new trade mode, firms B and C acquire their usual supply of packaging, but approximately 30% of it is without invoice from firm A.

5. Firms B and C supply their croissants to small food shops and cafés. They try to persuade them with similar proposal for a proportion of the monthly supply without invoices.

6. Firms A, B, C and D start to exist in an informal chain of interdependence and therefore apply constant pressure up and down their relevant supply chains.

An enterprise consuming a proportion of their business goods and/or services without invoices could genuinely support this trade status for various reasons amongst which:

1. Exercising a similar level of informality together with their supplier(s) and evading taxes at the preferred level to increase profits.

2. Unable to achieve market competitive product or service if they start to acquire all their products and services with invoices (i.e. higher prices) from alternative fully formal suppliers.

3. Not having the resources or willingness to find alternative fully formal suppliers prepared to always issue invoices.

The corporate tax in Bulgaria is 10%. It could be argued that enterprises, which

manage to acquire invoice for less than 80% of their business needs, would be strongly incentivised to pursue compensatory strategies. Based on variable Q18 and Q19 (see Summary statistics of individual variables), it is evident that from the 58.6% who experience purchasing of goods and services without invoices, 33.5% acquire invoices not exceeding 80% of their consumption as businesses (Table 14). For that reason, this group of enterprises could be considered to function in informality chain of interdependence with varying magnitude of tax concealment and degree of complexity. As a matter of relevant importance to the research instrument's design, it should be noted that that any business customer in Bulgaria is lawfully entitled to receive an invoice and associated fiscal receipt (where applicable). The business supplier is lawfully obliged to invoice every sale. Claiming eligible business expense without the relevant paperwork is not possible.

The third question (out of three for survey's block 5) is represented by variable Q20. This question seeks to identify an important characteristic of an informality chain of interdependence. In particular, this is the conditional probability with which a firm would endeavour towards a higher level of formalisation as a function of the involvedness' level with partly informal suppliers. To resolve this analytical task a Tree Augmented Naïve (TAN) Bayesian Network has been utilised – the modelling steps are presented below. According to the survey's display logic, question 20 was asked to all respondents who reported that they are not receiving invoice for more than 5% of their business needs. As it is evident from the summary statistics of variable Q20, 87.2% of the respondents to this question would prefer all of their suppliers to provide them with invoices. The conditional probability, that a firm would prefer to receive invoices for all the business goods and services it needs to trade, increases together with the share of properly invoiced expenditures (see Table 21 below). The higher the proportion of goods and services for which a firm receives invoices, the more likely it is for such firm to demand invoices from all its suppliers and vice versa. If a firm is managing to obtain invoices for 45.4%-63.6% of the consumed business goods and services, then the conditional probability is only 8%. The next band (63.6%-81.8%) represents the first significant increase – to 29% accordingly. Firms with more than 81.8% of their business needs properly invoiced, represent conditional probability of 56% to demand invoices from all of their suppliers (formalisation pressure on the supply chain). The observed causal correlation demonstrates the coercive power, which an informality chain of interdependence exercises over its enterprise members – up and down the supply chain. Effectively, clusters of firms, which exist in informality chain of interdependence, act as enablers of informalisation processes and simultaneously as barriers to formalisation to both inside and outside firms. Explanation of the modelling stages follows below.

Variable Q20_1 – endeavour to formalisation

1. **Step 1:**

Using the Auto Classifier node, 64 configurations of TAN and Markov Blanket Bayesian Networks have been tested and ranked accordingly. More details are available in "Auto Classifier – Training summary for Q20_1" in Appendix 34. The best estimated accuracy has been achieved by TAN Bayesian Network at 87.16%.

2. **Step 2:**

The model from step 1 has been evaluated using the analysis node. The predictive power of the model secures correct prediction in 85.23 percent of all cases. The performance evaluation is in Appendix 34.

3. **Step 3:**

Table 21 summarises the findings from TAN Bayesian Network modelling of variable Q20_1 as a function of variable Q19. The conditional probabilities are presented in the last row of the table (4% to 56%).

Table 21: Effort to formalise from within an informality chain of interdependence.

Target variable Q20_1	Input variable Q19				
	For what part of the goods and services, which you need as a business, do you usually receive an invoice?				
Would you prefer all of your suppliers to provide you with invoices?	<=27.2%	27.2% ~ 45.4%	45.4% ~ 63.6%	63.6% ~ 81.8%	> 81.8%
Yes.	4%	2%	8%	29%	56%

The next important line of inquiry is to establish the characteristics and attitudes of firms, which operate in an informal chain of interdependence, but are striving towards full formalisation mode. This analytical task has been achieved through TAN Bayesian Network deployment over a set of pre-selected special cases. These firms underreport their turnover to varying extent and simultaneously manage to acquire invoices for only up to 80% of their business needs. However, all of them would prefer their suppliers to provide invoices for all transactions. The strongest predictor of such informal chain of interdependence status is the firm's full disagreement with fair expenditure of tax money. The conditional probabilities for such firms to disagree or fully disagree that NRA has fair and professional attitude are 29% and 50% accordingly. This means that 79% of all firms with the studied status will have the discussed characteristics. Similar conditional probabilities of 25% (disagree) and 61% (fully disagree) accordingly exist in regards with fair and professional attitude from the state institutions. This means that 86% of all firms with the studied status will be strongly dissatisfied with the state institutions. Unsurprisingly, the firms which would like to fully formalise breaking free from informality chain of interdependence, possess conditional probabilities to agree (29%) or fully agree (57%) that tax evasion is the answer from the business to the poor tax and administrative policies. This is equal to 86% of all firms. Explanation of the modelling stages follows below.

Variable Q20_1 – characteristics of informal chain of interdependence' firms.

1. **Step 1:**

Using the Auto Classifier node, 256 configurations of TAN and Markov Blanket Bayesian Networks have been tested and ranked accordingly. More details are available in "Auto Classifier – Training summary for Q20_1" in Appendix 35. The best estimated accuracy has been achieved by TAN Bayesian Network at 69.62%.

2. **Step 2:**

The best model from step 1 was not sufficiently accurate and therefore further training has been implemented. The predictive power of the improved model reached correct prediction in 96.2 percent of all cases. The performance evaluation and coincidence matrix are presented in Appendix 35.

3. **Step 3:**

Table 22 summarises the findings from TAN Bayesian Network modelling of variable Q20_1 as a function of several variables. The conditional probabilities are presented in the last row of the table's sections.

Table 22: Informality chain of interdependence, institutional quality and fair use of public money.
This table summarises modelling with TAN Bayesian Network. Conditional probabilities are expressed in percentages.

Target variable Q20_1	**What is your opinion about correct and fair expenditure of tax money? Input variable Q13_1.**				
Would you prefer all of your suppliers to provide you with invoices?	Fully agree	Agree	Neither agree, nor disagree	Disagree	Fully disagree
Yes.	9%	6%	17%	28%	41%

Target variable Q20_1	**Input variable Q13_1**	**What is your opinion about fair, professional attitude and treatment by the NRA? Input variable Q13_2.**				
Would you prefer all of your suppliers to provide you with invoices?	**What is your opinion about correct and fair expenditure of tax money?**	Fully agree	Agree	Neither agree, nor disagree	Disagree	Fully disagree
Yes.	Fully disagree.	7%	0%	14%	29%	50%

Target variable Q20_1	**Input variable Q13_1**	**What is your opinion about fair, professional attitude and treatment by the state institutions? Input variable Q13_3.**				
Would you prefer all of your suppliers to provide you with invoices?	**What is your opinion about correct and fair expenditure of tax money?**	Fully agree	Agree	Neither agree, nor disagree	Disagree	Fully disagree
Yes.	Fully disagree.	0%	4%	11%	25%	61%

Target variable Q20_1	**Input variable Q13_1**	**Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria? Input variable Q14_2.**				
Would you prefer all of your suppliers to provide you with invoices?	**What is your opinion about correct and fair expenditure of tax money?**	Fully agree	Agree	Neither agree, nor disagree	Disagree	Fully disagree
Yes.	Fully disagree.	57%	29%	11%	4%	0%

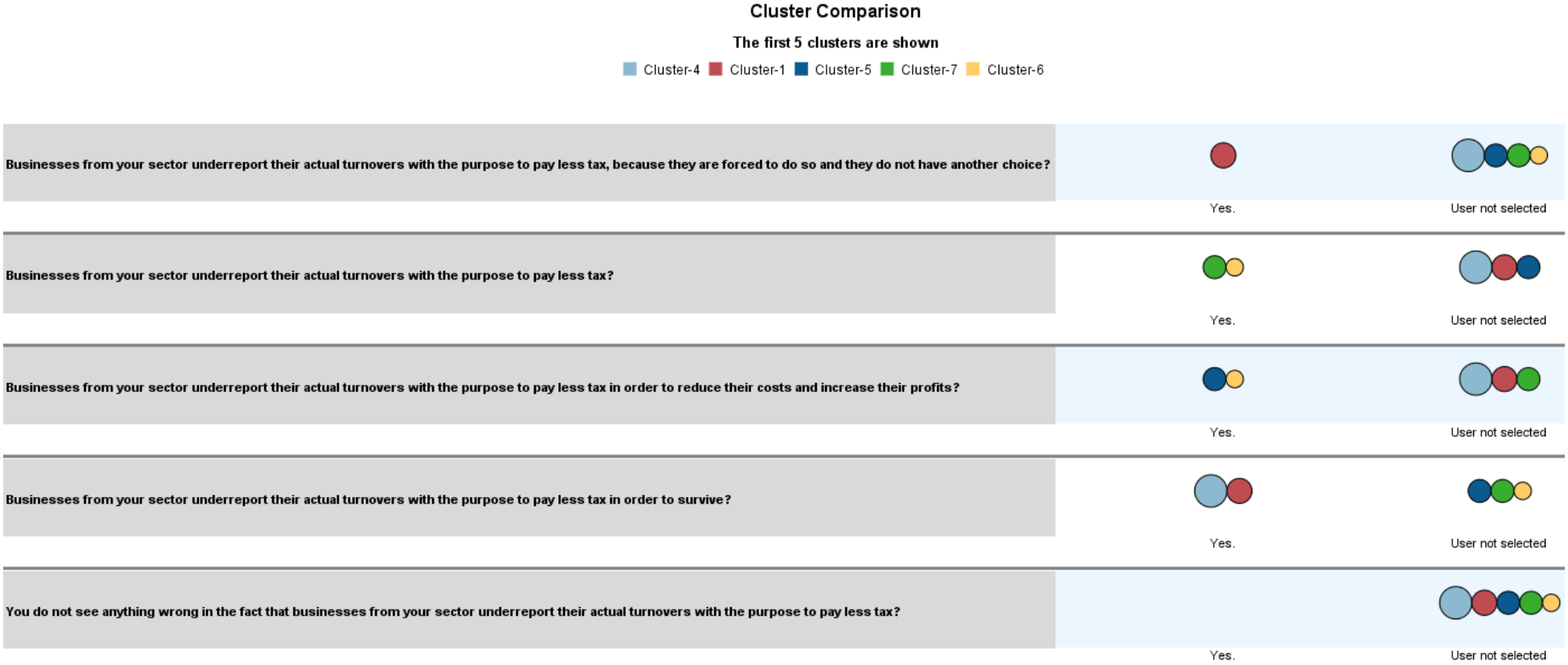
6.2.4 Unfair competition and turnover underreporting.

Block 7 from the survey (Appendix 18) contains two questions and is designed to detect consequences from unfair competition from informal businesses. Unfair competition constitutes a strong incentive for formal businesses to engage in informal compensatory strategies. Usually it results in turnover underreporting, which formal businesses might employ as a compensatory strategy against unfair competition. From an economic perspective unfair competition is simultaneously enabler of informality and barrier to formalisation. The enabling aspect is in the economic market force rendering a fully formal business unable to compete against one or more unfair competitors. To preserve competitiveness a formal business might need to informalise to a level corresponding to the unfair competitors. Unfair competition is also a barrier to formalisation due to the same inherent aspect. Assuming that a partly informal business operates as such mainly to cope with unfair competition, it could only fully formalise if a higher economic competitiveness in comparison with the unfair competitors is achievable. That economic competitiveness should be sufficiently higher to accommodate the full cost of formal mode of doing business. Understandably, it is a hardly attainable business target especially in typically low margin trades with non-specialised goods.
Evidently, from the individual summary statistics of variable Q31, 63.6% (263) of all firms (see Table 11) experience difficulties due to unfair competition from informal businesses. It was argued that 76.5% of all firms are functioning in partly informal mode. To establish a relationship between unfair competition and turnover underreporting it is important to find out how many firms of those experiencing difficulties due to unfair competition (N=263, 63.6% from N=413) also engage in turnover underreporting and by what motive. The first part of this analytical task is resolved by conditional case selection syntax via the Select Node. The modelling details are presented in Appendix 36. The second part is addressed via the use of a two-step cluster evaluation analysis based on log-likelihood as distance measure and Akaike Information Criterion (AIC). This particular cluster model was selected through appraisal of 504 possible model configurations from within Kohonen, K-Means and Two Step clustering algorithms. The modelling details are presented in Appendix 36. Seven clusters in total have been discovered and the model achieved average silhouette coefficient of 1. Considering that the silhouette coefficient is a “dimensionless quantity which is at most equal to 1” (Kaufman, Rousseeuw PJ, 1990, p. 87-88), evidently this cluster model is 100% accurate representation of the real patterns in the data. Figure 25 is an illustration of the cluster comparison. The upper section of the figure shows the patterns’ distribution between the variables in a graphical form – only the first 5 clusters are shown. The lower section of Figure 25 is a table quantification of the very same data, but includes data from all clusters. The interpretation of both modelling techniques is as follows:

1. 83.27% from all firms (N=263) which experience difficulties due to unfair competition from informal businesses engage in turnover underreporting with various motivations. This amounts to 53% from the entire studied population (N=413). Clearly, this finding confirms strong causal relationship between exposure to unfair competition and tax evasion via turnover underreporting.

2. Firms, which experience difficulties due to unfair competition and use turnover underreporting as a compensatory strategy, have the survival motive as the strongest determinant of their tax evasive behaviour (62.4%, N=219, see the lower section of Figure 25).
   Note: It should be noted that Question 16 from the business survey is a multiple choice multiple answer one and therefore respondents may select up to 3 possible answers. Hence, the percentages overlap in the cluster modelling.

Figure 25: Cluster analysis of unfair competition and turnover underreporting.



Variable	Variable label	Clusters	Members
Q16_2	Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to survive?	1, 2, 3, 4	111 = 62.4%
Q16_3	Businesses from your sector underreport their actual turnovers with the purpose to pay less tax in order to reduce their costs and increase their profits?	3, 5, 6	55 = 30.9%
Q16_1	Businesses from your sector underreport their actual turnovers with the purpose to pay less tax?	2, 6, 7	55 = 30.9%
Q16_5	Businesses from your sector underreport their actual turnovers with the purpose to pay less tax, because they are forced to do so and they do not have another choice?	1, 2	45 = 25.3%
Q16_4	You do not see anything wrong in the fact that businesses from your sector underreport their actual turnovers with the purpose to pay less tax?	0	0 = 0%

Chapter Seven

7. Discussion of findings and contributions.

Shadow economies are multi-faceted socio-economic structures, with a level of systemic dynamism and cultural embeddedness. For this reason, it is very difficult to conduct comprehensive studies of these economies and, consequently, current policy approaches to tackling shadow economies lack deep understanding of the root motivations of participants. It is widely accepted that the informal economy is not only a widespread phenomenon and growing relative to the formal sector in emerging markets, but also in Western societies (Thai et al., 2013, p. 14). Almost two thirds (1.8 billion) from all 3 billion workers in the world are, in this sense, informal (Jütting & Laiglesia, 2009, p.17-26), while informal commercial activities account for an estimated 30% of economic activity around the world (Thai et al., 2013). From a more regional perspective, the cumulative informal economy share amongst EU countries in recent years is estimated to be approximately 20% from their GDP by MIMIC method (see Figure 3). In light of the reported facts, governments across the world have strong incentives to tackle their informal economies in an attempt to foster economic development and international competitiveness.
However, at present policy approaches to tackling shadow economies experience three major obstacles: context appropriate definition, measurement of the size and exhaustive appraisal of the root causes. It has been argued that the development of adequate policy approaches is subservient on precise locally relevant definition, comprehensive understanding of the root causes of informality at a national level and accurate measurement. Still, one of the biggest obstacles in tackling the informal economy has often been the fact that it was being "conceptually framed as a "tame" problem (i.e. a problem that is complicated but easily solvable, often with a discrete response that can be replicated anywhere)" (Williams, 2014c, p. 3). Governments tend to underestimate how sophisticated this socio-economic phenomenon is and fail to appraise the full scale of micro and macro participatory motivations. This is so, due to the lack of regular deployment of appropriate empirical inquiries seeking to establish what a country's specific root causes of informality are. Furthermore, it has been argued that neither of the current informal economy explanatory paradigms are sufficiently adequate. A group of scholars started to address the need for a new holistic theorisation, suggesting the institutional asymmetry paradigm (Williams et al., 2015a; Williams et al., 2015d; Williams et al., 2014d). Adopting IAP, this study aims to address its empirical challenges – indeed, how to measure empirically the intrinsic circularity between tangible and intangible phenomena, affecting tax evasive behaviour. The contribution to the literature is via appraisal of these issues, proposing an economic model of partly informal entrepreneurship through innovative socio-economic profiling of business tax evaders. The provision of solid empirically verified answers for the main research questions has been offered from within what is considered the most distinctive for this type of study EU member state. Indeed, Bulgaria is uniquely characterised by the lowest overall index of social cohesion-OISC (Dragolov et al., 2013, p. 41) and the highest informal economy share (see Figure 2, 3 and 4).

The main research aim in this study was to develop a holistic account of the socio-economic influencers of the processes of business tax evasion, formalisation and informalisation and as a result to inform adequate policy responses. This task required emergence of two investigative levels – institutional and entrepreneurial. The first research level was attained with a qualitative method to achieve an in-depth understanding of the institutional context with a focus on tax policy management. In particular, revealing how the three major environmental pressures (institutional, economic and political) affect NRA's role in revenue raising policies and the

associated institutional asymmetry. The main qualitative research question pursued in this thesis can be expressed as follows: How does a national revenue agency try to tackle tax evasion in an institutional context that makes operating within the informal economy a practically-viable alternative to operating in the formal economy? To explore this central question, four lines of inquiry have been reported: discovery and exploration of agency strategies; agency operational approaches and the predominant type in the mix of policy measures; and the triggering procedures and methods of enforcing control and compliance – tax audits and inspections. The analysis and findings have been presented in chapter 5. The second investigative level was attained with a quantitative method (a business survey) deployed amongst entrepreneurs. The main quantitative research questions can be expressed as follows: What the strongest determining factors of business tax evasion are and whether the intrinsic circularity between tangible and intangible socio-economic phenomena is affecting tax evasive behaviour? This research task required the development of an empirical instrument capable to appraise the root causes for partial informality amongst micro and small businesses. In doing so, the employed questionnaire and associated analysis contribute in two main aspects. Firstly, it advances the so called predictive economic profiling of business tax evaders exploiting an innovative application of proven multi-disciplinary analytical methods. This approach enables detection of business informality with certain probability based on indirect questions (indicators). The strength of the approach is in its capability to reveal the strongest, root cause determinants of tax evasive behaviours. Secondly, the quantitative empirical instrument enriches the literature debates about informal economy measurement's approaches – still arguing about the superiority of either macro or micro-economic methods. This has been achieved with the reported measurement of informality, indicating that the proportion of GDP in the informal economy is much higher than current macro-measurement methods report for Bulgaria. The research design and particularly the quantitative method aimed to address the "extraordinary challenges", which this field faces (Slemrod & Weber, 2012, p. 31). Indeed, to avoid the weaknesses of macro-measurement approaches and measure the unmeasurable indirectly. Given the very sensitive research topic, great effort was invested in utilising the experience of previous studies with accent on gradual and indirect information extraction approach. The empirical strategy in this regard benefited from the incorporation of NRA's extensive experience in detecting and measuring informality. For maximum clarity, the relevant analysis and findings have been presented in chapter 6. However, the discussion in regards with the extant literature takes place in this last chapter 7 and so offers a holistic research account of the qualitative and quantitative findings.

Recent progress in deeper understanding of tax evasion and informal economy phenomena exposed their extremely complex dynamic nature. Indeed, that tax compliance decisions involve not only the socio-economic characteristics, but also the tax payers' morale and beliefs (Braithwaite, 2007; Braithwaite et al., 2007) as well as a degree of rational economic behaviour. Such an advanced level of knowledge drives the need of multi-disciplinary integrative research inquiries and this study attempts to address this necessity. The utilised analytical approaches relate to relatively new methodological paradigms in researching tax evasion and compliance, which are starting to gain popularity. Some notable examples are the use of agent-based modelling of tax compliance to explain the connection between governance modes of tax authorities and the resulting behaviour of taxpayers (Hashimzade, Myles, Page & Rablen, 2015; Hokamp, 2014; Pickhardt & Seibold, 2014). In similar fashion, the study of Warner et al. (2015) applies elements of computer science, economics and theoretical biology to tax compliance in the form of modelling tax evasion with genetic algorithms. In doing so, it offers unique insights to auditors seeking to reduce the US national tax gap, which is estimated to exceed $450 billion.

Utilising proven multi-disciplinary methods (Decision trees, Neural and Bayesian networks), this study advances predictive socio-economic profiling of business tax evaders. The findings are discussed in regards with the extant literature within the next 5 sub-sections.

7.1 **Tax evasion diagnostics and motivations.**

Firstly, it has been discovered that 76.5% of the surveyed entrepreneurs underreport their turnovers and as a result evade taxes with various motivations (see variable Q16 and Table 15). This amounts to a significant proportion of the country's GDP and suggests that the proportion of GDP in the informal economy is much higher than current macro-measurement (indirect) methods report for Bulgaria – 31.2% of GDP according to Schneider (2013). Secondly, the three most frequent motivations for underreporting turnovers are paying less tax in order to survive (29.3% of responses to a multi-set, 45.7% of all cases), reducing costs to increase profits (19% of responses, 29.7% of all cases) and just to pay less tax without a particular reason expressed (17.4% of responses, 27.3% of all cases). Xheneti, Smallbone and Welter (2013) argued that cross-border informal entrepreneurship developed in response to the regional disparities and social inequalities experienced during the post-communist transformation period. Where the limited effect of EU regional policies made this type of informal activities essential for individual survival. Given, that the most significant motivation for tax evasion in the capital of Bulgaria is also survival, this constitutes an intriguing research question. Indeed, to what extent these phenomena could correlate to EU policies linked to enlargement and regional development.

It is important to note that the reported tax evasion measurement was amongst the small businesses exercising partial mode of informality in Sofia. Due to more favourable economic environment there, it is argued that the level of partial informality would be higher in the rest of the country, because small businesses would be less profitable and under reduced compliance control. The evaluation stated in this study does not include the share of undeclared work in combination with envelope wages and the evasive footprint of fully informal businesses. For the described reasons, there are grounds for believing that the reported finding (76.5%) could be on the conservative continuum in comparison with a nationally representative study. Calculation to what proportion of the GDP all these relevant phenomena would cumulatively account to is beyond the scope of the present study.

However, the described findings are simultaneously alarmingly concerning and significant for this field of research. There is convincing evidence that higher informality is recognised to increase income inequality, lower average GDP growth and reduce international trade (ILO et al., 2009, p. 88). Therefore the suggested much bigger share of informal economy appear to be the biggest impediment for economic development of the poorest EU member state (Dragolov et al., 2013, p. 41). Obviously, one of the contributions of this thesis is to serve as a justification for the resurrection of the debate about measurement method's superiority and applicability to different countries. Appraising the current trends in measurement methods Williams (2014a) reports that there is "an emergent consensus to use indirect measurement methods to estimate the size of the shadow economy and direct survey methods for unravelling the nature of the shadow economy." It should be noted that the application of macro methods is considered even implausible for transition economies by some scholars (Palda, EconWPA, 2004). Based on the discovered research insights (both methodological and knowledge related) in this study, it is argued that indirect methods are better suited to advanced economies with relatively small share of informal economy. Direct measurement methods, such as the micro-economic one developed and utilised in this thesis, tend to be superior in transitional economic settings or

countries with bigger shares of informal economy.

Another worthy aspect of the findings of this thesis relates to the very core structure of informality in a given society. The main argument is that from policy tackling perspective, the most important is not so much the share of the informal economy, but its fine composition. Illustratively, two similar in size informal economies (in two different countries) might have completely different structures (explanation follows). It is argued that understanding the proportion and features of participatory characteristics is of paramount importance for the design of adequate policy responses to informality. To explain this, two new conceptual terms are being introduced:

Conceptual terms' definitions:

1. *Frequency of societal involvedness in the informal economy:* How many members of a given society are involved in exercising tax evasion of some form. An indicator, expressed as a percentage from the total number of members.
2. *Intensity of societal involvedness in the informal economy:* What is the average amount of tax evasion per member of society, amongst those who are involved in evasion.

For descriptive purpose – if we assume that the share of informal economy as a percentage of the GDP for a given society is X%, then it might have very different composition theoretically. The same share of informal economy as a percentage from a given country's GDP could be achieved in two possible scenarios (theoretical scenarios 1 and 2 below):

Theoretical scenario 1 for X% share of informal economy as a percent from the GDP
Higher *Frequency of societal involvedness in the informal economy* multiplied by Lower *Intensity of societal involvedness in the informal economy*

Theoretical scenario 2 for X% share of informal economy as a percent from the GDP
Lower *Frequency of societal involvedness in the informal economy* multiplied by Higher *Intensity of societal involvedness in the informal economy*

To utilise the new conceptual terms into the researched context, it is emphasized that Bulgaria has very large share of informal economy, characterised by:

High *Frequency of societal involvedness in the informal economy* 76.5% of formally registered entrepreneurs underreport their turnovers multiplied by **High *Intensity of societal involvedness in the informal economy*** Entrepreneurs conceal 37.8% of their turnovers on average (Mean = 37.8%)

Looking at the two possible scenarios above, it could be theorised that given the same X% share of informal economy, a status of institutional asymmetry applies to bigger

proportion of the society (scenario 1), while in the second it affects smaller number of social actors. Transition economies with large informal sectors (such as Bulgaria) are characterised by high *Frequency of societal involvedness in the informal economy* multiplied by high *Intensity of societal involvedness in the informal economy.* Therefore, the institutional asymmetry applies to big part of the society and to a much deeper degree in every instance. Unsurprisingly, informal economies with such characteristics are particularly challenging to tackle and in the case of Bulgaria, it is evident how so many consecutive governments failed. The theoretical implication in comparative appraisals of institutional asymmetry in two different economies, characterised by the same share of informality, becomes an important consideration from policy perspective. Indeed, depicting scenario 1, where fewer tax payers engage in tax evasive behaviours, to a much deeper degree though, requires targeting tackling efforts towards that social stratum. Aligning the moral values and beliefs of that social group to the formal institutional framework is a much easier in comparison with doing the same in scenario 2. Where, albeit that institutional asymmetry leads to lesser evasion per member of society, it affects much wider group of the same society. It has been already shown that generally the size of informal economies correlates with the overall index of social cohesion (OISC) for a given society (see Figure 4). Emerging from the new insights, it could be hypothesised (for future research) that countries with the same OISC, but bigger shadow economies have similar *Frequency of societal involvedness in the informal economy*, but higher *Intensity of societal involvedness in the informal economy.*

It has been demonstrated via multi-layered analysis (decision trees, neural networks and two-step cluster modelling), that unfair competition from informal businesses is the most frequent and strongest predictor of tax evasion (to be reviewed in a section below). Evidently, from the findings, it is highly likely that entrepreneurs who experience such competition would compensate their losses via turnover underreporting. In addition, entrepreneurs exercise tax evasion as a corresponding response to the poor tax and administrative policies, which cause them to spend more valuable resources than deemed acceptable. On some occasions, bribes demanding inspections from NRA, appear to be the cause of tax evasive responses. In general, when profits or potential profits are taken away from firms through regulation, taxation, or corruption (Johnson, Kaufmann, Shleifer, Goldman & Weitzman, 1997), an offsetting attitude through unofficial activity is highly likely. Emerging from the cluster modelling (see Figure 24, Appendix 30), it is argued that the entrepreneurs who have provided more than one motive for tax evasion simultaneously have some important coincidence of their answer patterns. Firstly, turnover underreporting is exercised, so that entrepreneurs can survive economically and secondly, this illicit activity is practiced because they do not have another choice. Thirdly, tax evasion is performed for the reduction of operation costs and increasing profits. This particular tripartite motivational pattern, suggests that the most distinctive cases of business tax evasion are being developed under a set of economic coercion forces, which are environmentally enforced (see Figure 26 for more details).

There are several theoretical and policy implications stemming from these findings. Various scholars (Round et al., 2008; Williams et al., 2010b; Williams et al., 2007b) proclaim that different explanatory theories of informal entrepreneurship applies more in some populations and countries' contexts than others. The findings of this thesis support their argument revealing some intriguing theory related aspects. Indeed, the presented empirical evidence suggests that the neo-liberal theory, with emphasis on the economic rationale behind tax evasive behaviours (Bayer & Sutter, 2009; Djankov et al., Belev B, 2003; Koettl et al., 2012; Perry et al., Arias O, Fajnzylber P, Maloney WF, et al., 2007), appear to be the most relevant one to transition economies. Bulgarian entrepreneurs experience very high cost of doing business according to the

rules and this has manifold basis. One aspect is rooted in the formal institutional framework: due to the poor tax and administrative policies which impose hefty financial burden on businesses. Another aspect is due to the so called informal chains of dependences, which have serious economic coercive powers over entrepreneurs inhibiting them with prohibitively high exit costs (towards full formalisation). Third aspect relates to the intense and wide spread unfair competition, clearly interconnected with institutional quality, procedural fairness and distributive justice. These phenomena will be discussed in greater detail in the following sections.

7.2 **Institutional quality, procedural fairness and distributive justice.**

Evidently from previous studies (Bitzenis et al., 2005; Valdez et al., 2013) macro-level institutional environment is significantly associated with entrepreneurial patterns in a given country's context. At the same time the likelihood of small firms to engage in corrupt activities is strongly influenced by low institutional efficiencies (Tonoyan et al., 2010) particularly in Central-Eastern Europe. Xheneti et al. (2012, p. 618) argued that entrepreneurs who successfully learn to weave their way through a corrupt institutional environment, acquire a competitive advantage over others and experience business growth as a result of corruption. Obviously, this illustrates very well one of the self-perpetuating aspects of large informal sectors, such as the Bulgarian one, in respect of poor formal institutional framework.

As already summarised (see 3 and 5.1), the Bulgarian socio-economic reality is the worst amongst EU member states in various aspects including GDP, OISC, GCI, informal economy share and institutional corruption level. It was discussed (see section 2.4) that institutional asymmetry perspective is a contemporary explanatory theory of tax evasion in societies with significant misalignment between formal and informal institutions. Therefore, the status of large institutional asymmetry is argued to be guaranteed by default in transition economies (such as Bulgaria), which logically leads to significant tax evasion. This study reveals the main determinants of entrepreneurs' tax evasive decisions emerging from the socio-economic environment in which the studied small firms operate. In doing so, it contributes by showing that institutional quality, procedural fairness and distributive justice all have tangible share in the *real cost of doing business* for entrepreneurs. Therefore, the high *real cost of doing business* due to the inadequate business environment is not surprising. It is discovered to be a significant driver for informality amongst enterprises (see section 4.4.1). The voice of the entrepreneurs sounds quite disturbingly: 60.6% of them either agree or strongly agree that doing business according to the rules is too expensive (Table 13). Simultaneously, 50.6% either agree or strongly agree that tax evasion is the answer from the business to the poor tax and administrative policies (Table 13). A viable public policy should be targeted at the creation and cultivation of trust in procedural fairness (Torgler & Schneider, 2009; Van Dijke et al., 2010) in order to influence positively the tax compliance will.
Decision tree modelling of entrepreneurs' perceptions of procedural fairness and institutional quality has been used in order to explore additional interdependence with evasive practices and associated phenomena (see section 6.2.2 and Table 20). The state institutions' services, including NRA, are deemed to be unfair and unprofessional from their perspective, which harms the perceived procedural fairness, and obstruct distributive justice accordingly. Most entrepreneurs spend on average more than 4 days per month to cope with administrative requirements imposed by various state regulations. Considering that the average cost of doing business per day for the studied entrepreneurs is 466.5BGN (Mode = 100, Mean = 466.5, see Table 14), this constitute a substantial monthly expense. There is a dual pattern of dissatisfaction amongst the entrepreneurs – simultaneously with the state institutions' quality,

including NRA and the way tax money is being spent. Evidently, from the decision tree modelling, entrepreneurs are likely to be pushed into evasive economic behaviour by the apparent poor quality of state institutions and NRA (percentages presented in the box below). The reported findings correspond with previous studies (Cummings et al., 2006; Hartner et al., 2008) such as the fact that tax compliance depends on individual perceptions that the tax system is fair and that the government is providing valued goods and services with the revenues. The decision tree modelling of entrepreneurs' perceptions of procedural fairness and institutional quality (Table 20, Appendix 31, 32 and 33) provides a conditional probability map to interested policy makers. Indeed, if a targeted policy intervention is applied to change attitudes, then it could be predicted what proportion of business tax payers it would affect and what this effect would be on their tax evasive motivations.

- 59.9% of the entrepreneurs either disagree or strongly disagree that the tax money is fairly spent (Table 13).
- 54.3% of the entrepreneurs either disagree or strongly disagree that the state institutions have fair professional attitude and treatment approach (Table 13).
- 41.4% either disagree or strongly disagree that NRA has fair professional attitude and treatment approach (Table 13).

7.3 **Informality chains of interdependence – enablers of informality and barriers to formalisation.**

Informality chains of interdependence (ICOI) constitute complex economic situation where an enterprise satisfies a proportion of their trade needs for goods and/or services without receiving invoices accordingly. In doing so, the participating firm cannot claim eligible expenses and develops strong motivation to offset the loss with a reciprocal action – not providing invoices to conceal taxable income. That is why some authors call them an "informality trap" (Institute for Market Economy, 2004, p. 61-70). An informality chain of interdependence could be initiated by economic pressure from other enterprises (including clients) on the supply or the demand side of the business, or by both at the same time. Section 6.2.3 provides an illustrative example how this works in practice. To further enhance the description, it should be noted that a cluster of firms might exist in an informal chain of interdependence, where different members could have a varying degree of informality, not excluding fully formal enterprises. A fully formal enterprise within such an economic dependency, which wants to remain tax compliant, would have a dualistic choice. That would be to either absorb the financial losses or exit the interdependence chain if they can afford that retaining commercial viability.

The corporate tax in Bulgaria is 10%. It is argued that enterprises, which manage to acquire invoice for less than 80% of their business needs, would be strongly incentivised to pursue compensatory strategies. That is to offset the loss due to inability to claim legitimate business expenses (≥20% of their cost of sales). Based on variable Q18 and Q19 (within section Summary statistics of individual variables), it is evident that from the 58.6% of firms which experience purchasing of goods and services without invoices, 33.5% acquire invoices not exceeding 80% of their consumption as businesses (Table 11 and Table 14). For that reason, this group of enterprises could be considered to function in informality chain of interdependence with varying magnitude of tax concealment and degree of complexity. Taking into consideration that 76.5% of all firms are partly informal (see variable Q16 and Table

15) this is rather conservative finding, which needs further research to be established precisely. The most frequent reason for firms to enter into such economic dependency is that cost of foreign informality takes over in the mix of real cost of doing business (see Figure 26). Therefore, some enterprises attempt to offset the increased real cost of doing business through reducing one of its components – the cost of foreign informality. One such strategy is to become partly informal, thus becoming members of ICOI. There are cases where fully informal enterprises are also part of informal chains of interdependence by their own choice or simply due to inability to develop sufficiently profitable business model from the very beginning. However, as elaborated earlier, this thesis is interested mainly in the most difficult to detect and tackle: the registered partly informal businesses exercising a varying degree of tax evasion. For that reason, when analysing informal chains of interdependence the focus is on partly formal enterprises and the associated effects for them, as well as for the fully compliant ones. In particular, the analytical effort here is to appraise the processes of informalisation and formalisation as a function of the so called ICOIs.

Informal chains of interdependence are sophisticated dynamic eco systems, which tend to attract new members through powerful economic coercion. One mechanism has been explained earlier in section 6.2.3. Another mechanism of such powerful economic coercion is via the strong unfair competition an informal chain of interdependence creates for other firms in any given industry sector. Given the fact that every ICOI involves a varying level of informality for its members, this inevitably provides them with an unfair advantage in comparison with outside fully formal businesses. Ultimately, the real cost of doing business and its structure (see Figure 26) is affected in contrary directions for both – members of an informal chain of interdependence and the immediate economic environment (other prospective businesses suppliers and clients from the sector). In that respect, the powerful forces of economic coercion act in multiple directions depending on the object: whether it is fully formal, partly formal or fully informal enterprise. To fully appraise the complexity of ICOI's coercive forces, it is important to review the inter-correlated nature of "cost of formality", "cost of informality" and *real cost of doing business* to firms. In particular, the economic compulsion towards informalisation and capacity to formalise accordingly. Such a discussion (offered below) is necessary to enlighten the economic behavioural patterns under pressure from ICOIs. It is followed by defining two of the toughest policy challenges and suggesting empirically informed solutions.

The concept *cost of doing business* should be reviewed in a multi-perspective fashion to reflect its complex underlying significance to entrepreneurs and enlighten turnover underreporting. One of the research aims of this study was to establish the relationship between tax evasive decisions as functions of cost of doing business. The first aspect of this relationship is whether high cost of doing business pushes entrepreneurs into partly informal modes in order to offset what they deem unacceptable. Another aspect is whether high cost of doing business serves as an impediment to those who would like to increase their level of formalisation and/or fully formalise their operations. However, the *real cost of doing business* could have different composition depending on the mode of business. An additional complication is the fact that actually "cost of formality" and "cost of informality", as De Soto defined them (1989, p. 132-177), both form the real cost of doing business. Figure 26 is developed to illustrate the complexity of the relationship between these concepts and a business entity, depending on its economic status. The aim of the figure is rather to illuminate and clarify, but not to descriptively exhaust all sophisticated connections and dependencies. For the purpose of this analysis and by mode of doing business, enterprises could exist in three forms: fully formal, partly formal and fully informal mode (Figure 26). The effects of *real cost of doing business* on firms' tax evasive

behavioural models are discussed below.

Fully formal enterprises: real cost of doing business and behavioural effects

The fully formal enterprises are registered, wholly compliant ones, reporting actual turnovers and paying the full amount of tax due. As such, they meet the full costs of doing business according to the rules. Consequently, such firms enjoy all privileges and benefits any legitimate business is entitled to utilise. These enterprises do not have to absorb any "cost of informality" in De Soto's sense, however they are exposed to cost of informality in the form of unfair competition and other related phenomena. In general, firms from this group could have two sub-modes of operation: with intent to remain fully formal or with intent to partly informalise. It is also theoretically possible to cease formal operations and convert to fully informal mode, but this is very rare and thus insignificant for the purpose of this discussion. The two possible sub-modes are of dynamic nature as the choice of any enterprise may change over different circumstances and time. This is also applicable to the sub-modes of operation observed in the other two groups (partly and fully informal ones). The two possible sub-modes of operation, could be exercised either by own choice of the firm or become environmentally imposed (see the first section of Figure 26 from the left). For instance, a firm might be subjected to unfair competition. In order to keep their market niche and remain fully formal, they will be forced to accept reduced profits in such circumstances. If the given firm is sufficiently profitable, they could afford such strategy by their own choice. However, the unfair competition could develop over time to a level, which is so severe that the formal enterprise could not overcome it with just surrendering their profits. In that case, the firm either should cease operations or choose a corresponding level of informalisation to keep its financial viability and remain in business. The latter option would be the so called environmentally imposed choice for the purpose of the analysis.

Returning to the concept of *real cost of doing business* and its composition (cost of formality and cost of informality), it should be noted that the latter component is actually two dimensional one. Indeed, that is the cost of foreign informality and the cost of own informality, but it only becomes visible when a firm is either partly or fully informal (reviewed below). To clarify for the reader, cost of foreign informality is for instance unfair competition or other forms of economic damages due to interaction with partly or fully informal enterprises. Cost of own informality is for instance the lack of proper commercial contract enforcement due to the unreported nature of such transactions. Or another example would be limited access to commercial credit due to business turnover underreporting and the associated inability to advertise it to lenders such as banks.

From the perspective of fully formal enterprise, their *real cost of doing business* is quite simple – it has only two sub-components. Indeed, that is the cost of formality (doing business according to the rules) and cost of foreign informality. Such enterprises would be willing to maintain their fully compliant mode of doing business as long as the cost of foreign informality is lower than the cost of formality (doing business according to the rules, see the first section of Figure 26 from the left). In such scenarios, an acceptable level of financial viability could be preserved. However, if there is an increased pressure from informal firms in the form or unfair competition or informal chain of interdependence, then the cost of foreign informality will take over the cost of formality. Therefore, the real cost of doing business may deprive the fully compliant firm to a level of lacking financial viability, unless the firm agrees to offset the losses via partial informalisation. So, such process would be obviously an environmentally imposed one.

Partly formal enterprises: real cost of doing business and behavioural effects

The partly formal enterprises are registered and operate exactly as the fully formal firms with the only difference that they underreport a proportion of their turnovers with the purpose to evade taxes. The underreporting motivations vary according to specific circumstances as it was shown in earlier parts of the thesis (see section 6.2.1). Given the fact that such firms are registered, they also enjoy almost all privileges and benefits such as legitimate businesses do. Illuminating the concept of *real cost of doing business* through the prism of partly formal enterprise, poses greater challenge in terms of depicting one more underlying business cost. While the fully formal enterprises need to pay the cost of formality and cost of foreign informality only, the partly formal enterprises need to absorb one more layer of costs – cost of own informality (see the middle part of Figure 26). The partly formal enterprises with intent to formalise achieve this state of will due to the cost of their own informality becoming the highest in the mix of real cost of doing business. Therefore, in order to sustain or realise the desired financial viability they need to fully formalise. For instance, their business environment (suppliers, clients, formal institutional environment) may require a full compliance mode in order to continue transacting business with them. An illustrative example of this process is provided in the box below:

A small manufacturing firm X is involved in relatively small turnover underreporting (≤25% from their turnover). Other than conducting this form of tax evasion, all other business aspects are in accordance with the rules. Some of the suppliers of the firm are also partly formal entities with similar evasive patterns and function together in "informality chain of interdependence" (see section 6.2.3 for detailed description). The operations of firm X become a gradual success and they start to participate effectively in tenders supplying local government structures. In doing so, the external business environment (suppliers, clients, formal institutional environment) starts to pose formalisation pressure. Firm X needs to demonstrate bigger turnover and detailed fully formal supply chain in order to qualify for more tendering positions (resulting in bigger sales). At this moment in time, the cost of own informality takes over the other two components of real cost of doing business (see the middle section of Figure 26). It happens that full formalisation would secure a greater financial viability for firm X even after losing its gain from current turnover underreporting. However, in practicable terms it is difficult for a firm in an "informality chain of interdependence" to secure a market position capable to cover the costs of exiting the chain.

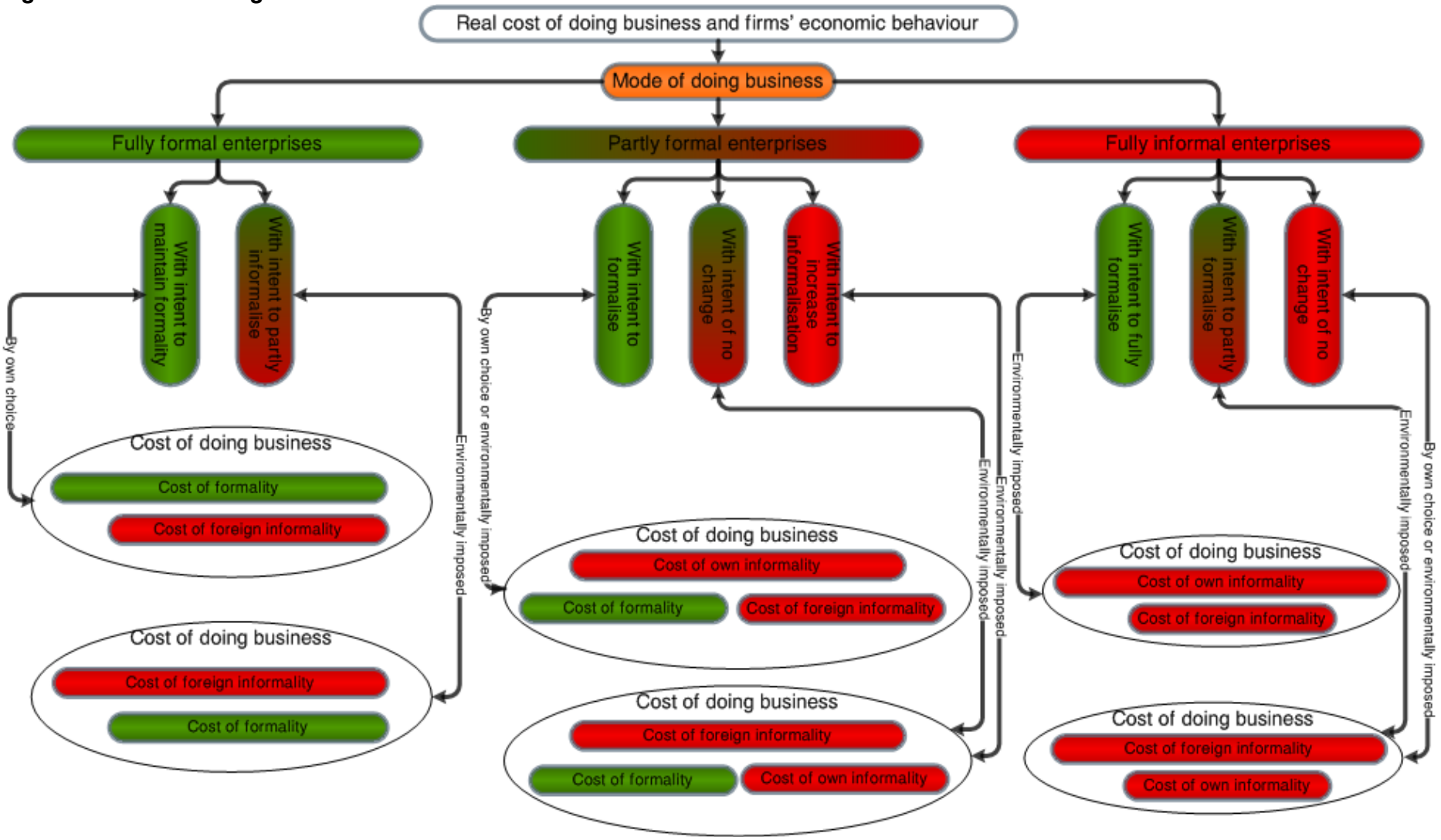
The next sub-group from the middle section of Figure 26 are the enterprises with the intent to keep their level of partial formality unchanged. In their mix of real cost of doing business, the cost of foreign informality is the highest from the 3 sub-components. Usually such firms are partly formal for the very same reason – offsetting the losses due to the foreign informality. As long as the cost of foreign informality remains constant, these enterprises are highly likely to keep their level of partial informality. This is so, because they do not perceive the formalisation benefits to be worthwhile (Thai et al., 2013). However, if there were further increase of the cost of foreign informality, then such firms would seek to compensate via increasing the level of their own informality. Obviously, this would be under environmental pressure towards further increase of informalisation.

Fully informal enterprises: real cost of doing business and behavioural effects

The fully informal enterprises are unregistered and not paying any tax and therefore bear the full cost of their own informality (see the right section of Figure 26). Usually

such enterprises start unregistered for various reasons. The real cost of doing business for them accommodates two sub-components: cost of own informality and cost of foreign informality. Fully informal firms still need to meet the cost of foreign informality, although they are much more capable to deal with it in comparison with fully or partly formal enterprise. As long as the cost of foreign informality exceeds the cost of own informality for a given enterprise, there is little chance it would want to formalise. It is only when cost of own informality takes over when there is sufficient formalisation incentive. Fully informal enterprises were not subject of this study and therefore are of limited interest in the discussion.

Figure 26: Cost of doing business.



The strong negative effects of unfair competition on legitimate businesses have been studied extensively (Riahi-Belkaoui, 2004; Schneider et al., 2013; Small Business Council, 2004; Williams, 2010b; Williams, 2014a) and there is agreement that appropriate tackling effort is required to address the consequences. Logically, it is important to establish what the effects of unfair competition are on turnover underreporting (i.e. tax evasion). To achieve this task, a multi-layered analysis (decision trees, neural networks and two-step cluster modelling) was employed (see section 6.2.4, Figure 25, Appendix 36). It was shown that unfair competition from informal businesses is the most frequent and strongest predictor of tax evasion. Evidently, from the findings, it is highly likely that firms, which experience such competition, would compensate their losses via turnover underreporting.

83.27% from all firms (N=263) which experience difficulties due to unfair competition from informal businesses engage in turnover underreporting with various motivations. This amounts to 53% from the entire studied population (N=413). Having argued that 76.5% of all firms evade taxes via turnover underreporting, it is clear that there is strong relationship in this respect. Indeed, that unfair competition causes tax evasion via turnover underreporting. To compare the reported finding with data available for an advanced EU economy, it should be noted that only 14% of UK small businesses feel negatively affected by informality and just 0.56% of all businesses view informal entrepreneurs as the primary threat to their business development (Webb, Bruton, Tihanyi & Ireland, 2012, p. 11). Empirical evidence based on data from 30 countries indicate that good competition laws are one of the significant factors affecting tax compliance internationally (Riahi-Belkaoui, 2004). At the same time, there is an increasing body of evidence that unfair competition is one of the most intense barriers to firm's performance in transition business environment (Krasniqi, 2007, p. 79).

Given the fact that unfair competition is highly damaging to legitimate businesses (Evans et al., 2006; Gërxhani, 2004a), it becomes important to reveal what the main motive behind turnover underreporting due to unfair competition is. This analytical task has been achieved utilising a two-step cluster evaluation analysis based on log-likelihood as distance measure and Akaike Information Criterion (see for details Figure 25 and Appendix 36). Considering that the silhouette coefficient achieved the maximum value possible (Kaufman, Rousseeuw PJ, 1990, p. 87-88), the cluster model is 100% accurate representation of the real patterns in the data. The significant finding is that firms, which experience difficulties due to unfair competition and use turnover underreporting as a compensatory strategy, have the survival motive as the strongest determinant of their tax evasive behaviour (62.4%, N=219, the lower section of Figure 25).

Unfair competition from informal businesses is effectively the cost of foreign informality (see Figure 26) for these firms, which are experiencing it. It could increase so much so that the *real cost of doing business* becomes unattainable unless firms engage in compensatory strategy in order to offset its effects and survive. From an economic perspective, unfair competition is simultaneously enabler of informality and barrier to formalisation. The enabling aspect is in the economic market force rendering a fully formal business unable to compete against one or more unfair competitors. To preserve competitiveness a formal business might need to informalise to a level corresponding to the unfair competitors. The same logic applies to partly informal businesses, which might need to deepen their level of informality in order to survive economically. Unfair competition is also a barrier to formalisation due to the same inherent aspect and has detrimental effect on legitimate businesses (Institute for Market Economy, 2004; Krasniqi, 2007; Oviedo, 2009; Pashev, 2005; Thai et al., 2013; Tonoyan et al., 2010). Assuming that a partly informal business operates as such mainly to cope with unfair competition, it could only fully formalise if a higher

economic competitiveness in comparison with the unfair competitors is achievable. That economic competitiveness should be sufficiently higher in order to accommodate the *real cost of doing business* (see Figure 26). Understandably, it is a hardly attainable business target especially in typically low margin trades with non-specialised goods within transition economies. The findings support the argument that similarly to other transition economies such as Albania (Bitzenis et al., 2005, p. 572) unfair competition is a major obstacle to sustaining healthy entrepreneurship environment with appropriate level of formalisation.

Unfair competition (due to unfair advantage) could have two forms when looked at from the perspective of informality. The first type (type one) is created by partly informal enterprises acting on their own without being part of ICOI. Obviously, this poses a problem for fully formal businesses, as they need to pay the full price of compliance and still earn a share of the market. The second type of unfair competition is the one created by an informal chain of interdependence – it is like type one, but illustratively speaking on steroids (i.e. much stronger). The example shown in section 6.2.3 draws a picture of fairly short supply chain, however it is evident how savings (due to partial informality/no invoicing) could transfer to upper members of the supply chain, which on this case is functioning as ICOI as well. Therefore, the final product (for instance a box of croissants, see section 6.2.3) already benefited from partial informality exercised at different stages of its production up to the delivery to the end user (for instance a local café). At this point of product manufacturing (sophistication), the unfair advantage takes the form of multiplied form of unfair competition. The production cost for this box of croissants (at the end of ICOI) is so low that very few fully compliant businesses could achieve a counter competitive offering. Evidently, the unfair competition from type two (via ICOI) is much stronger than unfair competition from type one (via partly informal firms not members of ICOI). As such, unfair competition of type two should be a priority target for tackling policies.

It has been already discussed that larger informal economies (for instance the Bulgarian one) tend to have self-perpetuating nature due to their degenerative effects on social cohesion, procedural fairness and distributive justice (see sections 2.5 and 2.5.1). One of the problematic circularity effects is that the individual cost of being honest in a corrupt society becomes significantly higher in comparison with societies where most individuals are honest (Andvig et al., 1990). This easily transfers to the world of small business, entrepreneurship and tax evasion. In particular, when the socio-economic environment is very difficult (see 3, Figure 14, 15, 16, Appendices 2 and 3) and most small businesses must exercise huge effort to just maintain economic viability. Informality chains of interdependence are therefore one of the significant self-perpetuating mechanisms of spreading informality and simultaneously obstructing formalisation. This turns them effectively into enablers of informality and barriers to formalisation, which as known in the literature (Johnson et al., 2000) could overlap particularly in transition economies. For that reason, it is obvious why tackling this double-edge negative effect of ICOIs is a significant policy challenge. While from small firms' perspective, resisting the joining force (informalisation pressure) emanated by ICOI could be very difficult financially, a potential exit is highly unlikely to be even feasible without an external possibly institutional help. Having revealed the mechanisms of economic coercion, which ICAOs generate, it becomes clear that there are two major policy challenges in this respect:

1. Policy challenge 1. Tackle formalisation resistance of the economic coercion forces – where an enterprise is already a member of ICAO and it seeks exit if financially feasible.
2. Policy challenge 2. Tackle the informalisation coercion forces – where ICAOs draw in new members or deepen informalisation levels of existing ones.

The following two sub-sections will review the policy challenges in question and suggest possible solutions based on the analysed empirical data.

Policy challenge 1 – tackle formalisation resistance

The strength and elegance of the proposed tackling approach is to indirectly manipulate the formalisation effort of each individual member of ICOI via reasonably affordable tax easement. The suggested policy response does not need to collect any more data different from what is already available in this study. The rationale behind the policy solution in question is presented below. It has been already shown (section 6.2.3) what the conditional probability with which a firm would endeavour towards a higher level of formalisation is as a function of the involvedness' level with partly informal suppliers (expanded further below). In practicable terms, this is very important characteristic of an informality chain of interdependence from policy perspective. The analysis conducted in section 6.2.3 reveals the connection between an individual firm's effort to formalise from within an ICOI and the dependence to partly informal suppliers. If a firm is member of ICOI and receives invoices only for part of its business needs how this is going to affect its will to formalise. To resolve this analytical task a Tree Augmented Naïve (TAN) Bayesian Network has been utilised. As it is evident from the summary statistics of variable Q20, 87.2% of the respondents to this question would prefer all of their suppliers to provide them with invoices (Table 11). The first important aspect of this finding signifies that big majority of ICOI members are willing to formalise should they have a chance to do so. The second important aspect is that most probably membership of ICOI was enforced through economic coercion and was not acquired by own choice.

The third important aspect is the conditional probability, under which a firm would prefer to receive invoices for all the business goods and services it needs to conduct business. The finding is that this probability increases together with the share of properly invoiced expenditures (see Table 21). The higher the proportion of goods and services for which a firm already receives invoices, the more likely it is for such firm to demand invoices from all its suppliers and vice versa. If a firm is managing to obtain invoices for 45.4%-63.6% of the consumed business goods and services (cost of sales), then the conditional probability is only 8%. The next band (63.6%-81.8%) represents the first significant increase – to 29% accordingly (see Table 21). Firms with more than 81.8% of their business needs properly invoiced, represent conditional probability of 56% to demand invoices from all of their suppliers. Therefore, such firms are likely to exercise formalisation pressure on the supply chain, which in this case is also ICOI. The formalisation pressure is expressed through the increased demand for invoices up and down the supply chain. The observed causal correlation demonstrates the economic coercive power, which an informality chain of interdependence exercises over its enterprise members. Effectively, clusters of firms, which exist in informality chain of interdependence, act as enablers of informalisation processes and simultaneously as barriers to formalisation to both inside and outside firms. Emerging from the analysis with Bayesian Network (summary of findings in Table 21 and section 6.2.3) a solution to policy challenge one is suggested. The development approach rests on the following assumptions:

1. A majority of ICOI members (87.2%) are willing to fully formalise should they have a chance to do so.

2. Most probably membership of ICOI was enforced through economic coercion and was not acquired by own choice, otherwise the observed formalisation effort (point 1 above), would not be so widely present.

3. Inability to claim legitimate business expenses due to lack of invoices is a strong predictor of reduced formalisation efforts – judging from the correlation between formalisation effort and properly invoiced business expenses (Table 21).

4. The potential policy solution should be an indirect one – aiming to deprive ICOIs of their formalisation resistance through initiation of formalisation effort from inside.

A tax allowance could be offered to any small firm for up to 20% of their reported business expenses. They should be allowed to claim business expenses 20% higher than what they are able to prove with available invoices. Effectively, it would mean that a revenue agency such as NRA in Bulgaria should offer to decrease the taxable base with 20% for small firms and therefore lose income from taxes in the short term. Such tax easement should be offered upon specific request and on "No questions asked" basis for a period of 1 to 2 years only, once in the lifetime of a firm. Every small firm should be enabled to claim the tax easement in front of the tax office with their choice of how much exactly (from >0% to ≤20%). The available empirical data and analysis contain the answer why precisely 20% business expenses allowance – it is elaborated below.

It has been already argued that 33.5% of the firms function in informality chain of interdependence. The predominant share of these firms (22.4%) receives invoices for only 60-80% of their business needs (Table 14). Suggesting 20% tax easement on business expenses achieves a couple of aims. Firstly, it addresses the predominant share of ICOI members who would like to fully formalise. Firms, which acquire invoices for only 60%-80% of their business needs, are given the chance to claim another 20% without being able to prove the expenses. Obviously, the same applies to all other firms (<60% and >80%). Secondly, the suggested tax easement doubles the formalisation pressure within ICOIs, because it effectively changes formalisation effort through the amount of claimable business expenses (in the context of the discussion and Table 21. From tax income revenue perspective and central government, the short term relatively small loss of taxes should be considered a worthwhile investment. It is going to deprive ICOIs nationwide from their formalisation resistance coercion to a significant degree. It is highly probable that the short term loss (tax income) would be recovered through significant positive impact on formalisation rates and the associated increased tax base. Intervening with such policy is effectively an indirect manipulation of the economic logic of every ICOI and as such, it means depriving that logic from the element of financial coercion. Consequently, many small firms would be able to escape their ICOI and find viable business existence with fully formal suppliers and clients during the grace period of 1-2 years. Furthermore, the tax office would acquire highly refined picture of the ICOIs in the country thanks to the small firms coming forward to take advantage of the tax easement. This would enable combination of indirect and direct policy approaches where applicable. It is expected that non genuine tax easement business claimants would be a minority as they are clearly aware of advertising themselves to the tax authorities and in doing so possibly becoming subject to a greater level of tax scrutiny.

Policy challenge 2 – tackle the informalisation coercion forces

The dynamic correlation between the formalisation effort of each individual member of ICOI and the ability to claim legitimate business expenses has been discussed above. Expressed differently, this is the conditional probability with which a firm would struggle towards a higher level of formalisation as a function of reduced involvedness' level with partly informal suppliers in ICOI. The discovered correlations and associated

findings provided rationale for the proposed policy solution above. This section advances the level of discussion, involving additional socio-economic characteristics. Indeed, the next important line of elevated inquiry is to appraise what the effects of institutional quality, procedural fairness and use of public money are on those ICOI enterprise members striving to formalise. This is an important group from policy perspective, because although engaged in varying degree of turnover underreporting, these firms would keenly work to fully formalise should an opportunity arise. Therefore, a policy effort in this direction is considered to be a rewarding investment.

It is important to acknowledge the existence of two types of informalisation coercion forces of ICOIs depending on their objects – existing ICOI's members and outside firms usually from the same sector or supply chain. In the first configuration, existing ICOI's member firms are drawn into increasing level of informalisation. In the second configuration, outside fully formal firms are exposed to all negative effects and forms of unfair competition as already discussed. Consequently, they might be enforced to accept a level of informalisation or otherwise perish economically. Although different by their objects, both types of informalisation coercion forces of ICOIs are similar in their compulsion methods and as a consequence require similar tackling methods from policy perspective.

The analytical base for this discussion stream has been achieved through TAN Bayesian Networks deployment over a set of pre-selected special cases – members of ICOIs, evading taxes, but willing to formalise (details in section 6.2.3, Table 22 and Appendix 35). The line of analysis deals with informalisation coercion forces within ICOIs. The strongest predictor of such firms' behaviour is their disagreement with fair expenditure of tax money. The next analytical step is to establish the attitudes of the firms in question in regards with institutional quality, procedural fairness and tax evasion behaviour. As evident (details in section 6.2.3, Table 22 and Appendix 35), the conditional probabilities for such firms to disagree or fully disagree that NRA has fair and professional attitude are 29% and 50% accordingly. This means that 79% of all firms willing to formalise from within ICOI will have the discussed characteristics – full disagreement with how tax money are spent and opinion that they are not treated in a fair and professional manner by NRA. Similar conditional probabilities of 25% (disagree) and 61% (fully disagree) accordingly exist in regards with fair and professional attitude from the state institutions. This means that 86% of all firms with the studied status will be strongly dissatisfied with the state institutions and expenditure of tax money. Unsurprisingly, the firms which would like to fully formalise breaking free from informality chain of interdependence, possess conditional probabilities to agree (29%) or fully agree (57%) that tax evasion is the answer from the business to the poor tax and administrative policies. This is equal to 86% of all firms with such status.

Evidently, this is a key finding in respect with the functional characteristics of ICOI and economic behavioural modelling of members willing to fully formalise from within. The poor tax and administrative policies in Bulgaria, lack of fair and professional attitude by the state institutions and most importantly unfair expenditure of tax money, appear to have very pronounced effect on tax evasion by enterprises. Small businesses utilise tax evasion as a compensatory strategy to offset their perceived financial losses due to the described phenomena. Emerging from the empirical data and analysis so far, it is theorised that low institutional quality, hampered procedural fairness and unfair use of public money have clearly financial expression in the real cost of doing business (cost of formality) in transition economies (such as Bulgaria).

The suggested solution to policy challenge one above was designed to tackle formalisation resistance of ICOI and while it would also have some positive effect on

informalisation coercion forces, it would not be sufficiently adequate. A more integrative solution should be employed to address policy challenge two. As evident from previous studies (Barone & Mocetti, 2011) public spending inefficiency shapes tax morale and the negative effect is even larger if there is lower level of spending in general. Transition economies have typically less resources to spend (in comparison with advanced ones), which are even further reduced by phenomena such as high institutional corruption, which is further fuelling evasion (Çule & Fulton, 2009; Johnson et al., 2000; Sandmo, 2012). In the case of Bulgaria, the connection between low tax morale and dissatisfaction with fair public expenditure is very pronounced. Further evidence (Williams et al., 2014c, pp. 16-17) reveals the three primary reasons for entrepreneurs having lower tax morality and resistance to the state. Indeed this is due to a lack of perceived tax fairness, redistributive and procedural justice. It has been already argued (section 6.2.3, Table 22, Appendix 35) that the strongest predictor for participation in ICOI is the disagreement with fair expenditure of tax money. 86% of all firms willing to formalise from within ICOI have full disagreement with how tax money is being spent and opinion that they are not treated in a fair and professional manner by the state institutions. Relating this to the relevant findings in this part of the thesis, an upgraded policy response should involve firstly, changing the perceptions of how tax money is being spent. Secondly, shifting the perceptions of professional and fair treatment by the state institutions including NRA lastly, but not least accordingly.

Evidently (Figure 16), key state institutions are being increasingly perceived as corrupt or extremely corrupt in Bulgaria. Predictably, this casts shadows on the trust that tax money is fairly invested in public goods and services, thus affecting tax compliance levels. Given the conclusive evidence that corruption levels are significantly higher in transition economies than those in mature market economies (Tonoyan et al., 2010), the tackling challenge is proportionally complicated. Reducing the perceived levels of state corruption would positively affect the belief that tax money is fairly spent and consequently, this is expected to increase voluntary tax compliance. In consideration of the extant literature on state capture and counteracting corruption, one particular legal approach could be a credible solution for largely corrupted economies. Indeed, the adoption of asymmetric legal design, shattering some of the confidence that corrupt favours will be reciprocated and fostering denunciation (Lambsdorff & Nell, Kreutner M, 2006). In practicable terms, this means that lawful sanctions for accepting forms of gifts or bribes by state officials should be low and those for illicitly providing favours in return high. Simultaneously, expected penalties for providing gifts by tax payers aimed at achieving influence and economic gains should be severe, while those for accepting illegal favours by state officials mild. This is supposed to disturb the dualistic logic of a corrupted transaction and decrease corruption in general.
The secondary aspect concerns changing the perceptions of professional and fair treatment by the state institutions. Two of the most viable methods are the so called “service and client” approach (Hofmann et al., 2008), as well as the “slippery slope framework” (Kirchler, Hoelzl & Wahl, 2008), within a responsive regulation response. The first method (“service and client”) cultivates collaborative, supportive and serving environment, where state institutions clarify, help and assist the tax payers to achieve what they want within the law. This way state institutions gain legitimacy and induce tax payers to respond with reciprocity and compliance. The latter method, clearly distinguishes the two methods of compliance namely voluntary and enforced compliance. It assumes that institutional power and trust are equally and dynamically important. They should be engaged accordingly giving advantage to the development of cooperative, voluntary compliance. Punitive, enforcement measures should be reserved only for cases where this is not possible. To facilitate perceptions for equal treatment, state institutions could provide greater publicity, particularly so about their enforcement activities.

7.4 **Contributions and conclusions.**

With almost a quarter of the regional (South-East Europe) national income (Schneider et al., 2013, p. 56) hidden from the authorities and a similar share of undeclared employment, Bulgaria was selected as a distinctively important case to study. The country has one of the lowest GDP in the EU (Appendices 2 and 3), which correlates simultaneously to the lowest overall index of social cohesion-OISC (Dragolov et al., 2013, p. 41) and the highest informal economy share (see Figure 2, 3 and 4). As a percent from the GDP by MIMIC method, the size of the informal economy is 12.7% to 13.8% higher in comparison with 27 EU countries' averages during 2003-2013 (Schneider, 2012; Schneider, 2013). The reported socio-economic factors and the intrinsic circularity between institutions and economic development (Casson et al., 2010) provided rationale for the necessity of this research in transition setting.

This thesis employed mixed methods in exploratory sequential design to achieve its research aims. A qualitative method (in-depth semi-structured interviews) has been used to collect primary data from key revenue officials. The key finding is that while operating in the most challenging EU economic environment, NRA experiences political influence for asymmetrical tax gap closure, which leads to unjustifiable regulative formalism within an outdated deterrence orientated approach. The collected intelligence via the interviews has been analysed to inform the creation of a quantitative method – a survey was used to collect data from entrepreneurs and small business owners. Researching tax evasion is particularly challenging and sensitive subject at both institutional and firm level. The institutional strategies and approaches are understandably protected by law and the required empirical access is notoriously difficult. Formally registered firms involved in tax evasive models could face severe penalties if exposed by the revenue authorities. Such administrative and financial punishments are very different in comparison with individual tax evaders – such as undeclared workers for instance. In consideration of these facts and to fulfil the aims of the thesis, an advanced empirical strategy and methods of analysis had to be utilised. As part of the development of the semi-structured questionnaire for NRA, a successful information access request was applied for and granted by BNAO (see section 4.3.1 and Appendices 10 and 11). It provided access to a special audit report of NRA and the related extensive documents' base (about 440 documents in excess of 55 000 words), which served for its inter-organisational creation. Some of the documents were protected by the so called institutional secret by law (Bulgarian National Assembly, 2013, article 14, paragraph 14). This special report was used to acquire expert understanding of NRA activities and strategies, which in turn shaped the interview questions. Firstly, the utilised empirical approach itself reveals a good practice for deep research in similar sensitive topics particularly in transition economy setting such as Bulgaria. Secondly, it enabled the development of the quantitative questionnaire, integrating established institutional tax evasion diagnostics with the best practices currently available in the literature. It is argued that the questionnaire is a methodological contribution in diagnosing and measuring tax evasion for Bulgaria. It could be easily adapted to other transition economies. As a result, this study acquired original and significant findings in the domain of tax evasion management, which in the view of the author could influence the management and research of informal economies.

When appraising tax evasion management from institutional perspective in countries, which transformed from socialism to capitalism, it is important to account for the inherently disturbed social contract. It is consequential to the fundamental conflict "with the common opinion and interests of the people leaving extensive distrust of the government" (Vihanto, 2000). Given the findings regarding the most difficult EU economic environment for NRA, accompanied with political pressure for asymmetrical

tax gap closure, it is obvious what fuels the self-perpetuating capacity of the largest informal sector. As evident, the institutional architecture and quality seems to be a key component in understanding the shadow economy (Torgler et al., 2009). Therefore, it is argued that the formal institutional behaviour should be studied, re-designed and guided as the symmetry's facilitator with the tax payers' informal norms and beliefs to reduce informality. The qualitative part of this thesis aimed to do exactly that – to provide a critical viewpoint on risk management and tackling tax non-compliance from government agency perspective. The main qualitative research question pursued in this study can be expressed as follows: "How does a national revenue agency try to tackle tax evasion in an institutional context that makes operating within the informal economy a practically-viable alternative to operating in the formal economy?" To explore this central question, four lines of inquiry have been reported: discovery and exploration of agency strategies; agency operational approaches and the predominant type in the mix of policy measures; and the triggering procedures and methods of enforcing control and compliance – tax audits and inspections.

Being at the forefront of tax collection and non-compliance management, NRA is in a very challenging position of playing catch up in a transitioning economy. It has quite limited resources to deliver on objectives and it appears that ruling governments are not willing to account these factors. As Calciolari et al. (2013) suggests public sector organisations are simultaneously subject to three types of environmental pressure: institutional, economic and political. The key qualitative contribution of this research is unveiling the three major environmental pressures upon NRA's strategic and operational approaches in tax policy management characterised by institutional incongruence status. While the first two – economic and institutional environments are amongst the worst in socio-economic terms within the EU, the political influence for asymmetrical tax gap closure as well as consequential misuse of collected revenue, appear to have the most significant impact on the agency work. In particular, this creates two major types of difficulties – firstly, this impedes the work of senior officials who want to modernise tax policy in order to secure greater level of operational independence and thus higher procedural fairness. Based on the in-depth interviews, some key officials are genuinely involved in "bottom to top" initiative for internal reformation striving to improve the agency performance through employment of modern and less subjective information technology solutions.

Secondly, my results indicate that current institutional practices disturb the social feeling for procedural fairness to such a degree that the institutional asymmetry and entrepreneurial informality are being progressively deepened. While reducing the asymmetry between the formal and informal institutions will improve the level of formalisation amongst SMEs (and generally the share of informal economy), this could be achieved only if revenue agencies in transition economies are granted sufficient political independence from ruling governments. It also became apparent that the described issues create the so called unjustifiable regulative formalism within the traditionally deterrence orientated approach of NRA.

Based on compelling literature evidence in the institutional regulatory studies with a focus on tax compliance and procedural fairness (Hartner et al., 2008; Murphy, 2005; Murphy, International Center for Public Policy AYSoPS, Georgia State University, 2007; Murphy, 2008; Murphy, 2016; Verboon et al., 2007), it is clear that deterrence-based enforcement strategies are counterproductive. However, only in recent years has there been a fundamental shift in tax compliance research with moving the accent from enforcement approaches to such ones inducing and nurturing voluntary cooperation (Alm, Kirchler & Muehlbacher, 2012). The later work of Williams (Williams, 2014a; Williams, 2014b; Williams, 2015b) is a notable example in this respect and the fact that the current debate is not so much over whether to use direct

or indirect controls to tackle informal economies, but rather which specific policy measures are most effective and in what combination. In his development of the European Platform on Tackling Undeclared Work, Williams (2016) suggests a holistic policy approach to member states, which "uses in a strategic and coordinated manner the full range of both the direct and indirect policy approaches and measures available to increase the power of, and trust in, authorities respectively". In order to develop such sophisticated approaches, the initial requirement is to acquire an in-depth understanding of the institutional context and interactions with the tax payers for every country. The qualitative part of this thesis was designed to serve this purpose for Bulgaria and study the key agency (NRA) in that regard.

In consideration of the substantial amount of interview data (in excess of 41 000 words) and for the purpose of achieving maximum validity of findings, transparency and objectivity, a dual software solution approach was employed with two of the industry benchmark solutions. In particular, IBM SPSS Modeler and NVivo as key representatives of CAQDAS, have been utilised from the perspective of being analytical research enhancer tools. In the context of this research, it was imperative to create "auditable 'footprint' of the progressive dialogue between the researcher and their data" (Sinkovics et al., 2012, p. 5). All findings and the software inquiries to derive them have been double and triple verified via software to software logic comparison and software to researcher's logic and knowledge simultaneous appraisal. The more traditional statistical techniques (NVivo) were amalgamated with the progressive natural language processing (NLP) algorithms (IBM SPSS Modeler) in the qualitative analysis (chapter 5). It is the author's view that this dual solution approach is in itself a methodological contribution to the way qualitative analysis might be conducted independently of the subject. It allows quicker, deeper and higher level analysis to be implemented in a very visible and trustworthy way.

The main research aim in this study was to develop a holistic account of the socio-economic influencers of the processes of business tax evasion, formalisation and informalisation and as a result to inform adequate policy responses. This task required emergence of two investigative levels – institutional (chapter 5) and entrepreneurial (chapter 6). The qualitative research question pursued in this thesis can be expressed as follows: "How does a national revenue agency try to tackle tax evasion in an institutional context that makes operating within the informal economy a practically-viable alternative to operating in the formal economy?" Operating in the most challenging EU economic environment, NRA experiences political influence for asymmetrical tax gap closure, which leads to unjustifiable regulative formalism within an outdated deterrence orientated approach. The results indicate that current state institutional practices disturb the social feeling for procedural fairness to such a degree that the institutional asymmetry and entrepreneurial informality are being progressively deepened. While reducing the asymmetry between the formal and informal institutions will improve the level of formalisation amongst SMEs (and generally the share of informal economy), this could be achieved only if revenue agencies in transition economies are granted sufficient political independence from ruling governments.

The quantitative part of this mixed method thesis appraised several interconnected phenomena (sections 6.2, 7.1, 7.2, 7.3), which affect entrepreneurs' tax evasive models of behaviour. The main quantitative research questions can be expressed as follows: "What are the strongest determining factors of business tax evasion and whether the intrinsic circularity between tangible and intangible socio-economic phenomena is affecting tax evasive behaviour." The research narrative aimed to build a holistic account of these socio-economic influencers discovered to affect the processes of tax evasion, formalisation and informalisation. To achieve this task some

innovative analytical methods have been used, hence enabling the so called predictive economic profiling of business tax evaders. It was discovered that 76.5% of the surveyed entrepreneurs are underreporting their turnovers with the purpose to evade taxes. This amount to a significant proportion of the country's GDP and suggests that the proportion of GDP in the informal economy is much higher than current macro-measurement methods report for Bulgaria (31.2% of GDP). Although there is no definitive evidence in support of the supremacy of a particular data collection technique (Williams, 2015a, p. 6) for evaluating the undeclared economy, the employed method here managed to yield a higher estimate in comparison with macro-economic measurements available so far. Micro-economic (direct) approaches are criticised that they lead to estimation of only part of the shadow activities, which then should be extrapolated to the entire population (Schneider, 2005, pp. 619-620). Capitalising on that particular measurement weakness and the utilisation of a few conservative assumptions in the deployment of the quantitative method, this study provides arguments that the actual share of the informal economy is alarmingly higher from what is currently known. It is argued that the indirect appraisal technique in the business survey, which also embeds NRA's experience, is a viable example of good measurement practice applicable to transition economies. Moreover, the findings of this approach serve as foundation for further research into which methods (micro or macro) are superior for detecting and measuring informality, depending on the deployment socio-economic context. Sadly (from authorship perspective), it also informs a research challenge to other scholars willing to discover the real share of the **formal economy** in Bulgaria. In particular, which is the scientifically correct question to ask "What is the share of the informal economy in Bulgaria?" or "What is the share of the formal economy in Bulgaria?"

The reported qualitative and quantitative findings enlightened some unknown characteristics of the Bulgarian informal sector and as such led to coining new descriptive terms of the very core structure of informality in a given society. Indeed, that transition economies with large informal sectors (such as Bulgaria) are characterised by high *Frequency of societal involvedness in the informal economy* multiplied by high *Intensity of societal involvedness in the informal economy* (see section 7.1). Therefore, it has been argued here that the institutional asymmetry applies to big part of the society and to a much deeper degree in every instance. Unsurprisingly, informal economies with such characteristics are particularly challenging to tackle from policy perspective and in the case of Bulgaria, it is evident how such efforts remain unsuccessful.

The literature appraisal conducted for the purpose of this thesis did not identify other studies, which use proven multi-disciplinary analytical methods (Decision trees, MLP Neural and TAN Bayesian Networks) in holistic modelling and predicting of tax evasive behaviour. Therefore, their integrative utilisation here could be considered as a qualitatively new approach. Decision trees are popular and powerful approach (Rokach, 2016; Rokach et al., 2008) in many disciplines, including structuring policy problems (Dunn, 2012, p. 25; Holzinger, 2015). Although decision trees are considered to retain superior accuracy in comparison with logistic regression (Coussement et al., 2014) the most important findings have been double verified with both methods. The predictive economic profiling capability for the given context emanates from the fact that both neural networks and decision trees have the functionality to discover the strongest influencers for entrepreneurs to engage in tax evasive activities. This is achieved with neural computation of relationships' significance between predictor and target variables (Neural Networks) and examination of all possible single or multi-way tree splits based on statistical significance tests (decision trees). The current quantitative data set was used to develop and train several models, which are now ready for predictive re-deployment

into longitudinal data sets or real world applications such as analysis of revenue offices' data flows. While the current models have been fine-tuned to transition economy context (namely Bulgaria), they could be re-configured to work in different countries simply by introducing relevant training data sets.

The quantitative analysis (based on the second investigative method) of this thesis advances an innovative predictive economic profiling of business tax evaders utilising decision trees, neural and TAN Bayesian networks. A variety of modelling methods taken from machine learning, artificial intelligence and statistics have been used. Emerging from IBM Cross-Industry Standard Process for Data Mining (CRISP-DM), a Modelling Workflow Framework (MWF) was specifically developed and applied to this management research inquiry. Consequently, this enabled structured automated evaluation and ranking of several thousand statistical and data mining algorithms against criteria for precision, accuracy and findings' meaningfulness (chapter 6). It is argued that the automated simultaneous appraisal of linear and non-linear analytical models against the available research data is a robust approach, which is particularly useful in the domain of tax evasion management. The main strengths of the approach are based on the discovery-orientated properties of the utilised techniques in contrast with verification-oriented methods such as conventional statistical tests. MLR has been used to verify the most important aspects of the quantitative findings. Furthermore, the software solution (IBM SPSS Modeler), but not the researcher, identifies any important features and relations of the data. This is possible due to the artificial intelligence algorithms. Any exposed insights have been then scrutinised by the knowledge of the investigator and adjustments had been made where necessary into the data mining strategies and techniques. The advantages of this approach are self-explanatory – maximum robustness and objectivity where the expert's knowledge is combined and enhanced with active analysis techniques supported by artificial (computer based) intelligence. This important capability allows the inductive interrogation of research data in a predictive manner. Facilitating economic profiling of tax evaders in this thesis is exactly that – it is achieved through deriving new meaningful research explanations with decision trees and neural networks, but not limited to just confirming statistically significant hypotheses. The automated interaction identification capabilities of decision trees enable efficient detection of significant relationships between variables, which is a powerful dimensional reduction technique. The relationships identified through this technique could be further explored (if necessary) in formal parametric models. It is argued that in highly sophisticated data sets such as the one used in this study (see section 4.4), the knowledge of the researcher greatly benefits from such analytical enhancing techniques.

A fundamental associated benefit to policy makers is the fact that all types of decision tree algorithms provide some form of significance of their conditional/branching structure. Their approach, which is similar to rule induction, is always visible in terms of reasoning – it is available from visually browsing the tree. In addition, the decision tree algorithms include only predictor variables, which contribute to the accuracy of the model; all others are excluded. As a result, decision-tree algorithms are invaluable instruments in appraising how a particular change of a certain condition will affect the dependent ones. In the context of this study, they have been used for predictive profiling of tax evaders utilising a business survey from within a European transition economy, namely Bulgaria. Therefore, the utilised level of automated modelling constitutes a qualitative change in how informal economy and tax evasive issues could be appraised and tackled via targeted public policy interventions. In particular, the modelling with logically clear and transparent decision trees and neural networks provides the opportunity for unproblematic and relatively cheap prospective policy evaluation in the domain of tax evasion. The modelling approach could be also integrated with appropriate tuning into the data streams of interested revenue offices

across the world and in consequence benefit the policy makers and local business environments simultaneously.

It was illustrated that Bulgarian entrepreneurs experience very high real cost of doing business according to the rules and this has manifold basis. One aspect is rooted in the formal institutional framework: due to the poor tax and administrative policies which impose hefty financial burden on businesses. Another aspect is due to phenomenon described by the newly coined term *informal chains of interdependency*. ICOIs have serious economic coercive powers over entrepreneurs, simultaneously attracting new members and inhibiting them with prohibitively high exit costs (towards formalisation). Third aspect relates to the intense and wide spread unfair competition (types one and two, see section 7.3), clearly interconnected with institutional quality, procedural fairness and distributive justice. A key argument in this discussion was that intangible phenomena (such as institutional quality, procedural fairness and distributive justice) transform into tangible economic impacts on the profitability of small businesses. The most frequent motive for tax evasion is to simply survive economically. To demonstrate the dynamics of the described socio-economic influencers of informality, a model was developed and discussed (Figure 26). It presented how the *real cost of doing business* affects firms' economic behaviour depending on their economic status with the view to advance an improved understanding amongst interested scholars and policy makers.

In terms of cost of doing business, there are four substantial drivers of tax evasion, which develops as a responsive-compensatory strategy at entrepreneurial level. Firstly, unfair competition from fully informal or partly informal businesses is the most frequent and strongest predictor of business tax evasion. The current institutional capacity to tackle this issue and provide a levelled business environment is clearly inadequate. Secondly, the problem with institutional corruption at all levels, which causes unfair and expensive public service delivery – the so called significant costs of "getting things done" from entrepreneurial perspective. Thirdly, the high cost of doing business according to the rules due to bureaucratic inefficiencies as a result of poor tax and administrative policies. It is important to note that development of enterprise policy in Bulgaria experienced the so called policy implementation gap – where consecutive governments failed to transfer and align working EU programs to the local context. Similarly to another transition economy, various programmes have been adopted for the wrong rationales (Xheneti, 2017; Xheneti et al., 2011) doing little in solving the problems of the business community. Lastly, there is a political influence for asymmetrical tax gap closure at macro policy level. This causes guided unequal enforcement of the existing tax laws and regulations, where many entrepreneurs are exposed to unjustified regulative formalism and tax payment scrutiny. At the same time, they observe exactly the opposite status for privileged businesses and attempt to compensate with turnover underreporting where possible.

Having identified the strongest influencers of tax evasion practices through the modelling with decision trees, neural networks and MLR regression, it was revealed where government efforts would be most fruitful. In order to increase tax compliance and reduce the share of the informal economy, a qualitatively new policy approach is required to reduce the institutional asymmetry. As evident, cultivating trust in the tax authorities (Kirchler et al., 2008; Kogler, Muehlbacher & Kirchler, 2015) appear to be the way forward to shift from the less efficient enforced compliance to much more financially and socially efficient voluntary compliance. The focus has to be moved from direct punitive and deterrence oriented controls to indirect, high trust, high commitment culture, so as to align the informal and formal institutions (Williams et al., 2015a). Effectively this is an establishment of the minimum preconditions of voluntary tax compliance (Hofmann et al., 2008). Given the poor status of the social contract

from entrepreneurs' perspective, the widely perceived institutional corruption in combination with poor performance, a transformation has to start with the formal institutions first. Increasing the perceived fairness of treatment would change the motivational posture on compliance (Hartner et al., 2008). To achieve that, a fundamental institutional reform is required in two aspects. Firstly, to achieve widely acceptable level of procedural fairness, namely equivalent treatment of everybody trough equal rules enforcement in the tax management domain. Secondly, that is to address the major deficiencies in the distributive justice mechanisms and in particular the perceived value of public goods and services as a function of the collected taxes. In order to achieve appropriately sustainable social and economic system, the taxpayer should feel that they bear a reasonable tax burden (Sandmo, 2012) within an acceptable real cost of existence in relation to others. Effective delivery at these two levels satisfies the prerequisites for addressing the institutional asymmetry via change of the formal institutions, hence achieving greater symmetry with the informal ones and consequently reducing tax evasion and informality.

7.5 **Limitations and future directions.**

While measurement of the informal economy share was not one of the primary research aims of this study, the developed and applied micro-economic approach contributes to the relevant literature debates in this respect. The empirical strategy and particularly the quantitative method aimed to address the "extraordinary challenges", which this field faces (Slemrod et al., 2012, p. 31). Indeed, to avoid the weaknesses of macro-measurement approaches and measure the unmeasurable indirectly. In doing so, one of the limitations is the fact that the reported extent of tax evasion was amongst the small businesses exercising partial mode of informality in Sofia. Therefore, this is not a nationally representative study, but it clearly justifies one in respect of the reported findings. Due to more favourable economic environment in the capital, it has been argued that the level of partial informality would be higher in the rest of the country. The evaluation stated in this study does not include the share of undeclared work in combination with envelope wages and the evasive footprint of fully informal businesses. For the described reasons, there is ground to believe that the reported finding (76.5%) could be on the conservative continuum in comparison with a nationally representative study. Calculation to what proportion of the GDP all these relevant phenomena would cumulatively account is definitely an important future research direction worth pursuing.

Some of the employed analytical approaches (Decision trees, Neural and Bayesian networks, NLP) bias this thesis towards relatively new methodological paradigms in researching tax evasion and compliance, which are starting to gain popularity. Some notable examples include the use of agent-based modelling of tax compliance (Hashimzade et al., 2015; Hokamp, 2014; Pickhardt et al., 2014), as well as the utilisation of genetic algorithms for tax evasion management (Warner et al., 2015). Because of the fact that this study advances an innovative predictive socio-economic profiling of business tax evaders in transition setting, the approach needs verification in multiple economic contexts. This is necessary to establish the robustness and validity of the applied methods in various environments. Nevertheless, if this research contributed to making the point for the inevitable utilisation of artificial intelligence in tax evasion management and more broadly in formulating and resolving public policy issues, then it has fulfilled its purpose.

At present, the use of artificial intelligence and co-evolutionary approaches could be broadly considered taking into two parallel, but independent directions. Approaches based on tax evasion modelling without mining primary data, but instead, using

simulations to generate and explore tax schemes and policy reactions (Hemberg, Rosen, Warner, Wijesinghe & O'Reilly, 2016). And approaches such as the one employed in this study. Indeed, utilising primary data and training artificial cognition algorithms on real world data, thus advancing predictive economic profiling of tax evaders. Currently, there is no clear cut answer, which type of approaches would prove superior in future research. It is the view of the author that the advocated methods in this study, might offer the better match to the needs of contemporary revenue agencies and policy makers.

8. Abbreviations.

This is alphabetical list by abbreviation with their corresponding meaning.

No	Abbreviation	Meaning
1.	ADP	Anomaly Detection Procedure
2.	AIC	Akaike Information Criterion
3.	ATO	Australian Taxation Office
4.	BEEPS	Business Environment and Enterprise Performance Survey
5.	BNAO	Bulgarian National Audit Office
6.	BSP	Bulgarian Socialist Party
7.	C&RT	Classification and Regression Trees
8.	CAQDAS	Computer aided qualitative data analysis software
9.	CAQDAS	Computer Assisted Qualitative Data Analysis Software
10.	CASM	Cognitive Aspects of Survey Methodology
11.	CHAID	Chi-squared Automatic Interaction Detector
12.	CITUB	Confederation of Independent Trade Unions in Bulgaria
13.	CLP	Confederation of Labour Podkrepa
14.	CRISP-DM	Cross-Industry Standard Process for Data Mining
15.	DIY	Do It Yourself
16.	DTF	Distance to frontier
17.	EBRD	European Bank for Reconstruction and Development
18.	E-CHAID	Exhaustive Chi-squared Automatic Interaction Detector
19.	EIB	European Investment Bank
20.	EIF	European Investment Fund
21.	EUROFOUND	European Foundation for the Improvement of Living and Working Conditions
22.	GCI	Global Competitiveness Index
23.	GDP	Gross Domestic Product
24.	GLI	General Labour Inspectorate
25.	GLI	General Labour Inspectorate
26.	GNP	Gross National Product
27.	IAP	Institutional asymmetry paradigm
28.	IBMWDC	IBM Watson Developer Cloud
29.	ICOI	Informal chain of interdependence
30.	ILD	Instituto Libertad Democracia
31.	ILO	International Labor Office
32.	IR	Information request
33.	IRS	Internal Revenue Service
34.	KNN	Nearest Neighbor Analysis
35.	MIMIC	Multiple Indicators and Multiple Causes
36.	MLP	Multilayer perceptron
37.	MLR	Multinomial Logistic Regression
38.	MWF	Modelling Workflow Framework
39.	NLP	Natural Language Processing

No	Abbreviation	Meaning
40.	NRA	National Revenue Agency
41.	NSSI	National Social Security Institute
42.	OB	Optimal Binning
43.	OBA	Open Book Accounting
44.	OISC	Overall Index of Social Cohesion
45.	PPS	Purchasing Power Standards
46.	QUEST	Quick, Unbiased, Efficient Statistical Tree
47.	SBC	Schwarz's Bayesian Criterion
48.	SLR	Systematic literature review
49.	SLS	Scoping Literature Screenings
50.	SMA	Strategic Management Accounting
51.	SVM	Support Vector Machine
52.	TAN	Tree Augmented Naïve – type of Bayesian network
53.	TAP	Text Analysis Package
54.	TCMP	Taxpayer Compliance Measurement Program
55.	TPDPVT	Technical protocol of data preparation, validation and transformations

9. List of figures.

10. List of tables.

11. Appendices.

11.1 Appendix 1[1] – Systematic Review (Petticrew and Roberts, 2006, p. 283-287).

Step 1. Define the question
Clearly specify the question that the review aims to answer. If it is a review of the effects of an intervention then specify the intervention, the population, the subpopulations, outcomes of interest, the time period within which you are interested, and the cultural or other context within which the intervention is delivered. Discuss the proposed review with stakeholders during this process.

Step 2. Consider drawing together a steering or advisory group
It can be helpful to appoint a steering group, chosen to represent a range of interests; for example a review of a healthcare intervention may include a practitioner who uses the intervention, a service manager who pays for it, a patient who has experience in its use, a researcher who has previously researched or perhaps evaluated the intervention, a statistician, an economist, and so on. They will also be able to advise on the review protocol.

Step 3. Write a protocol and have it reviewed
It is essential to write a protocol stating the review question, the methods to be used, the study types and designs which the reviewer intends to locate, and by what means, and how these studies will be appraised and synthesized.
It is good practice to have the protocol reviewed by people who are likely to know something about the topic area – for example, this could include topic experts as well as the intended users.

Step 4. Carry out the literature search
Having decided the question, and discussed it with your advisory group, you will know what sort of studies you need to answer the review question. The next step is to find them, most probably searching electronic databases, bibliographies, book chapters and conference proceedings, and contacts with experts (including your advisory group).

Step 5. Screen the references
The literature search retrieves hundreds, or thousands of references, often with abstracts. These need to be sifted to identify which ones are needed for further review.

Step 6. Assess the remaining studies against the inclusion/exclusion criteria
After the clearly irrelevant studies are excluded (keeping a detailed note of the number of studies included and excluded at each stage) there are still likely to be many studies left – sometimes several hundred. Some can be confidently excluded after further examination of the abstract, but full copies of the rest of the papers may need to be obtained. These are examined to determine whether they meet the review's inclusion and exclusion criteria.

Step 7. Data extraction
Systematic reviews adopt a formal, systematic approach to extracting relevant information from primary studies; this often involves developing a data extraction form, which the reviewer completes for every study in the review. This outlines the population, details of the intervention (if any), outcomes of interest, and relevant methodological and other information. This method is intended to ensure consistency and objectivity. Data extraction also involves drawing up a detailed table describing every study that is reviewed in detail (not every study that was located in the review – only those studies that meet all the inclusion criteria).

11.2 **Appendix 1[2] – Systematic Review (Petticrew and Roberts, 2006, p. 283-287).**

Step 8. Critical appraisal
Every study in the review that meets the inclusion criteria needs to be assessed with respect to its methodological soundness. This process helps to identify any important biases. It also helps the reader interpret the data. The results of the critical appraisal are used when synthesizing the results of the primary studies.

Step 9. Synthesis of the primary studies
The included studies need to be integrated, taking into account variations in population, intervention (if any), context and setting, study design, outcomes, and the degree to which they are affected by bias. This integration can be done statistically (meta-analysis), and/or narratively – by systematically describing, reporting, tabulating, and integrating the results of the studies. Graphical displays (such as Forest plots) of quantitative data are also helpful in achieving this synthesis.

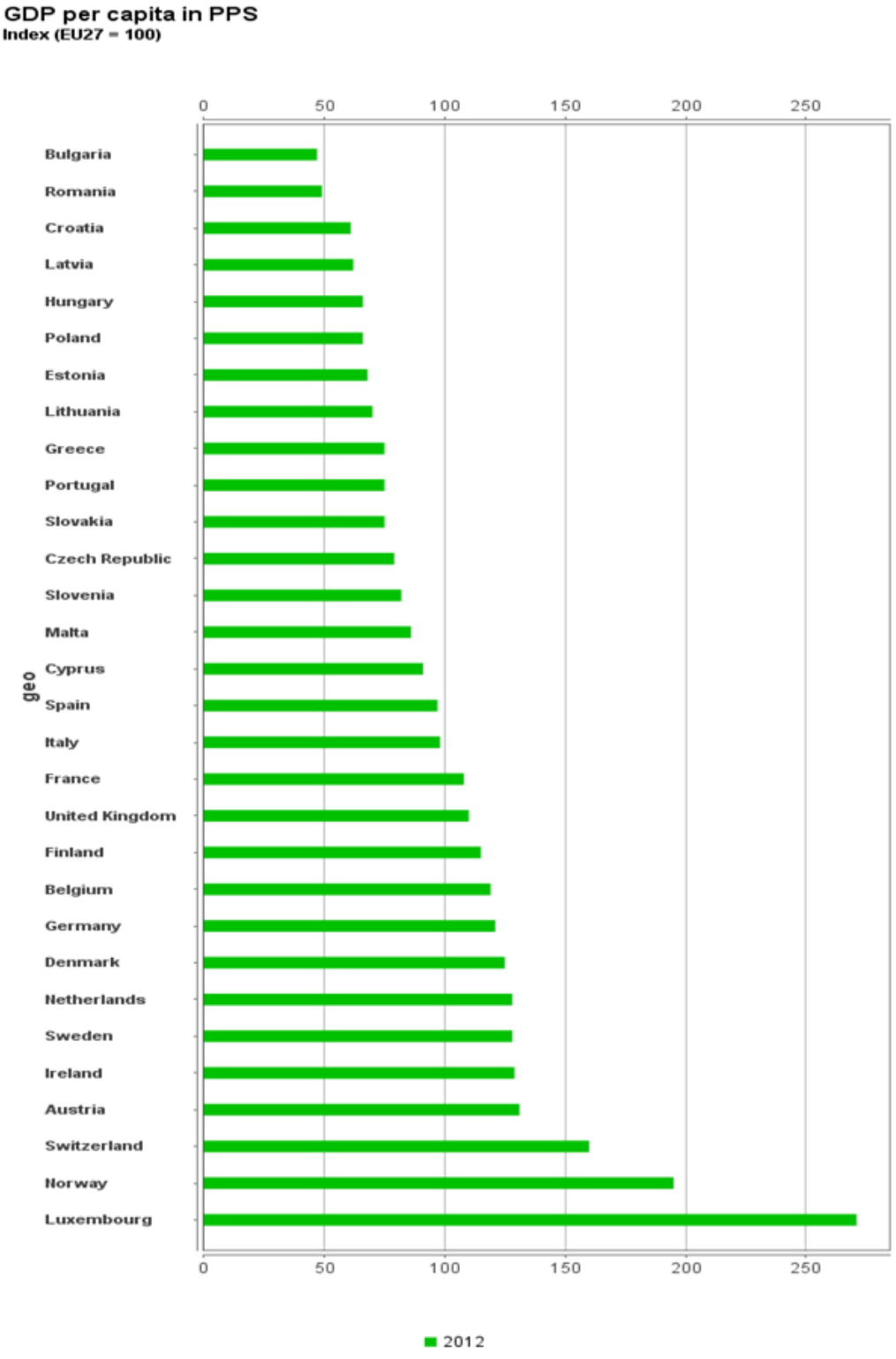
Step 10. Consider the effects of publication bias, and other internal and external biases
It is known that issues such as study size, study quality, source of funding, and publication bias can affect the results of primary studies. This can have a major impact on the conclusions of a systematic review of quantitative studies; at worst, a review may over-represent the true size of the effect in question. For quantitative studies, the effects of such a bias can be explored graphically (for example, using funnel plots) or narratively.

Step 11. Writing up the report
For many reviews, the final output is a report or journal article. In some cases it first involves producing an electronic version of the review (for example, Campbell or Cochrane reviews are made available on the Web and/or on CD). The final version of the review needs to include details of the full search, and the "flow" of studies through the review process, including how many studies were excluded at each stage, and why. Providing this information (for example, in a flow chart) is a prerequisite for publication in some journals. Some health journals also require the authors to follow the QUORUM (QUality Of Reporting Of Meta-analyses – see Appendix 3) or MOOSE (see Appendix 2) guidelines for reporting systematic reviews.

Step 12. Wider dissemination
At the start of the review, you will have defined clearly who you expect the audience for the review to be, and perhaps will have developed a plan in advance for disseminating your review's findings to them, and helping them to interpret and use them. Now you have to implement that plan. This may involve producing summaries or other versions of the review for decision makers and non-research audiences, but may also involve working with users to implement the results, and helping potential users to understand the implications of the review's findings for policy, practice, and future research, where appropriate. Conferences, briefings of groups and individuals, seminars, public meetings, public inquiries, the media, the Web, and many other outlets may play a role. You should also consider assessing the impact your review has on relevant outcomes. This may involve measuring either social or health outcomes, or process outcomes (such as how the review was perceived, and used, and whether it in fact made any contribution to decision-making). The ability of a reviewer to do any of these wider dissemination tasks is, of course, highly dependent on time, resources, and the length of their research contract.

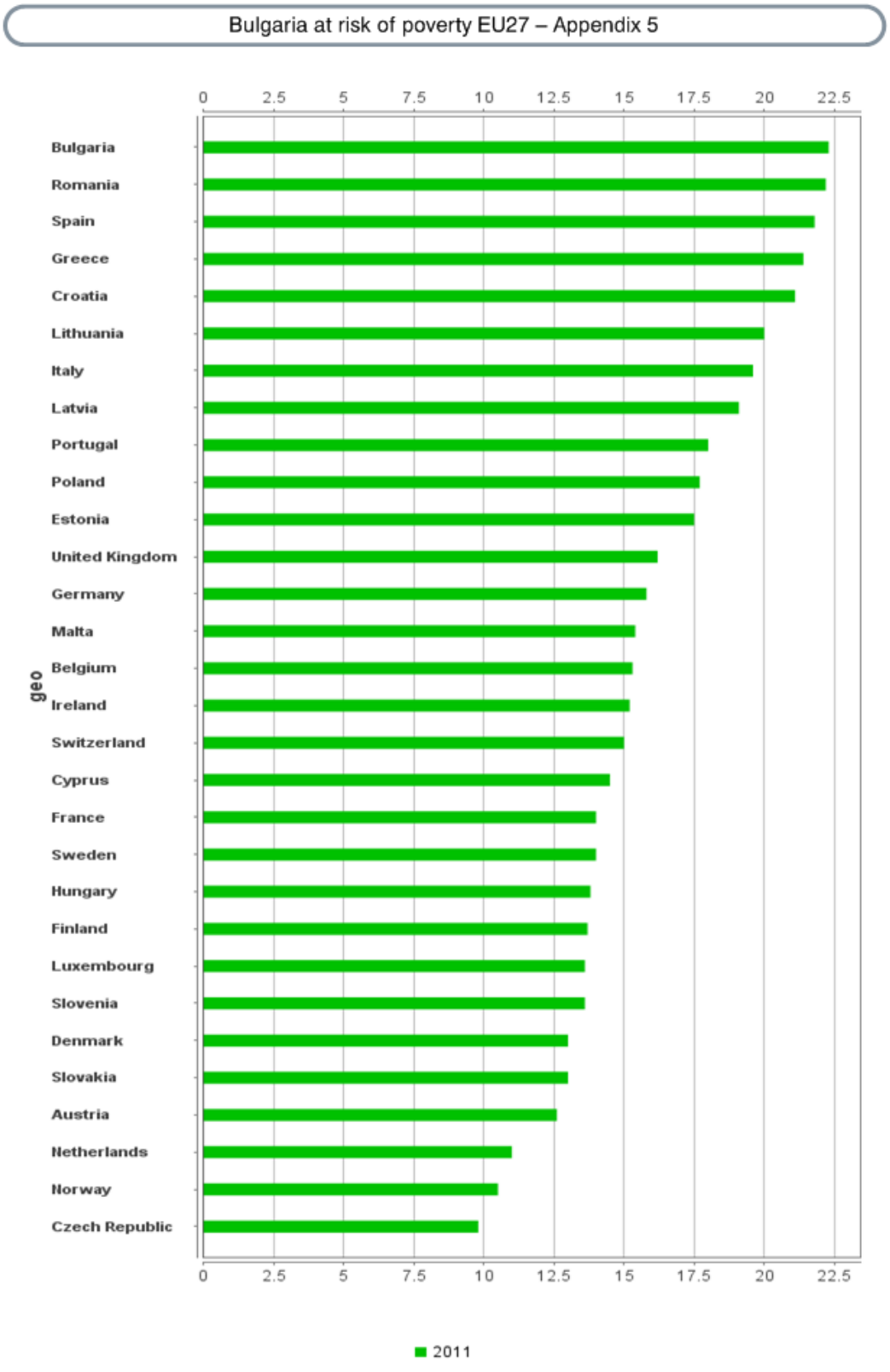
11.3 **Appendix 2 – The volume index of GDP per capita in Purchasing Power Standards EU27.**

The volume index of GDP per capita in Purchasing Power Standards EU27– Appendix 4



Source of Data Eurostat
Last update: 27.06.2013
Date of extraction: 12 Jul 2013 17:39:13 MEST
Hyperlink to the graph: http://epp.eurostat.ec.europa.eu/tgm./drawGraph.do&init=1&plugin=1&language=en&pcode=tec00114&toolbox=legend

11.4 **Appendix 3 – Bulgaria at risk of poverty EU27.**



Source of Data Eurostat
Last update: 03.07.2013
Date of extraction: 12 Jul 2013 17:49:58 MEST
Hyperlink to the graph: http://epp.eurostat.ec.europa.eu/tgm./drawGraph.do&init=1&plugin=1&language=en&pcode=tessi010&toolbox=legend

11.5 **Appendix 4 – Table 1.**

Advantages and disadvantages of systematic literature review (note – the table entries are listed in a non-particular order).

Advantages of SLR	**Disadvantages of SLR**
Reduced review bias due to rigorous criteria.	Could limit the review scope and omit important review themes.
Transparency to other readers to judge the review process and its results.	Requires significant planning and re-design with justification throughout the entire study.
Higher level or relevancy (more pragmatic).	Could reduce creativity of the review and affect cross-disciplinary perspective.
It is more evidence based compared to non SLR.	It is more complex to account for others' biases against each other.
Identifying of knowledge gaps is easier and more thorough.	Searches in electronic databases could omit important articles or books.
Could easily avoid duplication of knowledge creation.	

11.6 **Appendix 5 – Table 2.**

Table 2: Systematic literature review – key search words and expressions (alphabetical order)	
Narrower focus	**Broader focus**
Cost of formality Cost of informality Cost-benefit analysis of informality Dynamic Process Tracing Approach (DPTA) Informal entrepreneurs Informal entrepreneurship Tax evasion(excluding corporate)	Black economy Dynamic Process Tracing Approach (DPTA) Grey economy Informal economy Legal tax avoidance Optimal design of tax systems Optimal taxation Shadow economy Underground economy

11.7 **Appendix 6 – Table 3.**

List of peer-reviewed journals containing "key expressions" articles
The generated search results (journal articles) were screened to ensure that each included journal contain at least one thematically relevant publication, which has a discussion in regards to a key expression. Most of the journals contained several relevant papers. Some articles and journals accordingly, have been excluded, because the discussion of the topic was only marginal. The results are presented in the following tables.

Key expressions (alphabetical order)	**Master database**
Grey economy Informal economy Informal entrepreneurs Informal entrepreneurship Shadow economy Underground economy Black economy Tax evasion (excluding corporate)	Primo Central Full list of all the databases included in Primo central is available: (University of Sheffield, 2013)
	ScienceDirect

Table 3: List of peer-reviewed journals containing "key expressions" articles (alphabetical order).	
1. Academy of Management Perspectives 2. Accounting, Auditing and Accountability 3. Accounting, Organizations and Society 4. Africa 5. American Anthropologist 6. American Behavioral Scientist 7. Annals Of The American Academy Of Political And Social Science 8. Annals Of Tourism Research 9. Antipode 10. Applied Economics 11. Carnegie-Rochester Conference Series on Public Policy 12. Cities 13. Conservation Biology 14. Crime, Law And Social Change 15. Critical Perspectives on Accounting 16. Eastern European Economics 17. Ecological Economics 18. Economic Modelling 19. Economic Systems 20. Economics Letters 21. Economics of Transition 22. Empirical Economics 23. Engineering Economics 24. Entrepreneurship & Regional	51. International Tax and Public Finance 52. Journal Of Business Ethics 53. Journal of Business Venturing 54. Journal of Development Economics 55. Journal of Developmental Entrepreneurship 56. Journal of Economic Behavior & Organization 57. Journal of Economic Dynamics and Control 58. Journal of Economic Psychology 59. Journal of Economic Studies 60. Journal of Economics and Business 61. Journal of Financial Stability 62. Journal of Government Information 63. Journal Of International Business Studies 64. Journal of Macroeconomics 65. Journal of Management Studies 66. Journal Of Monetary Economics 67. Journal of Policy Modelling 68. Journal Of Public Economics 69. Journal of Socio-Economics 70. Journal Of Youth Studies 71. Labour Economics 72. Local Government Studies 73. New Political Economy 74. Physica A: Statistical Mechanics and its Applications

Development 25. European Economic Review 26. European Journal of Law and Economics 27. European Journal of Political Economy 28. European Journal Of Public Health 29. European Journal Of Social Work 30. European Management Journal 31. European Societies 32. Europe-Asia Studies 33. Expert Systems With Applications 34. Futures 35. Geoforum 36. Geographical Review 37. Geography Compass 38. German Economic Review 39. Habitat International 40. Health Policy 41. International Entrepreneurship And Management Journal 42. International Journal of Educational Development 43. International Journal of Entrepreneurial Behaviour & Research 44. International Journal of Human Resource Management 45. International Journal of Social Economics 46. International Journal of Urban and Regional Research 47. International Review of Economics & Finance 48. International Review of Law and Economics 49. International Small Business Journal 50. Journal of Ethnic and Migration Studies	75. Post-Communist Economies 76. Procedia - Economics and Finance 77. Procedia - Social and Behavioral Sciences 78. Progress In Development Studies 79. Public Choice 80. Regional Studies 81. Research in Economics 82. Research in Social Stratification and Mobility 83. Research Policy 84. Resources, Conservation and Recycling 85. Review Of African Political Economy 86. Revue Économique 87. Science 88. Security Dialogue 89. Social & Cultural Geography 90. Social Forces 91. Social Science & Medicine 92. Socio-economic Review 93. Sociological Perspectives 94. Sociological Perspectives 95. Structural Change and Economic Dynamics 96. Technological and Economic Development of Economy 97. Technological Forecasting and Social Change 98. Textual Practice 99. The economics of transition 100. The Journal of Modern African Studies 101. The Journal of Socio-Economics 102. Urban Forum 103. Women's Studies International Forum 104. World Development

11.8 **Appendix 7 EN – semi-structured interview with NRA officials template.**

Semi-structured interview with NRA officials – EN template

1. What are the main tasks and responsibilities of your department?

2. What challenges do you encounter in the implementation of your tasks and responsibilities?

3. One of NRA's strategic aims is to encourage voluntary tax compliance. What initiatives and/or activities do you undertake in that regard?

4. Does NRA have specific methods or systems in place with which you reveal the ways and means used by the tax payers to evade their tax duties or lie to you – for instance concealment of business turnover?

Special Block of Questions

5. Do you have a process for selection in regards to tax audits and inspections? If yes, then what types of selection do you exercise and to what extent are they automated.

6. What kind of events could initiate the selection processes – for example, a particular risky business behaviour, means/ways of business funding? In order to attract a selection the business subjects should possess certain type of characteristics – what are they?

7. Are your selection criteria dynamic or static? I mean whether they are preliminary defined without an option to change or could they change in accordance with the type and quality of gathered information about the tax subjects?

8. How these criteria and principles are being developed? Is it happening in your department?

9. If a high risk of tax non-compliance has been suspected or diagnosed, what are the next diagnostics and control activities?

10. Are the assignments and management of tax audits and inspections automated to some extent or not at this moment of time?

11. In your opinion, is there any subjectivism in the selection, delegation and management process of tax audits and inspections? Or is this avoided through a level of automation of those processes?

12. What are the different types of investigations, which you undertake when you suspect tax law non-compliance or simply you would like to find out if that is the case?

13. Is there a process for analysis of the possible measures, the implementation

expenses for them and the expected effect?

14. Do you cooperate with other institutions and agencies in regards with the tax audits and inspections, which are undertaken by NRA and if yes with whom?

15. What penalties are being imposed when you find out infringements?

16. How many risk subjects (firms) are identified and analysed approximately on average annually – for instance in years 2012 and 2013? How many of these identified firms are actually audited (checked)?

17. How is it decided into which areas of tax non-compliance to invest efforts? Said differently, which types of tax infringements attract most attention?

18. What is being aimed to achieve with the application of counter-activities against real or potential tax law offences? Are these measures directed towards specific groups breaking the law or against particular people – business owners or physical persons?

19. What are the most frequent reasons/motives for turnover underreporting by the Bulgarian businessmen (entrepreneurs)?

20. Is there any data collection and interest about consequential behavioural changes after applied control by NRA – tax audits, inspections and alike. If yes, what is being done in this connection?

21. If you have noticed change in the behaviour of firms and/or individuals as a result of counter evasive activities, is there any mechanism of changing any future counter activities and if yes how is this happening?

22. How is it being decided/analysed which tax law counter activities would have led to the behavioural change of the targeted groups or individuals? In other words, how is being evaluated, which measures have been effective, if such practices exist at this moment of time?

23. At this moment of time what methods or campaigns are being used to encourage voluntary tax compliance of firms and traders?

24. From your perspective, what are the characteristics of the firms and traders who allow themselves to declare lower tax base than the real one? Is there something particularly common between them or something which differentiate them from the other tax payers who are compliant?

25. In your view, what kinds of difficulties are being faced by the Bulgarian businessmen (entrepreneurs) in their efforts to be tax compliant?

26. Is there any information, which is of particular interest to you about the firms and traders evading taxes, which you would like to know?

11.9 **Appendix 8 – justification of each question – semi-structured interview with NRA officials.**

1. What are the main tasks and responsibilities of your department?

This question collects information about the main tasks and responsibilities of the department from the individual perspectives of the key people. It aims to discover the actual level of alignment in the different opinions and thus to build a picture of departmental role cohesion and potential related issues?

2. What challenges do you encounter in the implementation of your tasks and responsibilities?

This question aims to reveal what are the difficulties which are currently encountered by NRA officials. The question is asked in such a way to provoke the respondents to speak about their real professional challenges and thus share their impressions about things that needs addressing.

3. One of NRA's strategic aims is to encourage voluntary tax compliance. What initiatives and/or activities do you undertake in that regard?

The purpose of question 3 is to discover how NRA satisfies one of its strategic aims. Is "encouraging voluntary tax compliance" just an advertised strategic aim of NRA or it is genuinely pursued. Based on the collected information it could be determined to what degree indirect policy measures are being favoured by NRA.

4. Does NRA have specific methods or systems in place with which you reveal the ways and means used by the tax payers to evade their tax duties or lie to you – for instance concealment of business turnover?

Question 4 is designed to reveal facts about the current approach in tackling informal practices exercised by businesses. It is generally questioning whether there is some level of automation (systems) employed to tackle tax evasion. This is the introductory question to the so called Special Block of Questions.

This special section of the questionnaire asks a set of particular questions about:

- How the so called selection criteria work for discovery and diagnostics or tax evasion.
- The development principles of those criteria – this may provide information for analysis of the rational of the selection process and how it is institutionally installed – what is the connection with the real world and how potential feedback could affect the process.
- What is the level of subjectivism in the selection, delegation and management process of tax audits and inspections.

Special Block of Questions

5. Do you have a process for selection in regards to tax audits and inspections? If yes, then what types of selection do you exercise and to what extent are they automated.

Aims to discover whether there is a systematic selection process and any level of automation.

6. What kind of events could initiate the selection processes – for example, a particular risky business behaviour, means/ways of business funding? In order to attract a selection the business subjects should possess certain type of characteristics – what are they?

This question is asked to learn the triggering characteristics of a selection process. Collecting such information is beneficial for analysing the effectiveness and adequacy of such processes in comparison with contemporary evasive practices.

7. Are your selection criteria dynamic or static? I mean whether they are preliminary defined without an option to change or could they change in accordance with the type and quality of gathered information about the tax subjects?

The main purpose here is to reveal any feedback mechanism with the real world. To what extent NRA considers their tactics (in terms of selection process) based on genuine correlation with tax evasive practices they observe – adequacy of their response.

8. How these criteria and principles are being developed? Is it happening in your department?

Further elaboration on the above matters and discovery to what degree each of the two key departments are being involved in the criteria development.

9. If a high risk of tax non-compliance has been suspected or diagnosed, what are the next diagnostics and control activities?

This question's purpose is to reveal NRA response upon discovery of non-compliance, but most importantly how systematic it is in reality.

10. Are the assignments and management of tax audits and inspections automated to some extent or not at this moment of time?

This is a very topical question in regards to the high corruption level in Bulgaria – affecting institutions and administrations of all types. It actually aims to expose the level of subjectivism and opportunities to influence impending or on-going audits and inspections. For instance HMRC are having almost full automation of this process while this is not the case for NRA.

11. In your opinion, is there any subjectivism in the selection, delegation and management process of tax audits and inspections? Or is this avoided through a level of automation of those processes?

This last question from the special section is designed to collect and confirm personal opinion about any subjectivism in regards with tax audits and inspections.

12. What are the different types of investigations, which you undertake when you suspect tax law non-compliance or simply you would like to find out if that is the case?

This question relates to question 9 and being asked 3 questions later in the interview, it aims to confirm and reveal additional aspects of NRA tackling tactics – such ones that might not be disclosed during question 9.

13. Is there a process for analysis of the possible measures, the implementation expenses for them and the expected effect?

It is a well-known fact the most contemporary tax offices are deciding upon their engagements (such as audits and inspections) based on embedded cost benefit ratio procedures. So, this question is looking to find out how it is at NRA – operational efficiency and real world feedback.

14. Do you cooperate with other institutions and agencies in regards with the tax audits and inspections, which are undertaken by NRA and if yes with whom?

Implicitly, how well NRA cooperates with other institutions. There is an existing agreement between NRA, GLI and NSSI for cooperation, but it is a strategic level only. The question aims to provide inside information of what is actually being done and how.

15. What penalties are being imposed when you find out infringements?

It is rather technical question and its main purpose is to find out about the general approach – types of penalties for non-compliance.

16. How many risk subjects (firms) are identified and analysed approximately on average annually – for instance in years 2012 and 2013? How many of these identified firms are actually audited (checked)?

One of the pressing issues with NRA is the huge number of audits and inspections (based on publicly available data) they are conducting nationwide. This (apart from the huge administrative cost) has a repressive nuance speaking from a policy perspective. Thus, the question aims to find out an internal NRA way of work issue – the correlation between high risk firms and actually audited ones.

17. How is it decided into which areas of tax non-compliance to invest efforts? Said differently, which types of tax infringements attract most attention?

This question relates to two previous questions (13 and 16). It is part of the so called gradual approach in extracting related information. Again, it seeks information about a relationship between applied compliance measures and expected outcome (cost benefit relationship).

18. What is being aimed to achieve with the application of counter-activities against real or potential tax law offences? Are these measures directed towards specific groups breaking the law or against particular people – business owners or physical persons?

This is a two-fold question. Firstly, it aims to discover the accent on policy measures. Which one is more important from NRA perspective – investing efforts to prevent tax

evasion or to treat it. Secondly, does NRA project the measures in question towards specific groups with something in common or they are individually tailored to particular tax subjects. In other words, what is the scope of the tackling response.

19. What are the most frequent reasons/motives for turnover underreporting by the Bulgarian businessmen (entrepreneurs)?

Clearly, the question is after NRA opinion about the most frequent reasons/motives for turnover underreporting by the Bulgarian entrepreneurs. At a later stage, after the business survey, it would be possible to analyse the entrepreneurs' motives to evade taxes and how this corresponds to NRA' views.

20. Is there any data collection and interest about consequential behavioural changes after applied control by NRA – tax audits, inspections and alike. If yes, what is being done in this connection?

This question is a further elaboration about the efficiency of NRA operations the recourses they spend to undertake them. The question is designed to function together with questions 13, 16 and 17. Positioning all these questions (13, 16, 17, 20, 21 and 22) in that specific order and intervals during the interview serves two purposes:

- These questions need to be asked according to the research questions.
- The questions are being asked in a specific sequence and formulated in a manner, which allows information accumulation and cross-verification based on respondents' answers.

21. If you have noticed change in the behaviour of firms and/or individuals as a result of counter evasive activities, is there any mechanism of changing any future counter activities and if yes how is this happening?

Question 21 is connected with 13, 16, 17, 20, 21 and 22. In addition to the above explanation, it aims to reveal any institutional culture of responsiveness. This aspect will be important in discussing the results from the entrepreneurs' survey.

22. How is it being decided/analysed which tax law counter activities would have led to the behavioural change of the targeted groups or individuals? In other words, how is being evaluated, which measures have been effective, if such practices exist at this moment of time?

This is the last question from a special series (13, 16, 17, 20, 21 and 22). The purpose of the series has been explained above.

23. At this moment of time what methods or campaigns are being used to encourage voluntary tax compliance of firms and traders?

This question is extension of question 3 and aims to collect some more information in that regard.

24. From your perspective, what are the characteristics of the firms and traders who allow themselves to declare lower tax base than the real one? Is there something particularly common between them or something which differentiate them from the other tax payers who are compliant?

The main purpose of this question is to find out whether NRA observed specific features characterising the informal firm or these firms, which regularly exercise

partially informal practices. This information will be used in the cross-analysis of the business survey, which will aim to provide some meaningful insights in this regard.

25. In your view, what kinds of difficulties are being faced by the Bulgarian businessmen (entrepreneurs) in their efforts to be tax compliant?

Question 25 would like to discover the opinion of key NRA officials about administrative and other difficulties experienced by Bulgarian businessmen, which drive them into informal practices. At a later stage, this would provide interesting ground for analysis when the results from the business survey are available – the perspective of the entrepreneurs.

26. Is there any information, which is of particular interest to you about the firms and traders evading taxes, which you would like to know?

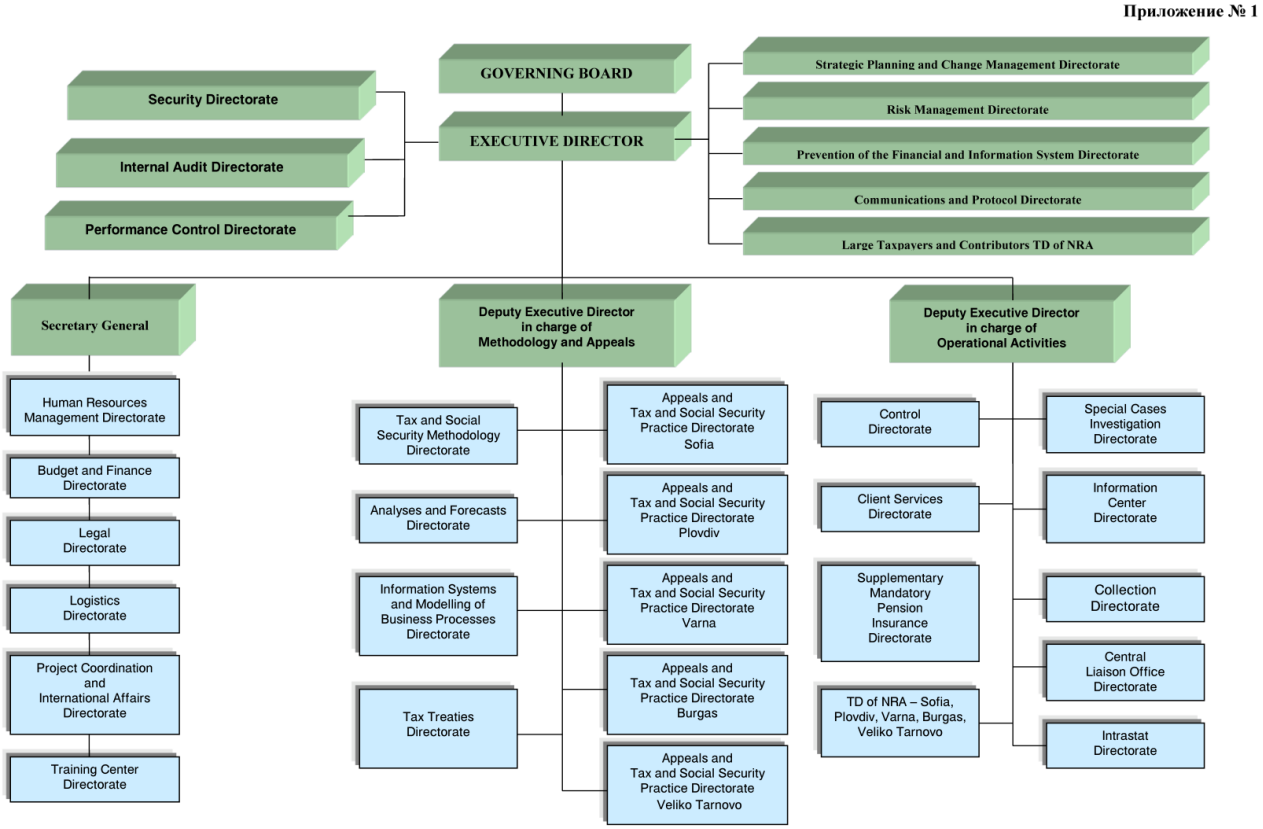
This last question has couple of implicit purposes:

- Firstly, it attempts to discover what kind of information NRA is after in regards to informal entrepreneurs (businesses). Thus, some conclusions about their approaches could be made.

- Secondly, the question aims to collect insights and inform the creation of question(s) for the business survey.

11.10 **Appendix 9 – NRA organisational structure.**

NRA Headquarters organigram retrieved on 26.09.14 from: http://www.nap.bg/en/document?id=135



11.11 **Appendix 10 – information access request 1 of 2.**

ДО
ПРЕДСЕДАТЕЛЯ НА СМЕТНАТА ПАЛАТА ПРОФ. ВАЛЕРИ ДИМИТРОВ

З А Я В Л Е Н И Е

ЗА ДОСТЪП ДО ОБЩЕСТВЕНА ИНФОРМАЦИЯ

от....Любомир Руменов Мишков.....
(трите имена на физическото лице, съотв. наименованието и седалището на юридическото лице и трите имена на неговия представител)

.....
.....

Актуален адрес за кореспонденция: Улица „Гео Милев“ 132, ет. 2, ап. 28, София

Служебен адрес: Management School Doctoral Centre, Room 8, 169-171 Northumberland Road, Sheffield, S10 1DF, UK

телефон за връзка: ..0887674609....., ел. поща ...Lrmishkov@sheffield.ac.uk.....

УВАЖАЕМИ ПРОФ. ДИМИТРОВ,

На основание Закона за достъп до обществена информация, моля да ми бъде предоставена следната информация (ако прецените, че мога да получа достъп до тази информация):

до долуизброените одитните доказателства по Одитен Доклад на Сметната палата (№ 0300000412 за извършен одит на изпълнението на контролните и принудителни действия за спазване на данъчното и осигурителното законодателство в Националната агенция за приходите за периода от 01.01.2009 г. до 31.12.2011 г.).....

Съобразно „ОПИС НА ОДИТНИТЕ ДОКАЗАТЕЛСТВА“ (страници 102-104 от Одитен Доклад № 0300000412)

1.4. Данъчно-осигурителна методология
1.7. Управление на риска
1.8. Разследване на особени случаи
1.10. Моделиране и проектиране на бизнес процесите
Точка 5 Годишна програма за спазване на законодателството и минимизиране на риска в НАП за 2010 г., 2011 г. и 2012 г.
Точка 6 Справки за извършените проверки за периода 2009-2011 г.

(включително за извършените оперативни проверки)
Точка 7 Стратегия за управление на риска в НАП от 2011 г.
Точка 8 Политика по информационна сигурност в НАП
Точка 9 Правила за правата и задълженията на потребителите, ползващи информационните активи в НАП
Точка 13 20 броя извършени анализи от дирекция УР за периода 2009-2011 г.
Точка 14 Анализ на оценката и управлението на риска в НАП
Точка 19 Анализ на резултатите от анкетното проучване. Препоръки и предложения за подобрение – 2010 г.
8: Работен документ за обобщаване на резултатите от проведената анкета № РД-И-40, с приложен диск с електронни таблици за обработка на данните
11: РД-И-7 от 25.04.2012 г. Протокол от проведено наблюдение на ПП Селекция, с приложени към него извлечения за част от основните функционалности
11: РД-И-7 от 25.04.2012 г. Протокол от проведено наблюдение на ПП Селекция, с приложени към него извлечения за част от основните функционалности

Причини за молбата ми: в момента правя докторантура във връзка със сивата икономика и избягване на данъци в Университета Шефийлд и се нуждая от тази информация във връзка с анализ, който искам да направя. Това са линкове към страници на университета, където има информация за мен, но ако е необходимо мога да поискам и представя официално писмо от Университета:
http://www.sheffield.ac.uk/management/doctoral_researchers/lyubomir_mishkov
http://www.shef.ac.uk/creed/crisp (член съм на Cluster for Research on the Informal Sector and Policy)

Моля да получа исканата информация в следната форма:
(отбелязва се предпочитаната форма)
1. Електронна поща или CD, а при невъзможност за технически носител то тогава на хартиен носител.

Дата……10.04.14……

Подпис……. **……**

11.12 **Appendix 11 – information access request 2 of 2.**

ДО
ПРЕДСЕДАТЕЛЯ НА СМЕТНАТА ПАЛАТА ЛИДИЯ РУМЕНОВА

З А Я В Л Е Н И Е

ЗА ДОСТЪП ДО ОБЩЕСТВЕНА ИНФОРМАЦИЯ

от....Любомир Руменов Мишков.....
(трите имена на физическото лице, съотв. наименованието и седалището на юридическото лице и трите имена на неговия представител)

.....

Актуален адрес за кореспонденция: Улица „Гео Милев“ 132, ет. 2, ап. 28, София

Служебен адрес: Management School Doctoral Centre, Room 8, 169-171 Northumberland Road, Sheffield, S10 1DF, UK

телефон за връзка: ..0887674609....., ел. поща ...Lrmishkov@sheffield.ac.uk.....

УВАЖАЕМА Г-ЖО ЛИДИЯ РУМЕНОВА,

Във връзка с предоставения ми достъп – преглед на информация, за което благодаря, моля да ми бъде предоставена следната информация в електронна форма:

до долуизброените одитните доказателства по Одитен Доклад на Сметната палата (№ 0300000412 за извършен одит на изпълнението на контролните и принудителни действия за спазване на данъчното и осигурителното законодателство в Националната агенция за приходите за периода от 01.01.2009 г. до 31.12.2011 г.).....

Съобразно „ОПИС НА ОДИТНИТЕ ДОКАЗАТЕЛСТВА“ (страници 102-104 от Одитен Доклад № 0300000412)

1.1 Контрол
1.2 Събиране
1.4. Данъчно-осигурителна методология
1.7. Управление на риска
1.8. Разследване на особени случаи
1.10. Моделиране и проектиране на бизнес процесите
Точка 5 Годишна програма за спазване на законодателството и

минимизиране на риска в НАП за 2010 г., 2011 г. и 2012 г.
Точка 6 Справки за извършените проверки за периода 2009-2011 г. (включително за извършените оперативни проверки)
Точка 7 Стратегия за управление на риска в НАП от 2011 г.
Точка 8 Политика по информационна сигурност в НАП
Точка 9 Правила за правата и задълженията на потребителите, ползващи информационните активи в НАП
Точка 11
Точка 13 20 броя извършени анализи от дирекция УР за периода 2009-2011 г.
Точка 14 Анализ на оценката и управлението на риска в НАП
Точка 15 Анализ на действията на публичните изпълнители от функция «Събиране»и на органите по приходите от функция «Контрол» при производството по налагане на предварителни обезпечителни мерки през 2010 г.
Точка 16 Анализ на процеса по установяване на нарушенията, образуване на административнонаказателни производства и издаване на наказателни постановления при извършване на проверки от органите по приходите от отдел «Оперативни проверки»
Точка 19 Анализ на резултатите от анкетното проучване. Препоръки и предложения за подобрение – 2010 г.
8: Работен документ за обобщаване на резултатите от проведената анкета № РД-И-40, с приложен диск с електронни таблици за обработка на данните
11: РД-И-7 от 25.04.2012 г. Протокол от проведено наблюдение на ПП Селекция, с приложени към него извлечения за част от основните функционалности

Съобразно моето предишно заявление вх. № 94-00-108/11.04.2014 съм получил достъп преглед на информация до всички гореизброени доказателства с изключение на тези по точки 1.1, 1.2, Точка 15 и Точка 16, които са допълнени в сегашната ми молба. Поради големия обем на изброените документи Ви моля любезно да получа достъп в електронна форма (в каквато форма Сметната Палата разполага и аз се запознах), а за точки 8 и 9 на хартиен носител (съответно по 15 и 9 страници).

Причини за молбата ми: в момента правя докторантура във връзка със сивата икономика и избягване на данъци в Университета Шефийлд и се нуждая от тази информация във връзка с анализ, който искам да направя. Това са линкове към страници на университета, където има информация за мен, но ако е необходимо мога да поискам и представя официално писмо от Университета:
http://www.sheffield.ac.uk/management/doctoral_researchers/lyubomir_mishkov
http://www.shef.ac.uk/creed/crisp (член съм на Cluster for Research on the Informal Sector and Policy)

Моля да получа исканата информация в следната форма:
(отбелязва се предпочитаната форма)
1. Електронен носител.

Дата……09.05.14……

Подпис…… **……**

11.13 **Appendix 12 – IBM SPSS Modeler categorisation techniques.**

Adapted from IBM (2014, p. 216-220)

Categorisation technique	**Explanation**
Concept Inclusion	The *concept inclusion* method uses algorithms to create categories by taking a term and finding other terms that include it. When determining inclusion, word order and the presence of such words as *in* or *of* are ignored. As an illustration, if you have the term *skill*, term inclusion will group terms such as *computer skills* and *skill set* in a *skill* category. The root term used to create the category can have words before it, after it, or both before and after (*computer skill set*). As an example of how word order is ignored, the term *advanced spanish course* will be included with *course in spanish*. Concept inclusion may give better results when the responses contain a lot of domain-specific terminology or jargon. This is especially true if you have tuned the dictionaries beforehand so that the special terms are extracted *and* grouped appropriately (with synonyms).
Semantic Networks	The *semantic networks* method creates categories using a semantic/lexical network based on WordNet, a linguistic project based at Princeton University. WordNet is a reference system of —English nouns, verbs, adjectives and adverb organized into synonym sets, each representing one underlying lexical concept.‖ WordNet uses the word —synonym‖ in a broader sense than the usual meaning (information on WordNet can be found at http://wordnet.princeton.edu). This technique begins by identifying extracted terms that are known synonyms (in the usual sense) and *hyponyms*. A hyponym is a word that is more specific than the category represented by a term. Thus, if *animal* is a term, *cat*, *dog* and *kangaroo* are hyponyms of *animal*; that is, things that are examples of an animal. Thus, if we extracted terms *cat*, *dog* and *kangaroo*, a category containing all these terms could be created representing *animals*. If we than had the terms of *oak tree* and *sunflower*, a second category could be created representing *plants*. These two categories could then be combined in another representing *living things*. Taken individually, many terms, especially single words, are ambiguous in meaning. For example, the term *buffet* can denote a type of meal or a piece of furniture. If a set of terms to be categorized includes *buffet, meal* and *furniture*, then the semantic network will have to choose between grouping *buffet* with one of the other two (because they are more generic). The choice made by the software may not be appropriate for your project, so results need to be reviewed carefully. The semantic network is mainly restricted to synonym and hyponym relationships. The technique can also group terms representing place names in part/whole relationships (terms of type *Location*). Thus, the technique will group U.S. states into a category representing the United States.

	The semantic network approach can yield better results than term inclusion with two types of data. First, when you expect to have terms that are related (*computer* and *hard drive*), and are interested in the relationships, the semantic network method is ideal. Second, when the open-ended responses are longer and contain more complex phrases, this method can often capture this information. The semantic network performs less well with highly technical or specialized terms.
Concept Derivation	The *concept derivation* method groups terms by looking at the endings (suffixes) of each component in a term and finding other terms that have corresponding components with a related ending (suffix). It uses a set of linguistic derivation rules to accomplish this. For example, there is a rule that says that a term component ending with the suffix —ical‖ might be derived from a term having the same stem and ending with the suffix —ic.‖ Using this rule, the algorithm would be able to group the terms *geological study* and *geologic studies*. Concept derivation ignores function words (such as *in* and *of*) and component order. Thus, by using another rule relating components ending in —y‖ to those ending in —ical,‖ the algorithm would also be able to group the terms *studies in geology* and *geological studies*. The set of component derivation rules has been chosen so that most of the terms grouped by this algorithm are synonyms. To increase completeness, there are some derivation rules that allow the algorithm to group terms that are situationally related. For example, the algorithm can group terms such as *career builder* and *career building*. You can use concept derivation on any sort of text. It produces fairly few categories, and each category tends to contain few terms. You may find it helpful to use this algorithm even if you are building categories manually; the synonyms it finds may be synonyms of those terms you are particularly interested in.

11.14 **Appendix 13 – full categorisation list of interviews' analysis with IBM SPSS Modeler.**

Top Level Category	Sub-level 1.1	Sub level 1.2	Sub level 1.3	
Risk Management	121 _risk analysis _difficulties _macro level _decision _evaluation _analysis _procedure _operational level _risk subjects _signal _part of risk _deal _aims _efforts _existence of the risk	8		
	automation	21 _automated selection _automation _computer _level of automation _computer system _new information system	8	
		information system	8 _information systems _database _module _oracle _new system _system _new information system _information system	8
		implementation	5 _implement | implementation _implement	8

Top Level Category	Sub-level 1.1	Sub level 1.2	Sub level 1.3	
			_implementation	
			_implementation expenses	
			_process of implementation	
		software	2	4
			_software product	
			_software	
	campaigns	15	8	
		_brochures		
		_information campaigns		
		_communication department		
		_campaigns		
		_information campaigns | informational campaigns		
		_communication		
		_initiatives		
		_media campaigns		
		_informational campaigns		
		letter	3	4
			_letter	
			_warning letter	
			_warning letters	
		calls	3	7
			_phone	
			_phone calls	
			_call	
	risks	11	8	
		_[risks + .]		
		_level of risk		
		_treatment of the risk		
		_methodology		
		_identification		
		_risks		
		_risk		
		_prioritization		
		_[existence of the risk + .]		
		_[risks + no]		

Top Level Category	Sub-level 1.1	Sub level 1.2	Sub level 1.3
	management	_existence of the risk 10 _diagnostics _deregistration _countermeasures _outcome _risk management _analysts _management process _management _tax law counter activities _combat	8
	individuals	7 _groups of individuals _liable individual _risky individuals _people _risky individual _group of individuals _individuals	8
	process	6 _process _business processes _passive _management process _selection processes _process of implementation	8
	approach	6 _prioritization of risk _approach _counter evasive activities _assessment _holistic approach _priority	8
	information	5	8

Top Level Category	Sub-level 1.1	Sub level 1.2	Sub level 1.3	
		_new information _information _search _external sources		
		research	1 _data	8
	criteria	5 _selection _principles _set criteria _criteria _selection criteria	8	
	data	3 _data _data collection _data mining	8	
	effect	3 _fiscal effect _effect _expected effect	8	
	practice	3 _institutional practices _practices exist _practice	8	
	strategies	3 _strategic level _strategies _goal	8	
	risk individuals	2 _risk individuals | risky individuals _risk individuals	6	
	indicators	2 _indicators | indication _indicators	5	
	project	2	4	

Top Level Category	Sub-level 1.1	Sub level 1.2	Sub level 1.3
		_pilot projects	
		_project	
	result	2	7
		_result	
		_good results	
Control	48	8	
	_control department		
	_development of control		
	_watch		
	_fiscal control		
	_applied control		
	_control actions		
	_tax people		
	_actions		
	_control		
	_subjectivism		
	_activities		
	_directorate control		
	_control activities		
	inspections	9	8
		_[labor inspection + .]	
		_[cross-inspections + .]	
		_[operational inspections + .]	
		_signal	
		_inspections	
		_[inspections + .]	
		_operational inspections	
		_cross-inspections	
		_labor inspection	
	audit	8	8
		_audit	
		_auditors	
		_[audit costs + .]	
		_audit costs	
		_[tax audits + .]	

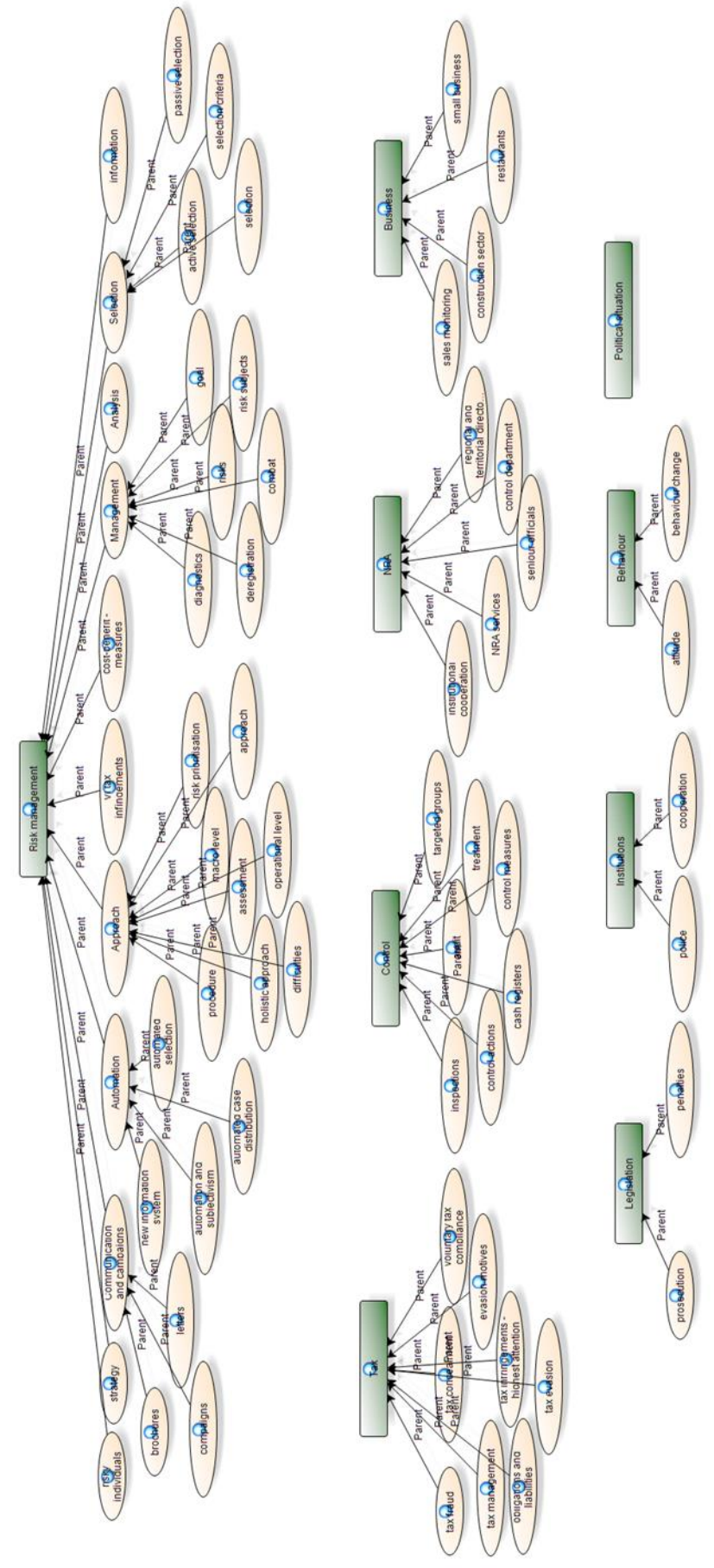
Top Level Category	Sub-level 1.1	Sub level 1.2	Sub level 1.3
		_tax audits	
		_[audit + .]	
		_check	
	measures	7	8
		_measures	
		_violent measures	
		_methods	
		_collection	
		_initiatives	
		_investigation of special cases	
		_treatment	
	selections	4	8
		_selections	
		_active selection	
		_selection criteria	
		_selection bodies	
	groups	3	8
		_groups of people	
		_targeted groups	
		_groups	
	management of tax	2	8
		_management of tax	
		_management of tax audits	
	monitor	2	4
		_monitor | monitoring	
		_monitor	
Tax	46	8	
	_tax legislation		
	_tax infringements		
	_subject		
	_tax authority		
	_tax evasion		
	_tax payment		
	_areas of tax		
	_tax subjects		

Top Level Category	Sub-level 1.1	Sub level 1.2	Sub level 1.3	
	_tax liabilities			
	_taxes			
	_tax people			
	_tax base			
	_persons			
	_tax duties			
	_tax obligations			
	_pay taxes			
	_people			
	_voluntary tax compliance			
	_business subjects			
	_tax payers			
	compliance	6	8	
		_compliance		
		_voluntary compliance		
		_obligations		
		_receipt		
		case	2	8
			_special cases	
			_case	
	tax law	4	8	
		_tax law		
		_potential tax law offences		
		_suspect tax law		
		_tax law counter		
	liabilities	4	7	
		_liabilities		
		_liable individual		
		_ignorance		
		_tax liabilities		
	concealment	3	8	
		_concealment		
		_concealment of turnovers		
		_concealment of business		
	vat	3	7	

11.15 **Appendix 14 – full refined categorisation – interviews' coding with NVivo 10.**

Name		Sources	References
Tax		0	0
voluntary tax compliance		9	38
tax management		9	27
tax infringements - highest attention		9	30
tax fraud		3	12
tax evasion		9	48
tax concealment		5	18
obligations and liabilities		7	19
evasion motives		9	43
Risk management		0	0
Approach		0	0
risk prioritisation		9	34
procedure		5	17
operational level		4	12
macro level		4	11
holistic approach		2	3
difficulties		9	20
assessment		9	45
approach		7	26
Analysis		5	26
VI tax infingements		9	35
strategy		9	20
Selection		0	0
selection criteria		9	42
selection		9	66
passive selection		3	3
active selection		9	32
risky individuals		5	13
Management		0	0
risks		8	23
risk subjects		9	19
goal		9	24
diagnostics		9	29
deregistration		2	2
combat		9	27
information		2	9
cost-benefit - measures		9	73
Communication and campaigns		0	0
letters		5	11
compaigns		9	48
brochures		6	18
Automation		1	1
new information system		9	50
automation and subjectivism		9	56
automated selection		9	58
automated case distribution		9	51
Political situation		1	2
NRA		9	27
seniour officials		4	9
regional and territorial directorates		1	3
NRA services		5	12
institutional cooperation		9	23
control department		6	23
Legislation		9	27
prosecution		9	24
penalties		9	21
Institutions		0	0
police		4	10
cooperation		9	23
Control		0	0
treatment		6	18
targeted groups		9	36
inspections		9	24
control measures		9	51
control actions		9	47
cash registers		4	10
audit		4	27
Business		0	0
small business		4	12
sales monitoring		3	7
restaurants		3	3
construction sector		3	13
Behaviour		0	0
behaviour change		6	17
attitude		3	3

11.16 **Appendix 15 – graphical categorisation – interviews' coding with NVivo 10.**



11.17 **Appendix 16 – ease of doing business.**

This table represents comparative data about ease of doing business, paying taxes and enforcing contracts. It is exclusively based on data available from World Bank (World Bank et al., 2015b).

Additional notes:

1. All Doing Business 2014 rankings have been recalculated to reflect changes to the methodology and revisions of data due to new information. For further details on changes, please refer to the data notes page (hyperlink embedded).
2. Please note that the data above may not reflect the full numeric precision of the data point, as the numbers are rounded. For the dataset containing full numeric precision, please download reform simulator file (hyperlink embedded).

			Starting a Business						Getting Credit						Paying Taxes								Enforcing Contracts				
Year	Ease of Doing Business Rank	Overall DTF	Rank	DTF	Procedures (number)	Time (days)	Cost (% of income per capita)	Paid-in min. capital (% of income per capita)	Rank	DTF	Strength of legal rights index (0-12)	Depth of credit information index (0-8)	Credit registry coverage (% of adults)	Credit bureau coverage (% of adults)	Rank	DTF	Payments (number per year)	Time (hours per year)	Profit tax (%)	Labor tax and contributions (%)	Other taxes (%)	Total tax rate (% profit)	Rank	DTF	Time (days)	Cost (% of claim)	Procedures (number)
DB2004	..	..	..	70.66	11	32	10.4	86.7	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	60.23	564	23.8	39
DB2005	..	..	..	70.99	11	32	10.3	81.6	..	56.25	9	0	1.3	0	..	..	..	..	..	..	..	..	..	60.23	564	23.8	39
DB2006	..	..	..	71.62	11	32	9.6	73	..	81.25	9	4	13.6	0	..	47.23	29	616	..	..	..	45.2	..	60.23	564	23.8	39
DB2007	..	..	..	75.34	9	32	7.9	63.9	..	81.25	9	4	20.7	3	..	49.01	29	616	..	..	..	41.6	..	60.23	564	23.8	39
DB2008	..	..	..	76	9	32	6.4	56.3	..	81.25	9	4	25.4	3	..	59.56	15	616	..	..	..	35.8	..	60.23	564	23.8	39
DB2009	..	..	..	80.17	4	49	2	47.8	..	93.75	9	6	30.7	5	..	60.48	15	616	..	..	..	33.9	..	61.27	564	23.8	38
DB2010	..	68.45	..	89.68	4	18	1.7	20.7	..	93.75	9	6	34.8	6.2	..	62.1	15	616	..	..	..	30.4	..	61.27	564	23.8	38
DB2011	..	68.58	..	90.99	4	18	1.6	0	..	93.75	9	6	37	13.1	..	63.24	15	616	..	..	..	28	..	61.27	564	23.8	38
DB2012	..	68.26	..	91	4	18	1.5	0	..	81.25	9	4	52.8	0	..	69.6	15	500	..	..	..	27.2	..	61.27	564	23.8	38
DB2013	..	68.76	..	91.05	4	18	1.1	0	..	81.25	9	4	56.3	0	..	72.83	13	454	..	..	..	27.7	..	61.27	564	23.8	38
DB2014	36	71.72	41	91.06	4	18	1	0	19	70	9	5	61	0	84	73.18	13	454	5	20.2	1.7	27	74	61.27	564	23.8	38
DB2015	38	71.8	49	91.09	4	18	0.8	0	23	70	9	5	62.9	0	89	73.18	13	454	5	20.2	1.7	27	75	61.27	564	23.8	38

11.18 **Appendix 17 – GCI index and institutional corruption levels for Bulgaria.**

The following table is produced using data from (World Economic Forum, Schwab K, Forum WE, 2015)

The Global Competitiveness Index rankings – Bulgaria for the period 2011-2015				
	2011-2012 (142 countries)	**2012-2013 (144 countries)**	**2013-2014 (148 countries)**	**2014-2015 (144 countries)**
GCI Rank	74	62	57	54
Score (1 of 7)	4.2	4.3	4.3	4.37

This following table is produced using data from (Transparency International, 2013)

Institutional corruption levels in Bulgaria		
Institution	**Corrupt/Extremely corrupt**	**Corrupt/Extremely corrupt**
	2010-11	**2013**
Parliament/legislature	21.40%	71.00%
Political parties	10.80%	76.00%
Judiciary	28.70%	86.00%
Police	19.50%	65.00%
Public officials and civil cervants	17.40%	64.00%
Business	35.90%	63.00%

11.19 **Appendix 18 – business survey questionnaire.**

Explanatory notes:

1. The business survey was performed using Qualtrics Research Suite.

2. This is a translation from Bulgarian to English. Best possible effort has been made to a meaningful and correct translation however, it has to be noted that certain questions differ in their actual cultural meaning when asked to Bulgarian respondents.

3. The exact design and outlook of the survey cannot be precisely replicated in Microsoft Word, however the best possible effort is made to make it closer to the original.

4. The business survey structure (set of blocks) was not visible to the respondents. It is entirely dedicated to support the aims of this academic research only. For clarity and transparency it is presented in the box below.

5. The questionnaire employed sophisticated display, skip and flow logic, which will be indicated in blue boxes nearby the questions, where relevant. This is to visualise the survey logic – what happens when a respondent selects a particular response.

6. If there were questions' instructions available to the respondents, they are presented in green text nearby the question they were intended for.

Business survey – structure (not visible to the respondents)

Block 1 – Informed consent.
Block 2 – General business profile and initial informality indicators.
Block 3 – Institutional quality and fair use of public money. Procedural fairness and distributive justice.
Block 4 – Diagnostics of informality, motivations and perceptions.
Block 5 – Interdependence of informal businesses (informality chain of interdependence).
Block 6 – External business environment and cost of doing business. Taxation, regulations and licensing framework.
Block 7 – Unfair competition from informal businesses.
Block 8 – End of the survey – free text box.

11.19.1 Block 1 – Informed consent.

1. Do you agree to participate in this business survey?
   a. Yes, I agree to participate.
      Display first question from Block 2.

   b. No, I do not agree to participate.
      Display the End of Survey page.

11.19.2 Block 2 – General business profile and initial informality indicators.

2. What is your main business subject?
   a. Trading with goods.
   b. Trading with services.
   c. Trading with goods and services.
   d. Something else – please specify in the free text box below.

Display logic: if the answer to question 2 is "Trading with goods", then display question 3 below.

3. Do you need to keep stock and in what quantities?
   a. Yes, the nature of our business requires maintenance of substantial stock availability.
   b. Yes, but not substantial stock availability.
   c. No, the nature of our business does not require maintenance of stock supplies.
   d. Something else – please specify in the free text box below.

4. How many people are working for your business? Please use the sliders to indicate.
   a. Full-time. (0 to 50)
   b. Part-time. (0 to 50)

5. Does your business have an office or you run it from home?
   a. Yes, there is an office.
   b. No, there is no office – it is run from home.
   c. Something else – please specify in the free text box below.

Display logic: if the answer to question 5 is "Yes, there is an office", then display question 6 below.

6. Is the office owned by the business or is it rented?
   a. The office is owned by the business/the company.
   b. The office is rented.
   c. Something else – please specify in the free text box below.

7. Have you formally registered your business since it began operating or this has happened at a later stage?
   a. Yes, from the very beginning of business operations.
   b. No, this has happened at a later stage.
   c. Something else – please specify in the free text box below.

Display logic: if the answer to question 7 is "No, this has happened at a later stage", then display question 8 below.

8. When was the business registered – please type the year of registration?

9. Are there any outstanding debts to your business and approximately what their size is?
   a. Yes, there are, but in small/insignificant amounts.
   b. Yes, there are in medium/acceptable amounts.

c. Yes, there are in substantial amounts.
d. No.
e. Something else – please specify in the free text box below.

Display logic: if the answer to question 9 is "Yes, there are in medium/acceptable amounts" or "Yes, there are in substantial amounts.", then display question 10 below.

10. Do you undertake debt recovery activities to collect the outstanding amounts?
a. Yes.
b. No.
c. Something else – please specify in the free text box below.

11. Do you use any forms of business finance/credit? You may select multiple options.
a. Business loans from banks.
b. Credit cards.
c. Business loans from non-bank institutions/organisations.
d. Loans from your business partners (suppliers or clients).
e. Loans from friends and family.
f. No external forms of funding.

Display logic: if the answer to question 11 is a, b, c, d or e, then display question 12 below.

12. What is approximately the lowest interest rate, which you have managed to achieve for the finance type you have used? Please use the sliders. It is a percentage scale from 0% to 100%. Please write 0 against these finance options, which you have not used.
a. Business loans from banks.
b. Credit cards.
c. Business loans from non-bank institutions/organisations.
d. Loans from your business partners (suppliers or clients).
e. Loans from friends and family.

11.19.3 Block 3 – Institutional quality and fair use of public money. Procedural fairness and distributive justice.

13. What is your opinion about:
I fully agree = 1
I fully disagree = 5
a. Correct and fair expenditure of tax money.
b. Fair, professional attitude and treatment by NRA.
c. Fair, professional attitude and treatment by the state institutions.

11.19.4 Block 4 – Diagnostics of informality, motivations and perceptions.

14. Your opinion about:
I fully agree = 1
I fully disagree = 5
a. Doing business entirely according to the rules is too expensive.
b. Tax evasion is the answer from the business to the poor tax and administrative policies in Bulgaria.
c. In the current business environment, tax evasion is an acceptable phenomenon.

d. In the current business environment, using bribes is an acceptable phenomenon.

15. In your view, what percentage of the turnover do other businesses from your sector are pushed to pay informally "to get things done" – such as administrative services, tenders, licensing, regulative permissions and alike? It is a percentage scale from 0% to 100%.

16. In your opinion, do businesses from your sector underreport their actual turnovers with the purpose to pay less tax? You may select up to three options below.
    a. Yes.
    b. Yes, in order to survive.
    c. Yes, they do it in order to reduce their costs and increase their profits.
    d. Yes and I do not see anything wrong in this.
    e. Yes, because they are forced to do so and they do not have another choice.
    f. No.
    g. I do not know.
    h. Something else.

Display logic: if the answer to question 16 is a, b, c, d or e, then display question 17 below.

17. According to your judgement, with what percentage usually the declared turnovers are lesser in comparison with the actual ones? Please use the slider – it is a percentage scale from 0% to 100%.

11.19.5 Block 5 – Interdependence of informal businesses (informality chain of interdependence).

18. Have you ever bought goods or services for your business without an invoice or from suppliers not formally registered?
    a. Yes.
    b. No.
    c. Something else – please specify in the free text box below.

Display logic: if the answer to question 18 is a, or c, then display question 19 below.

19. For what part of the goods and services, which you need as a business, do you usually receive an invoice? Please use the slider – it is a percentage scale from 0% to 100%.

Display logic: if the answer to question 19 is less or equal to 95%, then display question 20 below.

20. Would you prefer all of your suppliers to provide you with invoices?
    a. Yes.
    b. No.
    c. It does not matter.
    d. Something else – please specify in the free text box below.

11.19.6 Block 6 – External business environment and cost of doing business. Taxation, regulations and licensing framework.

21. How would you assess NRA's work (procedures) and in particular its administrative capacity in regards with your business?
    a. There is stimulating effect on the business.
    b. It does not hamper the business.
    c. There is small hampering effect.
    d. There is medium hampering effect.
    e. There is significant hampering effect.
    f. There is enormous hampering effect.
    g. Something else – please specify in the free text box below.

22. What is your perception about the way the current tax rates and social security contributions affect your business?
    a. They constitute small impediment to the business.
    b. They constitute medium impediment to the business.
    c. They constitute significant impediment to the business.
    d. They constitute enormous impediment to the business.
    e. They do not constitute impediment to the business.
    f. The tax rates have stimulating effect on the business.
    g. Something else – please specify in the free text box below.

23. What would be the acceptable level of tax on business profit? Please use the slider – it is a percentage scale from 0% to 100%.

24. Which three factors create the most difficulties for your business? You may select up to three options in descending order and drag them into the box on the right – the most obstructive ones on the top.
    a. National Revenue Agency (NRA).
    b. The tax rate.
    c. Unfair competition from the informal sector (the grey economy).
    d. The corruption in the state administration.
    e. Low efficiency of the state administration.
    f. The laws and regulations
    g. Licenses and regulative regimes.
    h. Difficult and expensive access to financing.
    i. The lack of adequately educated/skilled labour force.
    j. Something else – please specify in the free text box below.

25. Approximately how many days per month do you spend to cope with administrative requirements imposed by various state regulations (such as business permissions, tax return, communication with NRA, NSSI and others)? Please use the slider – the scale is in days.

26. Approximately what is your cost of doing business per day (all business expenses including cost of salaries and others). Please use the slider – the scale is from 0 to 3000BGN per day.

27. Have you ever been inspected by NRA tax officials?
    a. Yes.
    b. No
    c. Something else – please specify in the free text box below.

Display logic: if the answer to question 27 is a, then display question 28 below.

28. How much time approximately did you spend in this connection? Please use the slider – the scale is in days.

Display logic: if the answer to question 27 is a or c, then display question 29 below.

29. During these inspection(s) visits have you been expected or encouraged to provide any form of informal benefit to the official(s)?
    a. Yes.
    b. No.
    c. I prefer to leave this question unanswered.
    d. Something else – please specify in the free text box below.

11.19.7 Block 7 – Unfair competition from informal businesses.

30. Do you experience unfair competition from unregistered or informal businesses (grey economy businesses)?
    a. Yes.
    b. No.
    c. Something else – please specify in the free text box below.

Display logic: if the answer to question 30 is a, then display question 31 below.

31. What level of difficulties does your business experience due to unfair competition from the informal sector?
    a. There are no any difficulties being experienced.
    b. We experience only difficulties, which we manage to cope with.
    c. We experience significant difficulties.
    d. Something else – please specify in the free text box below.

11.19.8 Block 8 – End of the survey – free text submission.

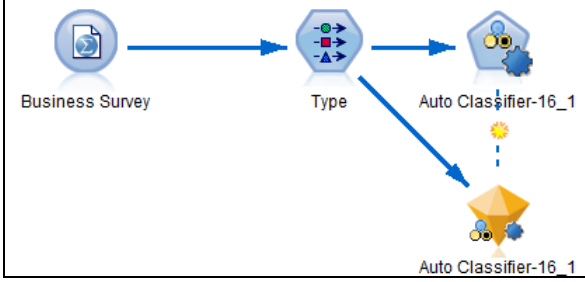
32. Thank you for taking part in this survey. If you want to share something, which you did not manage to do in the questionnaire, please do so in the free text box below.

11.20 **Appendix 19 – modelling of variable Q16_1.**

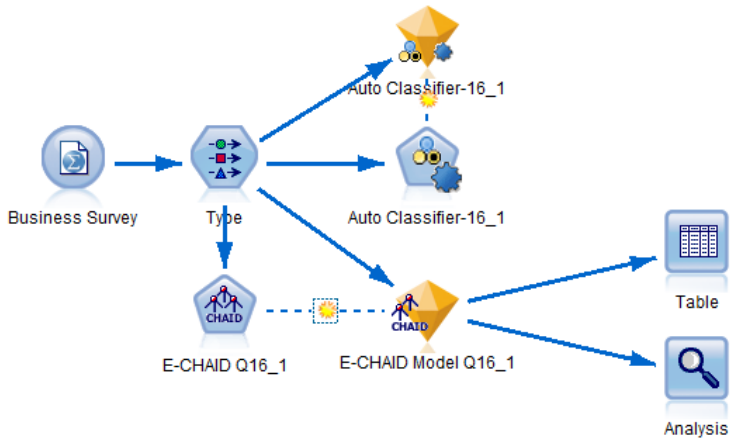
Click here to go to variable Q16_1 in the main body of the thesis.

11.20.1 Auto Classifier – training summary for Q16_1.

Auto Classifier – training summary for Q16_1	
Target	Q16_1
Inputs	All variables, excluding free text fields and Q16.
Planned models	432
Completed models	424
Retained models	100
Models discarded from final results	324
Models which failed to build or score	8



11.20.2 E-CHAID Model Q16_1 – modelling stream.



11.20.3 Performance evaluation and coincidence matrix for E-CHAID Model Q16_1.

1 = Yes, 2 = User not selected, 999 = User non response

Results for output field Q16_1

Comparing $R-Q16_1 with Q16_1

Correct	333	80.63%
Wrong	80	19.37%
Total	413	

Coincidence Matrix for $R-Q16_1 (rows show actuals)

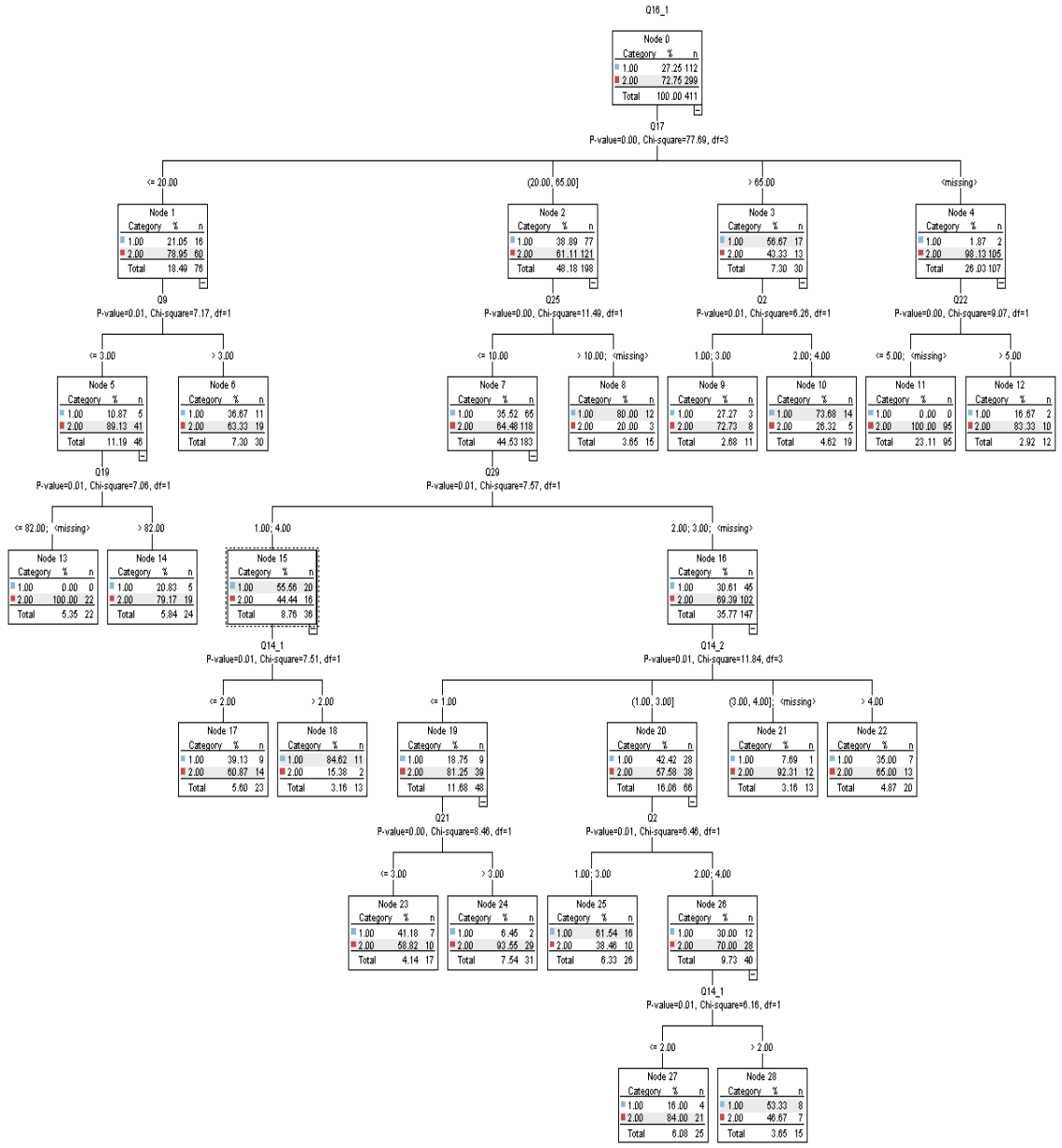
	1.000000	2.000000
1.000000	61	51
2.000000	27	272
999.000000	0	2

Performance Evaluation

1.000000	0.938
2.000000	0.145

11.20.4 E-CHAID tree model Q16_1 – visual representation.

As it obvious from the graphic image of this decision tree, it is too complex to be visualised properly on A4 page. The full detail (high-resolution) image is available as part of this PhD thesis data store to examiners and supervisors or otherwise upon an authorised request. Q17, Q9, Q25 and such alike, shown beneath nodes are related variables indicating decision tree splits by related significance.



11.20.5 Symbolic rule expression derived from E-CHAID Model Q16_1.

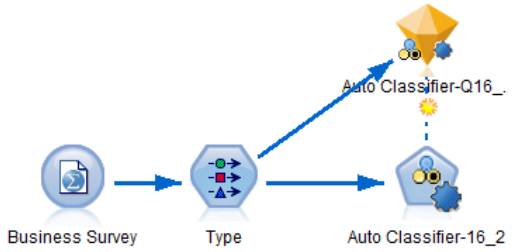
Rule 1 for “Yes” at Q16_1 (84.6% confidence)
if Q17 > 20
and Q17 <= 65
and Q25 <= 10
and Q29 = 1 or Q29 = 4
and Q14_1 > 2
then “Yes” at 16_1

11.21 **Appendix 20 – modelling of variable Q16_2.**

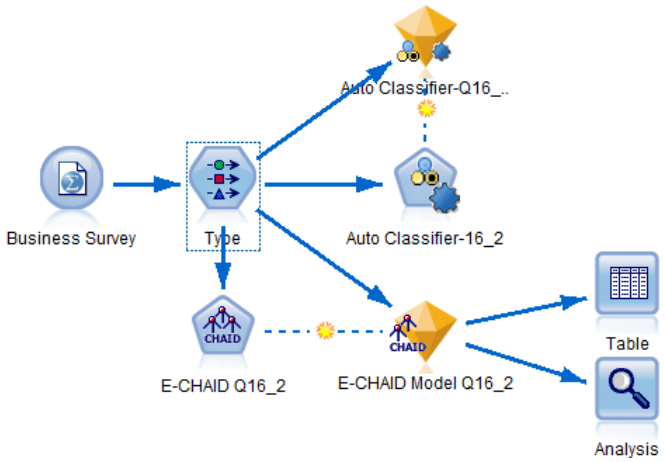
Click here to go to variable Q16_2 in the main body of the thesis.

11.21.1 Auto Classifier – training summary for Q16_2.

Auto Classifier – training summary for Q16_2	
Target	Q16_2
Inputs	All variables, excluding free text fields and Q16.
Planned models	570
Completed models	440
Retained models	100
Models discarded from final results	340
Models which failed to build or score	130



11.21.2 E-CHAID Model Q16_2 – modelling stream.



11.21.3 Performance evaluation and coincidence matrix for E-CHAID Model Q16_2.

1 = Yes, 2 = User not selected, 999 = User non response

Results for output field Q16_2

Comparing $R-Q16_2 with Q16_2

Correct	333	80.63%
Wrong	80	19.37%
Total	413	

Coincidence Matrix for $R-Q16_2 (rows show actuals)

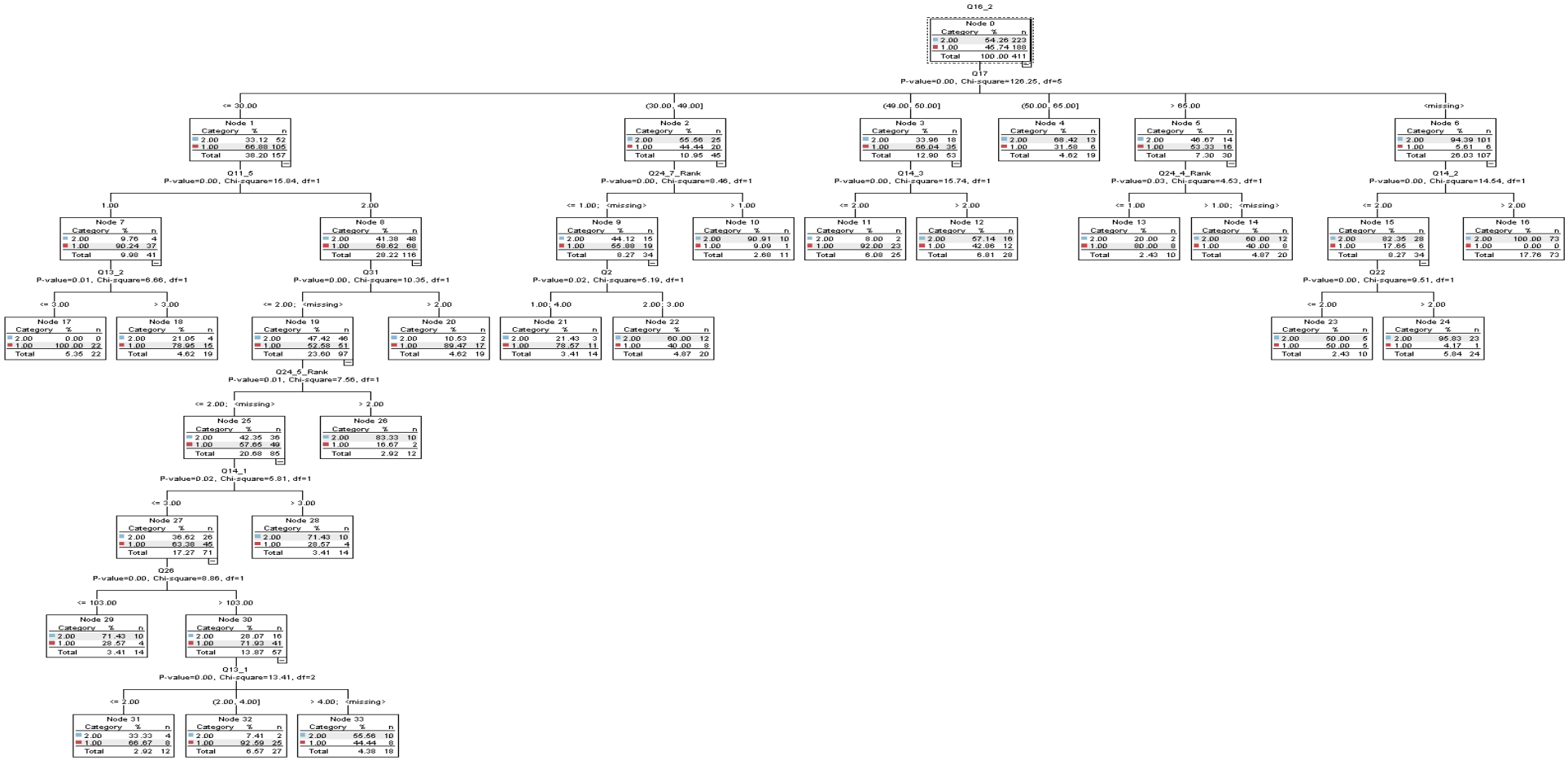
	1.000000	2.000000
1.000000	129	59
2.000000	19	204
999.000000	0	2

Performance Evaluation

1.000000	0.65
2.000000	0.355

11.21.4 E-CHAID tree model Q16_2 – visual representation.

As it obvious from the graphic image of this decision tree, it is too complex to be visualised properly on A4 page. The full detail (high-resolution) image is available as part of this PhD thesis data store to examiners and supervisors or otherwise upon an authorised request. Q17, Q11_5, Q24_7 and such alike, shown beneath nodes are related variables indicating decision tree splits by related significance.



11.21.5 Symbolic rule expression derived from E-CHAID Model Q16_2.

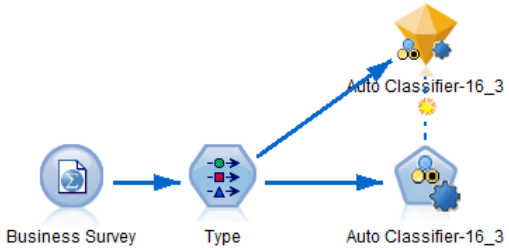
```
Rule set for 1.0 “Yes” at Q16_2 (92.6% confidence)
        if Q17 <= 30
        and Q11_5 = 2
        and Q31 <= 2 or Q31 IS MISSING
        and Q24_5_Rank <= 2 or Q24_5_Rank IS MISSING
        and Q14_1 <= 3
        and Q26 > 103
        and Q13_1 > 2
        and Q13_1 <= 4
        then 1.00 at Q16_2 (“Yes, in order to survive” on Q16)
```

11.22 **Appendix 21 – modelling of variable Q16_3.**

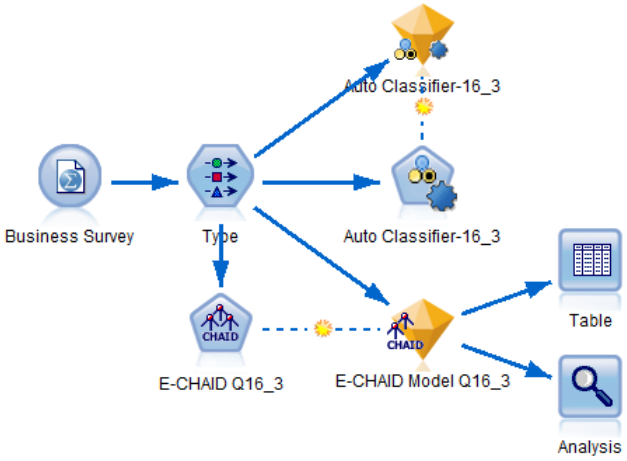
Click here to go to variable Q16_3 in the main body of the thesis.

11.22.1 Auto Classifier – training summary for Q16_3.

Auto Classifier – training summary for Q16_3	
Target	Q16_3
Inputs	All variables, excluding free text fields and Q16.
Planned models	570
Completed models	432
Retained models	100
Models discarded from final results	332
Models which failed to build or score	138



11.22.2 E-CHAID Model Q16_3 – modelling stream.



11.22.3 Performance evaluation and coincidence matrix for E-CHAID Model Q16_3.

1 = Yes, 2 = User not selected, 999 = User non response

Results for output field Q16_3

Comparing $R-Q16_3 with Q16_3

Correct	329	79.66%
Wrong	84	20.34%
Total	413	

Coincidence Matrix for $R-Q16_3 (rows show actuals)

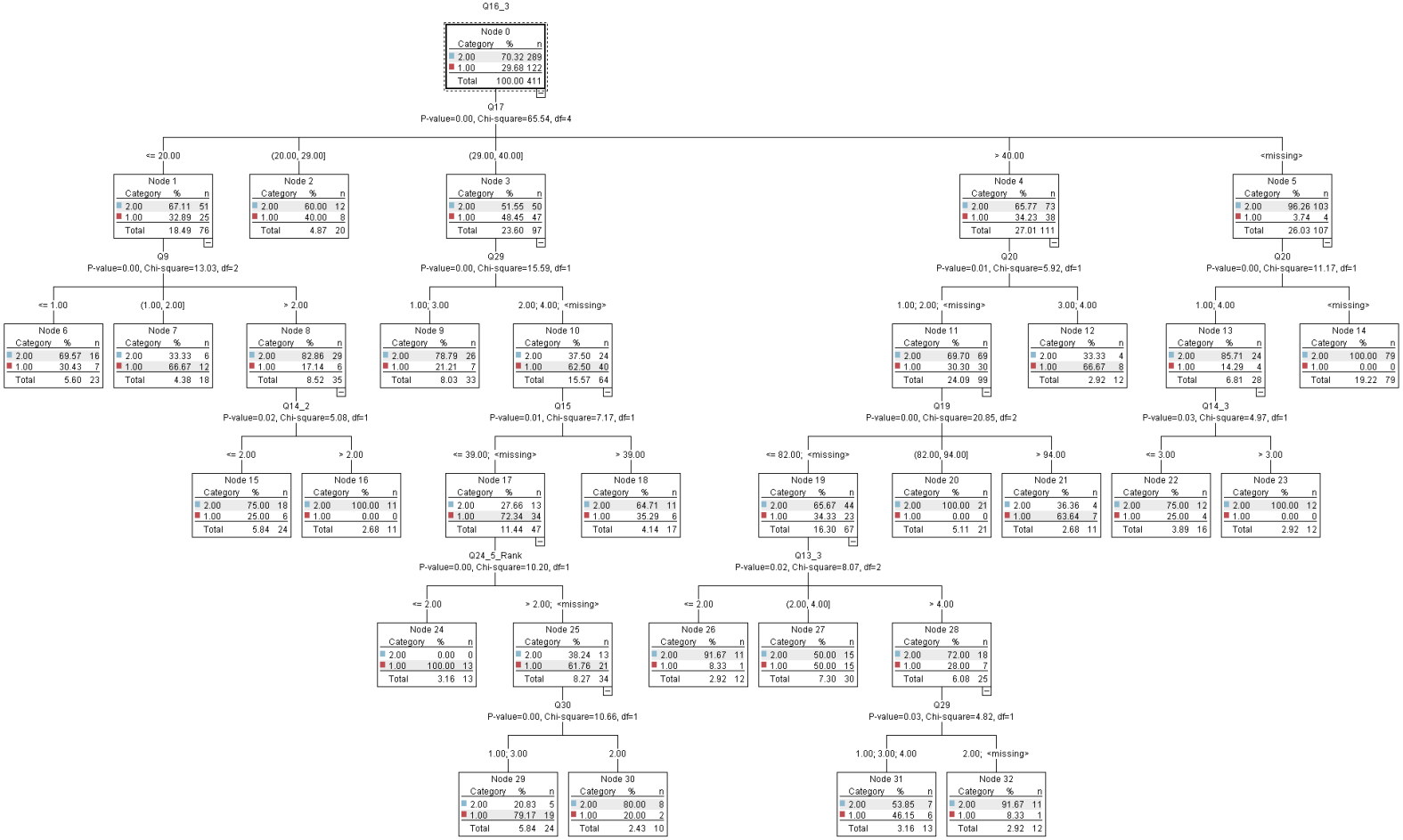
	1.000000	2.000000
1.000000	59	63
2.000000	19	270
999.000000	0	2

Performance Evaluation

1.000000	0.94
2.000000	0.141

11.22.4 E-CHAID tree model Q16_3 – visual representation.

A high-resolution image is available as part of this PhD thesis data store to examiners and supervisors or otherwise upon an authorised request. Q17, Q9, Q29 and such alike, shown beneath nodes are related variables indicating decision tree splits by related significance.



11.22.5 Symbolic rule expression derived from E-CHAID Model Q16_3.

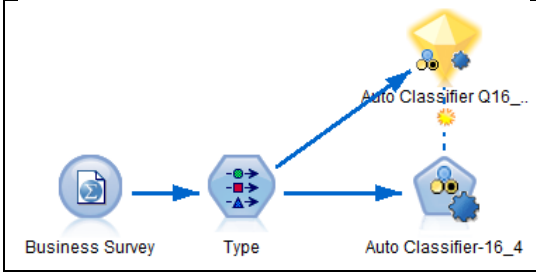
Rule set for 1.0 (Yes) at Q16_3 (79.2% confidence)
if Q17 > 29
and Q17 <= 40
and Q29 = 2 or Q29 = 4 or Q29 IS MISSING
and Q15 <= 39 or Q15 IS MISSING
and Q24_5_Rank > 2 or Q24_5_Rank IS MISSING
and Q30 = 1 or Q30 = 3
then 1.00 (Yes) at Q16_3 (“Yes, they do it in order to reduce their costs and increase their profits” on Q16)

11.23 **Appendix 22 – modelling of variable Q16_4.**

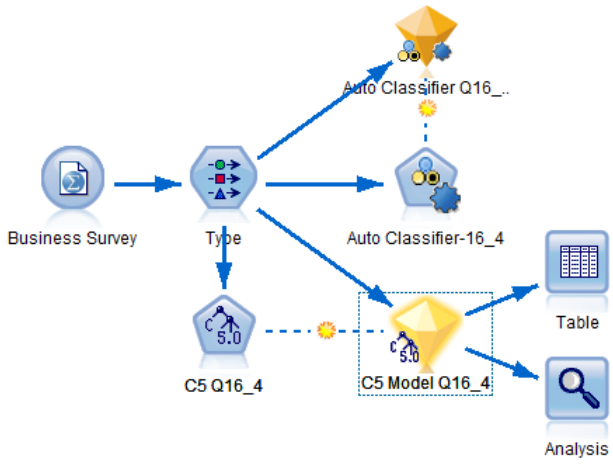
Click here to go to variable Q16_4 in the main body of the thesis.

11.23.1 Auto Classifier – training summary for Q16_4.

Auto Classifier – training summary for Q16_4	
Target	Q16_4
Inputs	All variables, excluding free text fields and Q16.
Planned models	570
Completed models	396
Retained models	100
Models discarded from final results	296
Models which failed to build or score	174



11.23.2 Q16_4 C5 – modelling stream.



11.23.3 Performance evaluation and coincidence matrix for Q16_4 C5 Model.

1 = Yes, 2 = User not selected, 999 = User non response

Results for output field Q16_4

Comparing $C-Q16_4 with Q16_4

Correct	409	99.03%
Wrong	4	0.97%
Total	413	

Coincidence Matrix for $C-Q16_4 (rows show actuals)

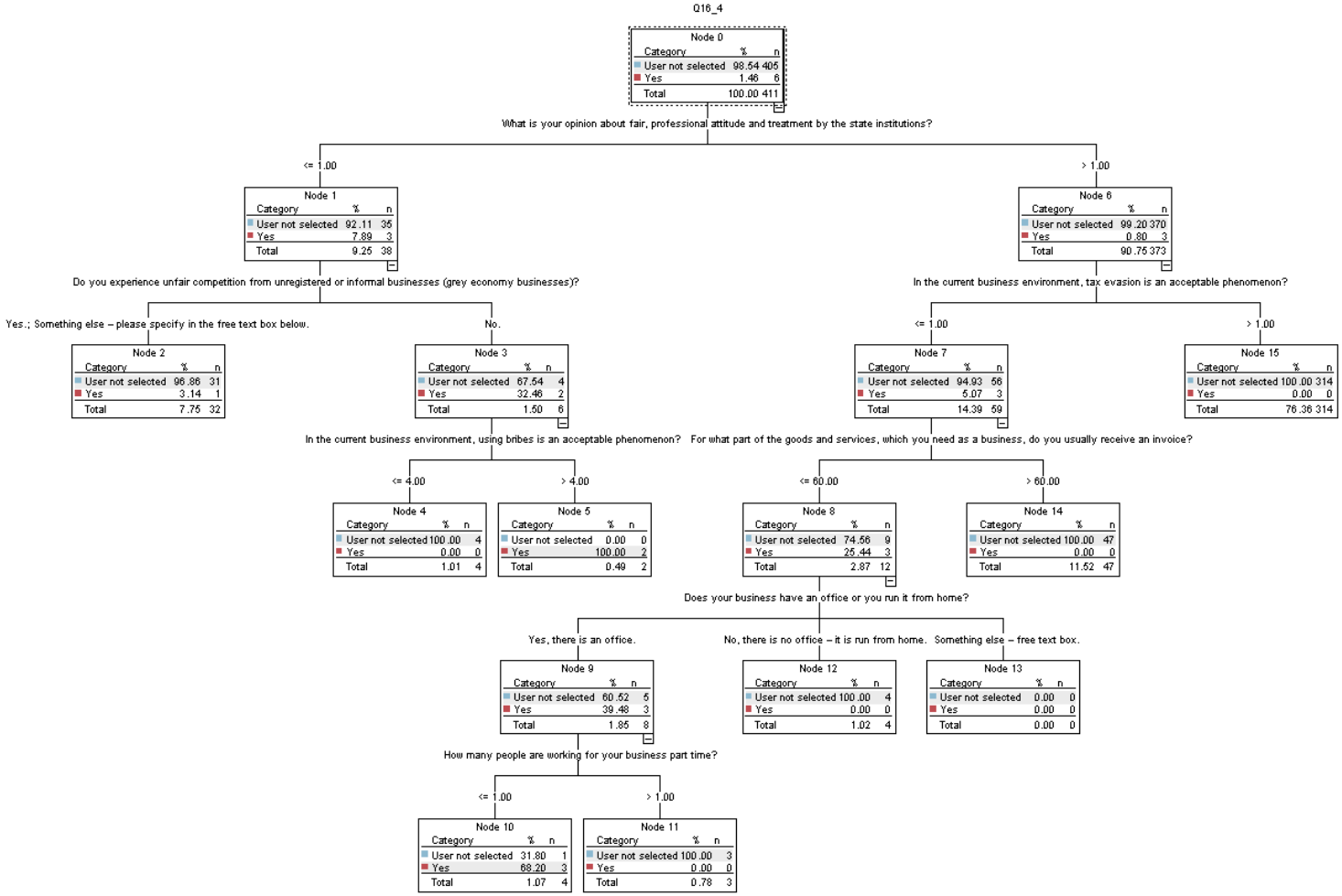
	1.000000	2.000000
1.000000	5	1
2.000000	1	404
999.000000	0	2

Performance Evaluation

1.000000	4.049
2.000000	0.012

11.23.4 Q16_4 C5 tree model – visual representation.

A high-resolution image is available as part of this PhD thesis data store to examiners and supervisors or otherwise upon an authorised request.



11.23.5 Symbolic rule expression derived from Q16_4 C5 model.

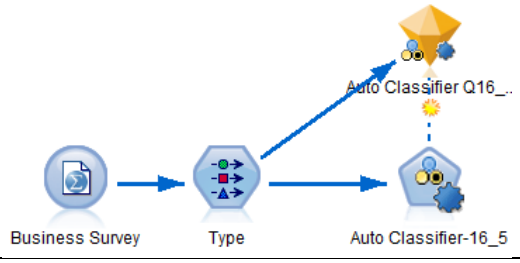
Rule set for 1.0 (Yes) at Q16_4 (91.3% confidence)
if Q4_1 > 1
and Q4_1 <= 4
and Q11_4 = 2.00
and Q14_4 <= 3
and Q19 <= 81
then 1.00 (Yes) at Q16_4 (“Yes and I do not see anything wrong in this” on Q16)

11.24 **Appendix 23 – modelling of variable Q16_5.**

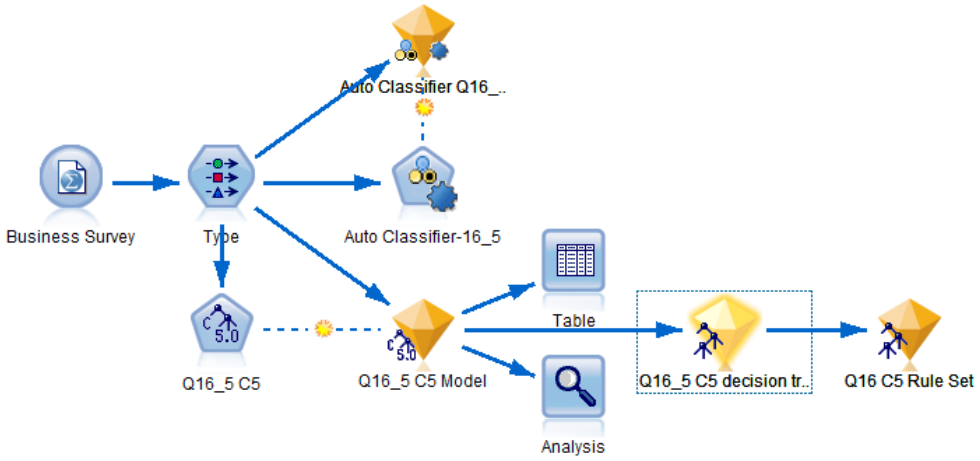
Click here to go to variable Q16_5 in the main body of the thesis.

11.24.1 Auto Classifier – training summary for Q16_5.

Auto Classifier – training summary for Q16_5	
Target	Q16_5
Inputs	All variables, excluding free text fields and Q16.
Planned models	570
Completed models	440
Retained models	100
Models discarded from final results	340
Models which failed to build or score	130



11.24.2 Q16_5 C5 – modelling stream.



11.24.3 Performance evaluation and coincidence matrix for Q16_5 C5 Model.

1 = Yes, 2 = User not selected, 999 = User non response

Results for output field Q16_5

Comparing $C-Q16_5 with Q16_5

Correct	395	95.64%
Wrong	18	4.36%
Total	413	

Coincidence Matrix for $C-Q16_5 (rows show actuals)

	1.000000	2.000000	$null$
1.000000	85	14	1
2.000000	0	310	1
999.000000	0	2	0

Performance Evaluation

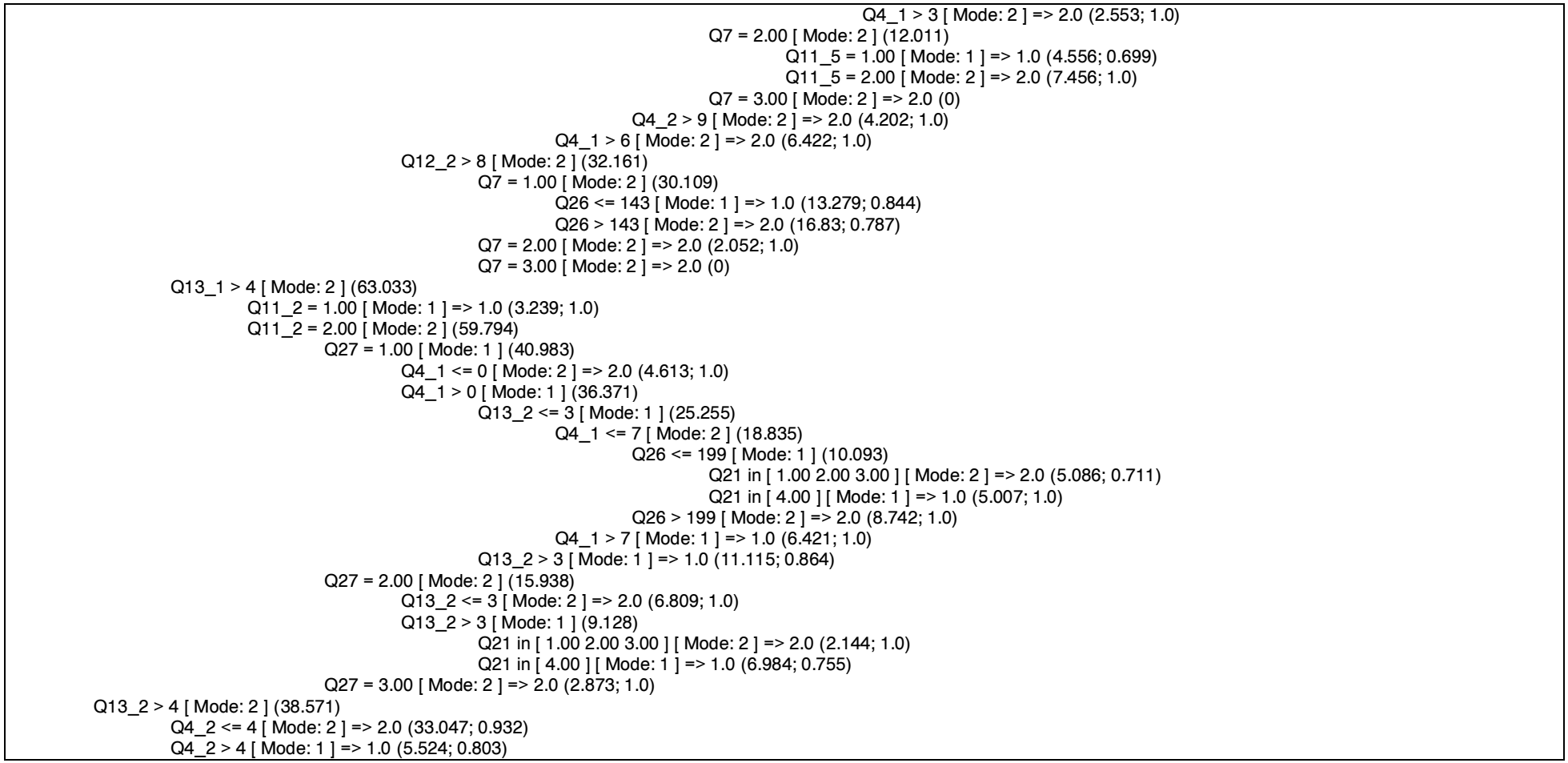
1.000000	1.418
2.000000	0.233

11.24.4 Q16_5 C5 tree model – visual representation.

The tree depth is 15 levels and therefore only the symbolic expression is presented below. A high-resolution tree image is available as part of this PhD thesis data store to examiners and supervisors or otherwise upon an authorised request.

```
Q21 in [ 5.00 6.00 7.00 ] [ Mode: 1 ] (122.014)
        Q21 in [ 5.00 6.00 ] [ Mode: 1 ] (115.554)
                Q7 = 1.00 [ Mode: 2 ] (103.835)
                        Q11_2 = 1.00 [ Mode: 1 ] => 1.0 (5.87; 0.861)
                        Q11_2 = 2.00 [ Mode: 2 ] (97.965)
                                Q4_2 <= 3 [ Mode: 1 ] (92.211)
                                        Q13_2 <= 3 [ Mode: 2 ] (14.025)
                                                Q11_5 = 1.00 [ Mode: 2 ] (8.55)
                                                        Q26 <= 300 [ Mode: 2 ] => 2.0 (4.475; 1.0)
                                                        Q26 > 300 [ Mode: 1 ] => 1.0 (4.075; 0.726)
                                                Q11_5 = 2.00 [ Mode: 2 ] => 2.0 (5.475; 1.0)
                                        Q13_2 > 3 [ Mode: 1 ] (78.186)
                                                Q27 = 1.00 [ Mode: 1 ] (52.742)
                                                        Q11_3 = 1.00 [ Mode: 1 ] => 1.0 (7.85; 0.749)
                                                        Q11_3 = 2.00 [ Mode: 2 ] (44.892)
                                                                Q4_2 <= 1 [ Mode: 1 ] (39.572)
                                                                        Q13_2 <= 4 [ Mode: 2 ] (13.686)
                                                                                Q21 in [ 5.00 ] [ Mode: 1 ] => 1.0 (5.629; 0.71)
                                                                                Q21 in [ 6.00 ] [ Mode: 2 ] (8.057)
                                                                                        Q4_1 <= 6 [ Mode: 2 ] => 2.0 (4.805; 1.0)
                                                                                        Q4_1 > 6 [ Mode: 1 ] => 1.0 (3.252; 0.586)
                                                                        Q13_2 > 4 [ Mode: 1 ] (25.886)
                                                                                Q21 in [ 5.00 ] [ Mode: 2 ] (10.326)
                                                                                        Q4_1 <= 3 [ Mode: 2 ] => 2.0 (6.325; 0.86)
                                                                                        Q4_1 > 3 [ Mode: 1 ] => 1.0 (4.001; 0.835)
                                                                                Q21 in [ 6.00 ] [ Mode: 1 ] => 1.0 (15.56; 0.756)
                                                                Q4_2 > 1 [ Mode: 2 ] => 2.0 (5.319; 1.0)
                                                Q27 = 2.00 [ Mode: 1 ] (23.223)
                                                        Q12_2 <= 10 [ Mode: 1 ] => 1.0 (9.062; 0.832)
                                                        Q12_2 > 10 [ Mode: 2 ] (14.16)
                                                                Q11_3 = 1.00 [ Mode: 2 ] => 2.0 (2.614; 1.0)
                                                                Q11_3 = 2.00 [ Mode: 1 ] (11.547)
                                                                        Q21 in [ 5.00 ] [ Mode: 1 ] => 1.0 (4.673; 1.0)
                                                                        Q21 in [ 6.00 ] [ Mode: 2 ] => 2.0 (6.874; 0.757)
                                                Q27 = 3.00 [ Mode: 1 ] => 1.0 (2.222; 1.0)
                                Q4_2 > 3 [ Mode: 2 ] => 2.0 (5.754; 1.0)
                Q7 = 2.00 [ Mode: 1 ] => 1.0 (8.733; 1.0)
                Q7 = 3.00 [ Mode: 1 ] => 1.0 (2.986; 1.0)
        Q21 in [ 7.00 ] [ Mode: 2 ] => 2.0 (6.461; 1.0)
```

```
Q21 in [ 1.00 2.00 3.00 4.00 ] [ Mode: 2 ] (278.987)
        Q13_2 <= 4 [ Mode: 2 ] (240.416)
                Q13_1 <= 4 [ Mode: 2 ] (177.383)
                        Q11_2 = 1.00 [ Mode: 2 ] => 2.0 (4.48; 1.0)
                        Q11_2 = 2.00 [ Mode: 2 ] (172.903)
                                Q11_3 = 1.00 [ Mode: 2 ] (25.863)
                                        Q7 = 1.00 [ Mode: 2 ] (23.033)
                                                Q27 = 1.00 [ Mode: 2 ] (16.469)
                                                        Q12_2 <= 15 [ Mode: 1 ] => 1.0 (5.824; 0.613)
                                                        Q12_2 > 15 [ Mode: 2 ] => 2.0 (10.644; 0.918)
                                                Q27 = 2.00 [ Mode: 2 ] => 2.0 (5.655; 0.869)
                                                Q27 = 3.00 [ Mode: 2 ] => 2.0 (0.909; 1.0)
                                        Q7 = 2.00 [ Mode: 2 ] => 2.0 (2.83; 1.0)
                                        Q7 = 3.00 [ Mode: 2 ] => 2.0 (0)
                                Q11_3 = 2.00 [ Mode: 2 ] (147.04)
                                        Q12_2 <= 8 [ Mode: 2 ] (114.879)
                                                Q13_2 <= 2 [ Mode: 2 ] (45.478)
                                                        Q26 <= 807 [ Mode: 2 ] => 2.0 (28.946; 0.967)
                                                        Q26 > 807 [ Mode: 2 ] (16.532)
                                                                Q21 in [ 1.00 2.00 3.00 ] [ Mode: 2 ] (12.177)
                                                                        Q26 <= 1,158 [ Mode: 1 ] => 1.0 (3.398; 0.885)
                                                                        Q26 > 1,158 [ Mode: 2 ] => 2.0 (8.779; 0.904)
                                                                Q21 in [ 4.00 ] [ Mode: 1 ] => 1.0 (4.355; 1.0)
                                                Q13_2 > 2 [ Mode: 2 ] (69.401)
                                                        Q4_1 <= 6 [ Mode: 2 ] (62.979)
                                                                Q4_2 <= 9 [ Mode: 2 ] (58.777)
                                                                        Q7 = 1.00 [ Mode: 1 ] (46.765)
                                                                                Q13_2 <= 3 [ Mode: 1 ] (20.784)
                                                                                        Q4_1 <= 3 [ Mode: 2 ] (14.297)
                                                                                                Q13_1 <= 2 [ Mode: 1 ] => 1.0 (2.935; 0.962)
                                                                                                Q13_1 > 2 [ Mode: 2 ] (11.362)
                                                                                                        Q11_5 = 1.00 [ Mode: 1 ] => 1.0 (3.91; 0.679)
                                                                                                        Q11_5 = 2.00 [ Mode: 2 ] => 2.0 (7.452; 1.0)
                                                                                        Q4_1 > 3 [ Mode: 1 ] => 1.0 (6.487; 0.905)
                                                                                Q13_2 > 3 [ Mode: 1 ] (25.981)
                                                                                        Q4_1 <= 3 [ Mode: 1 ] (23.429)
                                                                                                Q26 <= 227 [ Mode: 2 ] (17.827)
                                                                                                        Q21 in [ 1.00 2.00 3.00 ] [ Mode: 2 ] => 2.0 (7.058; 0.762)
                                                                                                        Q21 in [ 4.00 ] [ Mode: 1 ] (10.77)
                                                                                                                Q4_1 <= 2 [ Mode: 2 ] => 2.0 (7.553; 0.632)
                                                                                                                Q4_1 > 2 [ Mode: 1 ] => 1.0 (3.216; 1.0)
                                                                                                Q26 > 227 [ Mode: 1 ] => 1.0 (5.601; 1.0)
```


Q4_1 > 3 [Mode: 2] => 2.0 (2.553; 1.0)
Q7 = 2.00 [Mode: 2] (12.011)
Q11_5 = 1.00 [Mode: 1] => 1.0 (4.556; 0.699)
Q11_5 = 2.00 [Mode: 2] => 2.0 (7.456; 1.0)
Q7 = 3.00 [Mode: 2] => 2.0 (0)
Q4_2 > 9 [Mode: 2] => 2.0 (4.202; 1.0)
Q4_1 > 6 [Mode: 2] => 2.0 (6.422; 1.0)
Q12_2 > 8 [Mode: 2] (32.161)
Q7 = 1.00 [Mode: 2] (30.109)
Q26 <= 143 [Mode: 1] => 1.0 (13.279; 0.844)
Q26 > 143 [Mode: 2] => 2.0 (16.83; 0.787)
Q7 = 2.00 [Mode: 2] => 2.0 (2.052; 1.0)
Q7 = 3.00 [Mode: 2] => 2.0 (0)
Q13_1 > 4 [Mode: 2] (63.033)
Q11_2 = 1.00 [Mode: 1] => 1.0 (3.239; 1.0)
Q11_2 = 2.00 [Mode: 2] (59.794)
Q27 = 1.00 [Mode: 1] (40.983)
Q4_1 <= 0 [Mode: 2] => 2.0 (4.613; 1.0)
Q4_1 > 0 [Mode: 1] (36.371)
Q13_2 <= 3 [Mode: 1] (25.255)
Q4_1 <= 7 [Mode: 2] (18.835)
Q26 <= 199 [Mode: 1] (10.093)
Q21 in [1.00 2.00 3.00] [Mode: 2] => 2.0 (5.086; 0.711)
Q21 in [4.00] [Mode: 1] => 1.0 (5.007; 1.0)
Q26 > 199 [Mode: 2] => 2.0 (8.742; 1.0)
Q4_1 > 7 [Mode: 1] => 1.0 (6.421; 1.0)
Q13_2 > 3 [Mode: 1] => 1.0 (11.115; 0.864)
Q27 = 2.00 [Mode: 2] (15.938)
Q13_2 <= 3 [Mode: 2] => 2.0 (6.809; 1.0)
Q13_2 > 3 [Mode: 1] (9.128)
Q21 in [1.00 2.00 3.00] [Mode: 2] => 2.0 (2.144; 1.0)
Q21 in [4.00] [Mode: 1] => 1.0 (6.984; 0.755)
Q27 = 3.00 [Mode: 2] => 2.0 (2.873; 1.0)
Q13_2 > 4 [Mode: 2] (38.571)
Q4_2 <= 4 [Mode: 2] => 2.0 (33.047; 0.932)
Q4_2 > 4 [Mode: 1] => 1.0 (5.524; 0.803)

11.24.5 Symbolic rule expression derived from Q16_5 C5 rule set model.

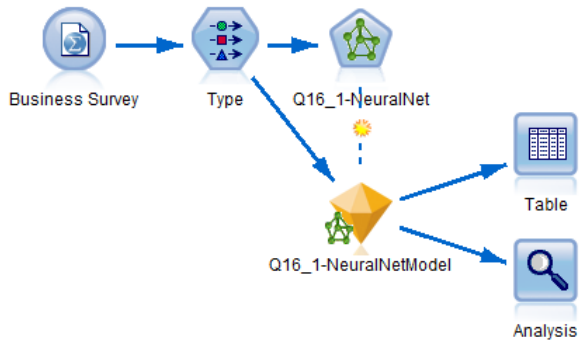
Rule set for 1.0 (Yes) at Q16_5 (86.4% confidence)
if Q13_2 <= 4
and Q13_1 > 4
and Q11_2 = 2.00
and Q27 = 1.00
and Q4_1 > 0
and Q13_2 > 3
and Q21 in [3.00 2.00 4.00]
then 1.00 (Yes) at Q16_5 (“Yes, because they are forced to do so and they do not have another choice”)

11.25 **Appendix 24 – decision trees' input variables for Q16_1 to Q16_5.**

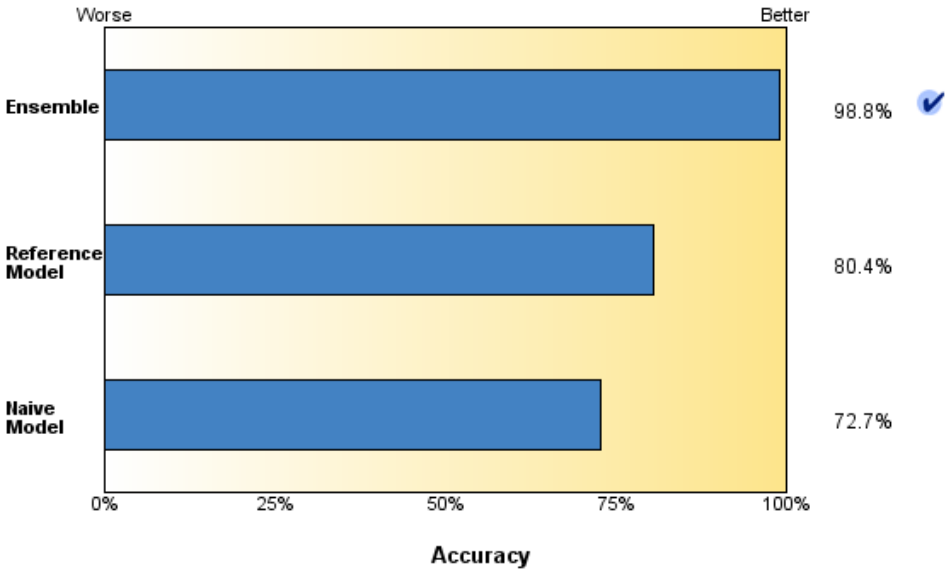
Q16_1	Q16_2	Q16_3	Q16_4	Q16_5
Q2 Q9 Q14_1 Q14_2 Q17 Q19 Q21 Q22 Q25 Q29	Q2 Q11_5 Q13_1 Q13_2 Q14_1 Q14_2 Q14_3 Q17 Q22 Q24_4_Rank Q24_5_Rank Q24_7_Rank Q26 Q31	Q9 Q13_3 Q14_2 Q14_3 Q15 Q17 Q19 Q20 Q24_5_Rank Q29 Q30	Q4_2 Q5 Q13_3 Q14_4 Q14_3 Q19 Q30	Q4_1 Q4_2 Q7 Q11_2 Q11_3 Q11_5 Q12_2 Q13_1 Q13_2 Q21 Q26 Q27

11.26 **Appendix 25 – neural network for Q16_1.**

11.26.1 Q16_1 neural network modelling stream.



11.26.2 Q16_1 neural network model – accuracy summary.



11.26.3 Q16_1 neural network performance evaluation and coincidence matrix.

Results for output field Q16_1

Comparing $N-Q16_1 with Q16_1

Correct	406	98.31%
Wrong	7	1.69%
Total	413	

Coincidence Matrix for $N-Q16_1 (rows show actuals)

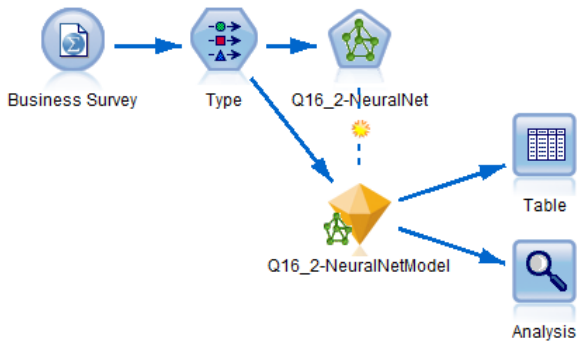
	1.000000	2.000000	$null$
1.000000	111	1	0
2.000000	1	295	3
999.000000	0	0	2

Performance Evaluation

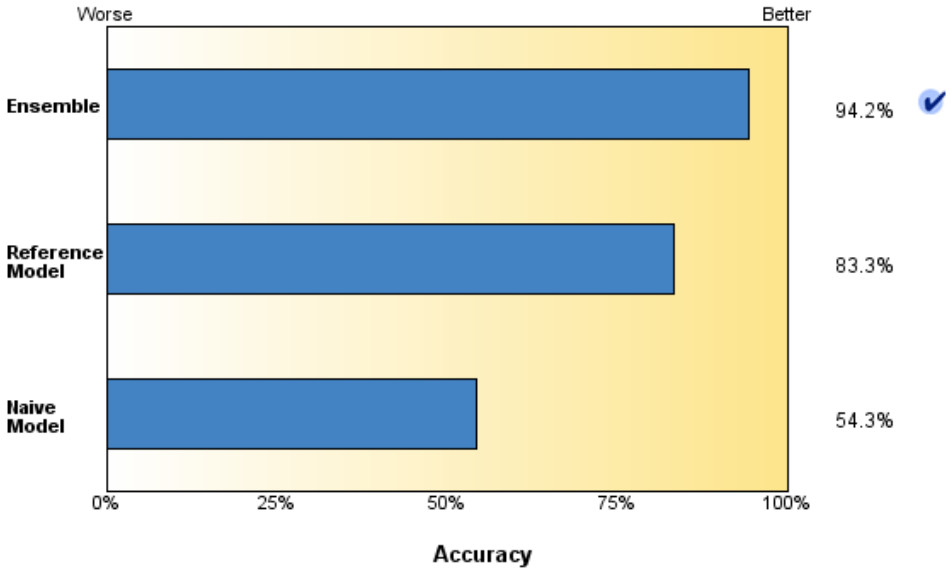
1.000000	1.296
2.000000	0.32

11.27 Appendix 26 – neural network for Q16_2.

11.27.1 Q16_2 neural network modelling stream.



11.27.2 Q16_2 neural network model – accuracy summary.



11.27.3 Q16_2 neural network performance evaluation and coincidence matrix.

Results for output field Q16_2

Comparing $N-Q16_2 with Q16_2

Correct	387	93.7%
Wrong	26	6.3%
Total	413	

Coincidence Matrix for $N-Q16_2 (rows show actuals)

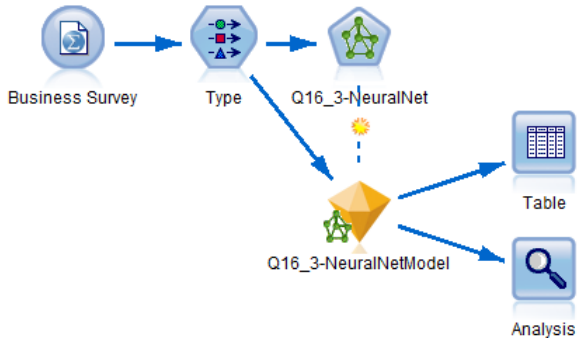
	1.000000	2.000000	$null$
1.000000	186	2	0
2.000000	19	201	3
999.000000	0	0	2

Performance Evaluation

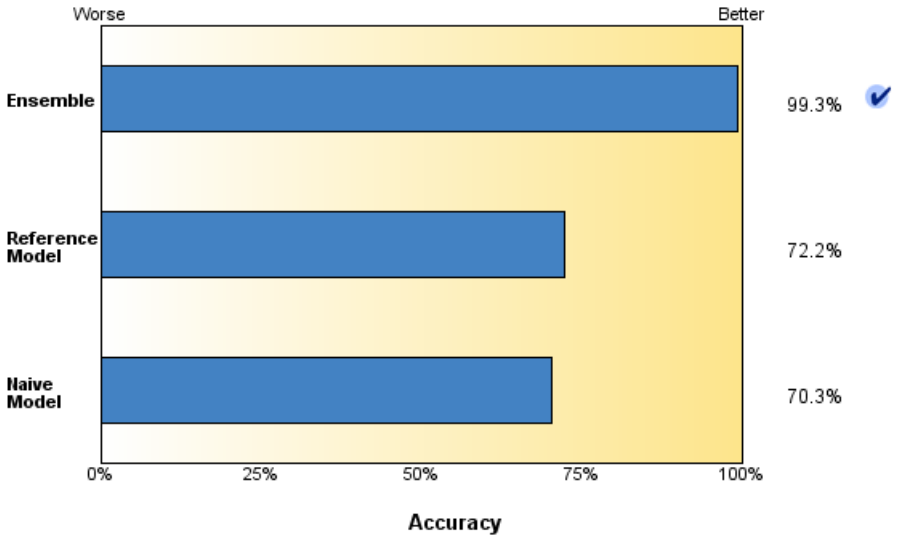
1.000000	0.69
2.000000	0.606

11.28 **Appendix 27 – neural network for Q16_3.**

11.28.1 Q16_3 neural network modelling stream.



11.28.2 Q16_3 neural network model – accuracy summary.



11.28.3 Q16_3 neural network performance evaluation and coincidence matrix.

Results for output field Q16_3

Comparing $N-Q16_3 with Q16_3

Correct	408	98.79%
Wrong	5	1.21%
Total	413	

Coincidence Matrix for $N-Q16_3 (rows show actuals)

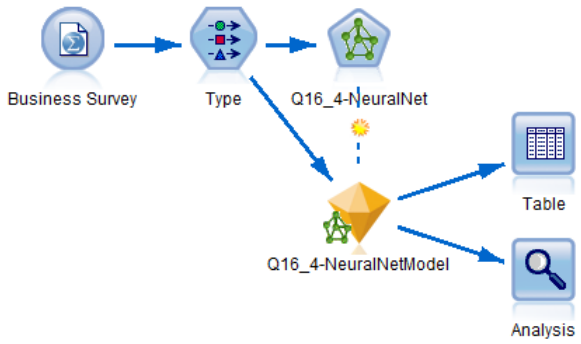
	1.000000	2.000000	$null$
1.000000	122	0	0
2.000000	0	286	3
999.000000	0	0	2

Performance Evaluation

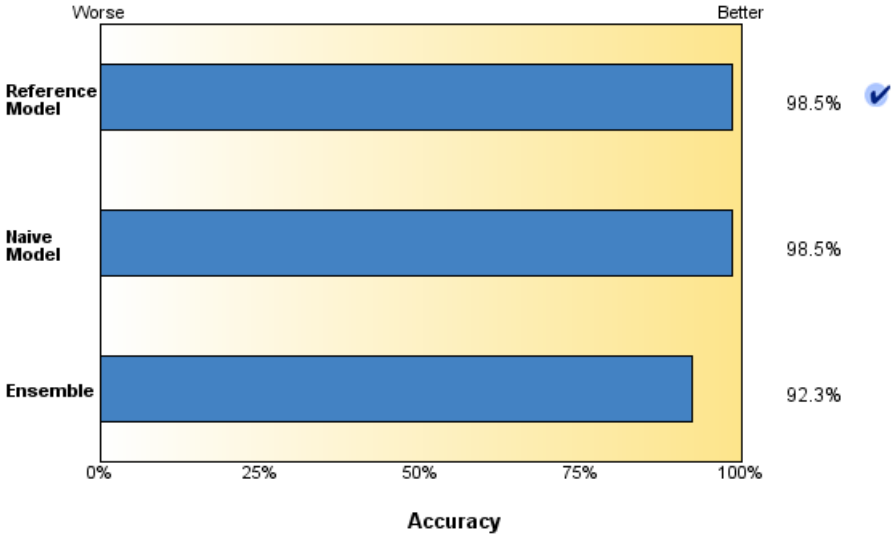
1.000000	1.219
2.000000	0.357

11.29 **Appendix 28 – neural network for Q16_4.**

11.29.1 Q16_4 neural network modelling stream.



11.29.2 Q16_4 neural network model – accuracy summary.



11.29.3 Q16_4 neural network performance evaluation and coincidence matrix.

Results for output field Q16_4

Comparing $N-Q16_4 with Q16_4

Correct	405	98.06%
Wrong	8	1.94%
Total	413	

Coincidence Matrix for $N-Q16_4 (rows show actuals)

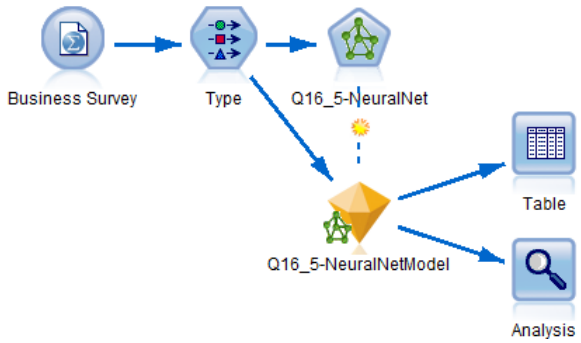
	1.000000	2.000000	$null$
1.000000	6	0	0
2.000000	3	399	3
999.000000	0	0	2

Performance Evaluation

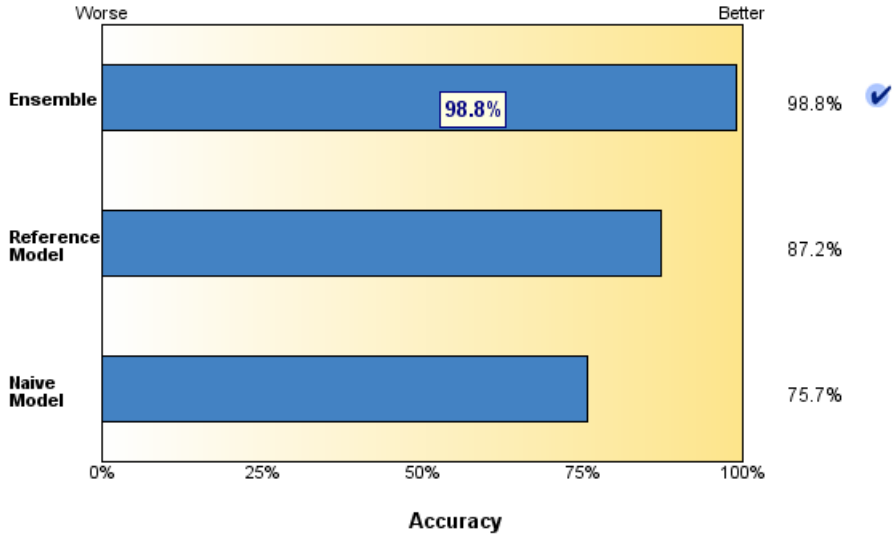
1.000000	3.826
2.000000	0.02

11.30 Appendix 29 – neural network for Q16_5.

11.30.1 Q16_5 neural network modelling stream.



11.30.2 Q16_5 neural network model – accuracy summary.



11.30.3 Q16_5 neural network performance evaluation and coincidence matrix.

Results for output field Q16_5

Comparing $N-Q16_5 with Q16_5

Correct	406	98.31%
Wrong	7	1.69%
Total	413	

Coincidence Matrix for $N-Q16_5 (rows show actuals)

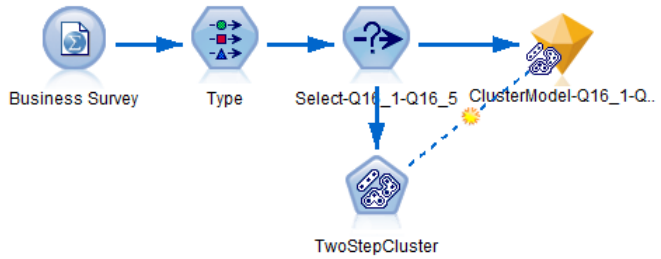
	1.000000	2.000000	$null$
1.000000	100	0	0
2.000000	2	306	3
999.000000	0	0	2

Performance Evaluation

1.000000	1.398
2.000000	0.284

11.31 Appendix 30 – two step clustering for variables Q16_1 to Q16_5.

11.31.1 Q16_1 to Q16_5 modelling stream.

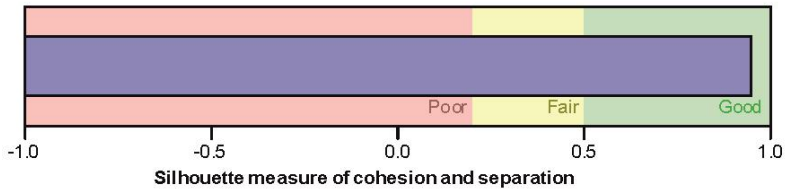


11.31.2 Model summary and evaluation.

Model Summary

Algorithm	TwoStep
Inputs	5
Clusters	9

Cluster Quality

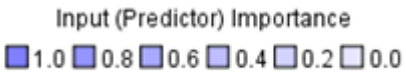


Average silhouette = 0.9

11.31.3 Clusters' characteristics.

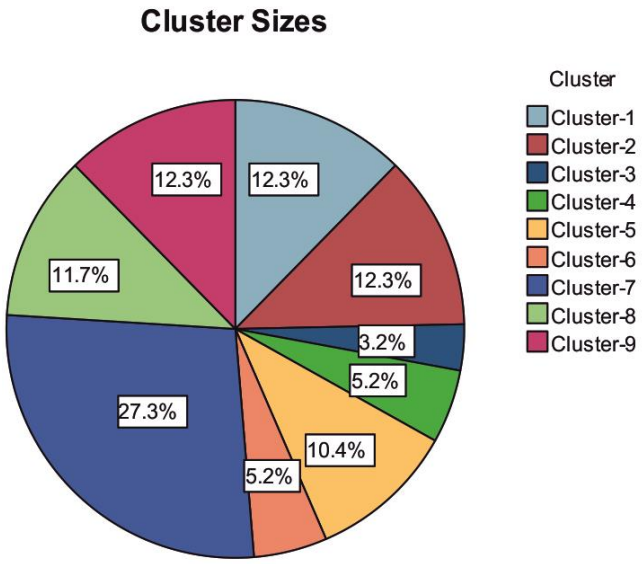
Inputs sorted in descending order by overall importance.

Clusters



Cluster	Cluster-7	Cluster-1	Cluster-2	Cluster-9	Cluster-8	Cluster-5	Cluster-4	Cluster-6	Cluster-3
Size	27.3%	12.3%	12.3%	12.3%	11.7%	10.4%	5.2%	5.2%	3.2%
Inputs	Q16_4	Q16_4	Q16_4	Q16_4	Q16_4	Q16_4	Q16_4	Q16_4	Q16_4
	Q16_5	Q16_5	Q16_5	Q16_5	Q16_5	Q16_5	Q16_5	Q16_5	Q16_5
	Q16_2	Q16_2	Q16_2	Q16_2	Q16_2	Q16_2	Q16_2	Q16_2	Q16_2
	Q16_1	Q16_1	Q16_1	Q16_1	Q16_1	Q16_1	Q16_1	Q16_1	Q16_1
	Q16_3	Q16_3	Q16_3	Q16_3	Q16_3	Q16_3	Q16_3	Q16_3	Q16_3

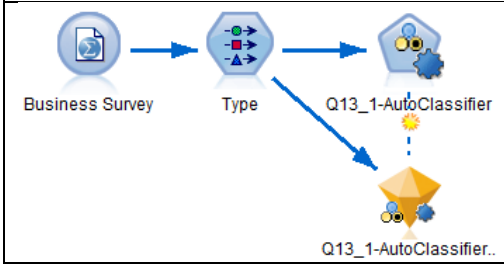
Cluster characteristics	Number of members	Percent
Size of smallest cluster	5	3.2%
Size of largest cluster	42	27.3%
Ratio of size (largest to smallest cluster)	8.4	


Cluster Sizes
Cluster
Cluster-1
Cluster-2
Cluster-3
Cluster-4
Cluster-5
Cluster-6
Cluster-7
Cluster-8
Cluster-9
12.3%
12.3%
12.3%
11.7%
3.2%
5.2%
27.3%
10.4%
5.2%

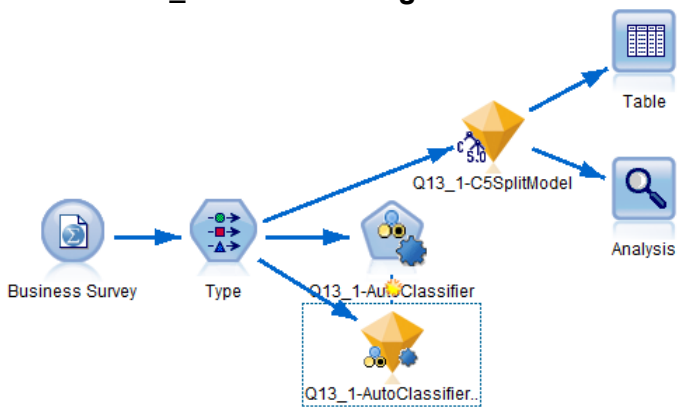
11.32 **Appendix 31 – modelling of variable Q13_1.**
Click here to go to variable Q13 in the main body of the thesis.

11.32.1 Auto Classifier – training summary for Q13_1.

Auto Classifier – training summary for Q13_1	
Target	Q13_1
Inputs	All variables, excluding free text fields and Q13_1
Planned models	3104
Completed models	2750
Retained models	100
Models discarded from final results	2650
Models which failed to build or score	354



11.32.2 Q13_1 C5 – modelling stream.



11.32.3 Performance evaluation and coincidence matrix for Q13_1 C5 Model.

1 = Fully agree, 2 = Agree, 3 = Neither agree, nor disagree, 4 = Disagree, 5 = Strongly disagree, 999 = User non response

Results for output field What is your opinion about the correct and fair expenditure of tax money?

Comparing $C-Q13_1 with What is your opinion about the correct and fair expenditure of tax money?

Correct	334	80.87%
Wrong	79	19.13%
Total	413	

Coincidence Matrix for $C-Q13_1 (rows show actuals)

	1	2	3	4	5
1	44	2	0	1	2
2	1	29	0	0	2
3	1	9	55	7	11
4	1	0	7	85	17
5	4	3	3	5	121
999	0	0	1	0	0
$null$	0	0	0	0	2

Performance Evaluation

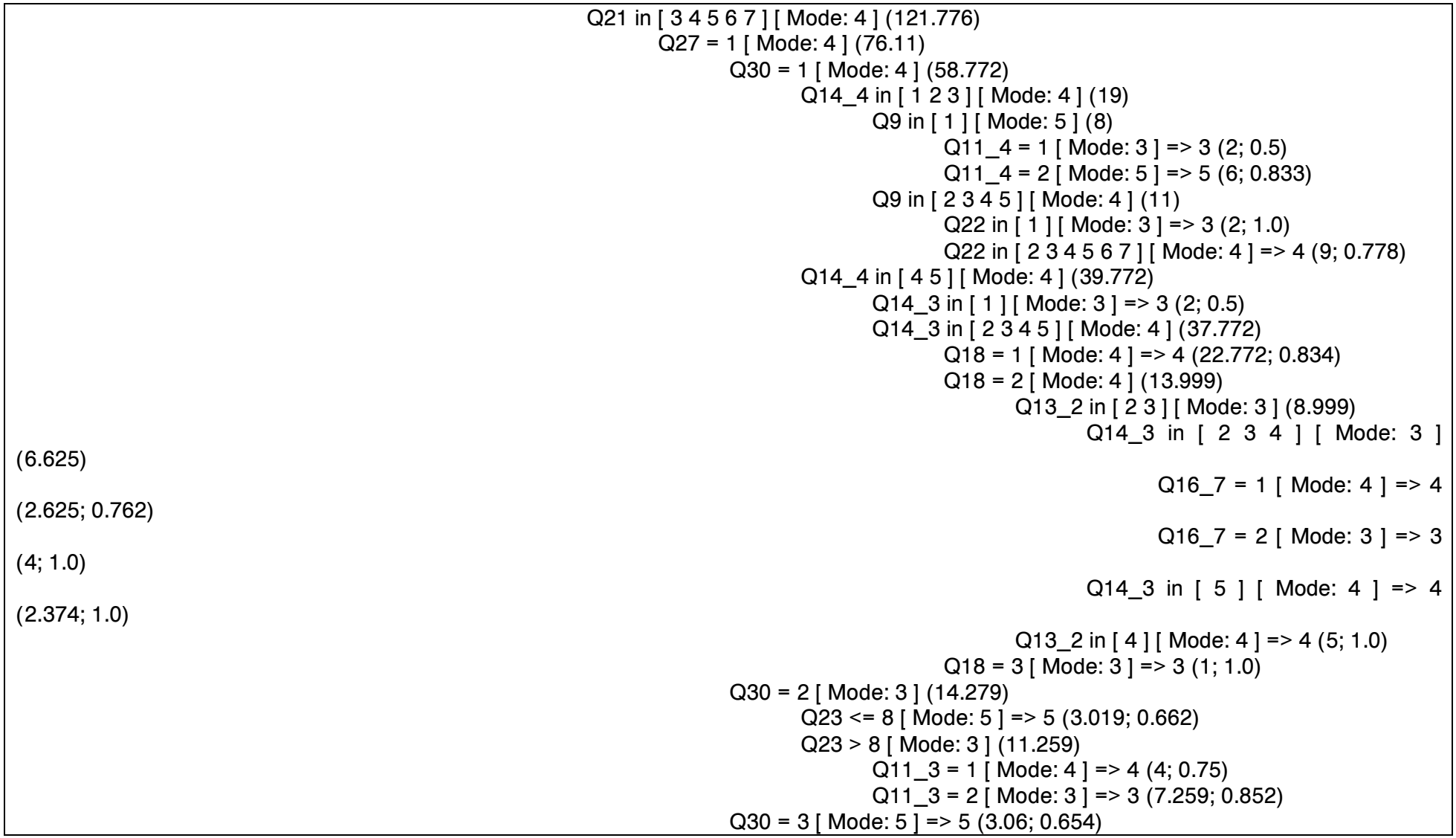
1	1.984
2	2.164
3	1.422
4	1.181
5	0.863

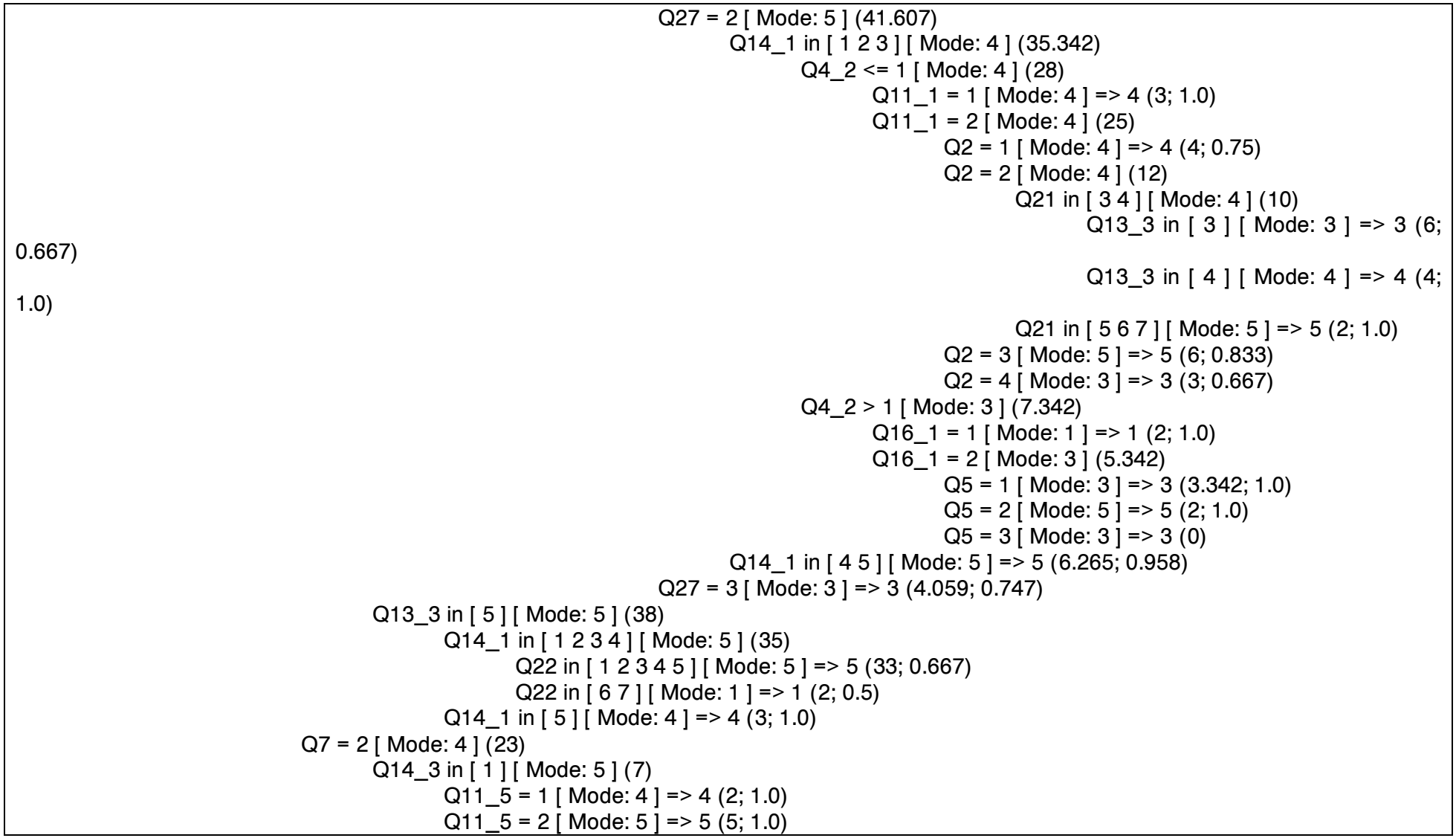
11.32.4 Q13_1 C5 tree model – symbolic representation.

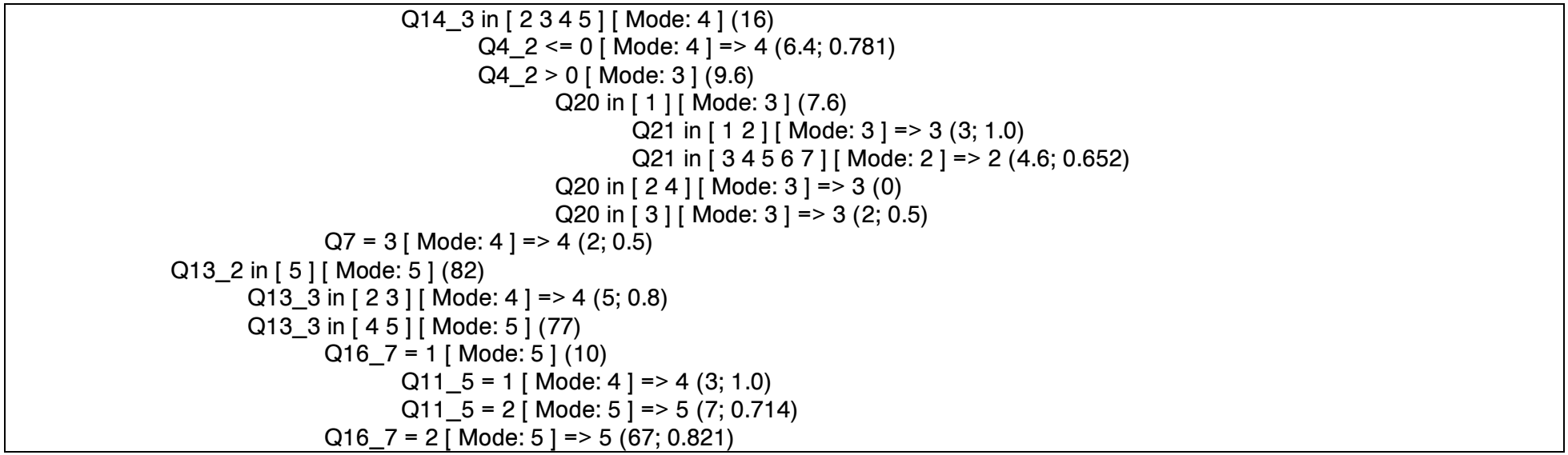
The tree depth is 17 levels and therefore only the symbolic expression is presented below. A high-resolution tree image is available as part of this PhD thesis data store to examiners and supervisors or otherwise upon an authorised request.

```
Rule 1 - estimated accuracy 81.5% [boost 39%]
        Q13_3 in [ 1 ] [ Mode: 1 ] => 1 (38; 0.921)
        Q13_3 in [ 2 3 4 5 ] [ Mode: 5 ] (372)
                Q13_2 in [ 1 2 3 4 ] [ Mode: 4 ] (290)
                        Q13_3 in [ 2 ] [ Mode: 2 ] (49)
                                Q4_1 <= 0 [ Mode: 4 ] (12)
                                        Q23 <= 10 [ Mode: 4 ] (10)
                                                Q21 in [ 1 2 3 ] [ Mode: 4 ] (8)
                                                        Q16_7 = 1 [ Mode: 4 ] => 4 (3; 1.0)
                                                        Q16_7 = 2 [ Mode: 4 ] (5)
                                                                Q14_1 in [ 1 2 ] [ Mode: 4 ] => 4 (2; 1.0)
                                                                Q14_1 in [ 3 4 5 ] [ Mode: 2 ] => 2 (3; 0.667)
                                                Q21 in [ 4 5 6 7 ] [ Mode: 1 ] => 1 (2; 1.0)
                                        Q23 > 10 [ Mode: 3 ] => 3 (2; 1.0)
                                Q4_1 > 0 [ Mode: 2 ] (37)
                                        Q13_2 in [ 1 2 ] [ Mode: 2 ] => 2 (30; 0.6)
                                        Q13_2 in [ 3 4 ] [ Mode: 3 ] (7)
                                                Q4_1 <= 2 [ Mode: 3 ] => 3 (3; 1.0)
                                                Q4_1 > 2 [ Mode: 5 ] => 5 (4; 0.75)
                        Q13_3 in [ 3 4 5 ] [ Mode: 4 ] (241)
                                Q7 = 1 [ Mode: 4 ] (216)
                                        Q13_3 in [ 3 4 ] [ Mode: 4 ] (178)
                                                Q16_6 = 1 [ Mode: 3 ] (9)
                                                        Q14_4 in [ 1 2 3 ] [ Mode: 1 ] => 1 (3; 0.333)
                                                        Q14_4 in [ 4 5 ] [ Mode: 3 ] => 3 (6; 0.833)
                                                Q16_6 = 2 [ Mode: 4 ] (169)
                                                        Q13_2 in [ 1 ] [ Mode: 5 ] (12)
                                                                Q14_4 in [ 1 2 3 ] [ Mode: 3 ] => 3 (2; 1.0)
                                                                Q14_4 in [ 4 5 ] [ Mode: 5 ] (10)
```

```
                                                                Q21 in [ 1 2 ] [ Mode: 5 ] (8)
                                                                        Q11_5 = 1 [ Mode: 5 ] => 5 (2; 0.5)
                                                                        Q11_5 = 2 [ Mode: 5 ] (6)
                                                                                Q30 = 1 [ Mode: 1 ] => 1 (4; 0.75)
                                                                                Q30 = 2 [ Mode: 5 ] => 5 (2; 1.0)
                                                                                Q30 = 3 [ Mode: 1 ] => 1 (0)
                                                                Q21 in [ 3 4 5 6 7 ] [ Mode: 4 ] => 4 (2; 1.0)
                                                        Q13_2 in [ 2 3 4 ] [ Mode: 4 ] (157)
                                                                Q21 in [ 1 2 ] [ Mode: 3 ] (35.224)
                                                                        Q9 in [ 1 2 ] [ Mode: 2 ] (20)
                                                                                Q25 <= 7 [ Mode: 2 ] (18)
                                                                                        Q16_5 = 1 [ Mode: 2 ] => 2 (2; 1.0)
                                                                                        Q16_5 = 2 [ Mode: 3 ] (16)
                                                                                                Q11_3 = 1 [ Mode: 4 ] => 4 (2; 1.0)
                                                                                                Q11_3 = 2 [ Mode: 3 ] (14)
                                                                                                        Q13_2 in [ 2 3 ] [ Mode: 3 ] (12)
                                                                                                                Q4_1 <= 3 [ Mode: 3 ] (8)
                                                                                                                        Q16_3 = 1 [ Mode: 5 ] => 5 (3;
0.667)
                                                                                                                        Q16_3 = 2 [ Mode: 3 ] => 3 (5; 0.8)
                                                                                                                Q4_1 > 3 [ Mode: 2 ] (4)
                                                                                                                        Q4_2 <= 2 [ Mode: 2 ] => 2 (2; 1.0)
                                                                                                                        Q4_2 > 2 [ Mode: 4 ] => 4 (2; 1.0)
                                                                                                        Q13_2 in [ 4 ] [ Mode: 2 ] => 2 (2; 1.0)
                                                                                Q25 > 7 [ Mode: 5 ] => 5 (2; 1.0)
                                                                Q9 in [ 3 4 5 ] [ Mode: 3 ] (15.224)
                                                                        Q11_6 = 1 [ Mode: 4 ] (12.224)
                                                                                Q23 <= 8 [ Mode: 3 ] => 3 (3.056; 0.982)
                                                                                Q23 > 8 [ Mode: 4 ] (9.168)
                                                                                        Q14_3 in [ 1 2 ] [ Mode: 3 ] => 3 (2; 1.0)
                                                                                        Q14_3 in [ 3 4 5 ] [ Mode: 4 ] => 4 (7.168; 1.0)
                                                                        Q11_6 = 2 [ Mode: 3 ] => 3 (3; 1.0)
```


Q21 in [3 4 5 6 7] [Mode: 4] (121.776)
Q27 = 1 [Mode: 4] (76.11)
Q30 = 1 [Mode: 4] (58.772)
Q14_4 in [1 2 3] [Mode: 4] (19)
Q9 in [1] [Mode: 5] (8)
Q11_4 = 1 [Mode: 3] => 3 (2; 0.5)
Q11_4 = 2 [Mode: 5] => 5 (6; 0.833)
Q9 in [2 3 4 5] [Mode: 4] (11)
Q22 in [1] [Mode: 3] => 3 (2; 1.0)
Q22 in [2 3 4 5 6 7] [Mode: 4] => 4 (9; 0.778)
Q14_4 in [4 5] [Mode: 4] (39.772)
Q14_3 in [1] [Mode: 3] => 3 (2; 0.5)
Q14_3 in [2 3 4 5] [Mode: 4] (37.772)
Q18 = 1 [Mode: 4] => 4 (22.772; 0.834)
Q18 = 2 [Mode: 4] (13.999)
Q13_2 in [2 3] [Mode: 3] (8.999)
Q14_3 in [2 3 4] [Mode: 3] (6.625)
Q16_7 = 1 [Mode: 4] => 4 (2.625; 0.762)
Q16_7 = 2 [Mode: 3] => 3 (4; 1.0)
Q14_3 in [5] [Mode: 4] => 4 (2.374; 1.0)
Q13_2 in [4] [Mode: 4] => 4 (5; 1.0)
Q18 = 3 [Mode: 3] => 3 (1; 1.0)
Q30 = 2 [Mode: 3] (14.279)
Q23 <= 8 [Mode: 5] => 5 (3.019; 0.662)
Q23 > 8 [Mode: 3] (11.259)
Q11_3 = 1 [Mode: 4] => 4 (4; 0.75)
Q11_3 = 2 [Mode: 3] => 3 (7.259; 0.852)
Q30 = 3 [Mode: 5] => 5 (3.06; 0.654)

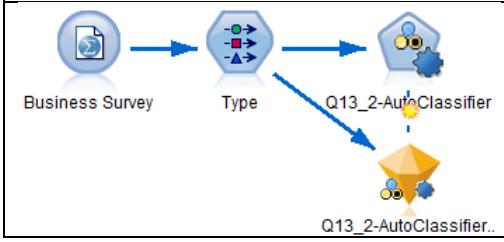

Q27 = 2 [Mode: 5] (41.607)
Q14_1 in [1 2 3] [Mode: 4] (35.342)
Q4_2 <= 1 [Mode: 4] (28)
Q11_1 = 1 [Mode: 4] => 4 (3; 1.0)
Q11_1 = 2 [Mode: 4] (25)
Q2 = 1 [Mode: 4] => 4 (4; 0.75)
Q2 = 2 [Mode: 4] (12)
Q21 in [3 4] [Mode: 4] (10)
Q13_3 in [3] [Mode: 3] => 3 (6; 0.667)
Q13_3 in [4] [Mode: 4] => 4 (4; 1.0)
Q21 in [5 6 7] [Mode: 5] => 5 (2; 1.0)
Q2 = 3 [Mode: 5] => 5 (6; 0.833)
Q2 = 4 [Mode: 3] => 3 (3; 0.667)
Q4_2 > 1 [Mode: 3] (7.342)
Q16_1 = 1 [Mode: 1] => 1 (2; 1.0)
Q16_1 = 2 [Mode: 3] (5.342)
Q5 = 1 [Mode: 3] => 3 (3.342; 1.0)
Q5 = 2 [Mode: 5] => 5 (2; 1.0)
Q5 = 3 [Mode: 3] => 3 (0)
Q14_1 in [4 5] [Mode: 5] => 5 (6.265; 0.958)
Q27 = 3 [Mode: 3] => 3 (4.059; 0.747)
Q13_3 in [5] [Mode: 5] (38)
Q14_1 in [1 2 3 4] [Mode: 5] (35)
Q22 in [1 2 3 4 5] [Mode: 5] => 5 (33; 0.667)
Q22 in [6 7] [Mode: 1] => 1 (2; 0.5)
Q14_1 in [5] [Mode: 4] => 4 (3; 1.0)
Q7 = 2 [Mode: 4] (23)
Q14_3 in [1] [Mode: 5] (7)
Q11_5 = 1 [Mode: 4] => 4 (2; 1.0)
Q11_5 = 2 [Mode: 5] => 5 (5; 1.0)


Q14_3 in [2 3 4 5] [Mode: 4] (16)
Q4_2 <= 0 [Mode: 4] => 4 (6.4; 0.781)
Q4_2 > 0 [Mode: 3] (9.6)
Q20 in [1] [Mode: 3] (7.6)
Q21 in [1 2] [Mode: 3] => 3 (3; 1.0)
Q21 in [3 4 5 6 7] [Mode: 2] => 2 (4.6; 0.652)
Q20 in [2 4] [Mode: 3] => 3 (0)
Q20 in [3] [Mode: 3] => 3 (2; 0.5)
Q7 = 3 [Mode: 4] => 4 (2; 0.5)
Q13_2 in [5] [Mode: 5] (82)
Q13_3 in [2 3] [Mode: 4] => 4 (5; 0.8)
Q13_3 in [4 5] [Mode: 5] (77)
Q16_7 = 1 [Mode: 5] (10)
Q11_5 = 1 [Mode: 4] => 4 (3; 1.0)
Q11_5 = 2 [Mode: 5] => 5 (7; 0.714)
Q16_7 = 2 [Mode: 5] => 5 (67; 0.821)

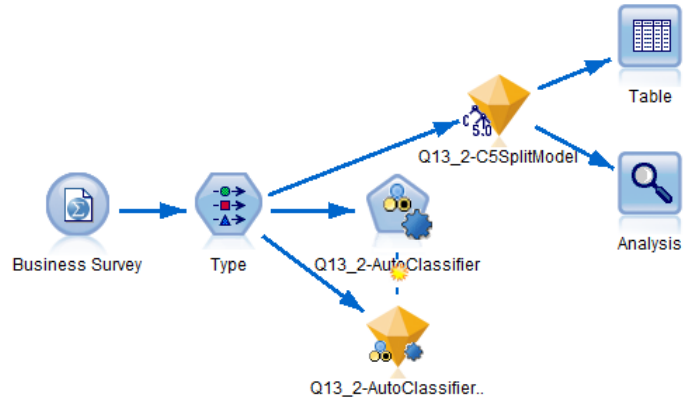
11.33 **Appendix 32 – modelling of variable Q13_2.**
Click here to go to variable Q13_2 in the main body of the thesis.

11.33.1 Auto Classifier – training summary for Q13_2.

Auto Classifier – training summary for Q13_2	
Target	Q13_2
Inputs	All variables, excluding free text fields and Q13_2
Planned models	3104
Completed models	2697
Retained models	100
Models discarded from final results	2597
Models which failed to build or score	407



11.33.2 Q13_2 C5 – modelling stream.



11.33.3 Performance evaluation and coincidence matrix for Q13_2 C5 Model.

1 = Fully agree, 2 = Agree, 3 = Neither agree, nor disagree, 4 = Disagree, 5 = Strongly disagree, 999 = User non response

Results for output field What is your opinion about fair, professional attitude and treatment by NRA?

Comparing $C-Q13_2 with What is your opinion about fair, professional attitude and treatment by NRA?

Correct	342	82.81%
Wrong	71	17.19%
Total	413	

Coincidence Matrix for $C-Q13_2 (rows show actuals)

	1	2	3	4	5
1	39	5	2	1	2
2	1	68	5	3	1
3	3	1	92	9	8
4	1	1	7	70	10
5	0	1	4	4	73
$null$	0	0	2	0	0

Performance Evaluation

1	2.011
2	1.556
3	1.099
4	1.317
5	1.364

11.33.4 Q13_2 C5 tree model – symbolic representation.

The tree depth is 14 levels and therefore only the symbolic expression is presented below. A high-resolution tree image is available as part of this PhD thesis data store to examiners and supervisors or otherwise upon an authorised request.

```
Rule 1 - estimated accuracy 74.21% [boost 41.5%]
        Q13_3 in [ 1 2 3 ] [ Mode: 3 ] (187)
                Q13_3 in [ 1 ] [ Mode: 1 ] (38)
                        Q7 in [ 1 3 ] [ Mode: 1 ] => 1 (35; 0.829)
                        Q7 in [ 2 ] [ Mode: 3 ] => 3 (3; 0.667)
                Q13_3 in [ 2 3 ] [ Mode: 3 ] (149)
                        Q13_3 in [ 2 ] [ Mode: 2 ] (49)
                                Q21 in [ 1 2 3 ] [ Mode: 2 ] (34)
                                        Q29 in [ 1 2 4 ] [ Mode: 2 ] => 2 (32.938; 0.909)
                                        Q29 in [ 3 ] [ Mode: 3 ] => 3 (1.062; 0.941)
                                Q21 in [ 4 5 6 7 ] [ Mode: 3 ] (15)
                                        Q27 in [ 1 ] [ Mode: 3 ] (13)
                                                Q9 in [ 1 ] [ Mode: 2 ] => 2 (7; 0.429)
                                                Q9 in [ 2 3 4 5 ] [ Mode: 3 ] => 3 (6; 0.833)
                                        Q27 in [ 2 ] [ Mode: 2 ] => 2 (2; 1.0)
                                        Q27 in [ 3 ] [ Mode: 3 ] => 3 (0)
                        Q13_3 in [ 3 ] [ Mode: 3 ] (100)
                                Q11_3 = 1 [ Mode: 4 ] (14)
                                        Q14_4 in [ 1 2 3 4 ] [ Mode: 3 ] => 3 (7; 0.714)
                                        Q14_4 in [ 5 ] [ Mode: 4 ] => 4 (7; 0.571)
                                Q11_3 = 2 [ Mode: 3 ] (86)
                                        Q16_7 = 1 [ Mode: 2 ] (21)
                                                Q25 <= 4 [ Mode: 3 ] (14)
                                                        Q2 in [ 2 ] [ Mode: 3 ] (10)
                                                                Q14_1 in [ 1 2 ] [ Mode: 1 ] => 1 (5; 0.4)
                                                                Q14_1 in [ 3 4 5 ] [ Mode: 3 ] => 3 (5; 1.0)
                                                        Q2 in [ 3 ] [ Mode: 1 ] => 1 (1; 1.0)
                                                        Q2 in [ 4 ] [ Mode: 2 ] => 2 (3; 0.667)
```

```
                                        Q2 in [ 1 ] [ Mode: 3 ] => 3 (0)
                                    Q25 > 4 [ Mode: 2 ] => 2 (7; 0.714)
                                Q16_7 = 2 [ Mode: 3 ] (65)
                                    Q18 in [ 1 ] [ Mode: 3 ] (43)
                                        Q19 <= 72 [ Mode: 2 ] => 2 (6; 0.5)
                                        Q19 > 72 [ Mode: 3 ] => 3 (37; 0.757)
                                    Q18 in [ 2 ] [ Mode: 3 ] (20)
                                        Q26 <= 103 [ Mode: 5 ] => 5 (5; 0.6)
                                        Q26 > 103 [ Mode: 3 ] (15)
                                            Q5 in [ 1 2 ] [ Mode: 3 ] => 3 (14; 0.786)
                                            Q5 in [ 3 ] [ Mode: 1 ] => 1 (1; 1.0)
                                    Q18 in [ 3 ] [ Mode: 3 ] => 3 (2; 1.0)
Q13_3 in [ 4 5 ] [ Mode: 5 ] (224)
    Q13_3 in [ 4 ] [ Mode: 4 ] (113)
        Q7 in [ 1 ] [ Mode: 4 ] (109)
            Q21 in [ 1 2 3 ] [ Mode: 4 ] (49.454)
                Q11_4 = 1 [ Mode: 2 ] => 2 (7; 0.429)
                Q11_4 = 2 [ Mode: 4 ] (42.454)
                    Q25 <= 7 [ Mode: 4 ] (36.389)
                        Q16_5 = 1 [ Mode: 4 ] => 4 (5; 0.6)
                        Q16_5 = 2 [ Mode: 3 ] (31.389)
                            Q23 <= 9 [ Mode: 3 ] => 3 (5.063; 0.605)
                            Q23 > 9 [ Mode: 4 ] (26.326)
                                Q14_3 in [ 1 2 3 ] [ Mode: 3 ] => 3 (14; 0.5)
                                Q14_3 in [ 4 5 ] [ Mode: 4 ] (12.326)
                                    Q14_3 in [ 4 ] [ Mode: 4 ] => 4 (7; 0.714)
                                    Q14_3 in [ 5 ] [ Mode: 2 ] => 2 (5.326; 0.563)
                    Q25 > 7 [ Mode: 4 ] => 4 (6.065; 0.824)
            Q21 in [ 4 5 6 7 ] [ Mode: 4 ] (59.546)
                Q23 <= 8 [ Mode: 5 ] (19.176)
                    Q14_4 in [ 1 2 ] [ Mode: 5 ] => 5 (5; 1.0)
                    Q14_4 in [ 3 4 5 ] [ Mode: 4 ] (14.176)
```

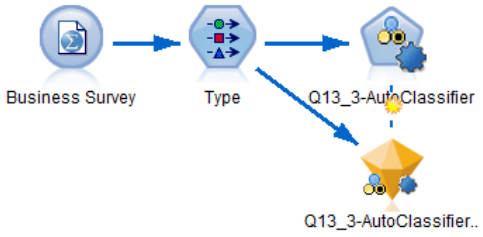
```
                                                Q13_1 in [ 1 2 3 4 ] [ Mode: 4 ] => 4 (7.176; 0.557)
                                                Q13_1 in [ 5 ] [ Mode: 5 ] => 5 (7; 0.571)
                                        Q23 > 8 [ Mode: 4 ] (40.37)
                                                Q23 <= 12 [ Mode: 4 ] (35.324)
                                                        Q16_1 = 1 [ Mode: 3 ] => 3 (9; 0.556)
                                                        Q16_1 = 2 [ Mode: 4 ] => 4 (26.324; 0.874)
                                                Q23 > 12 [ Mode: 4 ] => 4 (5.046; 0.594)
                        Q7 in [ 2 ] [ Mode: 4 ] => 4 (3; 1.0)
                        Q7 in [ 3 ] [ Mode: 5 ] => 5 (1; 1.0)
                Q13_3 in [ 5 ] [ Mode: 5 ] (111)
                        Q21 in [ 1 2 3 ] [ Mode: 5 ] (35)
                                Q16_8 = 1 [ Mode: 5 ] => 5 (5; 0.4)
                                Q16_8 = 2 [ Mode: 5 ] (30)
                                        Q14_1 in [ 1 ] [ Mode: 3 ] (13)
                                                Q15 in [ 0 1 2 3 4 5 6 7 8 9 10 12 13 14 15 17 18 19 20 21 23 ] [ Mode: 5 ] => 5 (6; 0.667)
                                                Q15 in [ 25 27 28 29 30 33 34 35 39 40 43 49 50 51 52 57 58 60 61 65 66 69 70 71 73 75 79 80 81 85
86 90 91 92 97 98 100 ] [ Mode: 3 ] => 3 (7; 1.0)
                                        Q14_1 in [ 2 3 4 5 ] [ Mode: 5 ] (17)
                                                Q13_1 in [ 1 2 3 4 ] [ Mode: 2 ] (10)
                                                        Q26 <= 213 [ Mode: 3 ] => 3 (5; 0.6)
                                                        Q26 > 213 [ Mode: 2 ] => 2 (5; 0.8)
                                                Q13_1 in [ 5 ] [ Mode: 5 ] => 5 (7; 0.714)
                        Q21 in [ 4 5 6 7 ] [ Mode: 5 ] (76)
                                Q21 in [ 4 5 6 ] [ Mode: 5 ] (71)
                                        Q21 in [ 4 ] [ Mode: 5 ] (25)
                                                Q4_2 <= 1 [ Mode: 4 ] (20)
                                                        Q30 in [ 1 ] [ Mode: 4 ] (17)
                                                                Q5 in [ 1 ] [ Mode: 4 ] (11)
                                                                        Q18 in [ 1 3 ] [ Mode: 4 ] => 4 (6; 1.0)
                                                                        Q18 in [ 2 ] [ Mode: 3 ] => 3 (5; 0.4)
                                                                Q5 in [ 2 ] [ Mode: 5 ] => 5 (5; 0.6)
                                                                Q5 in [ 3 ] [ Mode: 3 ] => 3 (1; 1.0)
```

Q30 in [2] [Mode: 5] => 5 (3; 0.667)
Q30 in [3] [Mode: 4] => 4 (0)
Q4_2 > 1 [Mode: 5] => 5 (5; 1.0)
Q21 in [5 6] [Mode: 5] => 5 (46; 0.804)
Q21 in [7] [Mode: 3] => 3 (5; 0.6)

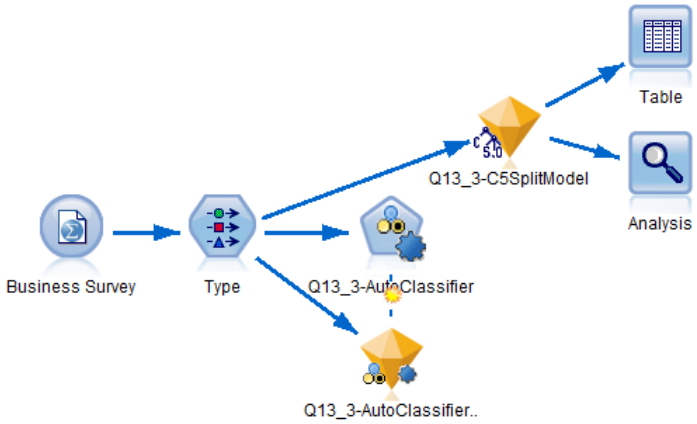
11.34 **Appendix 33 – modelling of variable Q13_3.**
Click here to go to variable Q13_3 in the main body of the thesis.

11.34.1 Auto Classifier – training summary for Q13_3.

Auto Classifier – training summary for Q13_3	
Target	Q13_3
Inputs	All variables, excluding free text fields and Q13_3
Planned models	3104
Completed models	2344
Retained models	100
Models discarded from final results	2244
Models which failed to build or score	840



11.34.2 Q13_3 C5 – modelling stream.



11.34.3 Performance evaluation and coincidence matrix for Q13_3 C5 Model.

1 = Fully agree, 2 = Agree, 3 = Neither agree, nor disagree, 4 = Disagree, 5 = Strongly disagree, 999 = User non response

Results for output field What is your opinion about fair, professional attitude and treatment by the state institutions?

Comparing $C-Q13_3 with What is your opinion about fair, professional attitude and treatment by the state institutions?

Correct	389	94.19%
Wrong	24	5.81%
Total	413	

Coincidence Matrix for $C-Q13_3 (rows show actuals)

	1	2	3	4	5
1	37	0	0	1	0
2	0	48	1	0	0
3	1	1	95	1	2
4	2	1	3	104	3
5	0	0	3	3	105
$null$	0	0	0	2	0

Performance Evaluation

1	2.308
2	2.091
3	1.347
4	1.231
5	1.267

11.34.4 Q13_3 C5 tree model – symbolic representation.

The tree depth is 17 levels and therefore only the symbolic expression is presented below for ruleset 1 (at 73.4% accuracy). A high-resolution tree image is available as part of this PhD thesis data store to examiners and supervisors or otherwise upon an authorised request.

```
Rule 1 - estimated accuracy 73.36% [boost 96.6%]
        Q13_1 in [ 1 ] [ Mode: 1 ] (49.12)
                Q13_2 in [ 1 ] [ Mode: 1 ] => 1 (33; 0.909)
                Q13_2 in [ 2 3 4 5 ] [ Mode: 1 ] (16.12)
                        Q27 in [ 1 ] [ Mode: 1 ] (11)
                                Q29 in [ 1 3 ] [ Mode: 5 ] => 5 (3; 1.0)
                                Q29 in [ 2 4 ] [ Mode: 1 ] => 1 (8; 0.375)
                        Q27 in [ 2 ] [ Mode: 3 ] => 3 (5.12; 0.586)
                        Q27 in [ 3 ] [ Mode: 1 ] => 1 (0)
        Q13_1 in [ 2 3 4 5 ] [ Mode: 4 ] (361.88)
                Q13_2 in [ 1 2 3 ] [ Mode: 3 ] (195.88)
                        Q13_1 in [ 2 ] [ Mode: 2 ] (29.131)
                                Q13_2 in [ 1 2 ] [ Mode: 2 ] => 2 (23; 0.826)
                                Q13_2 in [ 3 ] [ Mode: 3 ] => 3 (6.131; 0.652)
                        Q13_1 in [ 3 4 5 ] [ Mode: 3 ] (166.75)
                                Q7 in [ 1 ] [ Mode: 3 ] (151.75)
                                        Q25 <= 12 [ Mode: 3 ] (146.716)
                                                Q21 in [ 1 2 3 4 5 6 ] [ Mode: 3 ] (139.67)
                                                        Q16_6 = 1 [ Mode: 5 ] (12)
                                                                Q13_2 in [ 1 2 ] [ Mode: 4 ] => 4 (7; 0.429)
                                                                Q13_2 in [ 3 ] [ Mode: 5 ] => 5 (5; 0.6)
                                                        Q16_6 = 2 [ Mode: 3 ] (127.67)
                                                                Q2 in [ 1 ] [ Mode: 3 ] (17)
                                                                        Q11_5 = 1 [ Mode: 2 ] => 2 (5; 0.6)
                                                                        Q11_5 = 2 [ Mode: 3 ] (12)
                                                                                Q15 in [ 0 1 2 3 4 5 6 7 8 ] [ Mode: 2 ] => 2 (5; 0.4)
                                                                                Q15 in [ 9 10 12 13 14 15 17 18 19 20 21 23 25 27 28 29 30 33 34 35 39
40 43 49 50 51 52 57 58 60 61 65 66 69 70 71 73 75 79 80 81 85 86 90 91 92 97 98 100 ] [ Mode: 3 ] => 3 (7; 0.571)
```

```
                                                Q2 in [ 2 ] [ Mode: 3 ] (64.92)
                                                        Q24_6_Rank in [ 1 ] [ Mode: 4 ] (22.913)
                                                                Q13_1 in [ 3 ] [ Mode: 3 ] => 3 (9.118; 0.503)
                                                                Q13_1 in [ 4 5 ] [ Mode: 4 ] => 4 (13.795; 0.493)
                                                        Q24_6_Rank in [ 2 3 ] [ Mode: 3 ] (42.007)
                                                                Q13_1 in [ 3 ] [ Mode: 3 ] => 3 (16.882; 0.557)
                                                                Q13_1 in [ 4 5 ] [ Mode: 3 ] (25.125)
                                                                        Q13_1 in [ 4 ] [ Mode: 4 ] (14.89)
                                                                                Q22 in [ 1 2 ] [ Mode: 3 ] => 3 (7.108; 0.364)
                                                                                Q22 in [ 3 4 5 6 7 ] [ Mode: 4 ] => 4 (7.782; 0.622)
                                                                        Q13_1 in [ 5 ] [ Mode: 3 ] => 3 (10.235; 0.42)
                                                Q2 in [ 3 ] [ Mode: 4 ] (30)
                                                        Q14_1 in [ 1 2 ] [ Mode: 3 ] (23)
                                                                Q18 in [ 1 ] [ Mode: 4 ] (13)
                                                                        Q19 <= 82 [ Mode: 3 ] => 3 (7; 0.714)
                                                                        Q19 > 82 [ Mode: 4 ] => 4 (6; 0.833)
                                                                Q18 in [ 2 3 ] [ Mode: 3 ] => 3 (10; 0.5)
                                                        Q14_1 in [ 3 4 5 ] [ Mode: 4 ] => 4 (7; 0.714)
                                                Q2 in [ 4 ] [ Mode: 3 ] => 3 (15.75; 0.571)
                                        Q21 in [ 7 ] [ Mode: 3 ] => 3 (7.046; 0.71)
                                Q25 > 12 [ Mode: 3 ] => 3 (5.033; 0.993)
                        Q7 in [ 2 ] [ Mode: 3 ] (14)
                                Q18 in [ 1 3 ] [ Mode: 3 ] => 3 (12; 0.833)
                                Q18 in [ 2 ] [ Mode: 5 ] => 5 (2; 1.0)
                        Q7 in [ 3 ] [ Mode: 5 ] => 5 (1; 1.0)
Q13_2 in [ 4 5 ] [ Mode: 5 ] (166)
        Q13_2 in [ 4 ] [ Mode: 4 ] (85)
                Q30 in [ 1 ] [ Mode: 4 ] (63.75)
                        Q14_3 in [ 1 2 ] [ Mode: 5 ] (22)
                                Q9 in [ 1 ] [ Mode: 4 ] => 4 (8; 0.75)
                                Q9 in [ 2 3 4 5 ] [ Mode: 5 ] => 5 (14; 0.714)
                        Q14_3 in [ 3 4 5 ] [ Mode: 4 ] (41.75)
```

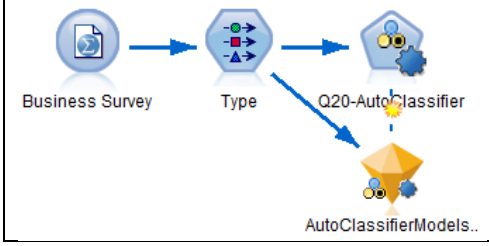
```
                        Q11_3 = 1 [ Mode: 3 ] => 3 (8; 0.75)
                        Q11_3 = 2 [ Mode: 4 ] (33.75)
                                Q27 in [ 1 ] [ Mode: 4 ] (22.75)
                                        Q25 <= 4 [ Mode: 3 ] => 3 (5; 0.8)
                                        Q25 > 4 [ Mode: 4 ] => 4 (17.75; 0.831)
                                Q27 in [ 2 3 ] [ Mode: 4 ] => 4 (11; 1.0)
        Q30 in [ 2 3 ] [ Mode: 4 ] => 4 (21.25; 0.859)
Q13_2 in [ 5 ] [ Mode: 5 ] (81)
        Q13_1 in [ 2 3 4 ] [ Mode: 5 ] (20)
                Q23 <= 7 [ Mode: 4 ] => 4 (5; 0.6)
                Q23 > 7 [ Mode: 5 ] (15)
                        Q7 in [ 1 3 ] [ Mode: 5 ] => 5 (12; 0.917)
                        Q7 in [ 2 ] [ Mode: 3 ] => 3 (3; 0.667)
        Q13_1 in [ 5 ] [ Mode: 5 ] (61)
                Q16_7 = 1 [ Mode: 5 ] => 5 (6; 0.667)
                Q16_7 = 2 [ Mode: 5 ] (55)
                        Q23 <= 9 [ Mode: 5 ] (22)
                                Q14_3 in [ 1 2 3 ] [ Mode: 5 ] (15)
                                        Q18 in [ 1 ] [ Mode: 5 ] (13)
                                                Q27 in [ 1 ] [ Mode: 4 ] => 4 (9; 0.556)
                                                Q27 in [ 2 3 ] [ Mode: 5 ] => 5 (4; 1.0)
                                        Q18 in [ 2 ] [ Mode: 4 ] => 4 (2; 1.0)
                                        Q18 in [ 3 ] [ Mode: 5 ] => 5 (0)
                                Q14_3 in [ 4 5 ] [ Mode: 5 ] => 5 (7; 1.0)
                        Q23 > 9 [ Mode: 5 ] => 5 (33; 0.97)
```

11.35 **Appendix 34 – endeavour to formalisation, variable Q20_1 with Q19.**

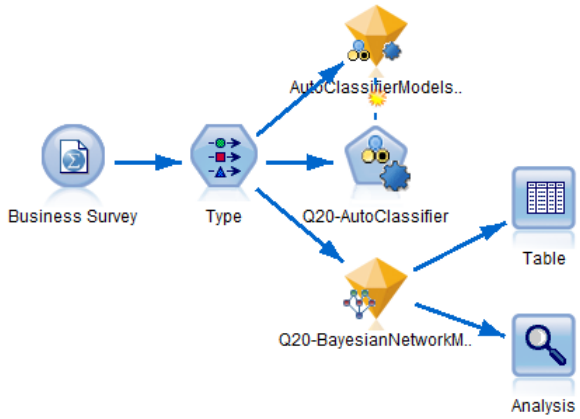
Click here to go to variable Q20_1 modelling section in the body text.

11.35.1 Auto Classifier – training summary for Q20_1.

Auto Classifier – training summary for Q20_1	
Target	Q20_1
Inputs	Q19
Planned models	64
Completed models	64
Retained models	64
Models discarded from final results	0
Models which failed to build or score	0



11.35.2 Q20_1 Bayesian Network – modelling stream.



11.35.3 Performance evaluation for Q20_1 Model.

Results for output field Would you prefer all of your suppliers to provide you with invoices?

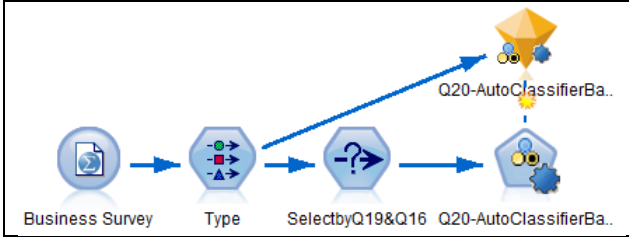
Comparing $B-Q20 with Would you prefer all of your suppliers to provide you with invoices?

Correct	352	85.23%
Wrong	61	14.77%
Total	413	

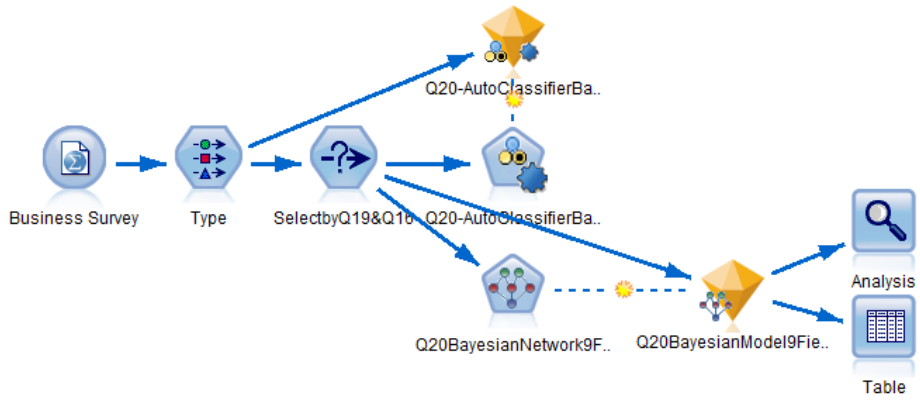
11.36 **Appendix 35 – characteristics of informal chain of interdependence' firms.**
Click here to go to variable Q20_1 modelling section in the body text.

11.36.1 Auto Classifier – training summary for Q20_1.

Auto Classifier – training summary for Q20_1	
Target	Q20_1
Initial set of input variables	Q7, Q13_1, Q13_1, Q13_2, Q13_3, Q14_1, Q14_2, Q14_3, Q14_4, Q15, Q16_1, Q16_2, Q16_3, Q16_4, Q16_5, Q19, Q29, Q30, Q31
Bayesian automated set of selected input variables	Q13_1, Q13_1, Q13_2, Q13_3, Q14_1, Q14_2, Q14_3, Q14_4, Q15, Q19, Q29, Q31
Select node conditional syntax	(Q19 <= 80) and (Q16_1 = 1 or Q16_2 = 1 or Q16_3 = 1 or Q16_4 = 1 or Q16_5 = 1)
Planned models	256
Completed models	256
Retained models	100
Models discarded from final results	156
Models which failed to build or score	0



11.36.2 Q20_1 Bayesian Network – modelling stream.



11.36.3 Performance evaluation for Q20_1 model.

Results for output field Would you prefer all of your suppliers to provide you with invoices?

Comparing $B-Q20 with Would you prefer all of your suppliers to provide you with invoices?

Correct	76	96.2%
Wrong	3	3.8%
Total	79	

Coincidence Matrix for $B-Q20 (rows show actuals)

	1	2	3	4	$null$
1	61	1	1	0	1
2	0	4	0	0	0
3	0	0	8	0	0
4	0	0	0	3	0

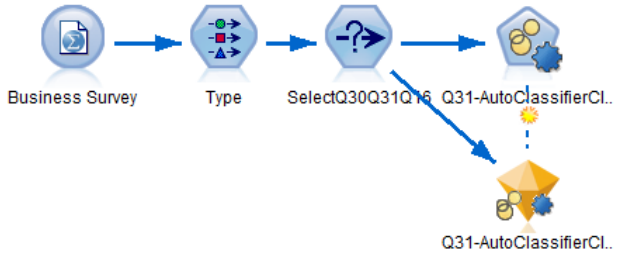
11.37 **Appendix 36 – unfair competition and turnover underreporting.**
Click here to go to the relevant section in the body text.

11.37.1 Identification of turnover underreporting due to unfair competition.

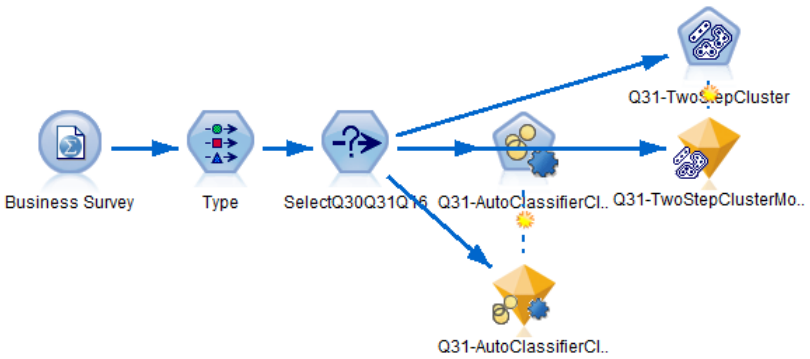
This modelling stream identifies firms which experience difficulties due to exposure to unfair competition and also engage in turnover underreporting.

Modelling stream	
Select node case selection syntax	(Q30 = 1 and Q31 = 2 or Q31 = 3) and (Q16_1 = 1 or Q16_2 = 1 or Q16_3 = 1 or Q16_4 = 1 or Q16_5 = 1)

11.37.2 Auto Cluster Classifier – training summary for Q31.

Auto Cluster Classifier – training summary for Q31	
Target (evaluation variable)	Q31
Input variables	Q16_1, Q16_2, Q16_3, Q16_4, Q16_5
Select node conditional syntax	(Q30 = 1 and Q31 = 2 or Q31 = 3) and (Q16_1 = 1 or Q16_2 = 1 or Q16_3 = 1 or Q16_4 = 1 or Q16_5 = 1)
Planned models	504
Completed models	496
Retained models	100
Models discarded from final results	396
Models which failed to build or score	8
	

11.37.3 Modelling stream – two step clustering.



11.37.4 Model summary and evaluation – two-step cluster.

Model Summary

Algorithm	TwoStep
Inputs	5
Clusters	7

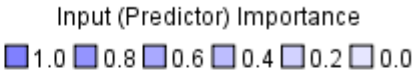
Cluster Quality

Average silhouette = 1

11.37.5 Clusters' characteristics.

Inputs sorted in descending order by overall importance.

Clusters



Cluster	Cluster-4	Cluster-1	Cluster-5	Cluster-7	Cluster-6	Cluster-2	Cluster-3
Size	29.2%	17.4%	14.6%	14.6%	8.4%	7.9%	7.9%
Inputs	Q16_5	Q16_5	Q16_5	Q16_5	Q16_5	Q16_5	Q16_5
	Q16_1	Q16_1	Q16_1	Q16_1	Q16_1	Q16_1	Q16_1
	Q16_3	Q16_3	Q16_3	Q16_3	Q16_3	Q16_3	Q16_3
	Q16_2	Q16_2	Q16_2	Q16_2	Q16_2	Q16_2	Q16_2
	Q16_4	Q16_4	Q16_4	Q16_4	Q16_4	Q16_4	Q16_4

Cluster characteristics	Number of members	Percent
Size of smallest cluster	14	7.9%
Size of largest cluster	52	29.2%
Ratio of size (largest to smallest cluster)	3.71	

11.38 **Appendix 37 – EUROFOUND database on policy measures.**

The figures, table and analysis in this appendix have been adapted from Marie Curie GREY project working paper 5 (Williams et al., 2014a). Data source EUROFOUND – retrieved on 10.07.2014 from http://www.eurofound.europa.eu)

The EUROFOUND database consists of 186 policy measures to tackle undeclared work in 33 countries. The majority of the policy measures are classified to more than one type (see the table at the end of the appendix). For instance a policy measure could be assigned to types: awareness raising, commitment to tax morality and prevention. Some of the policy initiatives do not have an assigned type. The following figures represent an overall picture for the distribution of policy initiatives by type and country. The countries are represented on the X-axis of each figure. The Y-axis indicates number of policy initiatives of any given type per country.

Summary:

- Policy initiatives which encourage commitment to tax morality are only 16 in total from within 186. Estonia and Italy are the countries, which have implemented more than 2 measures – 3 and 4 accordingly. Twenty two out of thirty three countries do not have any commitment to tax morality measures.

- Policy initiatives raising awareness are 27 in total and it is only Estonia where there more than 2 of this type.

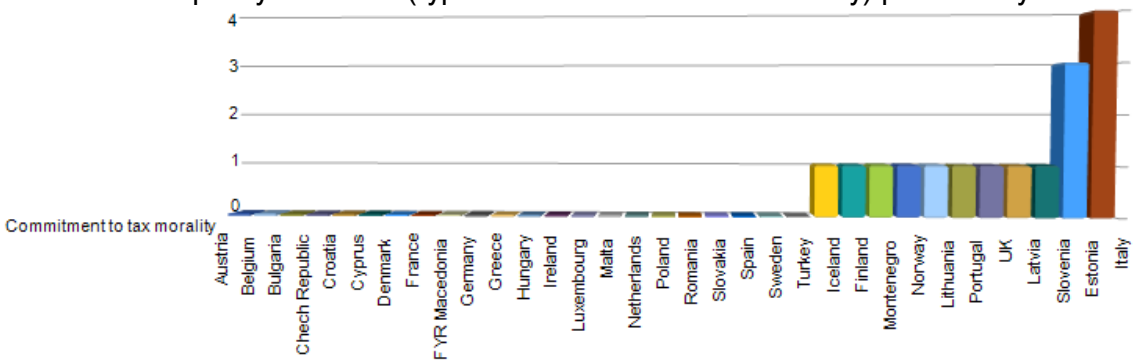
- Policy measures of preventative type are 64. In comparison with the other countries, Poland, Estonia and Italy are those with most of these measures available.

- Policy initiatives towards legitimising undeclared work are 45 in total and it is only Finland which has more that 3 – 7 measures in place. There 11 countries without any of this type.

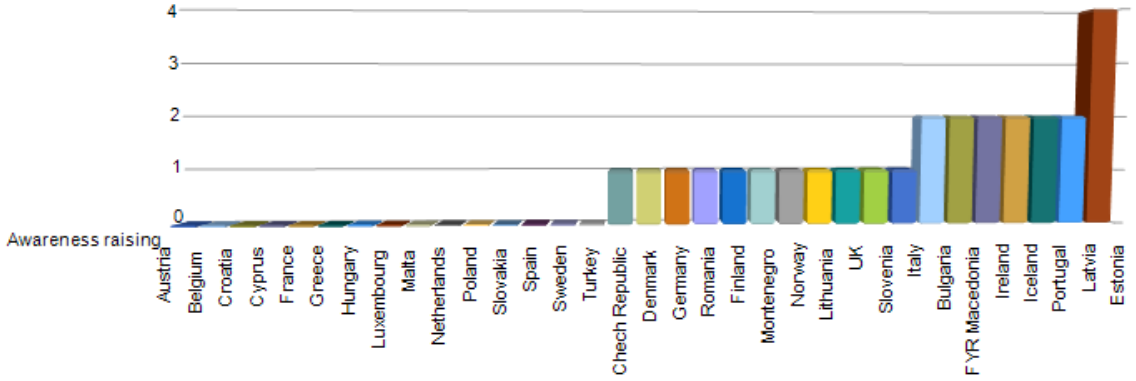
- Policy measures, which improve detection, are 56. Ireland, Germany, Sweden and Belgium have the highest number in comparison with the other countries.

- Policy initiatives, orientated towards increased penalties are 23. Sixteen countries are lacking this type at all and the highest number per country is 2.

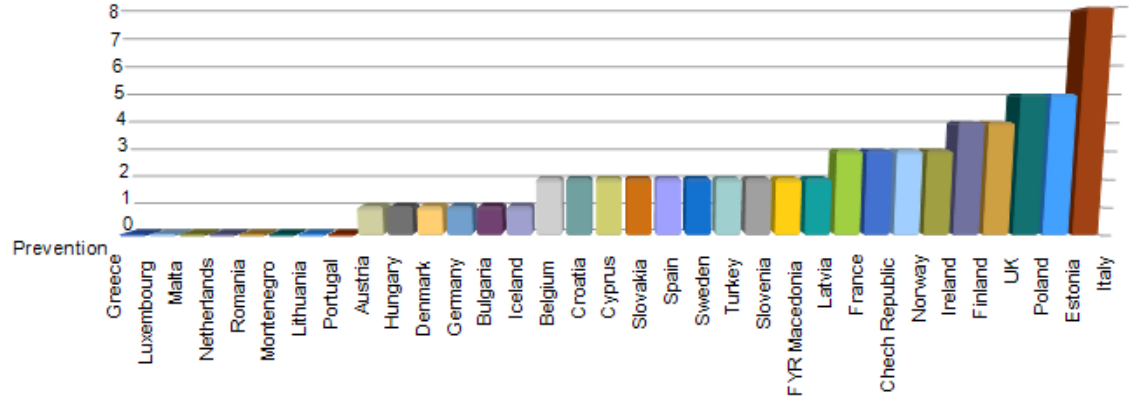
Distribution of policy initiatives (type = commitment to tax morality) per country.



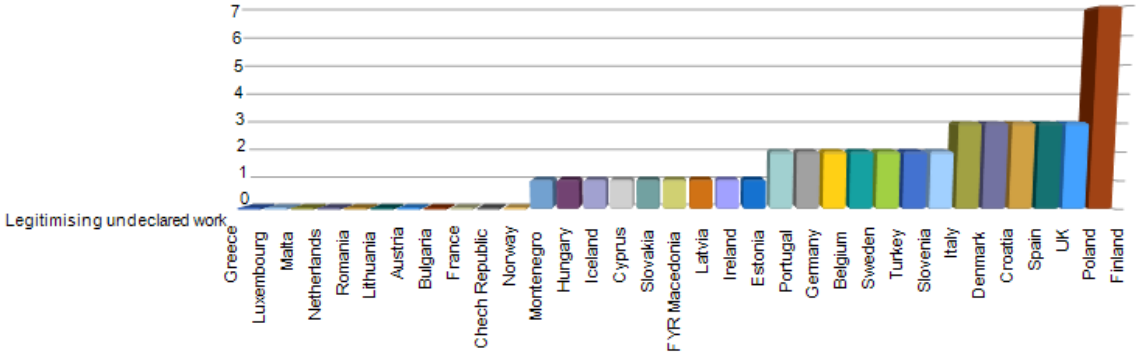
Distribution of policy initiatives (type = awareness raising) per country.



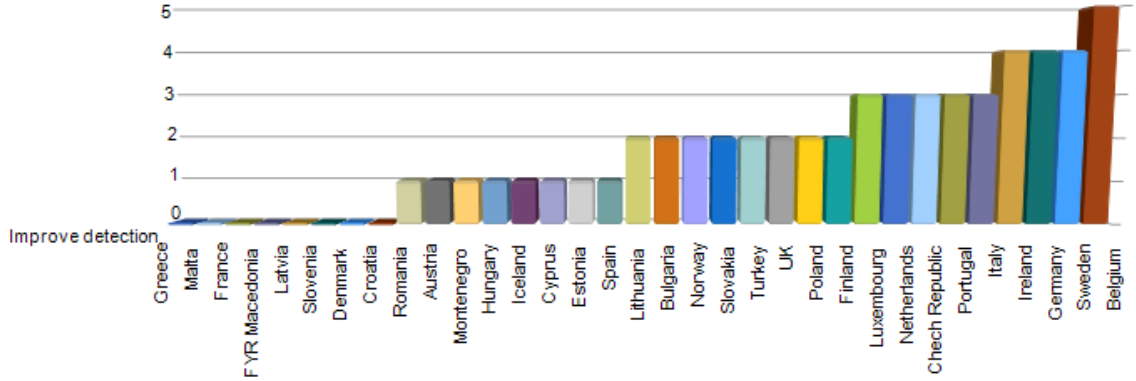
Distribution of policy initiatives (type = prevention) per country.



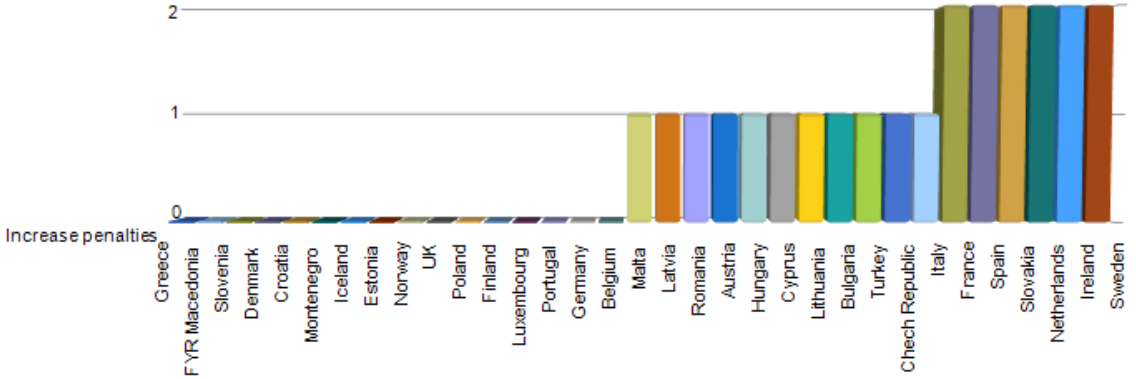
Distribution of policy initiatives (type = legitimising undeclared work) per country.



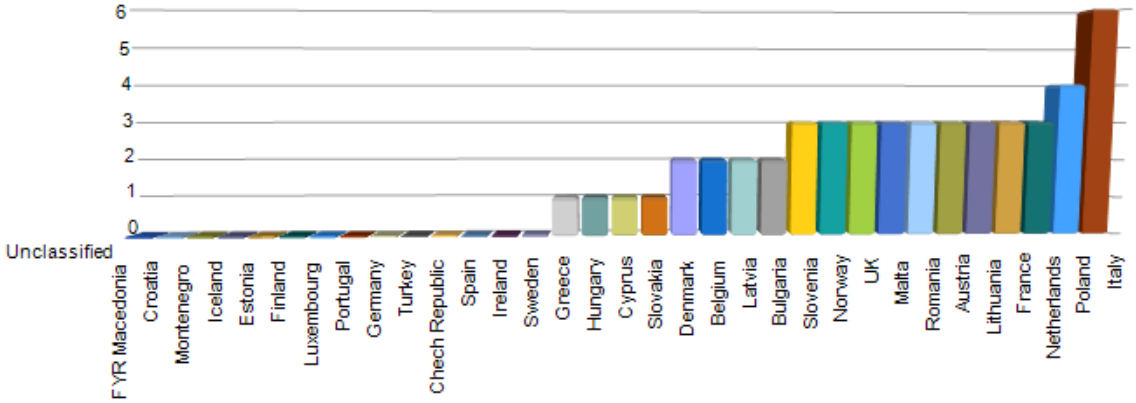
Distribution of policy initiatives (type = improve detection) per country.



Distribution of policy initiatives (type = increase penalties) per country.



Distribution of policy initiatives (type = unclassified) per country.



All types of policy initiatives per country using data from EUROFOUND (Retrieved on 10.07.2014 from http://www.eurofound.europa.eu).

	Commitment to tax morality	Awareness raising	Prevention	Legitimising undeclared work	Improve detection	Increase penalties	Unclassified
1: Austria	0	0	1	0	1	1	3
2: Belgium	0	0	2	2	5	0	2
3: Bulgaria	0	2	1	0	2	1	2
4: Czech Republic	0	1	3	0	3	1	0
5: Croatia	0	0	2	3	0	0	0
6: Cyprus	0	0	2	1	1	1	1
7: Denmark	0	1	1	3	0	0	2
8: France	0	0	3	0	0	2	3
9: FYR Macedonia	0	2	2	1	0	0	0
10: Germany	0	1	1	2	4	0	0
11: Greece	0	0	0	0	0	0	1
12: Hungary	0	0	1	1	1	1	1
13: Ireland	0	2	3	1	4	2	0
14: Luxembourg	0	0	0	0	3	0	0
15: Malta	0	0	0	0	0	1	3
16: Netherlands	0	0	0	0	3	2	3
17: Poland	0	0	5	3	2	0	4
18: Romania	0	1	0	0	1	1	3
19: Slovakia	0	0	2	1	2	2	1
20: Spain	0	0	2	3	1	2	0
21: Sweden	0	0	2	2	4	2	0
22: Turkey	0	0	2	2	2	1	0
23: Iceland	1	2	1	1	1	0	0
24: Finland	1	1	4	7	2	0	0
25: Montenegro	1	1	0	1	1	0	0
26: Norway	1	1	3	0	2	0	3
27: Lithuania	1	1	0	0	2	1	3
28: Portugal	1	2	0	2	3	0	0
29: UK	1	1	4	3	2	0	3
30: Latvia	1	2	2	1	0	1	2
31: Slovenia	1	1	2	2	0	0	3
32: Estonia	3	4	5	1	1	0	0
33: Italy	4	1	8	2	3	1	6
Total per type:	**16**	**27**	**64**	**45**	**56**	**23**	**49**

12. References.

1. Aberbach, J. D. and Christensen, T. (2007). The challenges of modernizing tax administration: putting customers first in coercive public organizations. *Public Policy and Administration, volume* 22 (issue: 2), 155-182. http://dx.doi.org/10.1177/0952076707071501
2. Alm, J., Kirchler, E. & Muehlbacher, S. (2012). Combining psychology and economics in the analysis of compliance: from enforcement to cooperation. *Economic Analysis and Policy, volume* 42 (issue: 2), 133-151. http://dx.doi.org/10.1016/S0313-5926(12)50016-0
3. Andrews, D., Sánchez, A. C. & Johansson, Å. (2011). Towards a Better Understanding of the Informal Economy. *OECD Economics Department Working Papers,* ECO/WKP(2011) *ECO/WKP(2011)* (42), Retrieved on 07.03.2013, from: http://dx.doi.org/10.1787/5kgb1mf88x28-en
4. Andvig, J. C. and Moene, K. O. (1990). How corruption may corrupt. *Journal of Economic Behavior & Organization, volume* 13 (issue: 1), 63-76. http://dx.doi.org/10.1016/0167-2681(90)90053-G
5. Azar, A. T. and El-Metwally, S. M. (2012). Decision tree classifiers for automated medical diagnosis. *Neural Computing and Applications, volume* 23 (issue: 7), 2387-2403. http://dx.doi.org/10.1007/s00521-012-1196-7
6. Barone, G. and Mocetti, S. (2011). Tax morale and public spending inefficiency. *International Tax and Public Finance, volume* 18 (issue: 6), 724-749. http://dx.doi.org/10.1007/s10797-011-9174-z
7. Batini, N., Levine, P. L., Kim, Y.-B. & Lotti, E. (2011). Informal Labour and Credit Markets: A Survey. IMF working Paper No. 2011-94. Retrieved on 11.05.14, from: http://www.nipfp.org.in/media/medialibrary/2013/04/wp_2011_94.pdf
8. Bayer, R.-C. and Sutter, M. (2009). The excess burden of tax evasion - an experimental detection–concealment contest. *European Economic Review, volume* 53 (issue: 5), 527-543. http://dx.doi.org/10.1016/j.euroecorev.2008.09.004
9. Benjamini, Y. and Maital, S. (1985). Optimal Tax Evasion & Optimal Tax Evasion Policy Behavioral Aspects. In: Gaertner W and Wenig A (eds) *The Economics of the Shadow Economy.* Berlin, Heidelberg: Springer.
10. Bhaskar, R. (1978). *A realist theory of science*, Hassocks, Sussex: Harvester Press.
11. Biggs, D., De Ville, B. & Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics, volume* 18 (issue: 1), 49-62. http://dx.doi.org/10.1080/02664769100000005
12. Bishop, C. M. (1995). *Neural networks for pattern recognition*, Oxford: Clarendon Press; Oxford University Press.
13. Bitzenis, A. and Nito, E. (2005). Obstacles to entrepreneurship in a transition business environment: The case of Albania. *Journal of Small Business and Enterprise Development, volume* 12 (issue: 4), 564-578. http://dx.doi.org/10.1108/14626000510628234
14. Braithwaite, V. (2007). Responsive regulation and taxation: introduction. *Law & Policy, volume* 29 (issue: 1), 3-10. http://dx.doi.org/10.1111/j.1467-9930.2007.00242.x
15. Braithwaite, V., Murphy, K. & Reinhart, M. (2007). Taxation threat, motivational postures and responsive regulation. *Law & Policy, volume* 29 (issue: 1), 137-158. http://dx.doi.org/10.1111/j.1467-9930.2007.00250.x
16. Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. (1984). *Classification and Regression Trees*, Taylor&Francis.

17. Bromley, R. (1978). Introduction - the urban informal sector: Why is it worth discussing? *WORLD DEVELOPMENT, volume* 6 (issue: 9–10), 1033-1039. http://dx.doi.org/10.1016/0305-750X(78)90061-X
18. Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative Research, volume* 6 (issue: 1), 97-113. http://dx.doi.org/10.1177/1468794106058877
19. Bryman, A. (2007). The Research Question in Social Research: What is its Role? *International Journal of Social Research Methodology, volume* 10 (issue: 1), 5-20. http://dx.doi.org/10.1080/13645570600655282
20. Buehn, A. and Schneider, F. (2011). Shadow economies around the world: novel insights, accepted knowledge, and new estimates. *International Tax and Public Finance, volume* 19 (issue: 1), 139-171. http://dx.doi.org/10.1007/s10797-011-9187-7
21. Bulgarian Industrial Capital Association. (2013). Комплексен аналитичен доклад за състоянието на неформалната икономика в България и извоци за нейното ограничаване и превенция. ISBN: 978-954-92331-9-3. Retrieved on 18.07.13, from: http://ikonomikanasvetlo.bg/c/document_library/get_file?uuid=5f2758d8-cd36-42ab-87c2-edc6ee92bf3e&groupId=55360
22. Bulgarian National Assembly. (2011). Law for access to public information. Retrieved on 07.10.14 from: http://lex.bg/laws/ldoc/2134929408
23. Bulgarian National Assembly. (2013). Law for National Revenue Agency. Retrieved on 25.09.14 from: http://www.lex.bg/bg/laws/ldoc/2135461320
24. Bulgarian National Audit Office. (2012). A performance audit of the control and enforcement actions undertaken by the National Revenue Agency to enforce compliance with the tax and social security legislation. No: 030000412 Retrieved on 25.09.14, from: http://www.bulnao.government.bg/bg/articles/download/7284/pk-nap-020414.doc.
25. Bulgarian National Audit Office. (2014). *Main task of the Bulgarian National Audit Office* Retrieved on 07.10.14, from: http://www.bulnao.government.bg/en/articles/about-us-1046
26. Cagan, P. (1958). The Demand for Currency Relative to the Total Money Supply. *Journal of Political Economy, volume* 66 (issue: 4), 303-328. http://dx.doi.org/10.1086/258056
27. Calciolari, S., Cristofoli, D. & Maccio, L. (2013). Explaining the reactions of Swiss municipalities to the 'amalgamation wave': at the crossroad of institutional, economic and political pressures. *Public Management Review, volume* 15 (issue: 4), 563-583. http://dx.doi.org/10.1080/14719037.2012.698852
28. Casson, M. C., Della Giusta, M. & Kambhampati, U. S. (2010). Formal and Informal Institutions and Development. *WORLD DEVELOPMENT, volume* 38 (issue: 2), 137-141. https://doi-org.sheffield.idm.oclc.org/10.1016/j.worlddev.2009.10.008
29. Cerase, F. P. and Farinella, D. (2009). Public Service Motivation: How Does it Relate to Management Reforms and Changes in the Working Situation of Public Organizations? A Case Study of the Italian Revenue Agency. *Public Policy and Administration, volume* 24 (issue: 3), 281-308. http://dx.doi.org/10.1177/0952076709103812
30. Chen, M. A. (2012). The Informal Economy: Definitions, Theories and Policies. WIEGO Working Paper No 1. Retrieved on 07.03.13, from: http://wiego.org/sites/wiego.org/files/publications/files/Chen_WIEGO_WP1.pdf
31. Cherkassky, V., Friedman, J. H. & Wechsler, H. (1994). *From statistics to neural networks: theory and pattern recognition applications*, Berlin: Springer-Verlag in cooperation with NATO Scientific Affairs Division.

32. Contini, B. (1981). Labor Market Segmentatation and the Development of the Parallel Economy - The Italian Experience. *Oxford Economic Papers, volume* 33 (issue: 3), 401-412. https://doi-org.sheffield.idm.oclc.org/10.1093/oxfordjournals.oep.a041515
33. Coussement, K., Van den Bossche, F. A. M. & De Bock, K. W. (2014). Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees. *Journal of Business Research, volume* 67 (issue: 1), 2751-2758. http://dx.doi.org/10.1016/j.jbusres.2012.09.024
34. Cowell, F. A. (1992). Tax evasion and inequity. *Journal of Economic Psychology, volume* 13 (issue: 4), 521-543. http://dx.doi.org/10.1016/0167-4870(92)90010-5
35. Çule, M. and Fulton, M. (2009). Business culture and tax evasion: Why corruption and the unofficial economy can persist. *Journal of Economic Behavior & Organization, volume* 72 (issue: 3), 811-822. http://doi.org/10.1016/j.jebo.2009.08.005
36. Cummings, R. G., Martinez-Vazquez, J., McKee, M. & Torgler, B. (2006). Effects of Tax Morale on Tax Compliance: Experimental and Survey Evidence. *Berkeley Program in Law and Economics, Working Paper Series,* 1-36. Retrieved on 14.06.2013, from: http://www.escholarship.org/uc/item/8sh2w9fp
37. Dabla-Norris, E., Gradstein, M. & Inchauste, G. (2008). What causes firms to hide output? The determinants of informality. *Journal of Development Economics, volume* 85 (issue: 1-2), 1-27. http://dx.doi.org/10.1016/j.jdeveco.2006.05.007
38. De Soto, H. (1989). *The other path: the economic answer to terrorism*, New York: BasicBooks.
39. Dečman, M., Stare, J. & Klun, M. (2010). E-government and cost-effectiveness: E-taxation in Slovenia. *Transylvanian Review of Administrative Sciences, volume* (issue: 31), 48-57.
40. Dell'Anno, R. (2003). Estimating the shadow economy in Italy: a structural equation approach. *Economics Working Papers,* Retrieved on 01.08.13, from: ftp://ftp.econ.au.dk/afn/wp/03/wp03_07.pdf
41. Dickerson, C. M. (2010). Informal-Sector Entrepreneurs, Development and Formal Law: A Functional Understanding of Business Law. *American Journal Of Comparative Law, volume* 59 (issue: 1), 179-226. http://dx.doi.org/10.5131/AJCL.2010.0011
42. Djankov, S., Lieberman, I., Mukherjee, J. & Nenova, T. 2003. In Belev B (Eds.), *The informal economy in the EU accession countries. Size, Scope, Trend and Challenges to the Process of EU Enlargement*. Sofia: CSD.
43. Dnevnik. (2014). NRA offers four new electronic services. *Dnevnik.* Sofia.
44. Dragolov, G., Ignácz, Z., Lorenz, J., Delhey, J. & Boehnke, K. (2013). An International Comparison of Social Cohesion. Retrieved on 04.08.13, from: http://www.bertelsmann-stiftung.de/cps/rde/xbcr/SID-E1F02C45-9D585A8E/bst_engl/xcms_bst_dms_38334_38335_2.pdf
45. Dunn, W. N. (2012). *Public policy analysis: an introduction*, (5th ed.). Boston: Pearson.
46. Dzhekova, R. and Williams, C. C. (2014). Tackling the Undeclared Economy in Bulgaria - a baseline report - working paper 1. Retrieved on 18.08.14, from: http://www.grey-project.group.shef.ac.uk/wp-content/uploads/2014/04/Tackling-the-Undeclared-Economy-in-Bulgaria-April-2014.pdf
47. Easterby-Smith, M., Thorpe, R. & Jackson, P. (2008). *Management research*, London: Sage.
48. Echambadi, R., Campbell, B. & Agarwal, R. (2006). Encouraging Best Practice in Quantitative Management Research: An Incomplete List of Opportunities. *Journal*

of Management Studies, volume 43 (issue: 8), 1801-1820. http://dx.doi.org/10.1111/j.1467-6486.2006.00660.x

49. ESRC. (2001). *ESRC Network for Evidence Based Policy and Practice (Evidence Network).* Retrieved on 16.05.13, from: http://www.sphsu.mrc.ac.uk/Evidence/About/EvidenceNet_MAIN.html
50. Eurobarometer. (2007). Undeclared Work in the European Union. Retrieved on from: http://ec.europa.eu/public_opinion/archives/ebs/ebs_284_en.pdf
51. Eurobarometer. (2014). Undeclared Work in the European Union. Retrieved on 09.05.14, from: http://ec.europa.eu/public_opinion/archives/ebs/ebs_402_en.pdf
52. Eurofound. (2013a). *Connecting fiscal devices to the National Revenue Agency - Bulgaria*. Retrieved on 03.08.13, from: https://eurofound.europa.eu/areas/labourmarket/tackling/cases/bg016.htm
53. Eurofound. (2013b). *Joint effective control on undeclared work - Bulgaria*. Retrieved on 03.08.13, from: https://eurofound.europa.eu/areas/labourmarket/tackling/cases/bg017.htm
54. Eurofund. (2009). *Mandatory registration of individual employment contracts - Bulgaria*. Retrieved on 03.08.13, from: https://eurofound.europa.eu/areas/labourmarket/tackling/cases/bg001.htm
55. European Commission. (2005). *The new SME definition - User guide and model declaration.* Retrieved on 10.07.12, from: http://ec.europa.eu/enterprise/policies/sme/files/sme_definition/sme_user_guide_en.pdf
56. European Commission. (2014). *Fact and figures about the EU's Small and Medium Enterprise (SME)*. Retrieved on 14.04.15, from: http://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition/index_en.htm
57. European Commission. (2016). *Small and medium-sized enterprises.* Retrieved on 21.09.16, from: http://ec.europa.eu/growth/smes
58. Eurostat. (2013). Basic figures on the EU - Summer 2013 edition. Retrieved on from: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-GL-13-002/EN/KS-GL-13-002-EN.PDF
59. Evans, M., Syrett, S. & Williams, C. (2006). Informal economic activities and deprived neighbourhoods. Retrieved on 07.03.13, from: http://www.eukn.org/E_library/Economy_Knowledge_Employment/Urban_Economy/Employment/Employment/Informal_economic_activities_and_deprived_neighbourhoods
60. Fallers, L. 1963. In Geertz C (Eds.), *Old societies and new states: the quest for modernity in Asia and Africa*. London: Collier-Macmillan.
61. Feige, E. L. (1979). How Big Is the Irregular Economy? *Challenge, volume* 22 (issue: 5), 5-13. http://dx.doi.org/10.1080/05775132.1979.11470559
62. Feige, E. L. (1986). A Re-Examination of the "Underground Economy" in the United States: A Comment on Tanzi. *Staff Papers - International Monetary Fund, volume* 33 (issue: 4), 768-781. doi: 10.2307/3867216
63. Feige, E. L. 1989. In (Eds.), *The Underground Economies: Tax Evasion and Information Distortion*. USA: University of Cambridge.
64. Feige, E. L. (1990). Defining and estimating underground and informal economies: The new institutional economics approach. *WORLD DEVELOPMENT, volume* 18 (issue: 7), 989-1002. http://dx.doi.org/10.1016/0305-750X(90)90081-8
65. Feige, E. L. 1996. In Pozo S (Eds.), *Exploring the underground economy: studies of illegal and unreported activity*. Kalamazoo: Mich.: W.E. Upjohn Institute for Employment Research.

66. Field, A. P. (2009). *Discovering statistics using SPSS*, (3rd ed.). London: SAGE.
67. Fleetwood, S. (1999). *Critical Realism in Economics: Development and Debate*, Taylor & Francis.
68. Frey, B. S. and Weck-Hanneman, H. (1984). The hidden economy as an 'unobserved' variable. *European Economic Review, volume* 26 (issue: 1–2), 33-53. http://dx.doi.org/10.1016/0014-2921(84)90020-5
69. Frey, B. S., Weck, H. & Pommerehne, W. W. (1982). Has the Shadow Economy Grown in Germany? An Exploratory Study. *Weltwirtschaftliches Archiv, volume* 118 (issue: 3), 499-524. http://dx.doi.org/10.1007/BF02706263
70. Fries, S., Lysenko, T. & Polanec, S. (2003). The 2002 Business Environment and Enterprise Performance Survey: Results from a survey of 6,100 firms. Working paper 84. Retrieved on 05.10.15, from: http://www.ebrd.com/downloads/research/economics/workingpapers/wp0084.pdf
71. Gallin, D. (2001). Propositions on trade unions and informal employment in times of globalisation. *antipode, volume* 33 (issue: 3), 531-549. http://dx.doi.org/10.1111/1467-8330.00197
72. General Labour Inspectorate. (2010). *Agreement No. 85*. Retrieved on 18.08.14, from: http://www.gli.government.bg/page.php?c=26
73. Gershuny, J. I. and Pahl, R. E. (1979). Work outside employment: some preliminary speculations. *Higher Education Quarterly, volume* 34 (issue: 1), 120-135. http://dx.doi.org/10.1111/j.1468-2273.1979.tb01252.x
74. Gërxhani, K. (2004a). The Informal Sector in Developed and Less Developed Countries: A Literature Survey. *Public Choice, volume* 120 (issue: 3-4), 267-300. http://dx.doi.org/10.1023/B:PUCH.0000044287.88147.5e
75. Gërxhani, K. (2004b). Tax evasion in transition: Outcome of an institutional clash? Testing Feige's conjecture in Albania. *European Economic Review, volume* 48 (issue: 4), 729-745. http://dx.doi.org/10.1016/j.euroecorev.2003.08.014
76. Gilbert, A. (1994). *The Latin American city*, London: London: Latin America Bureau.
77. Grilo, I. and Thurik, R. (2004). Determinants of Entrepreneurship in Europe. *Discussion Papers on Entrepreneurship, Growth and Public Policy,* Retrieved on 03.07.13, from: ftp://papers.econ.mpg.de/egp/discussionpapers/2004-30.pdf
78. Gutmann, P. M. (1977). The Subterranean Economy. *Financial Analysts Journal, volume* 33 (issue: 6), 26-34. doi: 10.2307/4478078
79. Hart, K. (1973). Informal Income Opportunities and Urban Employment in Ghana. *The Journal of Modern African Studies, volume* 11 (issue: 1), 61-89. https://doi.org/10.1017/S0022278X00008089
80. Hartner, M., Rechberger, S., Kirchler, E. & Schabmann, A. (2008). Procedural fairness and tax compliance. *Economic Analysis and Policy, volume* 38 (issue: 1), 137-152. http://dx.doi.org/10.1016/S0313-5926(08)50010-5
81. Hashimzade, N., Myles, G., Page, F. & Rablen, M. (2015). The use of agent-based modelling to investigate tax compliance. *Economics of Governance, volume* 16 (issue: 2), 143-164. http://dx.doi.org/10.1007/s10101-014-0151-8
82. Haykin, S. S. (2009). *Neural networks: a comprehensive foundation*, (3rd). Harlow: Prentice Hall.
83. Helmke, G. and Levitsky, S. (2004). Informal Institutions and Comparative Politics: A Research Agenda. *Perspectives on Politics, volume* 2 (issue: 4), 725-740. http://dx.doi.org/10.1017/S1537592704040472
84. Hemberg, E., Rosen, J., Warner, G., Wijesinghe, S. & O'Reilly, U.-M. (2016). Detecting tax evasion: a co-evolutionary approach. *Artificial Intelligence and Law, volume* 24 (issue: 2), 149-182. http://dx.doi.org/10.1007/s10506-016-9181-6

85. HMRC. (2012). Hidden Economy: Quantitative Feasibility Study. Retrieved on 02.10.13, from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/344826/report209.pdf
86. Hofmann, E., Hoelzl, E. & Kirchler, E. (2008). Preconditions of voluntary tax compliance: knowledge and evaluation of taxation, norms, fairness, and motivation to cooperate. *Zeitschrift für Psychologie, volume* 216 (issue: 4), 209-217. http://dx.doi.org/10.1027/0044-3409.216.4.209
87. Hokamp, S. (2014). Dynamics of tax evasion with back auditing, social norm updating, and public goods provision – an agent-based simulation. *Journal of Economic Psychology, volume* 40 (issue:), 187-199. https://doi.org/10.1016/j.joep.2013.01.006
88. Holzinger, A. (2015). Data Mining with Decision Trees: Theory and Applications. *Online Information Review, volume* 39 (issue: 3), 437-438. http://dx.doi.org/10.1108/OIR-04-2015-0121
89. Huang, P. C. C. (2009). China's Neglected Informal Economy Reality and Theory. *MODERN CHINA, volume* 35 (issue: 4), 405-438. http://dx.doi.org/10.1177/0097700409333158
90. IBM. (2010). Introduction to IBM SPSS Text Analytics for IBM SPSS Modeler (SP0A102). (0A102), Retrieved on 14.10.14, from: https://www-304.ibm.com/ibm/university/academic/pub/page/ban_predictive_analysis
91. IBM. (2014). Mastering new challenges in text analytics. Retrieved on 14.10.14, from: https://www.ibm.com/developerworks/community/files/form/anonymous/api/library/ef2b3bf0-86fe-4369-9c18-92bcddcac54b/document/cab990c3-f9ea-4b05-ba90-2811a6ee54cd/media/Mastering%20New%20Challenges%20in%20Text%20Analytics.pdf
92. IBM. (2015a). IBM SPSS Modeler 17 Algorithms Guide. Retrieved on 16.03.16, from: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/17.0/en/AlgorithmsGuide.pdf
93. IBM. (2015b). IBM SPSS Modeler 17 Modeling Nodes Guide. 1 *1* Retrieved on 16.03.16, from: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/17.0/en/ModelerModelingNodes.pdf
94. IBM. (2015c). IBM SPSS Neural Networks 23. Retrieved on 16.03.16, from: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/23.0/en/client/Manuals/IBM_SPSS_Neural_Network.pdf
95. IFC. (2013). World Bank Enterprise Survey - Survey Module 2013. Retrieved on 13.05.14, from: http://www.enterprisesurveys.org/~/media/GIAWB/EnterpriseSurveys/Documents/Methodology/ES_Services_Questionnaire.pdf
96. ILO and WTO. (2009). Globalization and Informal Jobs in Developing Countries. WTO ISBN: 978-92-870-3691-9. Retrieved on 11.03.13, from: http://www.wto.org/english/res_e/publications_e/jobs_devel_countries_e.htm
97. Institute for Market Economy. (2004). Study of Incentives, Characteristics and Strategies of Firms Operating in the Shadows. CT-2002-00131. Retrieved on from: http://www.ime.bg/pdf_docs/papers/Overall_Report_IME_final.doc
98. International Finance Corporation. (2014). Enterprise Surveys Indicator Descriptions. Retrieved on 13.05.14, from: http://www.enterprisesurveys.org/Data/ExploreEconomies/2013/~/media/GIAWB/

EnterpriseSurveys/Documents/Misc/Indicator-Descriptions.pdf

99. International Labour Office. (1972). Employment, incomes and equality: A strategy for increasing productive employment in Kenya. 72B09_608_eng. Retrieved on 08.01.13, from: https://search.proquest.com/docview/38218290?accountid=13828

100. International Labour Organization. (2012). Measuring informality: a Statistical Manual on the informal sector and informal employment. Retrieved on 06.02.13, from: http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms_182300.pdf

101. International Labour Organization. (2013). Measuring informality: a Statistical Manual on the informal sector and informal employment. Retrieved on 03.11.13, from: http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms_182300.pdf

102. Isachsen, A. J. and Strom, S. (1985). The size and growth of the hidden economy in Norway. *Review of Income & Wealth, volume* 31 (issue: 1), 21-38. http://dx.doi.org/10.1111/j.1475-4991.1985.tb00496.x

103. Jenkins, G. (1996). Information technology and innovation in tax administration. Retrieved on 17.04.15, from: http://EconPapers.repec.org/RePEc:qed:dpaper:123

104. Job, J., Stout, A. & Smith, R. (2007). Culture change in three taxation administrations: from command-and-control to responsive regulation. *Law & Policy, volume* 29 (issue: 1), 84-101. 10.1111/j.1467-9930.2007.00247.x

105. Johnson, P. and Duberley, J. (2003a). Reflexivity in Management Research. *Journal of Management Studies, volume* 40 (issue: 5), 1279-1303. 10.1111/1467-6486.00380

106. Johnson, P. and Duberley, J. (2003b). *Understanding management research: an introduction to epistemology*, London: SAGE, 2000.

107. Johnson, S., Kaufmann, D., McMillan, J. & Woodruff, C. (2000). Why do firms hide? Bribes and unofficial activity after communism. *Journal of Public Economics, volume* 76 (issue: 3), 495-520. http://dx.doi.org/10.1016/S0047-2727(99)00094-8

108. Johnson, S., Kaufmann, D., Shleifer, A., Goldman, M. I. & Weitzman, M. L. (1997). The Unofficial Economy in Transition. *Brookings Papers on Economic Activity, volume* 1997 (issue: 2), 159-239. http://dx.doi.org/10.2307/2534688

109. Jones, T., Ram, M. & Edwards, P. (2004). Illegal Immigrants and the Informal Economy: Worker and Employer Experiences in the Asian Underground Economy. *International Journal of Economic Development, volume* 6 (issue: 2), 98-119.

110. Jütting, J. R. and De Laiglesia, J. (2009). *Is informal normal? Towards more and better jobs in developing countries*, France: OECD Development Centre.

111. Kamleitner, B., Korunka, C. & Kirchler, E. (2012). Tax compliance of small business owners: A review. *International Journal of Entrepreneurial Behavior & Research, volume* 18 (issue: 3), 330-351. http://dx.doi.org/10.1108/13552551211227710

112. Kass, G. V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Journal of the Royal Statistical Society, volume* 29 (issue: 2), 119-127. http://dx.doi.org/10.2307/2986296

113. Kaufman, L. (1990). *Finding groups in data: an introduction to cluster analysis*, New York, Chichester: Wiley.

114. Kaufmann, D. and Kaliberda, A. (1996). Integrating the unofficial economy into the dynamics of post-socialist economies: a framework of analysis and evidence. WPS1691. Retrieved on 17.06.2013, from: http://dx.doi.org/10.1596/1813-9450-1691

115. Kazemier, B. (2014). Hidden workers and the hidden worker potential in the

Netherlands. *Economic Analysis and Policy, volume* 44 (issue: 1), 39-50. http://dx.doi.org/10.1016/j.eap.2014.01.005

116. Kazemier, B. and van Eck, R. (1992). Survey investigations of the hidden economy: Some methodological results. *Journal of Economic Psychology, volume* 13 (issue: 4), 569-587. http://dx.doi.org/10.1016/0167-4870(92)90012-V

117. Kirchler, E., Hoelzl, E. & Wahl, I. (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework. *Journal of Economic Psychology, volume* 29 (issue: 2), 210-225. http://dx.doi.org/10.1016/j.joep.2007.05.004

118. Koettl, J. and Weber, M. (2012). Does Formal Work Pay? The Role of Labor Taxation and Social Benefit Design in the New EU Member States. *Discussion Paper Series No. 6313,* Retrieved on 02.04.14, from: http://ftp.iza.org/dp6313.pdf

119. Kogler, C., Muehlbacher, S. & Kirchler, E. (2015). Testing the 'slippery slope framework' among self-employed taxpayers. *Economics of Governance, volume* 16 (issue: 2), 125-142. http://dx.doi.org/10.1007/s10101-015-0158-9

120. Krasniqi, A. A. (2007). Barriers to entrepreneurship and SME growth in transition: the case of Kosova. *Journal of Developmental Entrepreneurship, volume* 12 (issue: 01), 71-94. http://dx.doi.org/10.1142/s1084946707000563

121. Kudva, N. (2009). The everyday and the episodic: the spatial and political impacts of urban informality. *Environment and Planning A, volume* 41 (issue: 7), 1614-1628. http://dx.doi.org/10.1068/a41101

122. Lambsdorff, J. and Nell, M. (2006). Corruption - Where we stand and where to go. In: Kreutner M (ed) *The corruption monster: Ethik, Politik, und Korruption.* Vienna: Czernin.

123. LaRose, R. and Tsai, H.-y. S. (2014). Completion rates and non-response error in online surveys: Comparing sweepstakes and pre-paid cash incentives in studies of online behavior. *Computers in Human Behavior, volume* 34 (issue: 110-119. http://dx.doi.org/10.1016/j.chb.2014.01.017

124. Lewis, W. A. (1955). *The theory of economic growth*, London: Allen & Unwin.

125. Mead, D. C. and Morrisson, C. (2004). The informal sector elephant. *WORLD DEVELOPMENT, volume* 24 (issue: 10), 1611-1619. http://dx.doi.org/10.1016/0305-750X(96)00065-4

126. Miller, K. D. and Tsang, E. W. K. (2011). Testing management theories: critical realist philosophy and research methods. *Strategic Management Journal, volume* 32 (issue: 2), 139-158. http://dx.doi.org/10.1002/smj.868

127. Minard, C. (2009). Valuing entrepreneurship in the informal economy in Senegal. *Social Enterprise Journal, volume* 5 (issue: 3), 186-209. http://dx.doi.org/10.1108/17508610911004304

128. Mingers, J. (2006). A Critique of Statistical Modelling in Management Science from a Critical Realist Perspective: Its Role within Multimethodology. *The Journal of the Operational Research Society, volume* 57 (issue: 2), 202-219. http://dx.doi.org/10.1057/palgrave.jors.2601980

129. Mishkov, L. (2012). Business intelligence systems – employability and significance to management of small and medium size organisations. *Management School.* University of Sheffield: University of Sheffield.

130. Murphy, K. (2005). Regulating more effectively: the relationship between procedural justice, legitimacy, and tax non-compliance. *Journal of Law and Society, volume* 32 (issue: 4), 562-589. http://dx.doi.org/10.1111/j.1467-6478.2005.00338.x

131. Murphy, K. (2007). Procedural justice and the regulation of tax compliance behaviour: the moderating role of personal norms. Report number: 96/1. Retrieved on 19.01.17, from: http://icepp.gsu.edu/files/2015/03/ispwp0731.pdf

132. Murphy, K. (2008). Enforcing tax compliance: to punish or persuade? *Economic Analysis and Policy, volume* 38 (issue: 1), 113-135. http://dx.doi.org/10.1016/S0313-5926(08)50009-9

133. Murphy, K. (2016). Turning defiance into compliance with procedural justice: Understanding reactions to regulatory encounters through motivational posturing. *Regulation & Governance, volume* 10 (issue: 1), 93-109. http://dx.doi.org/10.1111/rego.12073

134. Nastav, B. and Bojnec, Š. (2007). Shadow Economy in Slovenia: The Labour Approach. *Managing Global Transitions, volume* 5 (issue: 2), 193-208.

135. North, D. C. (1990). *Institutions, institutional change and economic performance*, Cambridge: Cambridge University Press.

136. Noutcheva, G. and Bechev, D. (2008). The Successful Laggards: Bulgaria and Romania's Accession to the EU. *East European Politics and Societies, volume* 22 (issue: 1), 114-144. http://dx.doi.org/10.1177/0888325407311793

137. O'Mahoney, J., Vincent, S. & Edwards, P. (2014). *Studying organizations using critical realism: a practical guide*, Oxford: OUP Oxford.

138. O'Neill, D. M. (1983). *Growth of the underground economy, 1950-81: some evidence from the current population survey: a study*, Washington: U.S. G.P.O.

139. Oviedo, A. M. (2009). Economic Informality: Causes, Costs, and Policies. Retrieved on 21.07.13, from: http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2012/07/04/000386194_20120704032940/Rendered/PDF/705720ESW0P1070ACS0Econ0Informality.pdf

140. Palda, F. (2004). Mission implausible III: measuring the informal sector in a transition economy using macro methods. Retrieved on 13.04.17, from: http://econwpa.repec.org/eps/pe/papers/0404/0404002.pdf

141. Pandya, R. and Pandya, J. (2015). C5.0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning. *International Journal of Computer Applications, volume* 117 (issue: 16), 18-21. http://dx.doi.org/10.5120/20639-3318

142. Pashev, K. (2005). Tax Compliance of Small Business in Transition Economies: Lessons from Bulgaria. Retrieved on 10.05.17, from: http://EconPapers.repec.org/RePEc:ays:ispwps:paper0510

143. Peattie, L. (1987). An idea in good currency and how it grew: The informal sector. *WORLD DEVELOPMENT, volume* 15 (issue: 7), 851-860. http://dx.doi.org/10.1016/0305-750X(87)90038-6

144. Pedersen, S. (2003). *The shadow economy in Germany, Great Britain and Scandinavia: a measurement based on questionnaire surveys*, Copenhagen: Rockwool Foundation Research Unit.

145. Perry, G. E. and Maloney, W. F. 2007. In Arias O, Fajnzylber P, Maloney WF, et al. (Eds.), *Informality: Exit and Exclusion. World Bank Latin American and Caribbean Studies*. World Bank Group.

146. Petticrew, M. and Roberts, H. (2006). *Systematic reviews in the social sciences: a practical guide*, Oxford: Blackwell.

147. Pickhardt, M. and Seibold, G. (2014). Income tax evasion dynamics: Evidence from an agent-based econophysics model. *Journal of Economic Psychology, volume* 40 (issue:), 147-160. https://doi.org/10.1016/j.joep.2013.01.011

148. Presser, S. (2004). Methods for Testing and Evaluating Survey Questions. *Public Opinion Quarterly, volume* 68 (issue: 1), 109-130. http://dx.doi.org/10.1093/poq/nfh008

149. Putnins, T. J. and Sauka, A. (2014a). Measuring the shadow economy using

company managers. *Journal of Comparative Economics, volume* 43 (issue: 2), 471–490. http://dx.doi.org/10.1016/j.jce.2014.04.001

150. Putnins, T. J. and Sauka, A. (2014b). Shadow economy index for the Baltic countries - 2009-2013. Retrieved on 21.05.14, from: http://www.sseriga.edu/files/content/sseriga_baltic_shadow_economy_index_2009_2013_eng.pdf

151. Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning, volume* 1 (issue: 1), 81-106. http://dx.doi.org/10.1023/a:1022643204877

152. Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*, San Francisco: Morgan Kaufmann Publishers.

153. Rademaker, L., Grace, E. & Curda, S. (2012). Using Computer-assisted Qualitative Data Analysis Software (CAQDAS) to Re-examine Traditionally Analyzed Data: Expanding our Understanding of the Data and of Ourselves as Scholars. *The Qualitative Report, volume* 17 (issue: 22), 1-11.

154. Raskolnikov, A. (2009). Revealing choices: using taxpayer choice to target tax enforcement. *Columbia Law Review, volume* 109 (issue: 4), 689-754.

155. Riahi-Belkaoui, A. (2004). Relationship between tax compliance internationally and selected determinants of tax morale. *Journal of International Accounting, Auditing and Taxation, volume* 13 (issue: 2), 135-143. http://dx.doi.org/10.1016/j.intaccaudtax.2004.09.001

156. Ripley, B. D. (1996). *Pattern recognition and neural networks*, Cambridge: Cambridge University Press.

157. Rokach, L. (2016). Decision forest: Twenty years of research. *Information Fusion, volume* 27 (issue:), 111-125. http://dx.doi.org/10.1016/j.inffus.2015.06.005

158. Rokach, L. and Maimon, O. (2008). *Data Mining with Decision Trees: Theory and Applications*, Singapore: World Scientific Publishing Co Ltd.

159. Round, J., Williams, C. C. & Rodgers, P. (2008). Everyday tactics and spaces of power: the role of informal economies in post-Soviet Ukraine. *Social & Cultural Geography, volume* 9 (issue: 2), 171-185. http://dx.doi.org/10.1080/14649360701856110

160. Ryan, M. (2009). Making visible the coding process: Using qualitative data software in a post-structural study. *Issues in Educational Research, volume* 19 (issue: 2), 142-161.

161. Salzberg, S. (1994). Book Review: C4.5: Programs for Machine Learning by J. Ross Quinlan. New York: Morgan Kaufmann Publishers.

162. Sandmo, A. (2012). An evasive topic: theorizing about the hidden economy. *International Tax and Public Finance, volume* 19 (issue: 1), 5-24. http://dx.doi.org/10.1007/s10797-011-9185-9

163. Saunders, M., Lewis, P. & Thornhill, A. (2012). *Research methods for business students*, (6th ed.). Rotolito Bombarda, Italy: Prentice Hall.

164. Sayer, R. A. (2000). *Realism and Social Science*, London: SAGE Publications Ltd.

165. Schneider, F. (2005). Shadow economies around the world: what do we really know? *European Journal of Political Economy, volume* 21 (issue: 3), 598-642. http://dx.doi.org/10.1016/j.ejpoleco.2004.10.002

166. Schneider, F. (2012). Size and Development of the Shadow Economy of 31 European and 5 other OECD Countries from 2003 to 2012: Some New Facts. Retrieved on 17.06.2013, from: http://www.econ.jku.at/members/Schneider/files/publications/2012/ShadEcEurope31.pdf

167. Schneider, F. (2013). Size and Development of the Shadow Economy of 31 European and 5 other OECD Countries from 2003 to 2013 - A Further Decline. Retrieved on 28.11.14, from: http://www.econ.jku.at/members/Schneider/files/publications/2013/ShadEcEurope31_Jan2013.pdf

168. Schneider, F. and Enste, D. H. (2000). Shadow Economies: Size, Causes, and Consequences. *Journal of Economic Literature, volume* 38 (issue: 1), 77-114. http://dx.doi.org/10.1257/jel.38.1.77

169. Schneider, F. and Williams, C. C. (2013). *The Shadow Economy*, London, UK: The Institute of Economic Affairs.

170. Schonlau, M., Fricker, R. D., Jr. & Elliott, M. N. (2002). *Conducting Research Surveys via E-Mail and the Web*, Santa Monica, CA: RAND Corporation.

171. Siggel, E. (2010). The Indian informal sector: The impact of globalization and reform. *International Labour Review, volume* 149 (issue: 1), 93-105. http://dx.doi.org/10.1111/j.1564-913X.2010.00077.x

172. Sinkovics, R. R. and Alfoldi, E. A. (2012). *Facilitating the interaction between theory and data in qualitative research using CAQDAS*, Sage Publications, London.

173. Slemrod, J. and Weber, C. (2012). Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance, volume* 19 (issue: 1), 25-53. 10.1007/s10797-011-9181-0

174. Small Business Council. (2004). Small Businesses in the Informal Economy: making the transition to the formal economy. *Selected Works,* Retrieved on from: http://academia.edu/200717/Small_Businesses_in_the_informal_Economy

175. Snyder, K. A. (2004). Routes to the informal economy in New York's East Village: Crisis, economics, and identity. *Sociol. Perspect., volume* 47 (issue: 2), 215-240. http://dx.doi.org/10.1525/sop.2004.47.2.215

176. Stafford, T. F. and Turan, A. H. (2011). Online tax payment systems as an emergent aspect of governmental transformation. *European Journal of Information Systems, volume* 20 (issue: 3), 343-357. http://dx.doi.org/10.1057/ejis.2010.63

177. Tanzi, V. (1983). The Underground Economy in the United States: Annual Estimates, 1930-80. *Staff Papers - International Monetary Fund, volume* 30 (issue: 2), 283-305. doi: 10.2307/3867000

178. Tanzi, V. (1986). The Underground Economy in the United States: Reply to Comments by Feige, Thomas, and Zilberfarb. *Staff Papers - International Monetary Fund, volume* 33 (issue: 4), 799-811. doi: 10.2307/3867219

179. Teddlie, C. and Tashakkori, A. (2009). *Foundations of mixed methods research: integrating quantitative and qualitative techniques in the social and behavioral sciences*, London: SAGE.

180. Tedds, L. M. (2010). Keeping it off the books: an empirical investigation of firms that engage in tax evasion. *Applied Economics, volume* 42 (issue: 19), 2459-2473. http://dx.doi.org/10.1080/00036840701858141

181. Thai, M. T. T. and Turkina, E. 2013. In (Eds.), *Entrepreneurship in the Informal Economy: Models, Approaches and Prospects for Economic Development*. Routledge.

182. Thomas, J. (1999). Quantifying the Black Economy: `Measurement without Theory' Yet Again? *The Economic Journal, volume* 109 (issue: 456), F381-F389. http://dx.doi.org/10.1111/1468-0297.00441

183. Tonoyan, V., Strohmeyer, R., Habib, M. & Perlitz, M. (2010). Corruption and Entrepreneurship: How Formal and Informal Institutions Shape Small Firm Behavior in Transition and Mature Market Economies. *Entrepreneurship: Theory &*

Practice, volume 34 (issue: 5), 803-831. http://dx.doi.org/10.1111/j.1540-6520.2010.00394.x

184. Torgler, B. and Schneider, F. (2009). The impact of tax morale and institutional quality on the shadow economy. *Journal of Economic Psychology, volume* 30 (issue: 2), 228-245. http://dx.doi.org/10.1016/j.joep.2008.08.004

185. Tranfield, D., Denyer, D. & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management, volume* 14 (issue: 3), 220. http://dx.doi.org/10.1111/1467-8551.00375

186. Transparency International. (2013). *Global corruption barometer*. Retrieved on 13.04.15, from: http://www.transparency.org/gcb2013/country/?country=bulgaria

187. Tuck, P., Lamb, M. & Hoskin, K. (2011). Customers? The reconstruction of the 'taxpayer' in Inland Revenue discourse and practice. *Accounting and Business Research, volume* 41 (issue: 4), 357-374. http://dx.doi.org/10.1080/00014788.2011.566015

188. Tyler, T. R. (1988). What is Procedural Justice?: Criteria used by Citizens to Assess the Fairness of Legal Procedures. *Law & Society Review, volume* 22 (issue: 1), 103-135. http://dx.doi.org/10.2307/3053563

189. University of Sheffield. (2013). *Primo Central collection list*. Retrieved on 04.06.13, from: http://www.sheffield.ac.uk/library/resourcediscovery/primocentral

190. Valdez, M. E. and Richardson, J. (2013). Institutional Determinants of Macro-Level Entrepreneurship. *Entrepreneurship Theory and Practice, volume* 37 (issue: 5), 1149-1175. http://dx.doi.org/10.1111/etap.12000

191. Valenzuela, A. (2001). Day labourers as entrepreneurs? *Journal of Ethnic and Migration Studies, volume* 27 (issue: 2), 335-352. http://dx.doi.org/10.1080/13691830020041642

192. Van Dijke, M. and Verboon, P. (2010). Trust in authorities as a boundary condition to procedural fairness effects on tax compliance. *Journal of Economic Psychology, volume* 31 (issue: 1), 80-91. http://dx.doi.org/10.1016/j.joep.2009.10.005

193. Verboon, P. and van Dijke, M. (2007). A self-interest analysis of justice and tax compliance: How distributive justice moderates the effect of outcome favorability. *Journal of Economic Psychology, volume* 28 (issue: 6), 704-727. http://dx.doi.org/10.1016/j.joep.2007.09.004

194. Vihanto, M. (2000). Tax evasion in a transition from socialism to capitalism: the psychology of the social contract. *BOFIT discussion papers, volume* 6 (issue: 5-31.

195. Wallace, C. and Latcheva, R. (2006). Economic transformation outside the law: Corruption, trust in public institutions and the informal economy in transition countries of Central and Eastern Europe. *EUROPE-ASIA STUDIES, volume* 58 (issue: 1), 81-102. http://dx.doi.org/10.1080/09668130500401707

196. Waller, V. (2007). The challenge of institutional integrity in responsive regulation: field inspections by the Australian Taxation Office. *Law & Policy, volume* 29 (issue: 1), 67-83. http://dx.doi.org/10.1111/j.1467-9930.2007.00246.x

197. Warner, G., Wijesinghe, S., Marques, U., Badar, O., et al. (2015). Modeling tax evasion with genetic algorithms. *Economics of Governance, volume* 16 (issue: 2), 165-178. http://dx.doi.org/10.1007/s10101-014-0152-7

198. Webb, J. W., Bruton, G. D., Tihanyi, L. & Ireland, R. D. (2012). Research on entrepreneurship in the informal economy: Framing a research agenda. *Journal of Business Venturing, volume* 25 (issue: 5), 598-614. http://dx.doi.org/10.1016/j.jbusvent.2012.05.003

199. Wickham, M. and Woods, M. (2005). Reflecting on the Strategic Use of CAQDAS to Manage and Report on the Qualitative Research Process. *Qualitative Report, volume* 10 (issue: 4), 687-702.

200. Williams, C. (2001). Tackling the participation of the unemployed in paid informal work: a critical evaluation of the deterrence approach. *Environment and Planning C: Government and Policy, volume* 19 (issue: 5), 729-750.

201. Williams, C., Horodnic, I. & Windebank, J. (2015a). Explaining participation in the informal economy: An institutional incongruence perspective. *Int. Sociol., volume* 30 (issue: 3), 294-313. http://dx.doi.org/10.1177/0268580915578745

202. Williams, C. and Horodnic, I. A. (2015b). Explaining the prevalence of the informal economy in the Baltics: an institutional asymmetry perspective. *European Spatial Research and Policy, volume* 22 (issue: 2), pp. 127-145. http://dx.doi.org/10.1515/esrp-2015-0029

203. Williams, C. and Horodnic, I. A. (2016). Cross-country variations in the participation of small businesses in the informal economy: an institutional asymmetry explanation. *Journal of Small Business and Enterprise Development, volume* 23 (issue: 1), pp. 3-24. http://dx.doi.org/10.1108/JSBED-02-2015-0021

204. Williams, C., Nadin, S., Barbour, A. & Llanes, M. (2012a). *Enabling enterprise. Tackling the barriers to formalisation.*, Sheffield: Published by University of Sheffield and Community Links.

205. Williams, C. C. (2004). *Cash-In-Hand Work*, Basingstoke: Palgrave Macmillan.

206. Williams, C. C. (2006). Evaluating the magnitude of the shadow economy: a direct survey approach. *Journal of Economic Studies, volume* 33 (issue: 5), 369-385. http://dx.doi.org/10.1108/01443580610706591

207. Williams, C. C. (2007). Do Entrepreneurs Always Play by the Rulebook. *Management Online Review,* Retrieved on 03.07.13, from: http://www.morexpertise.com/download.php?id=61

208. Williams, C. C. (2008). Beyond ideal-type depictions of entrepreneurship: some lessons from the service sector in England. *Service Industries Journal, volume* 28 (issue: 7/8), 1041-1053. http://dx.doi.org/10.1080/02642060701846770

209. Williams, C. C. (2010a). Beyond the formal/informal jobs divide: evaluating the prevalence of hybrid 'under-declared' employment in south-eastern Europe. *The International Journal of Human Resource Management, volume* 21 (issue: 14), 2529-2546. http://dx.doi.org/10.1080/09585192.2010.523573

210. Williams, C. C. (2010b). Tackling undeclared work in southeast Europe: lessons from a 2007 Eurobarometer survey. *Southeast European and Black Sea Studies, volume* 10 (issue: 2), 123-145. http://dx.doi.org/10.1080/14683857.2010.486944

211. Williams, C. C. (2013). Evaluating cross-national variations in the extent and nature of informal employment in the European Union. *Industrial Relations Journal, volume* 44 (issue: 5-6), 479-494. http://dx.doi.org/10.1111/irj.12030

212. Williams, C. C. (2014a). *Confronting the Shadow Economy*, Cheltenham, UK: Edward Elgar Publishing.

213. Williams, C. C. (2014b). Out of the shadows: a classification of economies by the size and character of their informal sector. *Work, employment and society, volume* 28 (issue: 5), 735-753.

214. Williams, C. C. (2014c). Policy approaches towards undeclared work - a conceptual framework. Grey working paper 4. Retrieved on 18.08.14, from: http://www.grey-project.group.shef.ac.uk/wp-content/uploads/2014/06/WP4-Policy-Approaches-Towards-Undeclared-Work1.pdf

215. Williams, C. C. (2015a). Designing survey methods to evaluate the undeclared economy: a review of the options. Working paper 7. Retrieved on 29.09.15, from: http://dx.doi.org/10.13140/RG.2.1.3979.8885

216. Williams, C. C. (2015b). Tackling entrepreneurship in the informal sector: An overview of the policy options, approaches and measures. *Journal of Developmental Entrepreneurship, volume* 20 (issue: 01), http://dx.doi.org/10.1142/S1084946715500065

217. Williams, C. C. (2016). Developing a holistic approach for tackling undeclared work. Retrieved on 19.01.17, from: https://www.researchgate.net/publication/311354713_Developing_a_Holistic_Approach_for_Tackling_Undeclared_Work_background_paper

218. Williams, C. C., Dzhekova, R., Baric, M., Franic, J. & Mishkov, L. (2014a). Assessing the cross-national transferability of policy measures for tackling undeclared work. Retrieved on 01.10.14, from: http://grey-project.group.shef.ac.uk/wp-content/uploads/2014/09/WP5-Assessing-the-cross-national-transferability-of-policy-measures-for-tackling-undeclared-work.pdf

219. Williams, C. C., Franic, J. & Dzhekova, R. (2015c). Explaining the undeclared economy in Bulgaria: an institutional asymmetry perspective. *South East European Journal of Economics and Business, volume* 9 (issue: 2), 33-45. http://dx.doi.org/10.2478/jeb-2014-0008

220. Williams, C. C. and Horodnic, I. A. (2015d). Evaluating the prevalence of the undeclared economy in Central and Eastern Europe: An institutional asymmetry perspective. *European Journal of Industrial Relations, volume* 21 (issue: 4), http://dx.doi.org/10.1177/0143831x14568835

221. Williams, C. C. and Martinez, A. (2014b). Do small business start-ups test-trade in the informal economy? Evidence from a UK survey. *International Journal of Entrepreneurship and Small Business, volume* 22 (issue: 1), 1-16. http://dx.doi.org/10.1504/IJESB.2014.062127

222. Williams, C. C. and Nadin, S. (2010a). The commonality and character of off-the-books entrepreneurship: A comparison of deprived and affluent urban neighbourhoods. *Journal of Developmental Entrepreneurship, volume* 15 (issue: 03), 345-358. http://dx.doi.org/10.1142/S1084946710001592

223. Williams, C. C. and Nadin, S. (2010b). Entrepreneurship and the Informal Economy: An Overview. *Journal of Developmental Entrepreneurship, volume* 15 (issue: 04), 361-378. http://dx.doi.org/10.1142/s1084946710001683

224. Williams, C. C., Nadin, S., Newton, S., Rodgers, P. & Windebank, J. (2011). Explaining off-the-books entrepreneurship: a critical evaluation of competing perspectives. *International Entrepreneurship and Management Journal, volume* 9 (issue: 3), 447–463. http://dx.doi.org/10.1007/s11365-011-0185-0

225. Williams, C. C., Nadin, S. & Rodgers, P. (2012b). Evaluating competing theories of informal entrepreneurship: some lessons from Ukraine. *International Journal of Entrepreneurial Behaviour & Research, volume* 18 (issue: 5), 528-543. http://dx.doi.org/10.1108/13552551211253919

226. Williams, C. C. and Round, J. (2007a). Re-thinking the Nature of the Informal Economy: Some Lessons from Ukraine. *International Journal of Urban and Regional Research, volume* 31 (issue: 2), 425-441. http://dx.doi.org/10.1111/j.1468-2427.2007.00730.x

227. Williams, C. C. and Round, J. (2009). Evaluating informal entrepreneurs' motives: evidence from Moscow. *International Journal of Entrepreneurial Behaviour & Research, volume* 15 (issue: 1), 94-107. http://dx.doi.org/10.1108/13552550910934477

228. Williams, C. C., Round, J. & Rodgers, P. (2007b). Beyond the formal/informal

economy binary hierarchy. *International Journal of Social Economics, volume* 34 (issue: 6), 402-414. http://dx.doi.org/10.1108/03068290710751812

229. Williams, C. C. and Shahid, M. S. (2014c). Informal entrepreneurship and institutional theory - explaining the varying degrees of (in)formalization of entrepreneurs in Pakistan. *Entrepreneurship & Regional Development, volume* 28 (issue: 1-2), http://dx.doi.org/10.1080/08985626.2014.963889

230. Williams, C. C. and Windebank, J. (1998). *Informal employment in the advanced economies: implications for work and welfare*, London: Routledge.

231. Williams, C. C. and Windebank, J. (2002). The Uneven Geographies of Informal Economic Activities: a Case Study of Two British Cities. *Work, Employment & Society, volume* 16 (issue: 2), 231-250.

232. Williams, N. and Vorley, T. (2014d). Institutional asymmetry: How formal and informal institutions affect entrepreneurship in Bulgaria. *International Small Business Journal, volume* 16 (issue: 2), 1-22. http://dx.doi.org/10.1177/0266242614534280

233. Windebank, J. and Williams, C. (1997). What is to be done about the paid informal sector in the European union? A review of some policy options. *International Planning Studies, volume* 2 (issue: 3), 315-327. http://dx.doi.org/10.1080/13563479708721687

234. World Bank and IFC. (2015a). *Distance to Frontier - DTF*. Retrieved on 11.04.15, from: http://www.doingbusiness.org/data/distance-to-frontier

235. World Bank and IFC. (2015b). *Doing Business 2015 - historical data and trends for Bulgaria*. Retrieved on 11.04.15, from: http://www.doingbusiness.org/Custom-Query/bulgaria#Topics

236. World Economic Forum. (2015). The Global Competitiveness Report 2014 - 2015. Retrieved on 13.04.15, from: http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2014-15.pdf

237. Xheneti, M. (2017). Contexts of enterprise policy-making – an institutional perspective. *Entrepreneurship & Regional Development, volume* 29 (issue: 3-4), 317-339. http://dx.doi.org/10.1080/08985626.2016.1271021

238. Xheneti, M. and Bartlett, W. (2012). Institutional constraints and SME growth in post-communist Albania. *Journal of Small Business and Enterprise Development, volume* 19 (issue: 4), 607-626. http://dx.doi.org/10.1108/14626001211277424

239. Xheneti, M. and Kitching, J. (2011). From Discourse to Implementation: Enterprise Policy Development in Postcommunist Albania. *Environment and Planning C: Government and Policy, volume* 29 (issue: 6), 1018-1036. http://dx.doi.org/10.1068/c10193b

240. Xheneti, M., Smallbone, D. & Welter, F. (2013). EU enlargement effects on cross-border informal entrepreneurial activities. *European Urban and Regional Studies, volume* 20 (issue: 3), 314-328. http://dx.doi.org/10.1177/0969776411434849

241. Yalamov, T. (2012). Hiding, Circumvention, Public Procurement, and Shaping Laws. *Eastern European Economics, volume* 50 (issue: 5), 93-111. http://dx.doi.org/10.2753/EEE0012-8775500504