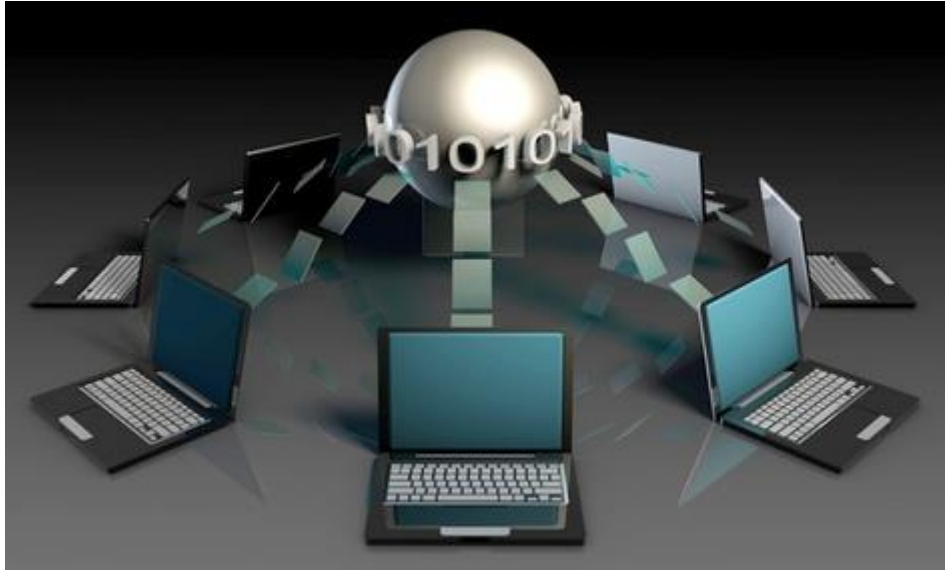# INTELLIGENT TAX FRAUD DETECTION

Bachelor Thesis Economie & Informatica
Tilburg University

Student:
Iris Schonenberg
ANR 684736
I.N.Schonenberg@uvt.nl

Supervisor:
Dr. H. Weigand

Abstract

Tax fraud is a problem every country in this world has to deal with and is responsible for losses of up to billions of dollars for some parties involved. Therefore, it is of great importance that tax fraud is reduced to a minimum and because of new technological developments, there are new ways to do so. This paper discusses both preventive (RTVAT, VLN and D-VAT) and detective techniques (data mining) to prevent and detect tax fraud. Furthermore, a case study regarding import/export taxes is presented.

Keywords: Tax fraud, Preventive Data Mining, Detective Data Mining

TABLE OF CONTENTS

1 Introduction

1.1 Introduction

The aim of this study is to investigate what technological methods can be used to detect tax fraud. Big data and data mining are two concepts that are widely discussed in various disciplines. Even though not much is known about data mining in relation to tax fraud, this is no reason to assume data mining is not applicable to the field of tax fraud. Technology is changing very rapidly and tax fraud should not have to be detected manually anymore. Therefore, this paper will focus on the automation of fraud detection and the various possibilities to do so.

This chapter will start with a problem description, followed by a problem statement and research questions. Each of these research questions will be the main topic of the chapters of this paper and are used to answer the problem statement.

1.2 Problem description

Tax fraud has been a problem throughout the ages. In 1947, Lewis D. Spencer wrote one of the earliest papers focussing on the proof of income tax fraud. Tax fraud is considered a felony and those who don't obey will be prone to receive legal penalties. However, penalties can only be given if the fraud is detected by the government and the fraudster has been prosecuted.

Unfortunately, tax fraud is a problem of huge proportion. It impacts the government, companies, investors, employees and all others involved and is known to add up to billions of dollars in some cases. More information about the size and impact of tax fraud can be found later on in this paper.

With the rise of the internet, mobile apps and the reducing costs of obtaining technology, a whole new spectrum of possibilities is open to be explored. Tax fraud no longer needs to be detected by manual labour, as technology is often capable of doing a better and faster job in selecting companies that might be guilty of committing tax fraud.

This thesis will focus on the various options that can be explored for fraud detection. Before doing so, it is important to have a sound understanding of what tax fraud exactly is and what forms of tax fraud exist.

1.3 Problem statement and research questions

The problem introduced in the former section will be formulated according to the following problem statement:

"What methods can be used to detect tax fraud within corporations? "

To answer this problem statement, the following research questions have to be answered:

1. What is tax fraud?
2. Which forms of tax fraud are there?
3. What are the consequences of tax fraud?
4. What can be done to reduce tax fraud?
5. Which technological solutions are available regarding tax fraud?

1.4   Scope

This paper discusses tax fraud within companies. This means it will discuss what can be done by companies and/or governments to prevent or decrease tax fraud. Regarding what can be done to reduce the amount of tax fraud committed by companies, this paper will describe possible solutions, technical and non-technical, and discuss possibilities for governments and organisations to reduce the amount of fraud committed.

1.5   Structure of the thesis

The problem stated in this chapter will be answered in the following manner. First, in the second chapter of this thesis, I will elaborate on what tax fraud is. The different forms of tax fraud will be discussed in chapter three. Then, the consequences of tax fraud are discussed for the individuals, the government and companies in chapter four. Chapter five will elaborate on the measures that could be implemented to decrease the amount of tax fraud. It also discusses some of the pros and cons of those measures. Chapter six will go deeper into some of the technical solutions and describe an implementation of one of those solutions. In chapter seven, a case study will be presented, which will look into the imports/exports of Hong Kong and China and how these can be monitored. The final chapter will discuss the findings and give some recommendation for further research.

2   Tax fraud

A previous study (Spencer, 1946) mentions the fine line between tax fraud, also known as tax evasion, and tax avoidance. Tax avoidance is meant when people try to keep their payable taxes at a minimum by planning their transactions, or in fact finding the loop holes in the law (Fisman & Wei, 2001). Tax evasion can be defined as a situation where people trespass the law and commit the felony of not paying their taxes within due able time. According to Spencer, tax fraud can be either intentionally or unknowingly. As people might be unaware of their actions and consequences, it is possible they get caught in a web and have trouble extracting themselves by explaining their unfortunate actions of tax evasion. This paper will focus on purposeful tax evasion, the consequences and solutions of this problem.

Sandmo (1981) concluded that the nature of tax fraud has been changing and that in some markets tax fraud is the rule rather than the exception. He tries to grasp the size of the 'hidden' economy, previously referred to as the 'shadow economy', and admits this as being a difficult task. The hidden economy probably holds around five to twenty-five percent of the Gross National Product (GNP). Even though the size is still uncertain, Sandmo (1981) stresses that tax fraud is an important problem and should not be overlooked.

Another study suggests that tax fraud only occurs if the tax rates are too high (Virmani, 1989). It provides evidence that as long as the tax rates remain below a certain tax threshold, a maximum that should be researched, firms are willing to pay their taxes. The possibility of tax evasion is significantly reduced, to such an extent that it could be ignored. Virmani (1989) also concludes that the amount of small firms in an economy increases the amount of tax evasion. According to this, an economy that holds more small businesses will be marked by a greater amount of tax fraud compared to an economy that consists of mostly large corporations.

The amount of companies that execute tax fraud is dependable on the penalty for tax fraud. The penalty should be dependable on the worth of concealed output. This can be done by making the sentence a percentage of the amount concealed. A high percentage of the concealed amount will lead to a decrease in firms committing tax fraud.

A Dutch study showed that non-evaders have a stronger belief that tax evasion behaviour will lead to apprehension and penalties than evaders (Weigel, Hessing, & Elffers, 1987). Weigel, Hessing and Elffers (1987) also found that tax evaders are likely to recommit fraud as they judge the risk of punishment being minimal. More so, tax evaders are less likely to see tax evasion as being wrong when compared to non-evaders.

While doing research about the Russian black market, Andrei Yakovlev used various models to vision tax evasion (2001). One of these models is a traditional western scheme, as is presented in figure 1. The model is based on the assumption that tax evasion occurs as tax-based tax evasion and often contain five main characteristics:

> ➢ The taxpayer gets a part of his/her receipts in cash.
>
> ➢ The model is typical for self-emloyed people or small, non-corporate businesses. The tax payer mentioned in this model is often an individual and thus not a corporation.
>
> ➢ A portion of the taxpayer's sales or work is illegal.
>
> ➢ Often the underreporting of revenue occurs.
>
> ➢ All transaction made are real.

As visualized, the tax evasion occurs when part of the operations are not- or underreported and can occur at both the business and personal level.
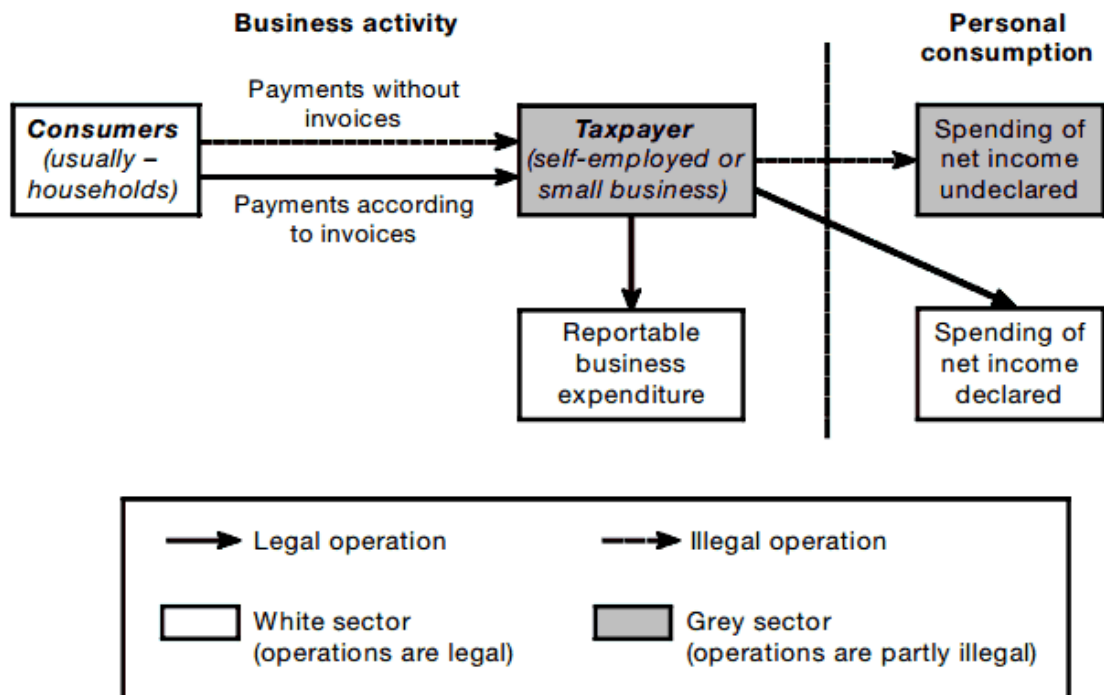
4

Figure 1. *The traditional western scheme of tax evasion.*

3 Different forms of tax fraud

3.1 Shadow economy tax evasion

According to Peterson, Thießen and Wohlleben (2010) the economy can be divided into three areas: the market economy, the shadow economy and the subsistence economy or household production. The subsistence economy is the primary source of human productive activities. The market economy is the economy people refer to and the shadow economy is everything in between the subsistence economy and the market economy. Unfortunately up on to this day the real size of the shadow economy is unknown. To fulfil their basic needs, people will be part of one or multiple of these areas.

Petersen, Thießen and Wohlleben (2010) believe tax evasion, a form of tax fraud, is most commonly found in the shadow economy. Tax fraud in the shadow economy continues to exist as it is often more rewarding than participating in the market economy or because this may give access to basic supplies that otherwise wouldn't be available. Another explanation to people's involvement in the shadow economy is because they do not trust their government and thus use this particular field to gain the ability to be independent and self-sufficient. They believe the burden of direct taxes and indirect taxes have created the incentive for people to find their way into the shadow economy.

3.2 Missing trader fraud (Value Added Tax (VAT) fraud)

It is said that Value Added Tax (VAT) is a very effective way of gaining tax revenues (Keen & Smith, 2007). Keen and Smith also mention VAT, like any other form of tax, is possibly being the subject of fraud or evasion. Nevertheless, VAT fraud or evasion is less likely to occur than any other form of tax fraud or evasion. Other researchers point out that VAT fraud gains twenty-five percent of total tax revenues and is therefore a major part of the revenues a country or state receives (Castellón González & Velásquez, 2013).

Ainsworth (2010) specifies the problem for VAT fraud in his research about Value Added Taxes allowing missing trader fraud. Missing trader fraud is most common with goods that have a high value and low volume, like technology, but can occur within both goods and services (Ainsworth, 2010). There are three types of taxable supplies: goods, tradable services and consumed services. This paper will only discuss two: goods and tradable services, as those are the only two that are susceptible to missing trader fraud.

Missing trader fraud is often referred to as missing trader intra-community goods as the fraud starts with a business-to-business transaction. Missing trader fraud happens when one company buys products or services without paying VAT, then sells the product to another buyer and subsequently collects the VAT. At last the tax 'disappears', for instance across the border, without paying the tax they are supposed to and being untraceable for the tax administer. Sometimes this kind of fraud is also referred to as carousel fraud. This happens when the same goods are used over and over again for missing trader fraud. Selling goods across borders multiple times can make one box of goods look like several percent's of the world supply just as what happened with a twenty-one year old selling one box of computer chips in UK-Irish trade (Ainsworth, 2010).

3.3 Insufficient tax payments

It is possible that an insufficient amount of tax is paid by shops, corporations or businesses. Another research states a variety of possible causes for this phenomenal (Keen & Smith, 2007).

➢ For example, a trader claims he sold less than he actually did, or fails to register the sales he made entirely. If this happens, there may or may not be invoices issued or there might not be an invoice at all which makes it hard to check the amount of goods sold. If the product is in its last stage and ends up at the consumer, the chances of VAT being paid is decreased to a minimum.

➢ Products can be misclassified regarding the percentage of applicable tax. Expensive products may be classified as having a low tax rate instead of a high tax rate and cheaper products may be allocated as high tax rates in return. This way, the net payable taxes can be strongly reduced.

> ➢ The owners or employees of an organisation may use its products or services for themselves. These products and services are taxable and therefore taxes should be paid either by the company of the owners/employees.

> ➢ Import taxes can be evaded. This leads to the opportunity of selling imported goods in the home market.

Another study suggests three forms of insufficient tax payment (Fisman & Wei, 2001), which partially correlate with those mentioned above. Namely the underreporting of the value of the units sold, the underreporting of the amount of taxable units and the reporting of units as having a lower taxable rate than determined by the law. It has proven difficult to detect underreporting of goods, where detecting mislabelling has a greater chance of succeeding as it leaves strong evidence. That is, if it has not been controlled yet, otherwise it is impossible to be found.

3.4   Indirect tax evasion

Even though most papers focus on direct tax evasion, indirect tax evasion is equally likely to occur (Virmani, 1989). Virmani (1989) focuses on the link between tax evasion and efficiency, the possible tax threshold for tax evasion and the characteristics of the industry of interest and investigates what is called 'indirect tax evasion'.

Indirect tax evasion may occur in situations where tax evasion is also possible. Even though people always thought direct tax evasion led to a higher efficiency, the opposite is true according to Virmani (1989). Direct tax evasion can lead to inefficient production, as a company's recourses are used to evade tax instead of used for running its operation. This leads to resource costs for the economy and therefore has negative effects on the economy that are not definable as direct taxes. Recourse costs are here defined as the costs that occur when concealing output, so this output will not be taxable. As the amount of concealed output increases, the amount of resource costs will also increase.

One of the believers of tax evasion leading to efficiency gains is Sandmo (1981). His research finds that tax evasion leads to a low effective tax rate for the elastic part of the labour supply. He argues tax authorities are reluctant to enforce tax laws in economies where they believe efficiency losses could be huge when no tax evasion would exist.

4 Consequences of tax fraud

4.1 Personal consequences

Those who commit tax fraud and get caught will logically experience negative consequences. Those consequences consist of legal sanctions and most likely social disapproval from people (Weigel et al., 1987). Even though gaining disapproval may seem like an unimportant aspect, this can have a significant impact when the person expressing this is of some significance to the fraudster. However, in case the fraudster has a reference group that is equally involved in tax fraud, it is likely that fraudulent behaviour will be encouraged.

If a company is associated with fraud, this may lead to stockholder losses, image problems or even bankruptcy (Whiting, Hansen, McDonald, Albrecht, & Albrecht, 2012). For those associated with the firm reduced income, a smaller amount of work, or no work at all may be one of the consequences.

4.2 Governmental consequences

Governments do not only bare the losses of those taxes that are unreported. They also pay tremendously for trying to discover the presence of tax evaders (Wu, Ou, Lin, Chang, & Yen, 2012). An absence in the ability to detect tax evasion leads to a government's inability to properly do public investments.

In 2009, it was estimated that Missing Trader fraud alone held a size of at least twenty-eight billion dollar in Europe (Ainsworth, 2010). The United Stated (U.S.) Internal Revenues Service (IRS) published some of its findings from careful and comprehensive estimates of the U.S. federal tax evasion in 2001 (Slemrod, 2007). This research defines tax evasion as every case in which a person unlawfully pays less tax than legally owed. The IRS found that he overall tax gap was around 345 billion dollar, which is 16.3 percent of the tax that should have been collected that year. Fortunately, the IRS expects to collect 55 billion dollars of taxes owed, which will result in a loss of 290 billion dollars for the U.S. in 2001. More specific numbers about the tax gap of the U.S. can be found in figure 2.

A comparison has been made between the tax gap of the U.S. and other high income countries. This research results in the conclusion that Sweden had a tax gap of 8 percent of total taxes in the year 2000 and the tax gap of the United Kingdom is estimated to have a similar size as the gap of Sweden and the U.S.

4.3 Consequences for companies

Other studies found that while companies are trying to avoid taxes, they are wasting valuable resources that could be used for their operations (Wu et al., 2012). This means that companies are more willing to spend their resources on avoiding taxations than to spend those same resources on creating more product and thus more profit.

Another negative result of fraud, financial fraud in this case, is that it decreases investor's confidence in a company's financial statement credibility (Gupta & Gill, 2012). This will then lead to a decrease in the amount of investments made in a company, a lower equity, less recourses and less budget for operations and growth. Stock markets can be used as a measurement for the impact of fraud, as the stock value of companies that have committed fraud is known to drop after such an incident (Whiting et al., 2012). Maybe more surprising is the fact that, when a traumatic event like the one on September 11, 2001 happens, companies that are associated with fraud have a sharper decline in the value of their stock than those who don't. Even to the point where some of those stocks become completely worthless. Eventually, fraudulent behaviour by a single company could affect the worldwide stock market.

8

	Tax gap ($billion)	Percentage of amount that should have been reported/filed/ paid/collected
Gross tax gap	345	16.3
Underreporting	**285**	**13***
Individual Income Tax	197	18
Underreported nonbusiness income	56	4
Wages and salaries	10	1
Net capital gains	11	12
Taxable pension annuities, IRA distributions	4	4
Taxable interest and dividends	3	4
Other	28	n.a.
Underreported business income	109	43
Nonfarm proprietor income	68	57
Partnership, S corporation, estate, and net trust income	22	18
Rent and royalty net income	13	51
Farm net income	6	72
Overreported Offsets to Income	15	4
Deductions	14	5
Exemptions	4	5
Other adjustments	−3	−21
Overreported Credits	17	26
Employment Tax	54	7
Self-employment tax	39	52*
FICA and unemployment taxes	15	2*
Corporation Income Tax	30	17*
Large (>$10 million assets) corporations	25	14*
Small (<$10 million assets) corporations	5	29*
Estate and Excise Taxes	4	4*
Nonfiling	**27**	**1***
Individual income tax	25	2*
Other	2	2*
Underpayment	**34**	**2***
Individual income tax	23	2*
Corporation income tax	2	1*
Other	9	1*
Enforced and other late payments	55	3*
Net tax gap (tax not collected)	290	13.7*

Figure 2.   Components of the U.S. federal tax gap.

5 Reducing tax fraud

5.1 Government measures

Slemrod (2007) stated that no government can introduce a tax system and then rely on taxpayers' honesty to pay what they owe the government. With that statement he formulated the problem the government faces in regard to taxation. It is the responsibility of the government to make sure taxes are collected and studies have suggested several methods to reduce the amount of tax evaders.

One study suggests that reducing the marginal tax rate would lead to a decrease in tax evasion, as a lot of fraudsters evade because of the high tax rates (Sandmo, 1981). This however, would discorporate one of the main advantages of having a tax system, namely the redistribution of income. Sandmo (1981) also discusses another option: an increase in the number of tax audits being performed and enforcing higher penalties for those who breach the law could, according to him, lead to a decrease in fraudulent behaviour.

Other research found that introducing a higher fine for tax evaders and increasing audits could eventually lead to a lower employment rate (Goerke, 2003). Goerke (2003) explains that increasing the fine for tax evasion could lead to a union's decision to demand higher wages. As employees are obligated to pay tax and their chance of getting caught as well as the height of their potential fine increases, they feel a higher need to support themselves without committing tax fraud. In their eyes, this can only be done by receiving higher wages. Having to pay people more for the same amount of work done, will lead to a company having to decrease the amount of employees, as more money is not available for salaries. Therefore, the amount of jobs available in the market will drop as higher penalties and probability of detection are introduced. A government's role in such a decision is to carefully weigh and evaluate both options. Not being able to collect the full amount of tax owed is not optimal, but having a higher unemployment rate could have even more negative consequences (for example: support from the government).

Sandmo (1981) also discusses the two fields regarding those options. Decreasing the tax rate is often studied as 'the optimum tax rate', where increasing audits and penalties belong to the literature of 'economics of crime and law enforcement'. However, he continues by stating that in order to make progress these two fields need to be brought together.

5.2 Company measures

As seen before, not only the government but also a company is better of when people obey tax laws. Hence, it is in the company's and its stakeholders own benefits to provide measures so as to decrease fraud. One of those measures could be a company auditing its financial statements during the year. Research has proven that fraud concealment measures differ significantly the first three quarters from the last quarter of the year (Whiting et al., 2012). As audit systematically is known to happen after the last quarter has taken place, fraud is easier detected during one of the first three quarters, when one does not expect to be audited and does a less thorough job of concealing the fraud.

6   Technical solutions

The bigger the proportion of output concealed, the higher the chance of a firm getting caught on tax fraud (Virmani, 1989). The following chapter will discuss the various technical methods of detecting or preventing tax fraud within firms.

6.1   Missing trader fraud solutions

As mentioned before, there exist two types of VAT fraud. The first is fraud being committed with goods and the second is fraud being committed with services (Ainsworth, 2010). Detecting missing trader fraud in services is most difficult, as the item disappears when used. A transaction and the accompanying theft of VAT can occur in only a few minutes and the funds used for the fraud will probably pass several international bank accounts. The return on which the trade is reported however could be due in over several months. As a result, the moment cash is withdrawn somewhere in the world, the fraud is no longer discoverable.

Ainsworth (2010) states that as missing trader fraud is made possible through the intensive use of technology, the most logical way to solve this fraud is by using technology. He offered three different technical options to solve this type of fraud are offered.

6.1.1   Real Time VAT (RTVAT)

The essence of Real Time VAT (RTVAT) is to move the moment of taxation from the date of invoice to the date of settlement, making it a settlement system. The RTVAT is a system that mandates any wireless or debit cards transfer of tax directly to the authorities. The introduction of RTVAT will result in a direct payment system and makes sure that a seller will never hold the VAT paid by the buyer. The system is designed to withhold VAT from the buyer's payments before the payment reaches the seller. The amount of VAT will be forwarded to the tax authorities real-time.

A disadvantage of the RTVAT system is that it only works when transactions of goods occur. As a result, it does not fulfil a solution for the missing trader fraud in the services department. Introducing the RTVAT will also have an impact on businesses that take advantage of cash flow opportunities, an introduction of RTVAT can lead to them partially financing the VAT. For instance when items are bought, but not easily resold. The buyer will then have pay the VAT and will not receive compensation until the products are sold.

6.1.2   VAT Locater Number system (VLN)

The VAT Locater Number system (VLN) is a system that is only looking for Missing Trader intra-Community fraud and possibilities to prevent such fraud. In order for the VLN to be successful, accountancy software, platforms have to make an automated VLN request to the centralized government system, which will then in turn either grant or deny a VLN. Each VLN is unique and depends on elements of the invoice and prior related VLNs from the transaction.

Once a business is either attempting to buy or sell a product/service, their system has to either secure a VLN (when selling) or validate the opposing trader's VLN (when buying). During the transaction the VLN system should require to have the encrypted VLN printed on the invoice. If the VAT Locater Number on the invoice is either invalid or not present at all, VAT return will be denied.

6.1.3   Digital VAT (D-VAT)

Similar to the VLN, the Digital VAT (D-VAT) proposes changes on the place of supply based on whether or not the business has the certified software needed for transactions. The solution of D-VAT is to implement software that is able to do the following:

➢ Determine the correct amount of tax for each transaction.

> ➢ Put the previously determined amount of tax on the accompanying invoice.

> ➢ Linking the VAT input or output to the correct VAT return.

> ➢ Completing the VAT accurately.

> ➢ Check if the buyer of trader one is doing business with, is also using certified software.

In principle the use of this software would be completely voluntary, however a government can choose to make the system mandatory for business that are known for missing trader fraud or all corporations in general. The D-VAT systems are useful for bot Business to Business transactions and Business to Consumers transactions.

Ainsworth (2010) mentions it would be logical to have the system automatically check the certified status of the systems of ones traders. The most secure way to do so would be by using a public key infrastructure (PKI).

6.2   Using data mining techniques to discover fraud

Texas was one of the first states in the United States of America to use data mining and Business Intelligence techniques to detect fraudulent tax cases (Wu et al., 2012). However, they only investigate about 1 percent of all tax reports each year. Because of this low rate, it is highly important that recourses are allocated well. Since the first application of data mining and Business Intelligence techniques over 362 billion US Dollars have been recovered.

As Kirkos, Spathis and Manolopoulos once define data mining as an iterative process within which progress is defined by discovery, this can be through manual or automatic methods (2007). Another descriptions is as follows: data mining is a way to discover hidden information within data.(Wu et al., 2012). Phua, Lee, Smith and Gayler (2010) define data mining as being about finding previously unknown insights that are statistically reliable and actionable from data. In the field of tax fraud data mining has the goal of enhancing an auditor's productivity while recovering tax losses. Kirkos et al. (2007) also researched the possibility of using data mining techniques to discover tax fraud or tax evasion. Even though these techniques have been broadly and successfully used in the financial industry, succesful cases remain sporadic within the field of fraud detection.

Data mining techniques can be combined with data warehousing. Data warehousing is a means to use data from different sources. By combining this with data mining, patterns and correlations can be detected from different data sets of different sources. By extracting the amount of sources and data, more patterns can be discovered and more suspicious cases can be extracted. This results in an auditor's higher productivity and a decrease in the amount of losses because of tax evasion.

It is believed that falsified financial statements can be discovered by creating a model and dividing statements into subcategories. Therefore, a number of methods have been developed. In the next paragraphs some of these methods statements will be discussed.

6.2.1   Techniques

Decision Trees

A decision tree can be described as a tree structure, where every node represents a test and every branch a possible outcome (Kirkos et al., 2007). The nodes are best used if they are capable of dividing the sample into exclusive subgroups, until no further deviation can be made that could have a significant statistical impact. There are several algorithms that can help splitting the sample, for example the Automatic Interaction Detection (AID).

When working with a large sample and then extracting multiple successful divisions, a large tree could be created. This may result in the need for 'tree pruning', the removal of splitting nodes without affecting the accuracy of the tree.

When an extra, earlier unseen object represents itself, the nodes and branches of the tree can be followed and will then lead to a class for the object. In the tree, the node represents the 'if' and the branch represents the 'then', making it an easy model for conducting if-then relationships.

Neural Networks

Another method that is used, is referred to as a Neural Network (NN). A Neural Network consists of an input layer, output layer and zero, one or multiple hidden layers in between. Each layer consists of various neurons which are connected to other neurons in the following layer, e.g. a neuron in the input layer is connected to the second, third and fifth neuron in the first hidden (second) layer. Every connection within a Neural Network is given a weight. This weight indicates the amount of influence one neuron has on the next connected neuron.

When the NN architecture is conducted, the network has to be thoroughly tested and adjusted until it has an acceptable deviation (the error) between the calculated output and the actual output. Once this acceptable error is reached, the Neural Network can be used on new and unknown data to detect actual fraud.

Neural Networks have several attributes that are highly desirable, namely: they do not make assumptions about independence of variables. They are capable of working with noised or inconsistent data and can be applied when algorithmic solutions are not suitable. Neural Networks are most applicable in detection of material errors and management fraud (Gupta & Gill, 2012).

Bayesian Belief Networks

Bayesian Belief Networks (BBN) distinguish themselves as they create the possibility to show dependencies among subsets of attributes. The network is set up with nodes and arrows, where each node represents an attribute and each arrow represents a dependency. These dependencies can also be viewed as a parent-child relationship.

The structure of a BBN can either be predefined or be built as one is working with the available data. Once the model is build and one of the nodes is set as a base, the probability of each and every other node appearing can be calculated.

6.2.2 Application of data mining techniques in fraud detection

Data mining can be divided into two sectors. The first is known as knowledge discovery techniques in database process, which is described as a series of processes to select the data from the database in question. The second is the actual data mining.

An association rule data mining technique finds correlations between data. Concerning VAT fraud, association rules can be used for detection by looking for patterns or relationships between the associated attributes of VAT evasion. When rules are created, there are two important aspects that can be used to validate a rule. Firstly, support (showing the statistical significance) and secondly confidence (showing the strength of the rule). For a rule to be validated, both support and confidence need to exceed a certain threshold.

When one is looking into tax data, it will soon be discovered that data is either numerical or categorical. The size of a numerical attribute domain can be very large. Schemes can then be used to reduce the size of such a domain. This will result in several intervals that will have no overlap.

Once all associations have been conducted, the formed rules can be used on a dataset to find actual tax evaders. An example of a rule is:

$$BSCD38 = [03] AND sales\_cap\_class = [01] AND sales\_class = [01]$$
$$\Rightarrow EVD\_TAX\_CLASS = [02]$$

Figure 3.   An example of a standard association rule

This rule states that if a company's business category is class 3, its sales capital ratio falls in class 1 and its sales amount is part of class 1, the company will be identified as possibly being a tax evader with an evasion amount of class 2. An auditor will then be assigned to this specific case to further investigate it and determine whether or not the company is truly trying to evade taxation.

It is also possible for two or more rules to be merged when both rules deliver the same result. Merging rules can lead to a higher support and confidence rate. An example of two rules that can be merged before and after merging:

$$sales\_cap\_class = [01] \ AND \ valued\_rate = [08] \Rightarrow EVD\_TAX\_CLASS$$
$$= [01]$$

and

$$sales\_cap\_class = [01] \ AND \ valued\_rate = [09]$$
$$\Rightarrow EVD\_TAX\_CLASS = [1]$$

Figure 4.   *An example of two standard association rules before merging*

$$sales\_cap\_class = [01] \ AND \ valued\_rate = [08 \sim 09]$$
$$\Rightarrow EVD\_TAX\_CLASS = [1]$$

Figure 5.   *An example of two standard association rules after merging*

6.2.3   An example of data mining techniques in financial fraud detection

In their research of prevention and detection of financial fraud by using data mining techniques, Gupta and Gill (2012) used a methodology that is displayed in figure 6. The first step in this methodology is 'feature selection'. The authors established which features would be subject to their investigation. In other words, the researchers selected sixty-two features of the company that were deemed useful for detecting fraud.

The second step of this process is to collect the company's data. Data of these features can be found in the financial statements of a company and are financials (cash, assets) and ratios (cash/total assets). They represent the characteristics of a company, as far as those can be expressed in numbers. For this study, data of 114 companies have been used. Of those 114 companies, 29 companies have been noted to be identified as having fraudulent statements by the U.S. Security and Exchange Commission. The remaining 85 companies have been identified as legitimate companies by absence of proof of fraud. It is important to note that the 85 companies that have been identified as non-fraudulent are not per se always fraud free, they might have slipped through and could commit fraud in the future.

Now that data is selected, the data has to be processed before the mining can begin. This step is called data pre-processing. Missing data has been, in the paper of Gupta and Gill (2012), replaced with the mean of the available data. Then an ANOVA test and descriptive data mining techniques have been conducted to select the variables that are considered informative and significant. 35 out of 62 features remain for further investigation.

Data pre-processing is followed by the selection of data mining techniques. Here, a descriptive data mining technique (association rules) is used as a prevention mechanism. A rule engine will generate association rules, which will then have to match a minimum support level of 0.95 and a minimal confidence level of 0.8 in order to be used for the model. Those rules are the basis for the rule monitor, a module that inserts the gathered company data in the rules and detects abnormalities. When an

14

abnormality is found within a company's data, this suggests the company is most likely committing fraud and should be subjected to further investigation by the company itself, or a third party.

In view of fraud prevention there are certain things a company can do. Fraud occurs when there is an opportunity, a motivation and rationalisation. For instance enforcing corporate governance within a company and thus taking away the opportunity to commit financial fraud, could decrease the amount of fraud tremendously. Making sure the presence of motive, opportunity and pressure is limited, is also an option to make sure the amount of fraud committed is minimized.

If it has proven impossible to prevent fraud with descriptive data mining, one can use predictive data mining techniques to look into fraud within the financial statement.

The final step of the framework is performance evaluation. A crucial step as the quality of the data mining methods used is evaluated. Amongst other things, Gupta and Gill (2012) focussed on the confidence, support and conviction rates of the rules generated by the rule engine to measure the performance of the data mining techniques they used.
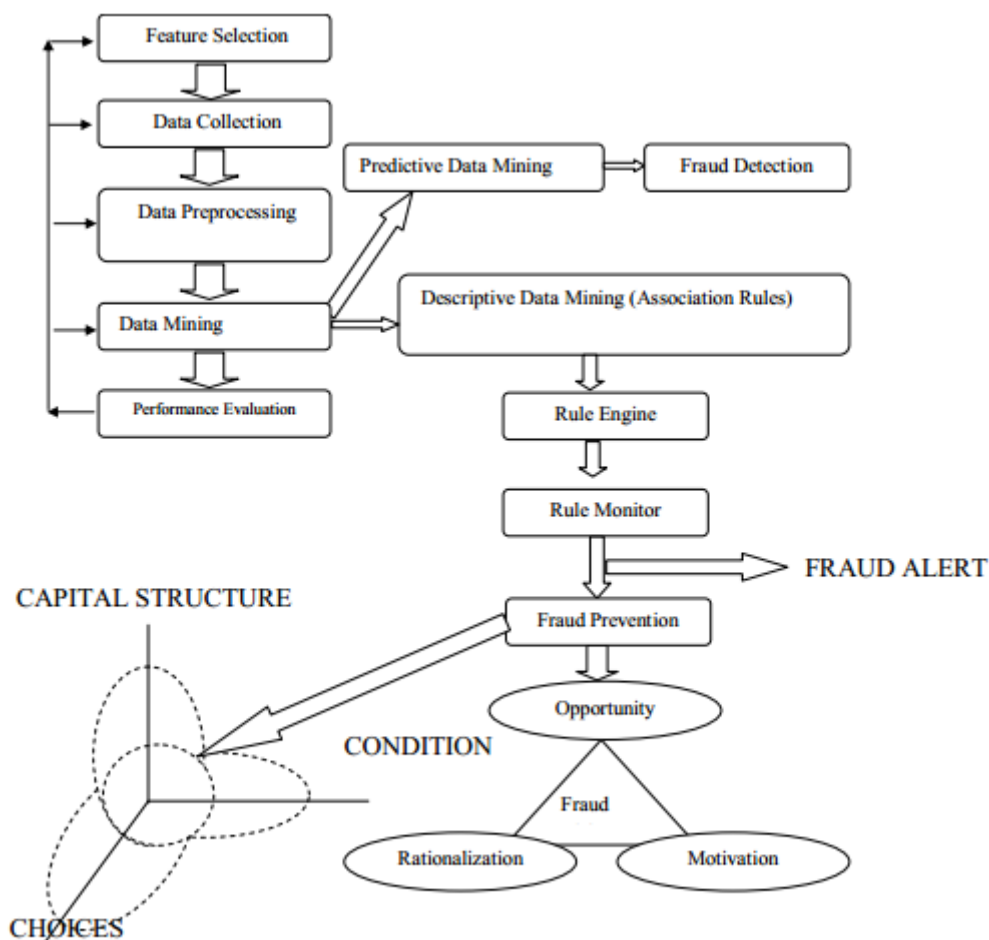

Figure 6.   A framework for prevention and detection of financial fraud

7 Case Study: "Missing imports from China"

7.1 Introduction

Fisman and Wei conducted a case study about the relationship between tax evasion and tax rates (2001). As previously mentioned, they believe that tax evasion is positively related to tax rates: the higher the rates the higher the rates of evasions. However, Fisman and Wei (2001) can not accept the lack of knowledge about the impact of an increase in tax rates has on the amount of tax fraud. To measure this effect, they performed a case study on missing exports from Hong Kong that should have been imported in China. As Fisman and Wei (2001) state, these missing imports have always been thought of as misscalculations instead of deliberate tax evasions. When taking a closer look at the missing imports, it can be concluded that goods with a high value and high tax rate were more often missing compared to their lower counterparts. In this paper, we will come to refer to this as 'Insufficient tax payments': a case where goods are underreported and the fraudsters have been smart enough to take advantage of the difference in tax rates when picking the goods that should go missing.

China's tax rates are a combination of a standard tariff and the VAT. The VAT being a rate assigned to a form of products or services, like alcohol or the service industry.

7.2 The imports

Hong Kong is known to be a former colony of China and making many imports and exports between these two countries likely. As this case focusses on the export of Hong Kong and import of China, I will from now on address all traffic between these countries as being exports from Hong Kong and import of China. Because of the high amount of traffic, it may not be surprising that the traffic can be classified into two categories: direct and indirect exports. Direct exports are those goods that traffic directly from Hong Kong to China. Indirect exports are exports from another country that travel through Hong Kong before finding their final destination in China, for example goods that have been produced in Taiwan are first shipped off to Hong Kong before continuing their transportation to China.

Even though it could be argued that indirect export could be a form of tax fraud, research suggests that this is not the case (Fisman & Wei, 2001). It is investigated that the amount of tax paid for a product that travels through different countries and a product that is transported directly to China is the same. Therefore, it would be highly unlikely if indirect transportation would be used to evade tax or is used to gain tax advantage for either the sender or receiver of the product.

However, as China does not make use of a system that reports indirect imports as being indirect imports, it labels imports as direct imports from the original country. This may lead to labels that hold wrong information about the country of a products origin. Updated labels are often seen, as some countries (like Taiwan) do not allow trade with China, those exports are therefore often labelled as having the destination of Hong Kong. In Hong Kong, these goods get a new label so they can be shipped to their final destination, China. It is logically found that, due to a difference in methodology, this may lead to a difference in imports from Hong Kong to China and the amount of direct exports from Hong Kong to China.

7.3 Conclusions

After conducting a model and executing some calculation, Fisman and Wei (2001) found that there is indeed a strong correlation between the gap of reported goods of the traffic between China and Hong Kong and the height of the tax rates present in China. Their research shows that very often it seems as though products with high tax rates 'accidentally' get lost during the export-import process. This leads to a high loss in tax revenues for the Chinese government. Maybe even more extraordinary, it is very likely that when Chinese authorities would decide to increase tax rates to increase their revenues, their tax revenues will actually drop. Fisman and Wei (2001) believe that the increase will not lead to a lower amount of imported goods, but a sudden increase in goods that will go 'missing', thus the difference between import and exports between China and Hong Kong will increase.

From this model, the authors also concluded that the missing products with high tax rates aren't the only cause of lack of revenue gained by tax payment. Part of the low revenue finds its origin in the mislabelling of products, thus the payment of low tax rates for a product that actually is found to have a high tax rate due. Another problem is also the underreporting of low value goods. Even though this may seem as a less important class, a huge amount of underreported low value, low tax rate goods are just as well capable of creating a gap in tax revenues.

8   Discussion

8.1   Conclusion

This thesis looked at tax fraud, its different forms and consequences. Tax fraud is a felony that exists in every country and is hard to detect. There are many different forms of tax fraud, for instance VAT fraud, income fraud and indirect tax fraud. Unfortunately, Tax fraud is harmful for the government, the economy and individuals and should be minimalized by all that are involved.

More importantly, this paper discussed various methods to minimize tax fraud. There are two forms of solutions when fraud is concerned, preventive techniques and detective techniques.

Preventive techniques discussed in this paper are applicable to VAT fraud. The three solutions offered are software based and would be most efficient if they would be made mandatory by the government. In short, RTVAT moves the moment of taxation from the date of invoice to the date of settlement. Making the moment of taxation instant when selling a product or service. The system ensures that tax will directly be paid to a government and is never owned by the seller. Therefore, RTVAT is also referred to as a settlement system.

VLN is a software system that is only useful when Missing Trader intra-Community fraud is a problem. The VLN system works with unique numbers that have to be requested at a central governments system. Transactions that are completed without a valid VLN on the invoice are not considered for a VAT return. In other words, transactions that are not legally registered, are not able to get a VAT return from the government DVAT and unpaid VAT can no longer be requested back.

D-VAT is the final method of preventive techniques discussed. This technique is similar to VLN, but differs in the place of certification. Where VLN certifies transactions midst a government system, D-VAT is based on the concept that the software used is certified. The software of a company would then establish a connection with the software of the buyer/seller and check automatically for the certified status. Furthermore, it handles invoicing, tax payable and VAT returns.

The detective technique discussed is suitable for all forms of tax fraud and tax evasion and is referred to as data mining. Data mining is a way of classifying data in such a way that new, valuable information can be derived. In order to create a model, three methods can be used: decision trees, Neural Networks and Bayesian Belief Networks.

Decision trees consist of nodes and branches, where every node represents a test and every branch is a possible outcome. The number of tests is dependent on the number of categories that can be formed within the data set, while still having a statistical relevance.

Neural Networks consist of an input layer, zero to multiple levels in the middle and an output layer. Each layer within a Neural Network is connected to the next level. Each connection is assigned a weight, which indicated the amount of influence one neuron has on the next.

Bayesian Belief Networks are useful when dependencies between subsets of attributes are asked for. These networks consist of nodes and arrows, each node represents an attribute and each arrow stands for a dependency.

Once a model has been conducted, association rules can be used to detect patterns within datasets. Association rules are tested and perfected before used for actual fraud detection. A rule needs to have an acceptable support and confidence rate in order to be labelled useful. When the final compilation of rules has been established, a data set can be subjected to the set of rules and a verdict can be made about the truthfulness of a company's financial statement.

8.2   Further Research

This paper discussed some options to prevent and discuss tax fraud and tax evasion. These methods have been discussed shortly and they could be subjected to further research. As most methods in this paper are based on only one or two previously written papers, extra or different information could be found

from different papers. These papers can be either from the same or different authors, where the latter is most likely to give a different view on the subject.

The VAT software described in chapter 6.1 is now based on the possibility of introducing such a system. It is very likely that one or more of those methods have been introduced in a country and the results of the use of the software can be measured. Also, the ease of use and functionality can be further investigated, which will then tell if the software can become a widely used method worldwide.

The methods discussed in chapter 6.2, the data mining methods, are the most commonly used methods. However, there are more methods than decision trees, Neural Networks and Bayesian Belief Networks. Those methods have the potential to be better or faster in creating a model for the actual data mining. Also, as now an example of data mining concerning financial statements is used, an elaboration of data mining concerning tax fraud could provide new insights and unknown results.

Furthermore, there are more methods of detecting tax fraud and tax evasion available and more can be detected in the coming years. For example, Benford's Law (Watrin, Struffert, & Ullmann, 2008) might provide useful information in regard to tax fraud detection.

A last recommendation for further research is the collaboration between a government and company, as both experience negative effects of fraud, they could combat fraud together. A study could suggest various methods to do so and discuss wether this is yields better results than a government or company can achieve on its own.

References

Ainsworth, R. T. (2010). VAT Fraud and Technological Solutions. *Boston University School of Law Working Paper*, *28*.

Castellón González, P., & Velásquez, J. D. (2013). Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Systems with Applications*, *40*(5), 1427–1436. doi:10.1016/j.eswa.2012.08.051

Fisman, R., & Wei, S.-J. (2001). *Tax rates and tax evasion: Evidence from" missing imports" in China*. National bureau of economic research.

Goerke, L. (2003). Income Tax Evasion in a Unionised Labour Market. *Australian Economic Papers*, *42*(2), 135–148.

Gupta, R., & Gill, N. S. (2012). Prevention and Detection of Financial Statement Fraud--An Implementation of Data Mining Framework. *International Journal of Advanced Computer Science & Applications*, *3*(8).

Keen, M., & Smith, S. (2007). VAT fraud and evasion what do we know, and what can be done? [Washington, D.C.]: International Monetary Fund. Retrieved from http://catalog.hathitrust.org/api/volumes/oclc/140215362.html

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, *32*(4), 995–1003. doi:10.1016/j.eswa.2006.02.016

Petersen, H.-G., Thießen, U., & Wohlleben, P. (2010). Shadow Economy, Tax Evasion, and Transfer Fraud – Definition, Measurement, and Data Problems. *International Economic Journal*, *24*(4), 421–441. doi:10.1080/10168737.2010.525973

Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv Preprint arXiv:1009.6119*.

Sandmo, A. (1981). Income tax evasion, labour supply, and the equity—efficiency tradeoff. *Journal of Public Economics*, *16*(3), 265–288.

Slemrod, J. (2007). Cheating ourselves: the economics of tax evasion. *The Journal of Economic Perspectives*, *21*(1), 25–48.

Spencer, L. D. (1946). Proof of Income Tax Fraud. *Tax L. Rev.*, *2*, 451.

Virmani, A. (1989). Indirect tax evasion and production efficiency. *Journal of Public Economics*, *39*(2), 223–237.

Watrin, C., Struffert, R., & Ullmann, R. (2008). Benford's Law: an instrument for selecting tax audit targets? *Review of Managerial Science*, *2*(3), 219–237.

Weigel, R. H., Hessing, D. J., & Elffers, H. (1987). Tax evasion research: A critical appraisal and theoretical model. *Journal of Economic Psychology*, *8*(2), 215–235.

Whiting, D. G., Hansen, J. V, McDonald, J. B., Albrecht, C., & Albrecht, W. S. (2012). Machine Learning Methods For Detecting Patterns Of Management Fraud. *Computational Intelligence*, *28*(4), 505–527.

Wu, R.-S., Ou, C.-S., Lin, H., Chang, S.-I., & Yen, D. C. (2012). Using data mining technique to enhance tax evasion detection performance. *Expert Systems with Applications*, *39*(10), 8769–8777.

Yakovlev, A. (2001). "Black Cash"Tax Evasion in Russia: Its Forms, Incentives and Consequences at Firm Level. *Europe-Asia Studies*, *53*(1), 33–55.