

## Tax Strategies Employed by Overseas American Individuals and SMEs

Dr. Alfred Howard Miller

Box 1626 Fujairah Women's College, Fujairah, UAE Higher Colleges of Technology, United Arab Emirates, Henley School of Business, Reading University, Greenlands, Henley-on-Thames, United Kingdom

### Abstract

A study of the tax behavior of overseas Americans, both individuals and small firms is proposed. The researcher aims to discover and model behavior, through text analysis of data collected from a wide range of sources using interviews, surveys, blog and forum postings, published reports as well as personal communications, to demonstrate and inform using the pattern matching method initially proposed by Trochim (1989). Text mining and modeling techniques, using unsupervised machine learning facilitate large-scale analysis, and have been widely deployed in a range of language-based studies, driven by human-machine interaction. Major multinational corporations are excluded, and the study focuses on individuals and the smaller-scale juristic persons such as small and medium enterprises (SME). Behavioral approaches to taxation will motivate a better understanding of the phenomenon tax avoidance and tax evasion, once quantitative modeled. Overseas Americans are taxable, no matter where they reside globally, on the basis of having American citizenship. Non-citizens with a USA connection may also be subject to US taxes. The range of US taxable entities operating overseas include corporations, individuals, estates and trusts, and many of the small businesses filing as flow-through entities under the individual code, namely S-corporations, sole proprietorships, and partnerships, will be included in the study. There are an estimated 9 million taxable overseas Americans corporations and business entities. The Congressional Research Service (Gravelle, 2015), reported that as many as 100 billion U.S. dollars may go uncollected, due to tax evasion and a similar tax shortfall figure of 100 billion dollars is due to tax avoidance. Avoidance tends to be attributed to U.S. origin, multinational corporations and evasion by the smaller entities. The tax collection is exacerbated by changes to the 2018 tax code, which encourages compliance through tax cuts to a fixed 21% rate for the corporate sector, and reduced taxes for individual, opening up new avenues for aggressive tax avoidance strategies. A gap in the literature is the uncertainty regarding changing of the U.S. tax code in 2018 and how it will affect overseas American tax entities.

**Keywords:** Tax; Evasion; Avoidance; Unsupervised machine learning; Text mining; GILTI; FBAR; FATCA.



CC BY: [Creative Commons Attribution License 4.0](https://creativecommons.org/licenses/by/4.0/)

### 1. Introduction

This paper refers to an early stage study of the tax behavior of overseas Americans, both individuals and small firms. The researcher aims to discover and model their behavior through text analysis of data collected from a wide range of sources such as interviews, surveys, blog and forum postings, published reports as well as personal communications, to demonstrate and inform using the pattern matching method initially proposed by Trochim (1989). Advances in technology—modern methods such as unsupervised machine learning utilize software and allow computing power to analyze text (Trochim, 2016). Text mining and modeling techniques have been widely deployed in range of language based studies and fall under the realm of human machine interaction. The focus is on tax behavior of United States of America (USA) taxable entities, physically located outside the USA. The study excludes the major multinational corporations, and instead focuses on individuals and the smaller juristic persons at the small and medium enterprise (SME) level of scale and below. Through the context of behavioral approaches to taxation; text mining of data coupled with machine learning techniques, offers a better understanding of the phenomenon tax avoidance and tax evasion, and obtained through quantitative modeling. Overseas Americans are taxable no matter where they reside globally, and are taxed on the basis of having American citizenship. In the case of non-citizens, with a USA connection they may also be subject to US taxes. The gamut of US taxable entities operating overseas, other than large multinational corporations, including smaller corporations, individuals, estates and trusts, and the many businesses filing as flow-through entities under the individual code, namely S-corporations, sole proprietorships and partnerships were included in the study.

#### 1.1. Problem Statement

Out of an estimated 9 million overseas Americans and corporations and business entities. The Congressional Research Service (Gravelle, 2015), reported that as many as 100 billion U.S. dollars may go uncollected due to tax evasion and a similar figure of 100 billion dollars, due to tax avoidance, mainly by U.S. origin, multinational corporations. Tax collection is exacerbated by changes to the 2018 tax code, which encourages compliance through tax cuts to a fixed 21% for the corporate sector. Taxpayers seeking reduced taxes will identify new avenues for aggressive tax avoidance strategies. A gap in the literature is thus revealed through the uncertainty surrounding the new U.S. tax code of 2018 and how it will effect overseas American tax entities.

## 2. Literature Review

The literature review consists of three sections. One section is a review of literature on the topic of taxation applicable to tax collection, avoidance and evasion by overseas Americans. A second topic is the unsupervised machine learning method. A third section explores the context of the study and reveals some of the tax planning, avoidance and evasion strategies, and mechanisms that are likely to be explored as this study develops.

### 2.1. Tax Collection, Avoidance and Evasion

The US Department of State, Bureau of Consular Affairs reports that the number of overseas Americans could be as high 9 million (CA by the Numbers, 2017). The overall size of the global diaspora is 244 million for 2017. While the US share of outgoing global diaspora is quite high, India, Mexico, Russian Federation, China, Bangladesh, Pakistan and Ukraine send more of their citizens abroad (UN, 2017). The U.S. taxpayer diaspora is one facet while the corporate sector is another. There is some overlap of the individual and corporate taxpaying sectors as many individuals choose either to incorporate or form pass through entities as part of their tax avoidance strategy. Corporate inversion is one phenomenon and is the practice of moving a corporation overseas or merging with a foreign company. Various tax planning strategies results in less taxes being remitted to the United States government and less jobs for Americans. US tax rates are reported to average 35%. Legislation has been introduced to make inversion more difficult, however, many US corporations continue to seek the tax advantages of a foreign domicile (Marshall, 2016).

A range of controls were either sought or introduced by the Obama administration in an attempt to collect more tax by closing tax loopholes. Corporations were more closely associated with tax avoidance whereas individuals were more closely aligned with tax evasion. Tax avoidance is considered legal, whereas tax evasion is illegal (Gravelle, 2015).

Tax avoidance schemes of shifting of income by multinational business entities is thought to result in losses of 100 billion dollars per year or more. By shifting debt to high tax jurisdictions profits can potentially be deferred indefinitely. Hybrid entities have been used to reduce passive income through exploiting varied treatment in different tax jurisdictions. Individual tax evasion is estimated to be a lower, yet still significant at 40-70 billion per year. By some estimates, this figure is as high as 100 billion. Individual tax evasion strategies include failure to report passive income from capital gains, interest and dividends originating from foreign income sources (Gravelle, 2015).

A European study purported that tax avoidance was typically seen alongside tax evasion; being options actively considered by taxpayers who weighed the consequences of both. Typically, the taxpayer deployed as much tax avoidance as possible, exhausting that opportunity, and then decided how much tax evasion they could get away with without significantly increasing the likelihood of being audited (degl'Innocenti and Rablen, 2017).

The strategy advocated by Gravelle (2015) for combatting individual tax evasion was to seek to collect more information about the foreign dealings overseas American taxpayers. The Foreign Account Tax Act (FATCA) requires the registration of accounts with US ownership affiliation. The effects are not known yet, but FATCA is being enforced to identify and combat interest paid to foreign recipients that is not taxed, as overseas American individuals often channel interest and funds into shell corporations and trusts located in foreign jurisdictions. New provisions recommended, include limits or repeal of deferral, limits on foreign tax credits, and introduction of an apportionment of income system.

Gravelle (2015), was a white paper produced by the Congressional Research Service. The purpose appeared to legislation strategy guidance by those in United States government seeking to collect more taxes by closing tax loopholes that could be exploited by overseas American taxable entities including, both individuals and corporations.

Apple avoided paying taxes on 44 billion in otherwise taxable off-shore income, over 2009-2013. Apple's tax strategy included structuring of three foreign subsidiaries for tax purposes that were not tax residents of any sovereign state. Apple claimed its tax strategy was legal and proper, being typical of similar firms such as Microsoft and Google (K@W, 2013).

The residential tax system of the US assesses taxes regardless of where profits are earned, provided the entity is identified as being a US taxpaying entity. Taxes paid by the entity to other countries, are subtracted from this tax calculation. Most countries employ the more typical, territorial tax system, upon which, taxes are paid on economic activities taking place within that country. The territorial tax system is the common system (K@W, 2013).

The US system does permit deferment of overseas income until that income is repatriated to the US. The system encourages US firms to incorporate in tax havens and shift their income using transfer pricing for goods, services and intellectual property sold overseas. The transfer pricing and income shifting can then occur across multiple subsidiary entities. For example, one arrangement called the 'Double Irish' is two Irish firms that tax each other; both are subsidiaries of a US firm, using advantageous transfer pricing to shift income. Examples of firms that deploy the *Double Irish* such as Adobe Systems, Eli Lilly, Facebook, G.E., Microsoft, Oracle, Pfizer and Starbucks. A variation sandwiches a Dutch firm between the two Irish firms (K@W, 2013).

Overseas American firms tend to defer tax indefinitely, then lobby lawmakers and administrators seeking a tax holiday; returning the cash to the US tax-free. When this repatriation of funds occurs, there is a subsequent jump in earnings and boost in share prices. The construction of such tax havens is costly in terms of consultants, accountants and lawyers, but ultimately cost effective due to the tax savings (K@W, 2013).

Analysts concede that US-based firms, competing globally are at a disadvantage given the 35% statutory tax rate. US firms operating abroad are forced to maintain complex tax entity structures in order to remain competitive globally. European's see these same tax structures employed by US multinational as reducing their ability to collect

European tax as well. Critics are calling for US corporate tax reform that cuts taxes to re-incentivize investing in the US economy. Recommendations range from more aggressive tax collection by closing loopholes and maintaining the 35% tax rate, in comparison to plans that promote business growth at home by cutting taxes to 20% and eliminating deferral as a tax planning tool (K@W, 2013).

The Internal Revenue Service (IRS) takes the position that taxpayers are not allowed to evade taxes by shifting liability to foreign tax entities. The IRS recognizes a range of schemes, which follow in the appendix, used by individuals to abuse the tax code and pay less taxes (IRS, 2017).

Tension between financial versus tax reporting is a highly researched topic. However, characteristics of aggressive tax avoiding firms and the effect of changes in the tax law on financial reporting is an area where a gap exists in the literature. Gupta *et al.* (2014), found a connection between mandatory financial disclosures and tax reporting and collections. They also found that multi-state active firms for tax purposes, showed a greater propensity for tax avoidance. This partly explains the drop in state level corporate tax revenue. Stricter disclosure requirements such as the FIN 48 regulation, of 2007, which required greater certainty in income reporting, encouraged greater collection of state income tax. It generally takes a period of years for planners to maximize exploitation of changes to the tax code.

The U.S. Tax code has recently changed, and as the tax code changes so do the approaches taken by tax planners. These planning tips or so-called hacks are described in an article in the New York Times by their inclusion in a category ranging from easy to difficult. Tax avoidance strategies generally revolve around three approaches, income shifting, timing of income and changing the character of income.

Under the easy category: Changes in the tax code to increase the Standard Deduction, by nearly doubling it provide an incentive to take charitable deductions now before the changes go into effect. A year one and two of a child's life parental leave tax credit will be offered to businesses allowing their employees to take this benefit (Bui and Sanger-Katz, 2017).

More difficult is use is the revised estate tax rules which doubles the tax free transfer exclusion limit to 11 million dollars until 2025. A 40% tax kicks-in for larger inheritances (Bui and Sanger-Katz, 2017). Only \$10,000 will be deductible for state and local taxes and this effect residents of high tax states such as New York, New Jersey and Connecticut, all with state and local taxes at over 12%, while helping those residents from lower tax states such as Alaska, Wyoming, South Dakota, Texas and New Hampshire, are taxed ranging from 6.5 to 7.9% (Tax Foundation, 2017). This encourages people in high tax states to shift as much income as possible forward to this year as well as reorganize the Traditional IRA as the Roth model. If a person is considering moving they should do it now as the moving expense tax credit is being repealed (Bui and Sanger-Katz, 2017).

Relaxed rules for 1031 capital gains swaps which where funds from the sale of one asset are swiftly rolled over to a new asset are being tightened to apply to real estate, but not other capital assets such as private jets. As such, non-real fully depreciated assets should be swapped before the tax code changes (Bui and Sanger-Katz, 2017).

The new tax bill will introduce a tax shield for private education expense of up to \$10,000 paid after passing through a Section 529 account. The new tax bill also favors pass-through tax enterprises where employees may declare a themselves as freelance consultants or self-employed thereby qualifying for a 21% tax rate for single earners making up to \$157,500 and \$315,000 for married taxpayers. There more stringent limits on income qualification when operating as a pass-through enterprise.

Another option is file as a C-Corporation and avail the 21% tax rate on all earnings. This would allow the deduction of state and local taxes at expenses of profit distributions to the owner(s) being taxed as dividends. Taking this concept, a step further, if an incorporated business is split into several entities, one entity could provide capital while another provides labor, working a achieve an optimal split of which firm claims how much in profits. Further ideas include operating one's personal brand as a pass through enterprise, or splitting company into enough dependent enterprises that everyone working at the firm qualifies for the pass through rate. Exchanges of assets for shares is non-taxable, the transferee can sell the asset taxable at a 21% rate, while shares can be passed on to heirs tax free.

Changes in the US tax code with the resulting favoritism toward the 21% corporate rate, opens up tax research and tax employment opportunities. It is believed that the new tax regime was designed to encourage repatriation of overseas US holdings due to more advantageous corporate tax rate structure back home.

Professors, Mark (Shuai) Ma of American University, Kogod School of Business and Wayne Thomas of the University of Oklahoma, see tax reform entirely differently. They provided empirical evidence that criminal prosecution including jail time for offending executives, was the best deterrent to tax evasion. Their methodology noted that states with stricter penalties and longer criminal jail sentences for executives of firms convicted of tax evasion, tended to have a higher effective tax rate at the state level indicating greater compliance and a less aggressive tax avoidance strategy. The authors found that the only time civil penalties were an effective deterrent was when management had a high rate of share ownership. Finally, Nevada and Delaware were identified as Domestic tax havens (Ma and Thomas, 2016). Computerized analysis methods are gaining in popularity due to high data volumes.

## 2.2. Unsupervised Machine Learning Method

With unsupervised machine learning method deployed for this study, human factors are largely divorced from the process. Data analysis is objectively performed by computer using an algorithm, some degree of interpretation is required to construct the evaluation model.

Trochim traces the advent of theory-driven approaches, and their application to program evaluation and creation of theoretical frameworks aiding conceptualization methodology (Trochim, 2016). Trochim has advocated establishing validity by matching the observational realm with the theoretical realm (1989). Pattern matching studies today can draw upon techniques borrowed from search engine optimization and search engine marketing.

One of the early advocates of analyzing the corpus or body of language was (Hunston, 2002). Hunston said, “The main argument in favour of using a corpus is that it is a more reliable guide to language use than native speaker intuition is” (p. 20). Hunston nevertheless valued intuition stating: “[Intuition] is an essential tool for extrapolating important generalizations from a mass of specific information in a corpus” (p. 22). The method is therefore a mixture of a computer analysis and researcher intuition.

The Handbook of Business Discourse was a useful reference for motivating content analysis in the context of business studies. Systematic replicable methods were described for working with large data sets of text, apportioning words into categories, motivated by uniformly applied computer coding rules. Intuition as well as objective means can be used to process the output. The approach has commonality with other data analysis methods, as a means to identify patterns, and trends. Content analysis methods are best applied when triangulated with other methods (Bargiela-Chiappini *et al.*, 2008). Yu *et al.* (2011), concurred. They stated that a computer aided quantitative approach to text-mining of natural language, was reliable and consistent, being compatible with grounded theory, provided the researcher remained neutral, as categories emerged from the data. Moreover, they justified content analysis because it could be used to validate evidence (Yu *et al.*, 2011). Trochim (2016), noted how widespread and interdisciplinary concept mapping methods had become.

Content analysis can work with smaller scale studies. Minami and Ohura (2015), used KH Coder for content analysis of 35 student survey responses. They focused on 20 key responses to Question 11, “How Student's Attitude Influences on Learning Achievement? An Analysis of Attitude-Representing Words Appearing in Looking-Back Evaluation Texts“. The author’s explored a potential link between learner’s attitude and achievement performance. The authors characterized their study as Education Data-mining (EDD) missed with Knowledge Discovery Data-mining (KDD). Low performers used language and words similar to high performers yet they were poorly implemented while the middle performance group had identifiable word differences. Their approach analyzing a limited number of student responses drew a parallel to the method deployed to analyze content—reflections in the case report on the SCIL model. Dissimilarities were use of a questionnaire, instead of discussion points and class focus on IT—information retrieval (Minami and Ohura, 2015).

KH-Coder has been accepted in the U.S. Court of Law (Posner, 2012; Smith, 2014). Text mining and content analysis has been used in over 600 studies, including studies from the Middle East and North Africa, such the mobile learning study of Pelet *et al.* (2014). In the tradition of Vygotsky, text mining and content analysis can be used to scaffold collaboration of proficient students helping weaker ones, through adoption of pedagogic processes of performance improvement to construct socio-cultural theories based on student responses from activities (Flowerdew, 2009). “Ultimately Algorithms are a set of instructions followed by computers to solve problems” (O’Neil, 2016). However, O’Neil challenges big data and says rogue algorithms are often with too small of sample sizes with only a few dozen students (2016). This study however will use the approach of Faul *et al.* (2009) to test for statistical power, where a sample size of approximately 500 respondents may ultimately be specified.

Measurement of student learning outcomes can be analyzed using trend data normally reported for quality assurance reports and content analysis techniques such as KH Coder. KH Coder permits coding-based text analysis and allows the researcher to deploy a range of techniques which include; word frequency analysis, hierarchical cluster analysis, co-occurrence network, multidimensional scaling and self-organizing maps.

A highly capable technology is MatLabs from Mathworks. In fact, computational thinking in business education is a boundary spanning technology when machine learning and numerical analysis capabilities of MATLAB are deployed outside of engineering sciences (Gross *et al.*, 2014). While KH Coder is open source and user friendly, MatLab a proprietary program available at the college from August 2017 onwards offers substantial computational and graphics capability. Both technologies permit user-coding and a user friendly interface. Times to run statistical procedures can be expected to take anywhere from a several minutes for factor analysis to several hours for self-organizing maps.

Assessing work readiness in accounting graduates via the SCIL-based model (Miller, 2016;2017b) is based on learners at Higher Colleges of Technology taking the Taxation course during either Semester 7 or 8 during Year 4 of the Bachelor’s degree. The sample size collected was  $n = 23$  (2015) and  $n = 28$  (2016),  $n = 9$  (2017) and  $n = 26$  (2017) in the latest study with additional data points analyzed beyond the initial study. The researcher deployed a corpus-based approach to ethnographic research that is collaborative and contextual. The research strategy was to collect a body of business discourse, namely student reflections from the Year 4 Accounting, Taxation course, and then performed a corpus-based computer aided linguistic analysis (Miller, 2017b;2017a).

Vesanto and Alhoniemi (2000), advocated the self-organizing map as being an excellent tool in exploratory data mining. They further explained the dendrogram supported nature of clustering using the self-organizing map (SOM). Furthermore, hierarchical cluster analysis uses dendrograms to specify clusters and as a technology can reinforce the SOM.

According to Berinato (2016), data visualization has become a must-have skill for all managers. This is because a visual abstraction is often times the only way to process the volume and velocity of data that arrives for processing. Furthermore, decision-making increasingly relies upon the ability to make-sense from and interpret of this voluminous amount of what is also known as big-data. Due to open source programs, the internet and proprietary,

yet affordable tools, visualization is becoming widely accessible. Access to tools without the deeper understanding of their application can result in producing charts that are inadequate or ineffective.

Berinato proposed that data managers and decision makers ask two questions. “Is the information conceptual or data-driven? and is the statement about the topic declarative or exploratory?” (Berinato, 2016). Berinato’s decision model identifies which of four types of visualization goals will be most effective, namely: “idea illustration, idea generation, visual discovery, or general data visualization” (Berinato, 2016). The implication of what has been stated by Berinato is guidance to the researcher conducting a study that is both quantitative and qualitative in nature.

Anzai and Matsuzawa (2013), used the content analysis methods of word frequency and co-occurrence networks to explore differences in the mission statements of pre-World War II and post-World War II Japanese universities. Their key finding was a greater focus on research in the pre-war university model compared to the post-conflict iterations. A moderating factor was the evolution of post-war universities through privatization to what is known as national university corporations and a greater focus on profitability.

The method to conduct data visualization with multidimensional scaling was explained by Buja *et al.* (2008). Tamura (2011), used word analysis, co-occurrence analysis, and categorization to examine a socio-cultural experience shared on an internet discussion forum. Tamura set word analysis limits and reported high frequency words appearing 40 or more times. Using co-occurrence networks and the Jaccard coefficient, Tamura (2011) was able to intuitively create categories.

The Jaccard coefficient is a word frequency algorithm. It divides the frequency of word intersection by the union of word appearance. For example, if the frequency of word a is 4, and frequency of word b is 3, then the frequency of words a and b is 2. As stated as a formula;  $2 / (4 + 3 - 2) = 0.4$  (Mori *et al.*, 2004). Tamura (2011), reported that computer coding allowed a researcher to objectively handle large amounts of data, and qualitative information could be easily represented numerically which served to decontextualize the data. Tamura (2011) considered de-contextualization to be an advantage via removing human factors from a crucial portion of data analysis.

### 2.3. Context of Study

The scope of data collection for pattern matching using approached such as multi-dimensional scaling and concept analysis is ideally governed by contextualism and multiplism. In “*Contextualism; we should articulate and observe the specifics of our object of interest, be it a program, measure, or participants in a context.*” (Trochim, 1989). This statement implies that data collected for analysis must be collected from overseas American entities who have a requirement to file taxes. Whereas *Multiplism* implies that a wide range of different types of entities such as sole proprietorships, partnerships, limited liability companies, limited liability partnerships, S-corporations, estates and trusts and C-corporations, will need to be polled, in order to articulate and engage these multiple manifestations of the gamut of contexts of tax avoidance and tax evasion.

*Relationalism*, is observing and discovering the relationship between these multiple manifestation of tax behaviors by the different types of taxpayer entities. It is important to assess how the groups being studied are interrelated. *Gradualism* is to apply a gradient in the analysis to gauge the level of similarity between entities, tax behaviors employed and deployment of approaches. Researchers, are encouraged to move away from the simple categorization objects and realize that tax behavior will vary multi-dimensionally. The concept of *Dualism* implies that patterns should be emergently identified at both the theoretical and observational level. While *Parallelism*, is the linkage between the theoretical and observed patterns. Pattern matching requires that there be correspondence between theoretical and observed structures. Finally, the degree of correspondence should be assessed using a statistical method (Trochim, 1989).

Trochim’s Concept mapping is a way to develop construct validity, using pattern matching, the Theory of Conceptualization, multidimensional scaling, cluster analysis, and bridging analysis to interpret the results of a concept map, where low values described a point location on the map while high values bridged between locations. Further methods include a go-zone plot, bibliometric analysis and multivariate statistical analysis.

Teixeira (2016), master’s thesis deployed factor analysis and explored two factors as they applied to tax evasion and tax avoidance. These factors were civic duty to pay tax and moral reasons to not pay tax. The fourteen item Likert scale survey was based on a previously validated instrument of Crowe (1944) and later updated by Alm (2012) certified as reliable at  $\alpha .909$  with a sample size of 71 banking employees. The survey considered inputs such as moral obligation, participation in the shadow economy, age, gender, religiosity, educational and occupational factors. To address this issue, a survey was designed, based on the tax ethics literature and previous studies, and administered to banking employees from a bank institution operating in Portugal. The survey addresses several issues like tax ethics, tax morals, tax evasion and tax compliance and the sample obtained consisted of 71 observations. The findings identified a strong shared ethical position against tax evasion but virtually no differences across the differentiating factors such as gender or religiosity (Teixeira, 2016).

An investigation into the problems of the shadow economy, tax evasion and tax avoidance in Greece’s economy was tackled by Tenidoua, Valsamidisa, Petasakisa, Mandilasa, in 2015. They sought to quantify the effectiveness of Elenxis a new algorithm for detecting audit candidates through an 80 sample factor analysis study using principle component analysis with varimax rotation. The sample consisted of accountants, namely public, private and freelance, while the data collection instrument was a 12 question, 4 factor Likert scale survey besides eight additional on demographic factors. The survey measured attitudes of respondents in relation to shadow economy, tax evasion and tax. The general consensus by 92.5% of those sample favored use of information systems as a way to combat tax evasion. The authors concluded that increased taxation led to an increase in tax evasion, that accountants

were jointly responsible for tax evasion with their customers, that the tax mechanism is itself was responsible for tax evasion and that there was laxity and a less constant attitude toward dealing with taxation. The authors identified that the sample was too small to be statistically meaningful, and findings limited as it was confined to only one country. Men tended to focus more on the joint responsibility and women on the tax mechanism. The authors concluded with a recommendation for a more comprehensive confirmatory factor analysis (Tenidoua *et al.*, 2015).

Recent research by Miller (2019) using Multidimensional Scaling (MDS), with a stress factor of 331, and using the Kruskal method with Jaccard distance formula. Yielded a six factor model. These six factors were 01 = Taxes Abroad, 02 = Data Analytics and Audit, 03 = Norms and Ethics, 04 = FBAR & FATCA, 05 = Impact of GILTI, and 06 = Concealed Assets Abroad. A two dimensional interpretation proposed Dimension 1, as the X-axis = Analytics on the left and Regulations on the right. Dimension 2, was the Y-axis, with new GILTI tax code at the top and Norms and Ethics at the bottom.

### 3. Methodology

A mixed methods study, using a secondary survey results of established validity and reliability, coupled with interview collected data from which keywords and distilled concepts using both human intelligence and computer interaction and with a third source of community-based question answering (Nie *et al.*, 2013) based on computer aided understanding of relevant web-based texts. The approach was implemented in conjunction with unsupervised machine learning, in a natural language processing context, coupled with content analysis, a quantitative method, to identify emergent themes toward understanding and modeling tax avoidance and tax evasion behavior of overseas American taxable entities. The modalities are a sharing of both qualitative and quantitative analysis, support answering the research questions through a triangulation of methods.

#### 3.1. Hypothesis

##### 3.1.1. Tax Evasion is Prevalent Among Overseas Americans as Individuals

##### 3.1.2. Tax Avoidance is Prevalent Among Overseas American Business Entities

##### 3.1.3. Research Question

**RQ 1.** What is the character of tax avoidance and tax evasion strategies deployed by overseas Americans and American entities?

Primary research thus included triangulation of three methods. 1) Survey by questionnaire, using a previously validated instrument, from a sampling frame representative of overseas American taxpayers, 2) Interviewing people from this same sampling frame, but not those same people as who participated in the survey. Sufficient collection will take place to achieve topic saturation, such that no more, new information is being introduced. 3) A third method targeted a web-based sampling frame with knowledge harvesting of text-based data collected from online communities, such as forums, blogs, and other web-based communications. An early stage sample of 52 unique texts were collected to achieve partial saturation of the tax topic.

Anonymity was preserved due to the ethical and moral nature of tax avoidance, evasion, and fraud which is a sensitive topic for those American citizens potentially in violation of the tax law, either involuntarily or cases actively working to circumvent the code, through exploitation of loopholes and other means. Since this research being primarily qualitative, with quantitative analysis in the form of the natural language processing of the text analysis, a larger sample size than the 52 text samples analyzed here is necessary to achieve saturation rather than a calculated statistical power.

Reliability will be obtained through confirmability, by using a uniform and repeated coding regime. The principal researcher (PI) was the only person working with coding the data set. Dependability will be achieved through member checking where transcribed notes from audio recordings are verified with interviewees to assure translation accuracy (Casey and Murphy, 2009). Reliability was also supported by triangulation within method, similarity of member responses, and established reliability of the survey instrument (Stevenson and Mahmut, 2013). Verification of participant answers, uniformity in their answers across responses, and triangulation across the methods provides a construct to test reliability of the survey, interview questions and corpus collected text from the online sources. Similarity in responses among the participants throughout the interview corroborates that research instrument and the accuracy of responses (Simon and Goes, 2016). A continuous member-checking feedback loop is reported by Harvey (2014) as a way to boost accuracy of the reliability process.

Validity can be assessed through the triangulation of the three primary data collection methods of survey, interview and knowledge harvesting from the Web. Trochim's pattern matching approach supports this verification of internal and external validity as it aligns with grounded theory, achievement of data saturation, and variation in means of participant selection. Data subjected to a quantitative approach will be examined using the output and interpretation of the machine learning algorithms.

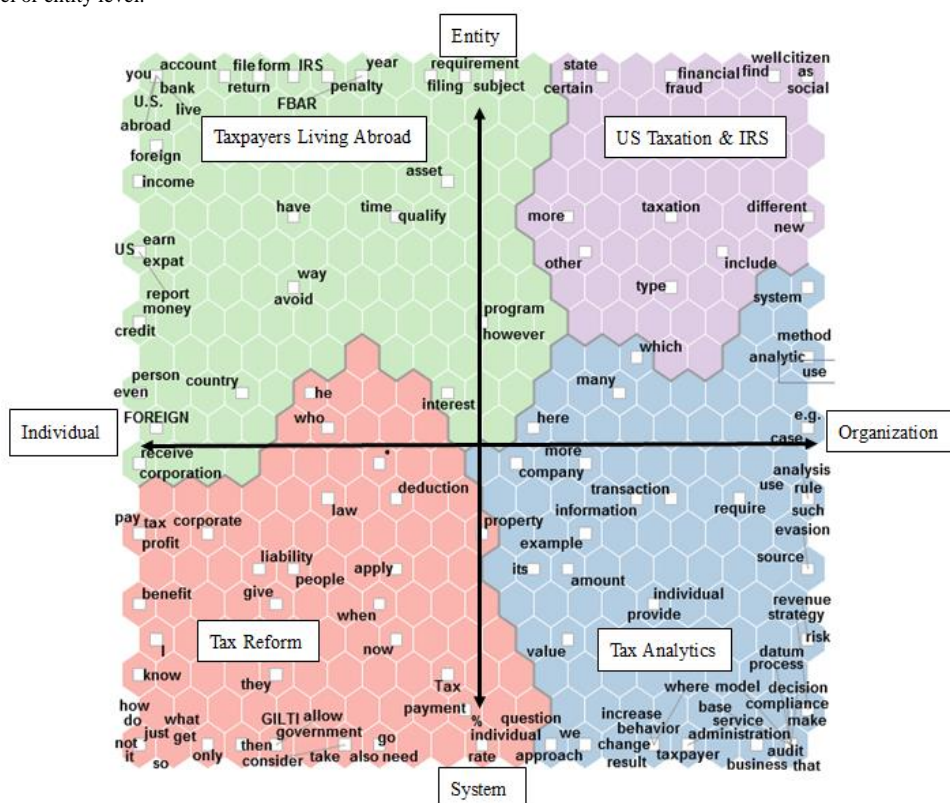
The primary researcher for this study, is the Applied Research Coordinator, for the Fujairah directorate encompassing two of the 17 HCT campuses and for the Business discipline, six campuses, also encompassing Sharjah and the Western Region sites. As an academic administrator the principal has access to a supportive research structure through Chairing the Business Division's system-wide Applied Research Committee, and a system-wide research circle. The campus based research resources include interdisciplinary expertise, experienced in machine learning from the Computer and Information Systems (CIS) division, as well as Media, Health Sciences, Engineering and Education. The principal investigator has and several publications using, coding, data mining with the open source KH Coder (Miller, 2017b;2017a;2018;2019). KH coder has been deployed in upwards of 700 studies for

content and text analysis, with natural language processing, often using the English language compatible Stanford POS tagger. The researcher has deployed relevant algorithms upon data collected from taxation learners to produce viable machine learning models (Miller, 2017b;2017a;2018;2019). The proposed research study offers an opportunity to deploy other machine learning technologies such as the proprietary RapidMiner, and Matlab and/or Python or R programming languages for the construction of suitable behavioral algorithms.

#### 4. Results

To test the hypothesis, the researcher used a self-organizing map taking 7 hours to construct using an I7 processor, to produce a comprehensive model that describes behavioral aspects the target groups. The model describes, output of tax evasion and tax avoidance strategies of both overseas Americans and American corporations operating abroad. The self-organizing map yielded a simpler four factor model compared to the six factor multi-dimensional scaling model, that resulted from the Miller, in process study. These findings will be disseminated in academic conferences, and scholarly journal publications. See Figure 1 on the following page for the four factor model described by *approach* and *interpretation*.

**Figure-1.**X Axis is Interpretation, either an individual or organizational perspective, while the Y Axis is Approach, either system level or entity level.



Using the Google search engine to analyze the context of the output from the self-organizing map four factors were derived. These factors were 1) Taxpayers Living Abroad, 2) US Taxation and IRS, 3) Tax Reform, and 4) Tax Analytics. Each axis, X Interpretation, and Y Approach had its respective extremes. For X or Interpretation, the model has Individual and Organizational perspectives as competing dynamics. For Y or Approach there is either a System or Entity level response. Key acronyms such as FBAR which is the foreign bank account reporting requirement is seen as being dealt with from an Individual and Entity level and located in the Taxpayers Living Abroad quadrant. GILTI stands for Global Intangible Low-taxed Income and is designed to capture tax on any investment income yielding a high rate of return. GILTI is aligned as an Individual and System level response and located within the Tax Reform quadrant. Tax Analytics being reflected by its own quadrant is demonstrative of the trend to use machine learning algorithms in the tax collection process. Tax Analytics is a System-level Organizational response. The final quadrant serves the Entity and Organization the IRS and US Taxation Code.

#### 5. Conclusion

With a global diaspora of 9 million overseas Americans and potentially \$100 billion dollars in tax evasion, primarily by individuals and another \$100 billion in tax avoidance, primarily by corporations, substantial taxes remain uncollected. The revised tax code of 2018 offers both an opportunity to collect additional revenue while at the same time opening up new avenues for pursuit of aggressive tax avoidance. The amount of uncollected revenue strongly supports the latest tax reform movement in the US and motivates exploration of further research into the various models of tax evasion, and aggressive tax avoidance. The outcome of this study included a graphic model that represents the tax environment based-on emergent theory supported by analysis conducted. While only early

stage research, it is anticipated that both tax avoidance and tax evasion will be defined as major problems in the forthcoming research from this approach as additional data is collected.

## 6. Practical Application

This study aimed to contribute to the existing body of knowledge about taxation with a contribution that will aid in understanding the problems of tax avoidance and tax evasion for overseas Americans and U.S. business entities. Potentially 9 million Americans are effectively involved and an estimated 200 billion dollars in uncollected tax revenue identifies this problem as being major in scope. By defining the scope of tax evasion and tax avoidance, for this important group of corporations and overseas Americans, through a robust investigation, and subsequent development of the relevant model, a contribution to the literature is assured. This study will be of use to those on both sides of the issue, those trying to minimize their tax and those seeking to combat fraud and collect full taxes owed. The academic community should appreciate the neutral perspective provided by the researcher.

## References

- Alm, J. (2012). Measuring, Explaining, And controlling tax evasion, lessons from theory, Experiments, And field studies. *International Tax and Public Finance*, 19(1): 54-77.
- Anzai, S. and Matsuzawa, C. (2013). Missions of the Japanese National University Corporations in the 21st century, Content analysis of mission statements. *Academic Journal of Interdisciplinary Studies*, 2(3): 197. Available: <http://www.palgraveconnect.com/pc/doi/10.1057/9780230627710>
- Bargiela-Chiappini, F., Nickerson, C. and Plancken, B. (2008). Business Discourse. Available: <http://www.palgraveconnect.com/pc/doi/10.1057/9780230627710>
- Berinato, S. (2016). Visualizations that really work. *Harvard Business Review*: Available: <https://hbr.org/2016/06/visualizations-that-really-work>
- Bui, Q. and Sanger-Katz, M. (2017). Hacking the tax plan, Ways to profit off the Republican tax bill. *The Upshot, The New York Times*: Available: <https://www.nytimes.com/interactive/2017/12/12/upshot/tax-hacks.html>
- Buja, A., Swayne, D. F., Littman, M. L., Dean, N., Hofman, H. and Chen, L. (2008). Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics*, 17(2): 444-72.
- CA by the Numbers (2017). Consular Affairs by the Numbers. US Department of State, Bureau of Consular Affairs. Available: [https://travel.state.gov/content/dam/travel/CA\\_By\\_the\\_Numbers.pdf](https://travel.state.gov/content/dam/travel/CA_By_the_Numbers.pdf)
- Casey, D. and Murphy, K. (2009). Issues in using methodological triangulation in research. *Nurse Researcher*, 16: 40-55. Available: <http://dx.doi.org/10.7748/nr2009.07.16.4.40.c7160>
- Crowe, M. T. (1944). *The Moral Obligation of Paying Just Taxes*. The Catholic university of America. Studies in sacred theology, No. 84.
- degl'Innocenti, D. G. and Rablen, M. (2017). Income Tax avoidance and evasion, A narrow bracketing approach. *Public Finance Review*, 45(6): Available: <https://doi.org/10.1177/1091142116676362>
- Faul, F., Erdfelder, E., Buchner, A. and Lang, A. G. (2009). Statistical power analyses using G\*Power 3.1, Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4): 1149-60.
- Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, 14(3): 393-417.
- Gravelle, J. G. (2015). *Tax Havens, International tax avoidance and evasion*. Congressional Research Service. <https://fas.org/sgp/crs/misc/R40623.pdf>
- Gross, S., Kim, M., Schlosser, J., Mohtadi, C., Lluch, D. and Schneider, D., 2014. "Fostering computational thinking in engineering education, Challenges, Examples, And best practices." In *IEEE Global Engineering Education Conference (EDUCON), Istanbul, 2014*. pp. 450-59.
- Gupta, S., Mills, L. F. and Towery, E. M. (2014). The effect of mandatory financial disclosures of tax uncertainty on tax reporting and collection, The case of fin 48 and multistate tax avoidance. *Journal of the American Taxation Association*, 36(2): 231-39.
- Harvey, L. (2014). Beyond member-checking, A dialogic approach to the research interview. *International Journal of Research & Method in Education*, 38(1): 23-38.
- Hunston, S. (2002). *Corpora in applied Linguistics*. Cambridge University Press.
- IRS (2017). Abusive offshore tax avoidance schemes-talking points. Available: <https://www.irs.gov/businesses/small-businesses-self-employed/abusive-offshore-tax-avoidance-schemes-talking-points>
- K@W (2013). *Corporate tax avoidance: Can the system be fixed?* Knowledge at Wharton. <http://knowledge.wharton.upenn.edu/article/corporate-tax-avoidance-can-the-system-be-fixed/#comments>
- Ma, M. S. and Thomas, W. B., 2016. "Legal environment and corporate tax avoidance, Evidence from state tax." In *American Tax Association 2016, Mid Year Conference*.
- Marshall, S. (2016). *Corporate inversion and the impact on the United States*. Business, Lagrange.edu.: [http://www.lagrange.edu/resources/pdf/citations/2016/11\\_Marshall\\_Business.pdf](http://www.lagrange.edu/resources/pdf/citations/2016/11_Marshall_Business.pdf)
- Miller, A. H., 2016. "Computer-aided content analysis of the corpus of business discourse, A comparison of Accounting and HR learners." In *NETs 2016 Osaka Japan*.
- Miller, A. H. (2017a). Preparing students for career success in Accounting, The SCIL-based model with a focus on content analysis. *Transnational Journal of Business*: Available: <http://www.acbsp.org/members/group.aspx?id=143359>



- Miller, A. H., 2017b. "A corpus-based computer aided linguistic analysis of taxation learning outcomes." In *International Conference on Business, Big-Data, and Decision Sciences (2017 ICBBDD) August 2-4, 2017 at Chulalongkorn University, Bangkok, Thailand*.
- Miller, A. H. (2018). Using unsupervised machine learning to model tax practice learning theory. *International Journal of Engineering and Technology*, 7(2.4): 109-16. Available: <http://dx..org/10.14419/ijet.v7i2.4.13019>
- Miller, A. H., 2019. "Data modeling and visualization of tax strategies employed by overseas American individuals and firms." In *International Conference on Emerging Internet, Data and Web Technologies (EIDWT-2019), Fujairah, UAE*.
- Minami, T. and Ohura, Y. (2015). How student's attitude influences on learning achievement? An analysis of attitude representing words appearing in looking back evaluation texts. *International Journal of Database Theory and Application*, 8(2): 129-44. Available: <http://dx.doi.org/10.14257/ijdt.2015.8.2.13>
- Mori, J., Matsuo, Y., Ishizuka, M. and Faltings, B., 2004. "Keyword extraction from the web for FOAF metadata." In *Workshop on Friend of a Friend, Social Networking and the Semantic Web, Galway, Ireland*.
- Nie, L., Wang, M., Gao, Y., Zha, Z. J. and Chua, T. S. (2013). Beyond text QA, Multimedia answer generation by harvesting WEB information. *Transactions on Multimedia*, 15(2): 426-41.
- O'Neil, C. (2016). *Rogue Algorithms' and the dark side of big data*. Knowledge@Wharton. University of Pennsylvania: Wharton. [http://knowledge.wharton.upenn.edu/article/rogue-algorithms-dark-side-bigdata/?utm\\_source=kw\\_newsletter&utm\\_medium=email&utm\\_campaign=2016-09-22](http://knowledge.wharton.upenn.edu/article/rogue-algorithms-dark-side-bigdata/?utm_source=kw_newsletter&utm_medium=email&utm_campaign=2016-09-22)
- Pelet, J. E., Khan, J., Papadopoulou, P. and Bernardin, E. (2014). *M-Learning: Exploring the Use of Mobile Devices and Social Media*, in Ed., Baporikar, N. *Handbook of Research on Higher Education in the MENA Region: Policy and Practice*, Hershey. PA: IGI Global. 261-96.
- Posner, R. (2012). *Opinion, United States of America, Plaintiff-Appellee v. Deanna L. Costello, Defendant Appellant, No. 11-291*. Court of Appeals Seventh Circuit Court: U.S.: <https://law.justia.com/cases/federal/appellate-courts/ca7/11-2917/11-2917-2012-01-31.html>
- Simon, M. and Goes, J. (2016). Reliability and validity in qualitative studies. Dissertation Recipes. Available: <http://www.dissertationrecipes.com/reliability-validity-qualitative-studies/>
- Smith, G. (2014). Data and Intuition. *The Conglomerate*: Available: <http://www.theconglomerate.org/corpuslinguistics/>
- Stevenson, R. J. and Mahmut, M. K. (2013). Using response consistency to probe olfactory knowledge. *Chemical Senses*, 38(3): 237-49. Available: <https://doi.org/10.1093/chemse/bjs139>
- Tamura, T. (2011). Application of text-mining methodology to sociological analysis of internet text in Japan. *CAJS Monograph Series*, 1: 77-97.
- Tax Foundation (2017). State-Local Tax Burden Rankings FY 2012. Available: <https://taxfoundation.org/publications/state-local-tax-burden-rankings/>
- Teixeira, D. P. (2016). Attitudes on the Ethics of Tax Evasion, A survey of Banking Employees, Master Thesis. <<https://www.iseg.ulisboa.pt/aquila/getFile.do?fileId=868711&method=getFile>>
- Tenidoua, E., Valsamidisa, S., Petasakisa, I. and Mandilasa, A. (2015). Elenxis, An effective tool for the war against tax avoidance and evasion 7th International Conference, The Economies of Balkan and Eastern Europe Countries in the changed world, EBEEC 2015, May 8-10, 2015. *Procedia Economics and Finance*, (33): 303-12. Available: [http://ebeec.teikav.edu.gr/ebeec2015/documents/proceedings\\_2015.pdf](http://ebeec.teikav.edu.gr/ebeec2015/documents/proceedings_2015.pdf)
- Trochim, W. M. (1989). Outcome program matching and pattern theory. *Evaluation and Program Planning*, 12: 355-66. Available: <https://www.socialresearchmethods.net/research/Outcome%20Pattern%20Matching%20and%20Program%20Theory.pdf>
- Trochim, W. M. (2016). Hindsight is 20/20, Reflections on the evolution of concept mapping. *Evaluation and Program Planning*:
- UN (2017). Sustainable development goals. Available: <http://www.un.org/sustainabledevelopment/blog/2016/01/244-million-international-migrants-living-abroad-worldwide-new-un-statistics-reveal/>
- Vesanto, J. and Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3): Available: <http://ftp.it.murdoch.edu.au/units/ICT219/Papers%20for%20transfer/papers%20on%20Clustering/Clusterin g%20SOM.p>
- Yu, C. H., Jannasch-Pennell, A. and DiGangi, S. (2011). Compatibility between text mining and qualitative research in the perspectives of grounded theory, Content analysis, And reliability. *The Qualitative Report*, 16(3): 730-44. Available: <http://nsuworks.nova.edu/tqr/vol16/iss3/6/>

## APPENDIX

The following potentially abusive structures make it appear that an unrelated foreign entity is the owner of certain assets, though they are actually controlled by a US national, whose ownership may be obscured.

- Foreign trusts
- Foreign corporations
- Foreign (offshore) partnerships, LLCs and LLPs
- International Business Companies (IBCs)

- Offshore private annuities
- Private banking (U.S. and offshore)
- Personal investment companies
- Captive insurance companies
- Offshore bank accounts and credit cards
- Related-party loans" ([IRS, 2017](#)).

There is a list of means to repatriate funds to the US, that takes advantage of US tax treatment of foreign persons or lie in a gray area that may be legally defensible ([IRS, 2017](#)). These means include the following:

- Credit cards which simply draw on the U.S. taxpayer's offshore account
- Loans from mystery offshore lenders
- Loans from domestic lenders in amounts beyond the taxpayer's apparent borrowing power (may be secured by offsetting deposits of offshore funds)
- Use of property titled to offshore entities at zero or below-market rental
- Bogus transactions designed simply to transfer funds to or from offshore entities, such as sales of property to offshore entities in jurisdictions where it is unlikely the property will actually be used or sold
- Gifts
- Scholarships for taxpayer's children
- Payable Through accounts ([IRS, 2017](#)).