Migration Letters

Volume: 21, No: S4 (2024), pp. 283-290

ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)

www.migrationletters.com

Modeling Expatriate Tax Evasion Using Unsupervised Machine Learning

*Alfred Howard Miller[1], Shaindra Sewbaran[2], Fatmah Alsereidi[1], Fatmah Kendi[1] and Saleimah Sebait

[1]

[2]

Abstract.

The level of tax noncompliance amounts to a very high sum globally. For the United States alone, there is a projected tax shortfall of $600 billion US dollars per year, which is $7 trillion over ten years, not adjusted for inflation [1]. Under the current system, it is estimated that 15% of individual taxes are uncollected [2]. This research aims to model the phenomenon of tax avoidance and evasion through the administration and collection of 67 survey responses and 22 expert interviews, collected by the researchers. While tax evasion is criminally illegal and tax avoidance is more about exploiting existing loopholes in the code, both result in less revenue being collected. Understanding the current state of tax evasion and avoidance for expatriates is important, given the global shortfall in tax revenue.

Keywords: *KH coder, unsupervised machine learning, self-organizing map, AI, text analytics, natural language processing, tax evasion, tax avoidance, expatriate.*

1 Introduction

1.1 Studies that have Addressed the Problem

Most articles written on tax evasion are published in tax practitioner journals and assume a practitioner or legal perspective. The random audit, used for investigating the practitioner and legal approach, is the academic gold standard for detecting tax evasion according to the classic model of Allingham and Sandmo (1972). Random sampling of tax returns for audit, coupled with the fear of getting caught, were a strong deterrent as long as the penalty was severe, and the chance of evasion being detected was high [3].

In a more recent study, [4] used an interpretation of the results of the defining issues test (DIT), comparing private tax practitioners to government tax practitioners and a non-tax control group. It was clear from the analysis that "private sector tax practitioners' reasoning in a tax context i[1]s different from that of government revenue tax practitioners and the non-tax control group". Tax practitioners faced a role conflict between duty to their client and duty to society as a whole. While this was an empirical study, a key limitation noted is the context, the study took place in Ireland and secondly, the useable sample size of 201 yielded 101 tax practitioners, 77 from the private sector, and 24 government revenue tax practitioners, with 100 respondents allocated to the non-tax control group by the authors—was a small sample size [4].

[1]Higher Colleges of Technology, Fujairah Women's College
[2]Higher Colleges of Technology, Abu Dhabi Men's College
* Corresponding Author

There is also a second competing perspective, based on approaches rooted in public finance or economics. Public finance is the role of government in the economy, and according to Stiglitz, former Chairman of the Council of Economic Advisers and Chief Economist of the World Bank, and a rigorous practitioner of the applied branch of economics. Drawing upon our knowledge of theoretical tax and expenditure policy, and advances in the economics of the public sector, the noted economist, Stiglitz [5] claims, we should know how to design better tax systems, reducing the dead weight loss, and use public sector market mechanisms to improve its efficiency and efficacy. However, we are faced with pervasive market failures in taxation, and complementary failure between the public and private sectors in public finance. This is because market failures in taxation have repeatedly had large systemic consequences affecting public finance.

Using a leaked data set of customer records from offshore financial institutions that was matched to administrative wealth records for Scandinavian countries, researchers demonstrated that offshore tax evasion was highly concentrated among the richest 0.01% households. The highly skewed distribution of offshore wealth saw an implied evasion of 25% of their taxes. However, in comparison to a stratified random audit, less than 5% of evasion by any group was detected throughout the entire distribution. This led the authors to conclude true wealth inequality failed to factor in accounting for unreported assets [6]. Key points raised by [7], is validity of behavioral approaches, [8] and [9], new theories will be outside mainstream economics and that the modeling individual behavior must shift to modeling of aggregated group behavior, and that it is important to conduct empirical field experiments [7]

A third approach for studies on tax compliance and tax evasion is from the perspective of ethics and morals. The ethics-based approach to taxation studies, while less common than the other two approaches, is nevertheless a widely deployed channel. One researcher, R. W. McGee [10] has authored or co-authored a stream of research of over 100 academic studies, with cumulatively thousands of citations.

The seminal approach of Crowe (1944) [11] and as further deployed by McGee and Smith [12] is to approach the moral and ethical issue by simultaneously investigating the three general contexts of moral and ethical judgement. These approaches are; 1) to examine the relationship of the individual to the state, 2) another is the relationship between the individual and the taxpaying community or a subset of this group and 3) the relationship of the individual to God. One advantage of the ethical and moral approach is the wide deployment across a range of contexts, allowing comparisons, many of which are US-based populations.

1.2 Deficiencies in Past Studies

With Allingham and Sandmo's (1972) model [3] there is an implicit assumption that audits lead to the detection of all tax evasion. In practice, however, random audits are not always successful in detecting tax evasion. Random audits are typically designed around detection of unreported self-employment income, exaggerated deductions and tax credit misuse. The Doyle, Frecknall-Hughes and Summers (2022) study [4], while current, demonstrated the limitation of not achieving a sufficient sample size in an empirical study, and targeting a different nationality than overseas Americans, preventing statistical generalizability of the results. Alstadsæter, Johannesen, and Zucman, of 2019 [6] identified a 25% evasion rate against a corresponding 5% detection rate, but based on leaked financial data which is not a repeatedly dependable source.

New research contends that current audit technology might fail to identify today's complex evasion tactics. This is because uncovering these tactics requires more information, greater

resources and more tax audit expertise than is available. It also requires more personnel than tax authorities have available for their random audit programs [13], and [14].

Alm [7] reports deficiencies in current studies. First, how to measure the extent of evasion, given it is a behavior that individuals and businesses will hide and obfuscate. The second paradigm, the economic and public finance perspective, recognizes that while enforcement plays a role, the tax administration functions as both a facilitator of tax collection, and a provider of services to the citizen- taxpayers. Alm's critique is that economic and public finance fails to account for the human behavioral factors of tax evasion. New theories will likely fall outside the mainstream of economics, and may indeed leverage social sciences such as psychology, sociology and anthropology to better understand naturally occurring features of taxation, and how they affect individual and business decisions.

Alm [7] cites the notion of "reciprocity" as an anthropological and sociological adherence to group norms. Using alternative perspectives on human behavior cannot help but expand our understanding of tax evasion behavior. The human factor, missing from the economics and public finance perspective is demonstrated most clearly in the Keynesian, governmental interventionist work of [8] and [9], to recognize the pervasive effects of confidence, fear, bad faith, corruption and concern for fairness.

2.1 Research Question

The researchers have proposed a research question upon which to gather data for quantitative and qualitative analysis. The purpose is to conduct sociological research on modeling tax evasion by expatriates. Representative applied studies in the literature review produced valid sociological inferences from similar data.

Past tax avoidance studies used the self-organizing map [15] to identify that business stakeholders, tax evasion and tax avoidance were tangentially separate from each other. A different study found evidence of gaming the system through effective tax planning using creative compliance [16]. Because these selected studies successfully integrated various types of data for analysis, such as the content of mass and social network communication, social research data, moral and ethical scaling, and government proceedings, it is surmised that relevant applied research models of expatriate taxpayers can be obtained by employing these applications. The goal of this study is to present models supported by facts and arguments for a better understanding of tax evasion and tax strategies employed by expatriates.

RQ1. What model(s) motivate practitioner and legal perspectives surrounding the pursuit of tax evasion and tax avoidance by expatriate taxpayers?

3.1 Research Approach

Research question 1 will be pursued concurrently using artificial intelligence (AI) as unsupervised machine learning, which yields a network modeling approach. To explore these first two perspectives, one is about the legal and practitioner perspective, the other is about the economic and public finance perspective, and interviews will be deployed to gain a qualitative overview from each person willing to respond about their direct involvement in expatriate taxation. This data will support a content analysis approach using natural language processing (NLP) [17], [18]. The content analysis will be motivated by natural language processing, using sentiment analysis, quantitative machine learning and artificial intelligence methodology, as text analytics software [18].

"Machine learning is defined as teaching machines to learn about something without explicit

programming." (personal communication, S. Shilbeyah, March 2, 2022).

The data set is built from the qualitative factors gathered from 22 expert interviews and 67 surveys, along with any outlier factors identified empirically from the survey results. Factors are coded from interview responses and the results of the open-ended survey question added to the Moral Obligation of Answering Just Taxes survey. The quantitative results from the survey identified four constructs that yielded a significantly above average response. These are:

1) Tax evasion is ethical if a significant portion of the money collected ends up in the pockets of corrupt politicians or their families and friends
2) Tax evasion is ethical if some of the proceeds go to support a war that I consider to be unjust
3) Tax evasion is ethical if I can't afford to pay
4) Tax evasion is ethical if the government imprisons people for their political opinions

The research technique is to manually code these constructs and combining them, where they are similar, into factors. It is not believed that data saturation was achieved with only 67 survey responses and 22 expert interviews, as new constructs relevant to tax evasion and tax avoidance, are continuing to be uncovered. For example, the sheer complexity of the US tax code which many US expatriates do not understand, and therefore they seek ways to avoid filling out their tax return entirely (Respondent email) and source [19].

The data from the data mining will be scrubbed for tags and non-text compatible features. Clean data will be preprocessed by a gatekeeper technology application prior to the text analytics. A text analytics processor can use unsupervised machine learning to analyze content of the data set. If data is too large to be analyzed effectively, given time constraints and the limitations of computer processing capability, a bootstrap sampling procedure will be used (K. Higuchi, personal communication, July 2020). A mainstream text analytics software program will be used to generate network diagrams and the investigator will study the output to determine what applies to answer each research question. Each type of network diagram features its own strengths. The main idea is to evaluate the models produced by the analytics software by their inherent criteria to determine the best tax evasion construct models.

Network diagrams will be interpreted qualitatively based on network structure using the proximity and physical arrangement of factors. Quantitative features such stress factors of model fit, outlier data points, factor proximity and communities will be noted and interpreted. Technologies available include hierarchical cluster analysis, correspondence analysis, co-occurrence network, multidimensional scaling, and the self-organizing map. By confirming the original data and explaining the visual results, the researchers propose to model the structural phenomena of tax evasion factors representative of expatriate taxpayers.

3.2 Self-Organizing Map

The self-organizing map (SOM) is widely deployed in academic studies. It has a neural network algorithm that can be used to map multidimensional data, to find a novel two-dimensional map of tax evasion factors using unsupervised learning,

[20]. SOM will allow the creation of a tax evasion and avoidance model through; 1) feature detection, 2) dimensionality reduction, 3) association, and 4) clustering.

Vesanto and Alhoniemi [21] advocated the self-organizing map (SOM) 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.

The self-organizing map motivates exploration of associations between words by creating a matrix output which resembles an abstract map, created from variables using their Euclidean distance in the data set. Each word is standardized and Euclidean distance is calculated using the same methodology as Multi-Dimensional Scaling. One drawback with the self-organizing map is using the default settings; it may take over an hour to process 70 words embedded in 1,200 documents, using a Core i7 CPU PC. For this reason, K. Higuchi (Personal Communication 2020) has advised utilizing a bootstrap sampling regime [18].

Some self-organizing map software programs offer several coloring options. In the mapping scheme review, pink color represents the most distant factor, and a solid pink line between nodes signifies vast distance between clusters. In contrast, blue nodes indicate that these words have a similar appearance patterns and form a cluster. [22], and [18].

4.1 Discussion

To interpret the self-organizing map shown in Figure 1, pink color or a close shade of color such as red or orange indicate distance, that there is a large difference in vectors of the neighboring nodes. The constructs shaded blue, purple and green are proximally related. Those constructs that are gray or close to gray are relatively neutral.

The Self-organizing map achieved a best fit with a 6-factor deployment. The model was specified on the basis of constructs. Factors identified include;
1) Tax Challenges, 2) Tax Evasion, 3) Building a Strategy, 4) Ethical Issues, 5) Education, Income and Wealth, and 6) Special Taxation Training. Notice how special Taxation training is separated or even physically blocked from the other construct by Education Income and Wealth. This phenomenon is new and not uncovered during earlier iterations of this research stream such as [23], Data Modeling and Visualization of Tax Strategies Employed by Overseas American Individuals and Firms. This finding implies that access to special tax training is only available based on education, income and wealth. The ethical issues construct is significant as there is a currently a strong push toward ESG, ethics, sustainability and governance.

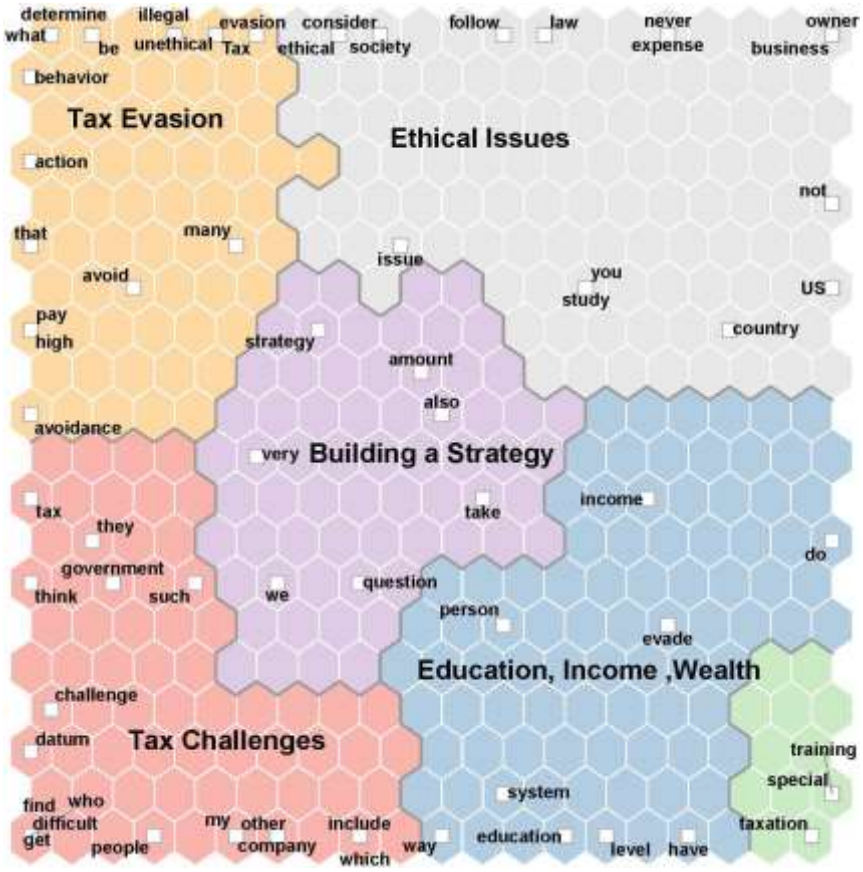

Fig. 1. Self-organizing map for tax behavior in the expatriate community

5.1 Conclusion

While the self-organizing map achieved adequate fit, and relevant analytics were performed; the sample size was too small and would not meet traditional quantitative methods for determining sample size. Sample size can be expanded in three ways. 1) additional surveys, 2) further expert interviews and 3) deployment of a web scraper to automate big data collection. A web scraper would enable tapping into a wide range of sites such as Reddit, Twitter, and tax specific websites, and forum discussions, via each sites API. While the original data collection regime had undergone two rounds of successful ethical approval by the IRB at Higher Colleges of Technology, continuation of the study by the researchers would require an additional IRB proposal for continued data collection.

References

1. DeBacker, J., Heim, B., Tran, A., & Yuskavage, A. (2020). Tax Noncompliance and Measures of Income Inequality. Tax Notes. Available from https://www.taxnotes.com/tax-notes- federal/compliance/tax-noncompliance-and-measures-income-inequality/2020/02/17/2c3y5

2. Sarin, N, (2021). The Case for a Robust Attack on the Tax Gap. US Department of the Treasury, Retrieved from https://home.treasury.gov/news/featured-stories/the-case-for-a-robust-attack-on-the-tax-gap?ftag=YHF4eb9d17

3. Allingham, M. G., Sandmo, A. (1972). Income tax evasion: a theoretical analysis. Income tax evasion: a theoretical analysis. Journal of Public Economics, (1)3–4, 323-338,

https://doi.org/10.1016/0047-
2727(72)90010-2.

4. Doyle, E., Frecknall-Hughes, J., & Summers, B. (2022). Ethical reasoning in tax practice: Law or is there more? Journal of International Accounting, Auditing and Taxation, 48. Doi: 10.1016/j.intaccaudtax.2022.100483.

5. Stiglitz, J. E., (2002). New perspectives on public finance: recent achievements and future challenges, Journal of Public Economics, 86(3), pages 341-36

6. Alstadsæter, A., Johannesen, N. & Zucman, G., (2019). Tax Evasion and inequality. American Economic Review, 109 (6): 2073-2103. doi: 10.1257/aer.20172043

7. 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.

8. Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. The Quarterly Journal of Economics, 115 (3), 715-753.

9. Akerlof, G. A., & Shiller, R. J. (2009). Animal Spirits: How Human Psychology Drives the Economy and Why It Matters. Princeton, NJ: Princeton University Press

10. McGee, R. W. (Ed.). (2011). The ethics of tax evasion: Perspectives in theory and practice. Springer Science & Business Media.

11. Crowe, M. T. (1944). The Moral Obligation of Paying Just Taxes. The Catholic University of America. Studies in Sacred Theology, No. 84.

12. McGee, R. W. and Smith, S. R. (2008). Opinions on the Ethics of Tax Evasion: A Comparative Study of Utah and New Jersey. SSRN Available at https://ssrn.com/abstract=1118140 or http://dx.doi.org/10.2139/ssrn.1118140

13. OECD (2021a). Tax administrations continue to accelerate their digital transformation. Retrieved from https://www.oecd.org/ctp/administration/tax-administrations-continue-to-accelerate-their-digital-transformation.htm#:~:text=Artificial%20intelligence%20and%20machine%20learning,be%20deployed%20to%20other%20areas.

14. OECD (2021b). Tax Inspectors Without Borders continues to significantly boost domestic revenue mobilisation in spite of COVID-19 crisis. OECD Tax. Retrieved from: https://www.oecd.org/tax/tax- inspectors-without-borders-continues-to-significantly-boost-domestic- revenue-mobilisation-in-spite-of-covid-19-crisis.htm

15. Miller, A. H., (2017). A corpus-based computer aided linguistic analysis of taxation learning outcomes. In International Conference on Business, Big-Data, and Decision Sciences (2017 ICBBD) August 2-4, 2017 at Chulalongkorn University, Bangkok, Thailand.

16. Onu, D., Oats, L. and Kirchler, E. (2019), The Dynamics of Internalised and Extrinsic Motivation in the Ethical Decision-Making of Small Business Owners. Applied Psychology, 68: 177-201. https://doi.org/10.1111/apps.12151

17. Meena, S. (2022). Statistics for analytics and data science: Hypothesis testing and Z-test vs. T-test. Analytics Vidhya. Retrieved October 29 2022 from: https://eur02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.analyticsvidhya.com%2Fblog%2F2020%2F06%2Fstatistics- analytics-hypothesis-testing-z-test-t- test%2F&data=05%7C01%7Camiller%40hct.ac.ae%7Cbf605780ffec4106a0ab08dab9871aa7%7C55488759d4c94a95ae92ada1488c4053%7C0%7C0%7C638026287292271227%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJ

XVCI6Mn0%3D%7C2000%7C%7C%7C&sdata=Z%2F7x5gzAGJl
d4CbzyAZskeySmBOPnWyax26gNNwkMn4%3D&reserved=0

18. Higuchi, K. (2016). KH Coder 3 reference manual. Kioto (Japan): Ritsumeikan University.

19. Graffy, C. (October 10, 2015). The law that makes U.S. expats toxic: A measure targeting tax evasion pushes U.S. Americans out of bank accounts--and jobs--abroad. Wall Street Journal, Opinion.

20. Kallio, M. & Back, B. (2011). The self-organizing map in selecting companies for tax audit. Contributions to Accounting, Auditing and Internal Control. Essays in Honour of Professor Teija Laitinen. Acta Wasaensia 234(8), 45–59. Eds Jokipii, A. & Miettinen, J. (8) The Self-Organizing Map in Selecting Companies for Tax Audit. Available from: https://econpapers.repec.org/bookchap/
sprsprchp/978-3-7908-2739-2_5f27.htm

21. 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/pap
ers%20on%20Clustering/Clusterin g%20SOM.p

22. Yin, H. (2008). The Self-Organizing Maps: Background, Theories, Extensions and Applications, Studies in Computational Intelligence (SCI) 115, 715–762.

23. Miller, A. H. (2019). Data modeling and visualization of tax strategies employed by overseas American individuals and firms. In International Conference on Emerging Internetworking, Data & Web Technologies, 309-321. Springer, Cham.