eJournal of Tax Research (2025) vol. 23, no. 1, pp. 9-31

How CHAID analysis can assist revenue authorities in analysing tax compliance costs

Karen Stark,* Sharon Smulders** and Elza Odendaal***

Abstract

Tax compliance costs represent an economic burden to society and can result in reduced tax compliance behaviour. Traditional techniques used to establish the determinants of tax compliance costs include, inter alia, regression and simple descriptive statistics. This article explains how a Chi-square automatic interaction detection (CHAID) analysis, a decision tree modelling technique, was used to analyse the tax compliance costs of 10,260 individual taxpayers in South Africa. CHAID analysis provided granular insights beyond traditional techniques to enable a better understanding of the determinants which could lead to targeted support to enhance taxpayer compliance and reduce government collection costs.

Keywords: tax compliance costs, personal income tax, determinants, Chi-square automatic interaction detection (CHAID)

* Senior Lecturer, Department of Taxation, University of Pretoria, South Africa. Email: karen.stark@up.ac.za. Research was conducted as a doctoral student at the University of South Africa (UNISA). This work was supported by: the South African Institute of Chartered Accountants that provided a research grant that initiated this study; the Women in Research grant from UNISA, which also made this research possible and the South African Revenue Service that assisted with input into and distribution of the research questionnaire.

** Professor Extraordinarius, Department of Financial Intelligence, University of South Africa (UNISA), South Africa.

*** Professor, Department of Auditing, University of South Africa (UNISA), South Africa.

1. INTRODUCTION AND MOTIVATION

The ultimate goal of tax policies is to raise sufficient revenue in a manner that minimises collection and enforcement costs (Alm, 1996). Tax compliance costs reduce the economic resources of taxpayers without providing any increase in the government's fiscal budget and are thus an economic problem for society (Eichfelder & Hechtner, 2018). Tax compliance costs also negatively affect entrepreneurship (Djankov et al., 2002) and impair the overall compliance within a tax system (Alm et al., 2010).

A review by Bruce-Twum and Schutte (2021) of 33 tax compliance cost studies published between 1992 and 2018 indicated that the primary technique used in analysing tax compliance cost data was regression (67%), while descriptive statistics were utilised in 33% of the studies and correlation in 18%.[1] Other analysis techniques used included the Individual Taxpayer Burden Model, Panel Data Analysis, Extrapolation, Simulation and Matrix Analysis (Bruce-Twum & Schutte, 2021). None were found using the Chi-square automatic interaction detection (CHAID) technique. CHAID offers a number of advantages over several more commonly used analysis techniques (such as regression) because it is non-parametric and non-linear (Önder & Uyar, 2017) and thus accommodates categorical variables without requiring data normality. It also segments data into subgroups based on statistically significant characteristics and then reveals unique patterns across different taxpayer groups, enhancing understanding of tax compliance cost determinants.

For this study, both multiple linear regression analysis, using forward stepwise modelling, and the CHAID technique were used to establish the determinants of tax compliance costs for individual taxpayers in South Africa for the 2018 year of assessment. The CHAID technique enabled the determination of patterns among the categorical variables considered as independent variables. This not only assisted in identifying the characteristics that will influence tax compliance costs but also in identifying different groups of taxpayers with different characteristics which have different levels of tax compliance costs. Knowing the determinants of tax compliance costs for the different taxpayer groups could assist with tax policy decisions and pinpoint areas that could alleviate the taxpayers' compliance burden.

The next section explains the methodology adopted in collecting the data and calculating the tax compliance costs; this is followed by a detailed discussion of the analyses performed, first on the scale items (for possible additional determinants), then the multiple linear regression. This is followed by the CHAID technique to explore linear and non-linear determinants of tax compliance costs. The findings of the regression and CHAID analysis are discussed and the conclusion is presented thereafter.

2. METHODOLOGY

A positivist research philosophy, functionalist paradigm and quantitative design were followed. An online questionnaire was used to collect the survey data, enabling various statistical techniques to be performed to analyse the data and quantify the tax compliance costs.

[1] The total percentages added to 161% since the coding was exhaustive.

2.1 The survey instrument

The online questionnaire was distributed by the South African Revenue Service (SARS) during March and April 2019 to a stratified random sample of individual taxpayers. The questionnaire included all the elements necessary to calculate tax compliance costs and to explore the determinants thereof. All tax-related activities of taxpayers were included as 'legitimate' tax compliance activities, in line with Tran-Nam, Evans and Lignier (2014), which included post-filing activities such as fulfilling verification and audit requests; lodging an objection; utilising the alternative dispute resolution process; and further litigation.

The data relevant to the determinants were obtained from responses regarding: Employment status (Q2), Who completed the tax return (Q4.1), Demographic questions (Q10.1-Q10.5) and Scale items (Q11.1-Q11.7). Some of the scale items have not previously been used in the context of tax compliance cost studies but were included in the questionnaire to test possible determinants and are explained in more detail later in the article.

Pilot testing of the questionnaire was performed and necessary adjustments were made. Lastly, the necessary ethical clearance was obtained and research ethics principles were adhered to during the research.

2.2 Response rate

In terms of the researchers' agreement with SARS, the questionnaire was delivered and administered by SARS. It was thus necessary for the researchers to rely on information and reports received from SARS relating to the distribution of the questionnaire. Despite this challenge, 14,100 responses were received and compared to the 145,037 persons who opened the email sent by SARS; the response rate equates to 9.7%. After the data cleaning process, the response rate drops to 7.07% (10,260 fully completed questionnaires). Even though the response rate is lower than the 29% obtained by Sapiei and Abdullah (2008) and 13.4% obtained by Tran-Nam et al. (2014), it is higher than the 2.36% obtained by Chattopadhyay and Das-Gupta (2002) and is thus deemed acceptable.

A wave analysis was conducted and assuming that late respondents can be used as a proxy for non-respondents, no non-response bias was detected for the survey.

2.3 Representativeness

The representativeness of the 10,260 responses collected was assessed against the population of individual taxpayers who submitted income tax returns for the 2018 year of assessment. The sample was found to be a good representation based on publicly available demographic characteristics such as taxpayer's location, age, gender and level of income (National Treasury & SARS, 2019) and thus no weighting was regarded necessary when determining the tax compliance costs.

2.4 Valuation of respondents' time

The valuation of a respondent's time was considered by requesting full-time employees to provide their gross monthly salaries (given that they are paid monthly) and asking for the (before-tax) hourly wage amounts from the other respondents (namely part-time employed, self-employed, retired or unemployed respondents). A follow-up question

regarding the respondent's income tax bracket helped to identify the marginal tax rate, to allow for using more than one valuation method of a respondent's time. Various valuation methods of valuing taxpayers' time have been recognised in the literature (Pope 1989, 1995); for example, each individual's valuation of their time or the same hourly value could be used for all respondents. These values could then be adjusted for tax and/or subjected to certain limitations. Six valuation methods were used to obtain a lower- and upper-bound estimate of tax compliance costs[2] but for the purposes of this article the tax compliance costs based on the weighted average after-tax hourly rates will be used.

3. SCALE ITEMS

The focus of this research was to examine taxpayer demographics, income, employment status and how SARS interactions influence tax compliance costs. The last factor, how SARS interactions influence tax compliance costs, was determined by including rating questions (scale items) in the questionnaire. These questions established whether or not taxpayers perceived the tax legislation and tax guidance issued by SARS as complex since the complexity of tax legislation has been identified as a determinant of tax compliance costs (Vaillancourt, 1989; Evans, 2003; Steyn, 2011). Another reason for including these questions was to consider the service orientation of SARS because Eichfelder and Kegels (2014) have found empirical evidence that the service orientation of a tax authority (or lack thereof) is significantly associated with higher tax compliance costs. From the perspective of the 'slippery slope' framework, the service orientation of tax authorities is not only relevant for the trust dimension of compliance (voluntary compliance) but also for the power dimension of compliance (enforced compliance): it can result in a cost advantage (for a customer-friendly administration) or a cost disadvantage (for a control-oriented administration) (Kirchler et al., 2008; Eichfelder & Kegels, 2014).

The questions in the questionnaire using scale items therefore related to the perceptions of the survey respondents regarding reasons for tax compliance (Q11.1); complexity (Q11.2); SARS's service quality (Q11.4); appeal procedures (Q11.5); assessments, audits and penalties (Q11.6); and SARS's consultation and communication regarding changes to its system (Q11.7). A 5-point Likert-style rating (1 = strongly disagree, 5 = strongly agree) was used to measure these constructs.

The first part of the analysis of the data discussed next provides a high-level overview of the extent to which the respondents agreed with the scale items provided. This is followed by an exploratory factor analysis to determine if the underlying factors could be seen as a unidimensional construct, which can then be used in the regression analyses. The identified factors that displayed satisfactory reliability, based on their Cronbach Alpha values, were then analysed further.

3.1 High-level overview of respondents' agreement with scale items

The first two questions containing scale items (Q11.1 and Q11.2) and the last question (Q11.7) were open to all respondents, but Q11.4 was only open to respondents who answered 'yes' to the following question (Q11.3): 'Did you have any interaction (via telephone, e-mail or in person) with a SARS official in the last 12 months?'. This was

[2] Refer to detailed discussion in sections 3.3 and 5.7 of Stark (2020).

done to ensure that only those taxpayers who had dealt with SARS directly provided answers to this question, thereby preventing respondents from answering based on hearsay. For the same reason, Q11.5 could only be answered by respondents who had used the appeal procedures, and Q11.6 by respondents who had been audited.

Table 1 sets out the mean of each scale item to provide an overview of the extent to which the respondents agreed with the items. For each item the minimum awarded was 1 (strongly disagree), 3 was neutral, and the maximum was 5 (strongly agree).

Table 1: Mean Ratings of Each Scale Item

	N	MEAN
Q11.1		
I pay my taxes to support the state and other citizens.		3.59
I pay my taxes without spending a long time thinking how I could reduce my tax.	9277	3.49
I pay my taxes reluctantly because the tax revenue is not spent appropriately.	9354	3.88
I pay my taxes even though I know that others do not.	9145	3.87
I pay my taxes because the punishment for tax evasion is severe.	9217	3.79
Q11.2		
Tax legislation is easy to read.	9244	2.94
Tax legislation is easy to understand.	9143	2.86
Guides issued by SARS are easy to read.	9250	3.30
Guides issued by SARS are easy to understand.	9217	3.25
It is easy to find the provisions in the tax legislation that apply to me.	8948	3.00
Q11.4		
SARS officials gave me precise answers (i.e. not vague).	4149	3.44
The information obtained from SARS corresponded to my needs.	4116	3.46
The answers provided by SARS were consistent regardless of who provided them (for example staff at different branches).	4104	3.30
SARS officials approached their jobs with professionalism and dedication.	4110	3.45
SARS officials are very capable to perform their jobs.	4105	3.40
Q11.5		

I have been able to express my views during those procedures.	237	2.73
I have had influence over the outcomes arrived at by those procedures.	233	2.42
Those procedures have been free of bias.	232	2.81
Those procedures have been fair and transparent.	231	2.47
Q11.6		
SARS investigates until it finds something.	3031	3.33
SARS primarily aims to punish.	3065	2.92
SARS's penalties are too severe relative to the offence.	2841	3.22
Q11.7		
SARS consults widely about how they might change things to make it easier for taxpayers to meet their obligations.	8619	3.04
SARS goes to great lengths to consult with the community over changes to its system.	8598	2.74
SARS communicates changes to its system clearly and effectively.	8850	2.91

From Table 1 it is clear that all items for Q11.1 and Q11.4 obtained a mean rating higher than 3, indicating for Q11.1 that there was a tendency to agree with coercive and persuasive reasons for tax compliance. For Q11.4 there was a tendency to agree with positive perceptions of SARS's service quality. The mean ratings of all the items for Q11.5 were lower than 3, indicating a tendency among respondents towards negative perceptions of their experiences during appeal procedures. The mean rating for the items in the other questions (Q11.2, Q11.6 and Q11.7) did not consistently tend towards either agreement or disagreement. Overall, the item that obtained the lowest mean rating (strongest disagreement) related to the appeal procedures – it stated: '*I have had influence over the outcomes arrived at by those procedures*'. The item that obtained the highest mean rating (strongest agreement) was '*I pay my taxes reluctantly because the tax revenue is not spent appropriately*'.

3.2 Exploratory factor analysis

An exploratory factor analysis was performed[3] in order to determine whether the items under each question constituted a unidimensional construct. The following six

[3] Principal axis factoring was used as the extraction method and Promax as the rotation method, as described by Costello and Osborne (2005). The appropriateness of conducting factor analysis was tested using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of Sphericity (Pallant, 2007). The KMO was equal to or higher than the acceptable threshold of 0.5 and the Bartlett's Test of Sphericity was statistically significant ($p < 0.001$) for the items in all the constructs (Pallant, 2007), indicating that a factor analysis was indeed appropriate for the constructs. For the '*reasons for tax compliance*' construct, two factors were identified based on the eigenvalue criterion (an eigenvalue greater than 1), thus indicating

unidimensional constructs, '*tax legislation complexity*' (Q11.2A), '*complexity of SARS guides*' (Q11.2B), '*SARS's service quality*' (Q11.4), '*appeal procedures*' (Q11.5), '*assessments, audits and penalties*' (Q11.6) and '*SARS's consultation and communication regarding changes to its system*' (Q11.7) were confirmed using the eigenvalue larger than 1 criterion and because the Cronbach Alpha values were all above the acknowledged exploratory threshold of 0.6 (as described by Field (2013)), the reliability was considered satisfactory.

The next section ascertains the determinants of the tax compliance costs of individual taxpayers based on the respondents' ratings of these six constructs and other variables (possible determinants). Results from the multiple linear regression analysis, using forward stepwise modelling, are presented first, followed by a CHAID decision tree modelling analysis to establish and interpret the most significant determinants of the tax compliance costs.

4. MULTIPLE LINEAR REGRESSION

Multiple linear regression modelling was conducted on the data to determine the determinants of tax compliance costs. Multiple regression modelling enables the identification of statistically significant predictors as well, enabling the comparison of the strength of the effect of each individual independent variable to the dependent variable by studying the standardised beta coefficients of each predictor. As it is known that large sample sizes do increase statistical significance, it was important to study the standardised beta coefficients in conjunction with the significance level indicated. The key determinants identified in the regression analysis were: self-employment (as compared to full-time employment), taxpayers who used paid help to submit their tax returns (in comparison to those who filed their own returns), and taxpayers in the higher tax brackets, specifically (in ZAR) 'R708,311 to R1,500,000' and 'Exceeding R1,500,000,' when compared to the reference tax bracket of 'Not exceeding R189,880'.

The regression analysis was followed by a CHAID analysis, which not only allowed for linear relationships but also non-linear relationships. It enabled the determination of patterns among the categorical variables considered as independent variables. This not only assisted in identifying the characteristics that will influence tax compliance costs but also in identifying different groups of taxpayers with different characteristics that have different levels of tax compliance costs.

---

that the construct was not unidimensional. The items that clustered on the same factor (Factor 1) appear to imply coercive compliance while the items of Factor 2 appear to imply persuasive compliance. However, as the Cronbach Alpha coefficient values for these factors were only 0.517 and 0.443 respectively, the reliability of these items was not considered acceptable. Hence, this construct was not considered further, in line with Pallant (2007). For all the remaining constructs the analysis identified only one factor in each instance based on the eigenvalue criterion (an eigenvalue greater than 1). The Cronbach Alpha values for the unidimensional constructs, '*tax legislation complexity*' (Q11.2A), '*complexity of SARS guides*' (Q11.2B), '*SARS's service quality*' (Q11.4), '*appeal procedures*' (Q11.5), '*assessments, audits and penalties*' (Q11.6) and '*SARS's consultation and communication regarding changes to its system*' (Q11.7) were all above the acknowledged exploratory threshold of 0.6 and were thus considered satisfactory, as described by Field (2013). Two of the constructs (*tax legislation complexity* and *complexity of SARS guides*) consist of two items each. Two-item constructs are more prone to estimation problems, especially when the sample size is small (Kline, 2005; Eisinga et al., 2013), but the large sample in this study and the very high correlation between the two items (> 0.7) made these constructs admissible.

5. CHAID DECISION TREES

The CHAID technique offers a number of advantages over several more commonly used statistical analysis techniques (such as regression) because it is non-parametric and non-linear (Önder & Uyar, 2017). A CHAID analysis starts by finding independent variables that have a statistically significant association with or best explain the dependent or target variable. In this research, the dependent variable was tax compliance costs. The possible determinants of tax compliance costs were the independent variables. It then assesses the category groupings, or interval breaks in the case of continuous variables, to pick the most significant combination – the independent variables homogeneous with respect to the dependent variable (Thomas & Galambos, 2004). These subgroups of the independent variables continue to generate more subgroups for significant variables (determinants), making the relationships between the subgroups visible. In other words, CHAID analysis shows the determinants and their interactions with the dependent variable via a tree diagram using the chi-square or F statistics, Bonferroni method and category merger algorithm (Kass, 1980; Önder & Uyar, 2017). The independent variable with the strongest association with tax compliance costs (the best determinant) becomes the first branch in a tree with a leaf for each category that is significantly different relative, in this case, to tax compliance costs (Thomas & Galambos, 2004).

Identifying the potential determinants of tax compliance costs to be specified as independent variables for this CHAID decision tree modelling analysis was thus very important. Guyton et al. (2003) divided the determinants of tax compliance costs into three groups, namely those relating to taxpayer characteristics, tax compliance methods and tax return complexity. Tran-Nam et al. (2014) seemed to combine elements of the last two groups by stating that the use of paid tax agents is an indicator of tax complexity and therefore an important determinant of tax compliance costs.

Based on the data obtained in this research, the *taxpayer characteristics* included as independent variables were employment status, taxable income, location, gender, age, education level and tax knowledge. The *tax compliance method* was represented by the variable on the type of assistance ('self', 'free help', 'paid help' or 'SARS employee'). Information regarding *complexity* was obtained from respondents' ratings of the constructs '*tax legislation complexity*' and '*SARS guides complexity*' and these factor-based variables were therefore also included as independent variables. It is accepted that the finding by Eichfelder and Kegels (2014) that a (lack of) *service orientation* by a tax authority is significantly connected with higher tax compliance costs, results in a fourth group of determinants of tax compliance costs. The respondents' ratings of the remaining four constructs (namely SARS's service quality, appeal procedures, assessments, audits and penalties, and SARS's consultation and communication regarding changes to its system) relate to SARS's service orientation. These were therefore also included as independent variables.

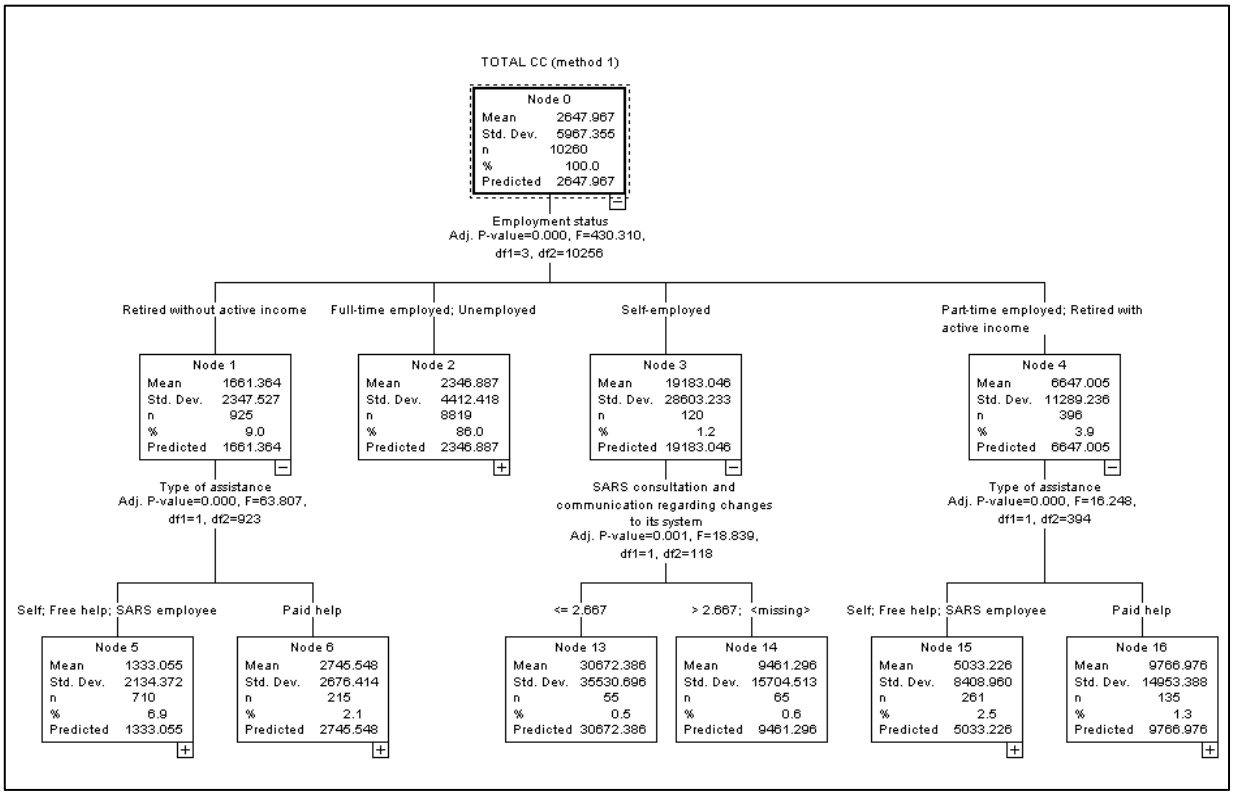
The results of the CHAID tree analyses will be presented next.

5.1 CHAID application to tax compliance costs

An example of the CHAID output from SPSS is provided in Figure 1 while subsequent Figures will present relevant information in a more user-friendly way. In Figure 1, Node 0 (also known as the 'trunk' of the tree) indicates the actual mean of the tax compliance costs of the 10,260 respondents using the weighted average after-tax hourly

rates valuation method[4] (abbreviated as 'TOTAL CC (method 1)' in the tree), namely R2,647.97 (Stark et al., 2024, p. 80).

Fig. 1: SPSS CHAID Output Example



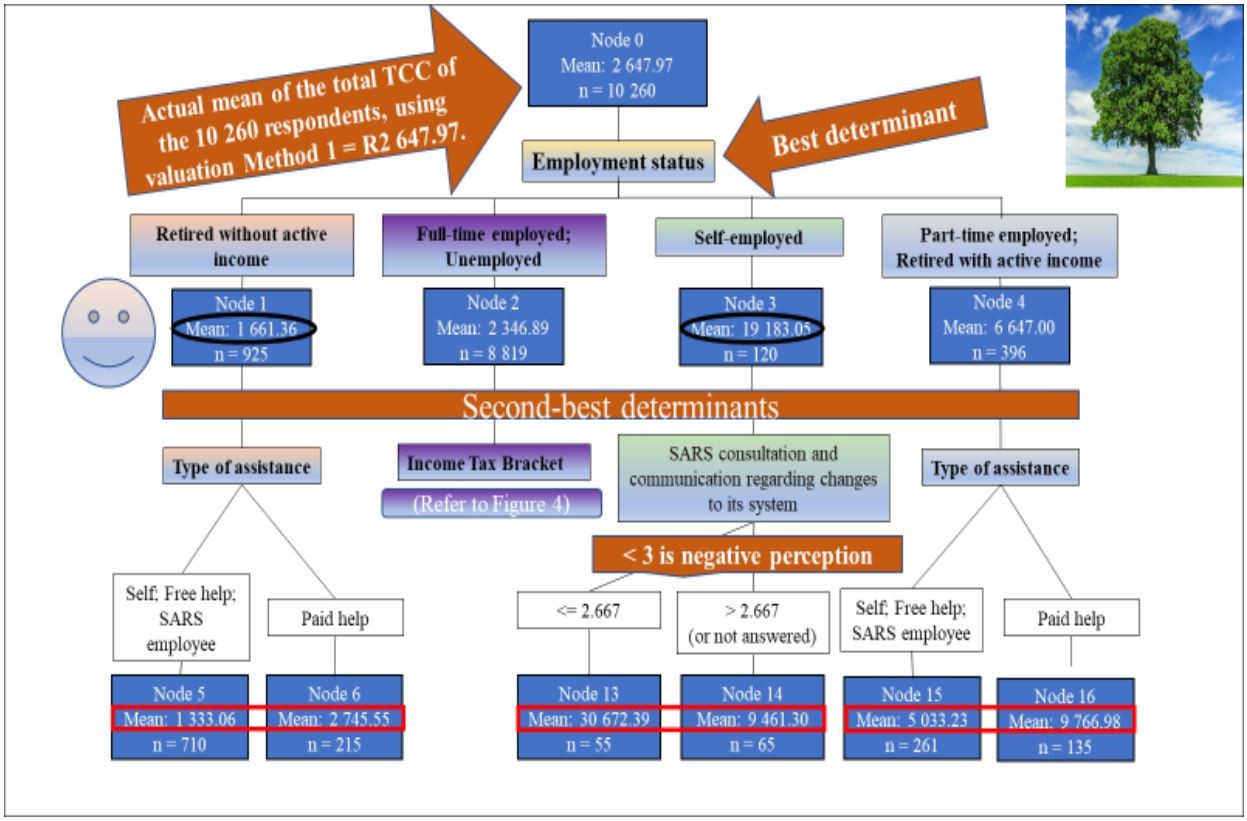
Source: authors' data and SPSS output.

From Figure 1, it is clear that the best determinant of the respondents' tax compliance costs was their *employment status* ($F=430.31$, $p<0.001$), indicated as the first layer of 'branches'. It is also clear that branches could further be expanded using the '+' indicators. To increase readability the visible branches will be managed in the discussion of the CHAID diagrams in the next figures.

Figure 2 presents the same information as contained in Figure 1 in a different format with additional clarifications.

[4] The first of the six methods to value an hour spent by respondents, as described in section 2.4.

Fig. 2: Tax Compliance Costs CHAID Tree with First and Second Layer of Branches Explained



Source: authors' data and diagram based on SPSS output.

As already pointed out in Figure 1 and indicated by the arrow on the right in Figure 2 the best determinant of the respondents' tax compliance costs was their *employment status*. The retired respondents without active income had the lowest average tax compliance costs (R1,661.36), and self-employed respondents' average tax compliance costs were the highest (R19,183.05). This corresponds to the findings of other studies[5] that employment status is a determinant of tax compliance costs and that self-employed individuals tend to have higher tax compliance costs than full-time employed individuals.

The second-best determinant of these respondents' tax compliance costs (the next layer of 'branches') depended on their employment status. For self-employed respondents, the second-best determinant was their perceptions relating to *SARS's consultation and communication of changes to its system*,[6] and *type of assistance* was the second-best

[5] For example, those by Allers (1994), Guyton et al. (2003), Lopes et al. (2012), Sandford et al. (1989) and Slemrod and Sorum (1984).
[6] $F=18.8$, $p<0.01$.

determinant for respondents who were retired without active income[7] or retired with active income/part-time employed.[8]

It is also clear from these second-best determinants in Figure 2 that respondents who used paid help[9] had on average almost double[10] the tax compliance costs of those who submitted income tax returns themselves, obtained free help from a family member or friend, or received assistance from a SARS employee. This finding that the use of paid assistance was a determinant of tax compliance costs corresponds to the literature.[11]

Figure 2 also illustrates that the tax compliance costs of self-employed respondents with a negative perception[12] of *SARS's consultation and communication on changes to its system* were on average three times[13] the costs of respondents who had a less negative perception.[14] This finding is in line with the finding by Eichfelder and Kegels (2014) that the lack of service orientation of a tax authority increases tax compliance costs.

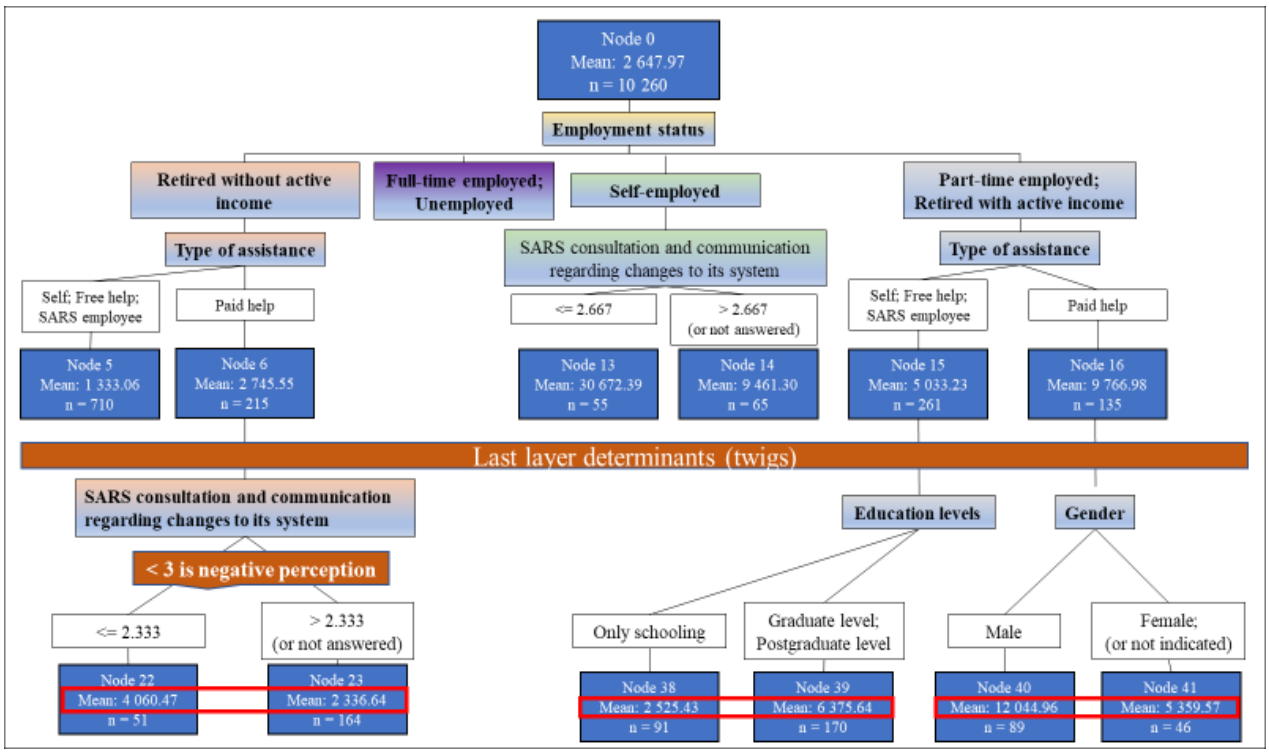
When the CHAID tree was expanded to the last layer[15] (also known as 'twigs'), three determinants emerged, namely *SARS's consultation and communication of changes to its system*,[16] *education level*[17] and *gender*[18] (refer to Figure 3). However, no further layer for self-employed respondents was formed, given the small number of respondents with this employment status.[19]

Considering the last layer of determinants presented in Figure 3, it was found that the tax compliance costs of retired respondents without active income and who used paid assistance, were further influenced by their perceptions of *SARS's consultation and communication regarding changes to its system*. A negative perception rating[20] resulted in almost double[21] the tax compliance costs than a less negative[22] (or no) perception. Improving or finding alternative ways to consult and communicate changes to the SARS systems to these individuals is thus warranted. Considering this is an older generation, using electronic means of communication may not be ideal but outreach in the form of paper pamphlets may be more suitable.

---

[7] F=63.8, p<0.001.
[8] F=16.2, p<0.001.
[9] Specifically, the retired (with and without active income) and part-time employed respondents.
[10] Comparing Nodes 5 and 6 (R1,333.06 and R2,745.55) and Nodes 15 and 16 (R5,033.23 and R9,766.98).
[11] For example, the studies by Blaufus et al. (2019) and Guyton et al. (2003).
[12] That was an average rating of 2.667 or less.
[13] Comparing Nodes 13 and 14 (R30,672.39 and R9,461.30).
[14] Keeping in mind that a rating of 3 was neutral, a rating above 2.667 may still be negative. Respondents who did not answer these questions were grouped with the respondents who had a rating of more than 2.667.
[15] For purpose of this analysis, the minimum number of cases for a new node was set at 40, which resulted in a maximum of three layers.
[16] F=17.3, p<0.01.
[17] F=13.0, p<0.01.
[18] F=6.3, p<0.05.
[19] The second layer of nodes (Nodes 13 and 14) consisted of only 55 and 65 respondents each, and the minimum number of cases (refer n 15, above) for a new node was not met. The second layer was thus closed in Figure 3 to enhance readability.
[20] A rating of 2.333 or less.
[21] Comparing Nodes 22 and 23 (R4,060.47 and R2,336.64).
[22] Keeping in mind that a rating of 3 was neutral, and therefore a rating above 2.333 may still be negative.

Fig. 3: Tax Compliance Costs CHAID Tree with Third-Layer Determinants



Source: authors' data and diagram based on SPSS output.

Figure 3 also illustrates that the tax compliance costs of respondents who were part-time employed or retired with active income and who did *not* use paid assistance, were further influenced by their *education* level. The tax compliance costs of respondents with graduate or postgraduate levels of education were on average two-and-a-half times[23] the tax compliance costs of those who only had a schooling level of education. Level of education was also reported as a determinant of tax compliance costs, for example, by Allers (1994), Klun (2004) and Mathieu et al. (2010). These findings confirm that the more educated a person is, the more time they tend to spend on their tax compliance activities.

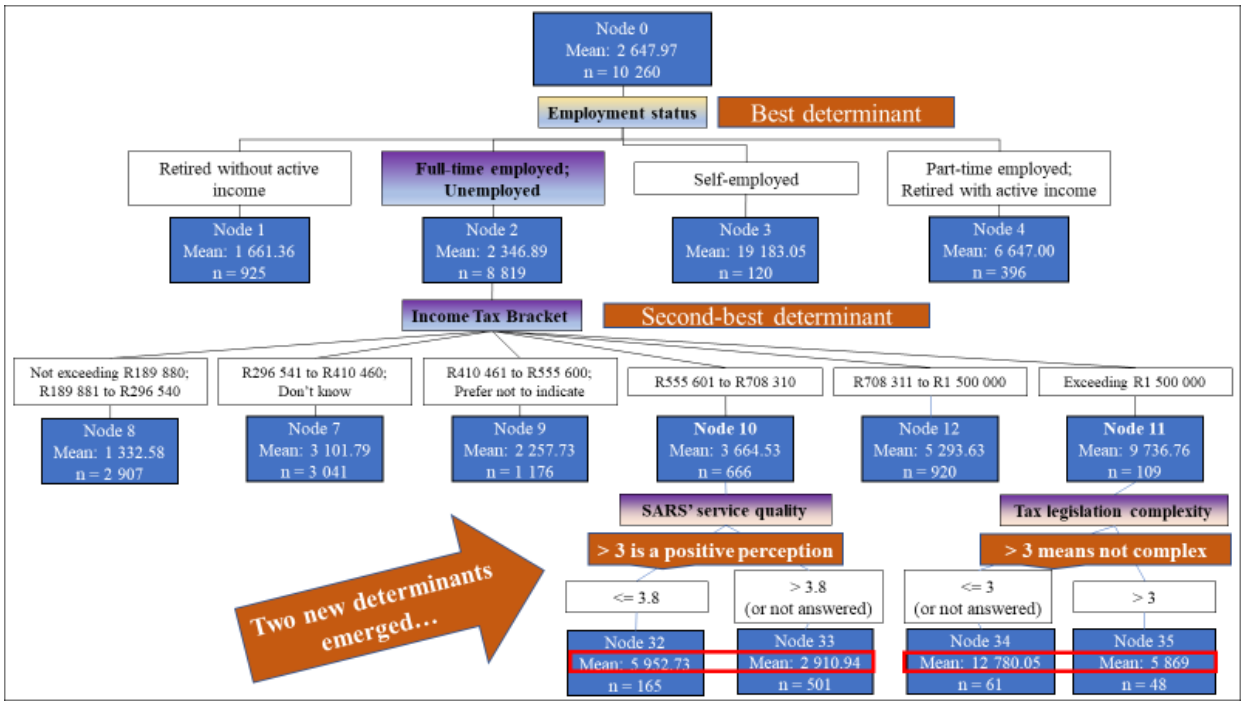
The tax compliance costs of respondents who were part-time employed or retired with active income and who used paid assistance, were further influenced by their *gender*. The tax compliance costs of male respondents was on average more than double[24] the tax compliance costs of the female respondents.[25] Allers (1994) and Blaufus et al. (2014) also reported that men spent more time completing their income tax returns than women. Reasons for this would require further research before determining what reforms could be put in place to reduce their tax compliance costs.

[23] Comparing Nodes 38 and 39 (R2,525.43 and R6,375.64).
[24] Comparing Nodes 40 and 41 (R12,044.96 and R5,359.57).
[25] Respondents who preferred not to indicate their gender were grouped with the female respondents.

The second-best determinant of the tax compliance costs of full-time employed (or unemployed) respondents, namely their *income tax bracket*,[26] was not considered in Figure 2 to enhance the readability of the CHAID tree and is now illustrated in Figure 4.

It should be noted that the groupings, in this case full-time employed and unemployed taxpayers, are not done manually. Category groupings or interval breaks in the case of continuous variables are done employing the CHAID technique, which picks the most significant combination – the independent variables homogeneous with respect to the dependent variable (Thomas & Galambos, 2004). A possible explanation for grouping these respondents together is that it is reasonable to expect that their main source of income would be indicated on an employee's tax certificate. If a now-unemployed person has worked for part of the year, that information pulls through to their tax return. Furthermore, with regard to income levels, it is possible that a now-unemployed person may have received a lump sum on retrenchment or loss of employment that increased their taxable income. However, these possible reasons are speculations and the researchers do not have enough information to substantiate these claims.

Fig. 4: Tax Compliance Costs CHAID Tree for Full-Time Employed (or Unemployed) Respondents



Source: authors' data and diagram based on SPSS output.

[26] F=215, p<0.001.

From Figure 4 it is clear that the average tax compliance costs increased from the lower tax brackets to the higher tax brackets,[27] namely from R1,332.58 for respondents whose taxable income fell in the lowest two tax brackets to R9,736.76 for respondents in the highest tax bracket. This can probably be partly attributed to higher hourly rates used by persons in the higher tax brackets to value their time spent on tax-related activities. It also corresponds to the findings of other studies that showed that a taxpayer's income level is a determinant of tax compliance costs, for example, the studies by Sandford et al. (1989), Pope and Fayle (1990), Allers (1994), Blaufus et al. (2019) and Blažić (2004).

Expanding the CHAID tree to the last layer, focusing on those respondents who were full-time employed (or unemployed), two of the previously discussed determinants[28] and two new determinants, *SARS's service quality*[29] and *tax legislation complexity*[30] emerged.[31] The tax compliance costs of respondents in this category of employment whose taxable income was '*R555,601 to R708,310*' were also influenced by their perceptions of SARS's service quality. The tax compliance costs of respondents with a very positive perception of SARS's service quality[32] were approximately half the cost of respondents with a less positive perception.[33] This finding supports the finding by Eichfelder and Kegels (2014) that a lack of service orientation of a tax authority increases tax compliance costs. It is therefore critical that SARS continues to strive to improve its services offered to taxpayers, especially to full-time employed and unemployed taxpayers as good service appears to reduce tax compliance costs and can improve tax compliance behaviour. Preparing guidance, in the form of a brochure for example, to those individuals that have become unemployed during the year could guide these taxpayers with their tax affairs and thus reduce their tax compliance costs.

Lastly, the tax compliance costs of respondents in this category of employment[34] whose taxable income '*exceeded R1,500,000*' were further influenced by their perceptions of the complexity of legislation. The tax compliance costs of the respondents who had a rating above 3 (and therefore did not regard the legislation as complex) were just less than half[35] the tax compliance costs of respondents who gave a neutral rating or considered the legislation complex.[36] This finding confirmed that tax legislation complexity is a determinant of tax compliance costs (Evans 2003), especially for high-income individuals in a South African context.

---

[27] The order of the output nodes has been changed to reflect the tax brackets in increasing sequence from left to right.
[28] *SARS's consultation and communication regarding changes to its system* and *type of assistance.* These related to nodes 7, 8, 9 and 12 and were not discussed again because similar conclusions were found than were discussed earlier.
[29] $F=39.9$, $p<0.001$.
[30] $F=8.9$, $p<0.05$.
[31] Refer to Nodes 10 and 11 in Figure 4.
[32] Namely an average rating of 3.8.
[33] Keeping in mind that a rating of less than 3.8 is still positive until it drops below 3.
[34] Full-time employed or unemployed.
[35] Comparing Nodes 34 and 35 (R12,780.05 and R5,869.25).
[36] Respondents who did not provide a rating were grouped with respondents giving a rating of 3 or less.

5.2 Gains table for nodes

Table 2 provides an overview of the spread of the CHAID decision tree in the form of a gains table, sorted from the highest mean tax compliance costs (highlighted in orange) to the lowest (highlighted in green).

Table 2: Gains in Tax Compliance Costs for Nodes

Node	N	Percentage	Mean
13	55	0.5%	30 672.39
34	61	0.6%	12 780.05
40	89	0.9%	12 044.96
14	65	0.6%	9 461.30
37	269	2.6%	7 712.93
39	170	1.7%	6 375.64
32	165	1.6%	5 952.73
35	48	0.5%	5 869.25
41	46	0.4%	5 359.57
29	99	1.0%	5 219.42
36	651	6.3%	4 293.95
22	51	0.5%	4 060.47
25	521	5.1%	3 581.02
17	42	0.4%	3 529.74
28	303	3.0%	3 014.32
33	501	4.9%	2 910.93
30	244	2.4%	2 864.71
38	91	0.9%	2 525.42
23	164	1.6%	2 336.64
31	833	8.1%	1 727.95
18	164	1.6%	1 680.80
24	2 520	24.6%[37]	1 559.74
27	1 020	9.9%	1 284.11
20	259	2.5%	1 220.05
26	1 584	15.4%	1 042.10
19	144	1.4%	917.57
21	101	1.0%	737.08
	10 260	**100.0%**	**2 647.97**

Source: authors' data; SPSS output.

[37] Node 24 represents the highest percentage of the respondents (namely 24.6%) with a predicted tax compliance costs of R1,559.74. This node branched off from Node 7 (see Figure 4) and represents those respondents who did not obtain paid assistance.

Node 13 (the highest mean tax compliance costs of R30 672) represented the *self-employed* respondents with a negative perception [38] of *SARS's consultation and communication regarding changes to its system* (refer to Figure 2). This aspect is under the control of SARS and efforts could be made to improve transparency, engagement and clarity in its interactions, especially with self-employed taxpayers.

Retired respondents without active income who did *not* use paid assistance and had a positive perception[39] of *SARS's consultation and communication regarding changes to its system* (Node 21[40]) had the lowest mean tax compliance costs, namely R737. The changes to SARS's system, for example, pre-population of third-party information such as investment income, that apply to these taxpayers appear to be making it easier for these individuals to comply. SARS must, however, continue to strive to empower these older taxpayers *not* having to use paid assistance by providing easy-to-understand information in a familiar format.

5.3 Isolated independent variables

To gain insight into the impact of location, age, tax knowledge and appeal procedures on tax compliance costs and the associated groups, a separate CHAID analysis was conducted with these independent variables.

The resultant CHAID tree indicated that the tax compliance costs of respondents with a negative experience of the *appeal procedures*[41] (an average rating of 2 or less out of 5) were on average double the costs of respondents with a less negative experience (an average rating of more than 2 out of 5) and approximately five times the costs of respondents who did not appeal their income tax assessment using the alternative dispute resolution process. This finding is in line with the finding by Blaufus et al. (2019) that appeal procedures increase tax compliance costs.

For the 2023-24 financial year, SARS reported a 76% improvement in appeals being resolved, which provided greater certainty to many taxpayers and enhanced the reputation of the dispute framework (SARS, 2024, p. 40). In this report, SARS also confirmed its commitment to resolve disputes as quickly as possible and without resorting to unnecessary litigation.[42]

With regard to *age*,[43] three subgroups were formed among the respondents who did not use the alternative dispute resolution process. Respondents who were 34 years or younger (or did not indicate their age) had the lowest tax compliance costs. The 35 to 44-year-old respondents on average had slightly higher tax compliance costs than the younger age group while respondents who were 45 years and older on average had the highest tax compliance costs. This is in line with Blažić's (2004) finding that older persons spend more time on tax compliance activities. Lopes et al. (2012) also found that the elderly have higher psychological costs caused by anxiety and stress than younger taxpayers.

---

[38] An average rating of 2.667 or less.
[39] An average rating of more than 3.667.
[40] Node 21 branched off from Node 5 in Figure 3, but was not discussed further.
[41] $F=146.6$, $p<0.001$.
[42] Ninety-seven per cent of appeals were resolved using the Alternative Dispute Resolution (ADR) process, which is an improvement from 95% in the 2022/23 financial year (SARS, 2024, p. 40).
[43] $F=31.9$, $p<0.001$.

Since the study, SARS has introduced new service offerings such as the MobiApp and self-service terminals. Older taxpayers who are not comfortable using the SARS MobiApp on their cell phones to, for example, submit their tax returns, make payments and submit supporting documents, can make use of the self-service terminals. These machines were deployed to migrate even more taxpayers to digital platforms where employees not only assisted taxpayers in fulfilling their tax obligations but also used these engagements to encourage migration to digital channels for future self-service (SARS, 2024, p. 39). A further introduction by SARS since the study was conducted that could also assist older taxpayers was auto-assessments.[44] However, it is suggested that it should be communicated more clearly that the auto-assessment must be edited if the taxpayer is entitled to claim additional medical expenses which is more probable for older persons.

The last layer in this CHAID tree contained the effect of *tax knowledge*[45] and *location*[46] on tax compliance costs. A relationship between the level of tax knowledge and tax compliance costs was only apparent for respondents 45 years and older. For these respondents, as their tax knowledge increased so did their tax compliance costs. Lastly, a relationship between location and tax compliance costs was only visible for respondents who were 34 to 44 years old. The tax compliance costs of these respondents residing in Gauteng, the Northern Cape, the Western Cape, Mpumalanga, or Limpopo, or not living in South Africa at all, were on average approximately 50% more than the tax compliance costs of respondents residing in the Eastern Cape, KwaZulu Natal, North West and the Free State province. Further research should be conducted to determine why this is the case as initiatives (if any) used in the provinces with lower tax compliance costs could be expanded to other provinces if not already implemented.

A few prior studies have found that sources of income and type of tax return were determinants of tax compliance costs, for example, Sandford et al. (1989), Pope and Fayle (1990), Allers (1994), Guyton et al. (2003) and Blaufus et al. (2019). Including *employment status* as one of the original 14 independent variables to a large extent already catered for the sources of income and type of tax return (for example, business income). However, it was decided to explore the sources of income and types of expenditure further since some respondents who were employed full-time could also have a more complex return as a result of, for example, a claim against a travel allowance which would not have been ascertained as a determinant of tax compliance costs based on employment status.

5.4 Sources of income and types of expenditure

A potential concern regarding multicollinearity between the different sources of income and types of expenditure was first addressed by running a multicollinearity test. Based on a variance inflation factor threshold of 10, as recommended by Pallant (2007), no multicollinearity was detected between the set of independent variables. Therefore, for this CHAID analysis, the 20 different sources of income provided in Q3.1 and the 11 types of expenses provided in Q3.2 were used as the independent variables and the tax compliance costs was selected as the dependent variable.

---

[44] The population of taxpayers who were automatically assessed is, on average, more than 2.5 million per year for the first two years and increased to 3.8 million in the 2023/24 financial year (SARS, 2024, p. 40).
[45] $F=17.7$, $p<0.001$.
[46] $F=16.6$, $p<0.05$.

In the resultant CHAID tree, the best determinant from all sources of income and types of expenditure was '*expenditure relating to other business income*'[47] (other than farming and rental) which aligns well with the fact that employment status was found to be the best determinant of the tax compliance costs (refer to Figure 1). The tax compliance costs for respondents with this type of expenditure was significantly higher than for respondents who did not have such expenditure.

For the respondents who did not have '*expenditure relating to other business income*', the second-best determinant was '*travel expenses*'.[48] Respondents who incurred travel expenses had almost three times the tax compliance costs of respondents who did not incur travel expenses. Having travel expenses necessitates keeping a logbook and results in a more complex tax return because additional information is required. Furthermore, the tax compliance costs of respondents who incurred travel expenses without receiving a salary (for example, earning commission income) was slightly more than double the tax compliance costs of respondents earning a salary.

Lastly, for respondents who incurred neither '*expenditure relating to other business income*' nor '*travel expenses*', '*expenditure relating to rental income*'[49] was a determinant of tax compliance costs. The tax compliance costs of respondents who incurred expenditure relating to rental income was more than three times the tax compliance costs of respondents who did not incur such costs. Declaring expenditure relating to rental income also resulted in a more complex tax return because additional information must be declared.

Additional record-keeping in all the above scenarios is inevitable but the extent of the increase in tax compliance costs could maybe be mitigated if SARS creates an online platform where taxpayers can record certain expenses in real-time throughout the year which can then be pre-populated on their tax return.

The last consideration in ascertaining the determinants of tax compliance costs related to post-filing activities since *post-filing* tax compliance costs may place a significant burden on (typically only a few) taxpayers (Eichfelder & Vaillancourt, 2014).

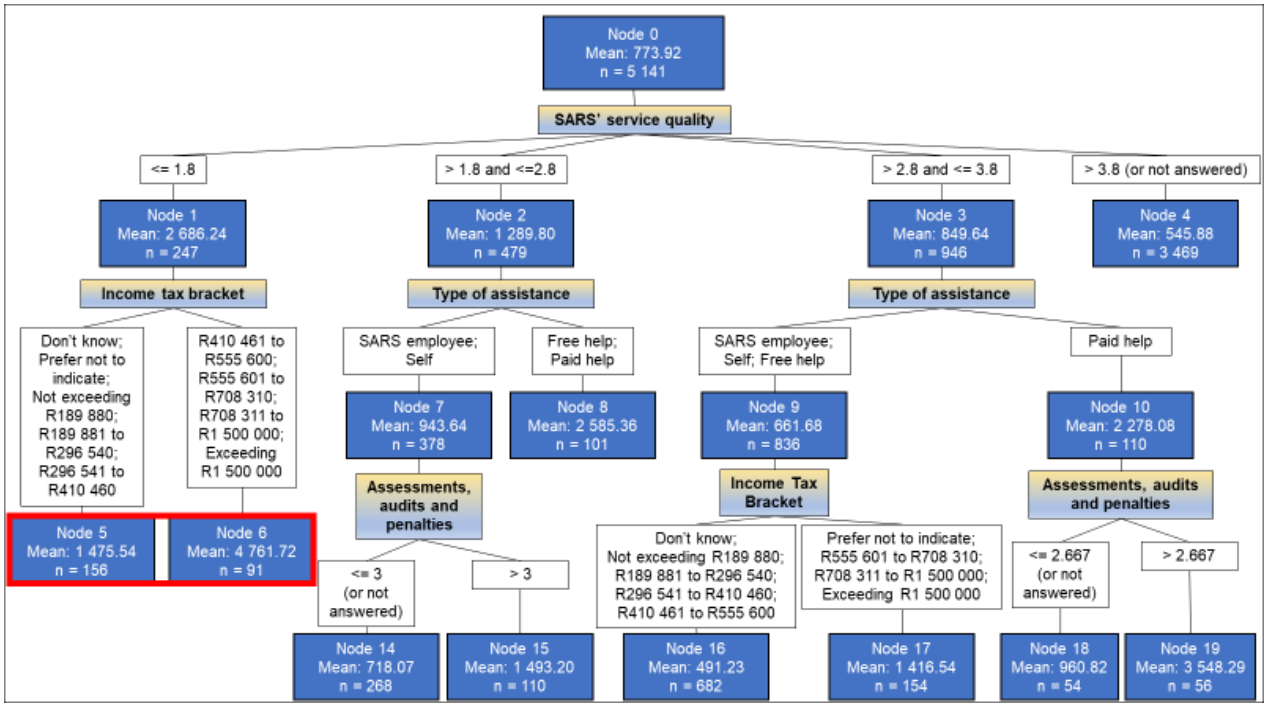
5.5 Post-filing tax compliance costs

It was found that the best determinant of post-filing tax compliance costs was *SARS's service quality*.[50] The CHAID tree diagram for post-filing activities is shown in Figure 5. For the sake of readability, Node 4 representing respondents with the most positive rating of SARS's service quality is not expanded further in the diagram. Figure 5 illustrates that the mean post-filing tax compliance costs of respondents with the most negative rating (<=1.8) of SARS's service quality and whose taxable income exceeded R410,460 (Node 6) were more than three times the costs of respondents with the same rating of SARS's service quality but with a taxable income of R410,460 or less (Node 5). Furthermore, comparing the mean post-filing tax compliance costs of the highest terminal node (Node 6) with the mean post-filing tax compliance costs of respondents with a rating of SARS's service quality higher than 3.8, the costs were almost nine times

[47] F=700.1, p<0.001.
[48] F=780.4, p<0.001.
[49] F=240.9, p<0.001.
[50] F=97.4, p<0.001.

higher. The other branches confirmed that the use of paid help and a more negative rating of assessments, audits and penalties increased (post-filing) tax compliance costs.

Fig. 5: Post-Filing Tax Compliance Costs CHAID Tree with All Three Layers and Indicating the Highest Terminal Node



Source: authors' data and diagram based on SPSS output.

6. CONCLUSION

The CHAID decision tree modelling technique was used to ascertain and segment the determinants of the tax compliance costs of individual taxpayers in South Africa. This was done by considering the respondents' ratings of six constructs ('*tax legislation complexity*', '*complexity of SARS guides*', '*SARS's service quality*', '*appeal procedures*', '*assessments, audits and penalties*' and '*SARS's consultation and communication regarding changes to its system*') as well as other characteristics of the respondents and the type of assistance they used to comply with their tax obligations as possible determinants. All of these possible determinants were the independent variables in the CHAID decision tree modelling technique. This technique showed the 'best' determinants (up to three levels) based on the independent variables that had the strongest association with tax compliance costs and their effects on tax compliance costs. The technique enabled the identification of specific groups of taxpayers associated with distinct ranges of these independent variables.

It was found that *employment status* and *income tax bracket* were the best determinants of tax compliance costs while *SARS's service quality* was the best determinant of *post-filing* tax compliance costs. Other determinants of tax compliance costs were the type

of assistance, gender, education level, tax legislation complexity, the complexity of SARS guides, SARS's service quality, assessments, audits and penalties, and SARS's consultation and communication regarding changes to its system. *Location*, *age*, *tax knowledge* and *appeal procedures* did not emerge as determinants of tax compliance costs using the CHAID technique due to other variables having a stronger relationship with costs but were isolated to gain insight into their impact on tax compliance costs and the associated groups.

By exploring sources of income and types of expenditure, it was found that '*expenditure relating to other business income*', '*travel expenses*', '*expenditure relating to rental income*' and '*salary income*' were determinants of tax compliance costs. Therefore, the determinants that emerged from the CHAID analyses in this research matched determinants from the three groups identified by Guyton et al. (2003), namely taxpayer characteristics, tax compliance methods and tax complexity. The determinants also related to the proposed fourth group, namely the service orientation of SARS – customer-friendly versus control-oriented as identified by Eichfelder and Kegels (2014).

This article showed how the CHAID technique could be used to provide a further dimension to ascertaining determinants of tax compliance costs, in addition to regression analysis used in several tax compliance cost studies, for example, by Slemrod and Sorum (1984), Vaillancourt et al. (2013) and Blaufus et al. (2019). This technique offers a number of advantages over several more commonly used analysis techniques because it is non-parametric and non-linear (Önder and Uyar, 2017). Furthermore, missing data do not present a problem, and normality and homogeneity assumptions of the data are not required. Linear relations between variables are neither assumed nor necessary. Moreover, the technique can be applied to continuous or discrete dependent and independent variables and its output is highly visual and easy to interpret with multiple trees (You et al., 2015; Önder and Uyar, 2017).

The CHAID technique outcomes thus identify the determinants, as with regression, however, they also determine the breakdown of individuals into specific groups according to the determinants that statistically significantly predict the dependent variable. This enables a better understanding of the influence of the continuous determinants' specific values (for example, SARS's service quality rating) and the categorical determinants (for example, education level and employment status) on tax compliance costs.

The CHAID technique thus provided an additional level of insight not possible with previously used data analysis techniques and could be a useful tool in future tax compliance cost studies. A more detailed level of patterns is visible which are powerful enough to enable specification and prioritisation of tax compliance cost initiatives by SARS.

7. REFERENCES

Allers, M A 1994, *Administrative and compliance costs of taxation and public transfers in the Netherlands*, Wolters-Noordhof, Groningen.

Alm, J 1996, 'What is an "optimal" tax system?', *National Tax Journal*, vol. 49, no. 1, pp. 117-133.

Alm, J, Cherry, T, Jones M & McKee, M 2010, 'Taxpayer information assistance services and tax compliance behavior', *Journal of Economic Psychology*, vol. 31, no. 4, pp. 577-586.

Blaufus, K, Eichfelder, S & Hundsdoerfer, J 2014, 'Income tax compliance costs of working individuals: Empirical evidence from Germany', *Public Finance Review*, vol. 42, no. 6, pp. 800-829.

Blaufus, K, Hechtner, F & Jarzembski, J K 2019, 'The income tax compliance costs of private households: Empirical evidence from Germany', *Public Finance Review*, vol. 47, no. 5, pp. 925-966.

Blažić, H 2004, 'Personal income tax compliance costs at an individual level in Croatia', *Australian Tax Forum*, vol. 19, no. 3, pp. 325-352.

Bruce-Twum, E & Schutte, D 2021, 'Tax compliance cost: A review of methodologies of recent studies', *Academy of Accounting and Financial Studies Journal*, vol. 25, no. 4, pp. 1-10.

Chattopadhyay, S & Das-Gupta, A 2002, *The compliance cost of the personal income tax and its determinants*, National Institute of Public Finance and Policy, New Delhi.

Costello, A B & Osborne, J W 2005, 'Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis', *Practical Assessment, Research, and Evaluation*, vol. 10, no. 7.

Djankov, S, La Porta, R, Lopez-de-Silanes, F & Shleifer, A 2002, 'The regulation of entry', *Quarterly Journal of Economics*, vol. 117, no. 1, pp. 1-37.

Eichfelder, S & Hechtner, F 2018, 'Tax compliance costs: Cost burden and cost reliability', *Public Finance Review*, vol. 46, no. 5, pp. 764-792.

Eichfelder, S & Kegels, C 2014, 'Compliance costs caused by agency action? Empirical evidence and implications for tax compliance', *Journal of Economic Psychology*, vol. 40, pp. 200-219.

Eichfelder, S & Vaillancourt, F 2014, 'Tax compliance costs: A review of cost burdens and cost structures', *Hacienda Pública Española/Review of Public Economics*, vol. 210, no. 3, pp. 111-148.

Eisinga, R, te Grotenhuis, M & Pelzer, B 2013, 'The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?', *International Journal of Public Health*, vol. 58, no. 4, pp. 637-642.

Evans, C 2003, 'Studying the studies: An overview of recent research into taxation operating costs', *eJournal of Tax Research*, vol. 1, no. 1, pp. 64-92.

Field, A 2013, *Discovering statistics using IBM SPSS Statistics*, 4th edn, Sage Publications, London.

Guyton, J L, O'Hare, J F, Stavrianos, M P & Toder, E J 2003, 'Estimating the compliance cost of the US individual income tax', *National Tax Journal*, vol. 56, no. 3, pp. 673-688.

Kass, G V 1980, 'An exploratory technique for investigating large quantities of categorical data', *Applied Statistics*, vol. 29, no. 2, pp. 119-127.

Kirchler, E, Hoelzl, E & Wahl, I 2008, 'Enforced versus voluntary tax compliance: The "slippery slope" framework', *Journal of Economic Psychology*, vol. 29, no. 2, pp. 210-225.

Kline, T J B 2005, *Psychological testing: A practical approach to design and evaluation*, Sage Publications, Thousand Oaks, CA, pp. 167-84.

Klun, M 2004, 'Compliance costs for personal income tax in a transition country: The case of Slovenia', *Fiscal Studies*, vol. 25, no. 1, pp. 93-104.

Lopes, C M, de Basto, J X & Martins, A 2012, 'Compliance costs of individual income taxation: Some empirical evidence from Portugal', *Journal of Higher Education Theory and Practice*, vol. 12, no. 4, pp. 151-164.

Mathieu, L, Waddams Price, C & Antwi, F 2010, 'The distribution of UK personal income tax compliance costs', *Applied Economics*, vol. 42, no. 3, pp. 351-368.

National Treasury & South African Revenue Service (SARS) 2019, *2019 tax statistics*, Pretoria. Available at: https://www.sars.gov.za/wp-content/uploads/Docs/TaxStats/2019/Tax-Stats-2019-Full-doc.pdf (accessed 12 January 2020).

Önder, E & Uyar, S 2017, 'CHAID analysis to determine socioeconomic variables that explain students' academic success', *Universal Journal of Educational Research*, vol. 5, no. 4, pp. 608-619.

Pallant, J 2007, *SPSS survival manual: A step by step guide to data analysis using SPSS*, 3rd edn, Open University Press, Maidenhead.

Pope, J 1989, 'The compliance costs of personal income taxation – a review of the issues', *Australian Tax Forum*, vol. 6, no. 2, pp. 125-142.

Pope, J 1995, 'The compliance costs of major taxes in Australia', in Sandford, C (ed.), *Tax compliance costs: Measurement and policy*, Fiscal Publications, Bath, UK, pp. 101-125.

Pope, J & Fayle, R 1990, 'The compliance costs of personal income taxation in Australia 1986/87: Empirical results', *Australian Tax Forum*, vol. 7, no. 1, pp. 85-126.

Sandford, C, Godwin, M & Hardwick, P 1989, *Administrative and compliance costs of taxation*, Fiscal Publications, Bath.

Sapiei, N S & Abdullah, M 2008, 'The compliance costs of the personal income taxation in Malaysia', *International Review of Business Research Papers*, vol. 4, no. 5, pp. 219-230.

Slemrod, J & Sorum, N 1984, 'The compliance cost of the US individual income tax system', *National Tax Journal*, vol. 37, no. 4, 461-474.

South African Revenue Service (SARS) 2024, *Annual report 2023/24*, https://www.sars.gov.za/wp-content/uploads/Docs/StratAnnualPerfplans/SARS-AR-29-Annual-Report-2023-2024.pdf (accessed 7 December 2024).

Stark, K 2020, 'An assessment of the tax compliance costs of individual taxpayers in South Africa', PhD Thesis, University of South Africa. Available at: https://uir.unisa.ac.za/bitstream/handle/10500/27270/thesis_stark_k.pdf?sequence=1&isAllowed=y.

Stark, K, Smulders, S & Odendaal, E 2024, 'Quantifying the tax compliance costs of individual taxpayers in South Africa', *South African Journal of Accounting Research*, vol. 38, no. 1, pp. 65-89.

Steyn, F 2011, 'An exploratory study of tax compliance costs for individuals as taxpayers in South Africa', Unpublished Master's mini-dissertation, University of Pretoria.

Thomas, E H & Galambos, N 2004, 'What satisfies students? Mining student-opinion data with regression and decision tree analysis', *Research in Higher Education*, vol. 45, pp. 251-269.

Tran-Nam, B, Evans, C & Lignier, P 2014, 'Personal taxpayer compliance costs: Recent evidence from Australia', *Australian Tax Forum*, vol. 29, no. 1, pp. 137-171.

Vaillancourt, F 1989, *The administrative and compliance costs of the personal income tax and payroll tax system in Canada, 1986*, Canadian Tax Foundation, Toronto.

Vaillancourt, F, Roy-César, É & Barros, M S 2013, *The compliance and administrative costs of taxation in Canada, 2013*, Fraser Institute, Vancouver.

You, Z, Si, Y-W, Zhang, D, Zeng, X, Leung, S C H & Li, T 2015, 'A decision-making framework for precision marketing', *Expert Systems with Applications*, vol. 42, no. 7, pp. 3357-3367.