# eJournal of Tax Research

Volume 9, Number 1 July 2011



Atax

UNSW
THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY • AUSTRALIA

eJournal of Tax Research (2011) vol. 9, no. 1, pp. 116-137

# Towards Effective and Efficient Identification of Potential Tax Agent Compliance Risk: A Stratified Random Sampling Approach

Ying Yang, Esther Ge, Ross Barns

## Abstract

We propose to use a stratified random sampling approach to identify whether a tax agent's return preparation behaviour is significantly different from its industry norm. Given a tax agent, TA, our approach creates a statistically sufficient number of notional peers for it. These peers comprise a reference group for TA, and the expectation for TA's tax return behaviour can be derived there from. By comparing TA's actual behaviour against its expected behaviour, one can infer whether TA behaves abnormally and to what degree TA incurs potential compliance risk. The novelty and advantage of our approach includes (1) effective and efficient risk identification, (2) an easy-to-understand methodology, (3) easy-to-explain results, (4) no need for any pre-defined threshold values and hence less able to be undermined by "game players" who seek to make claims just under the threshold, and (5) low cost of identification as our approach conducts unsupervised learning that does not demand a supply of labelled tax agents[1] as training data.

## 1. INTRODUCTION

Individual income tax is a major revenue source for the Australian government. Over

A definitive solution to tax agent compliance risk identification is to check every single tax return lodged by every single tax agent and then reach a conclusive statement. However such a solution is neither practical nor sustainable due to resource

## 2. How to Create Peers for a Tax Agent

Given a tax agent T A, our approach creates a statistically sufficient number of peers for T A. These peers comprise a reference group (the industry norm) against which T A is compared. This section first introduces the definition of a peer and then proposes how to create peers.

### 2.1 Definition of a peer

For a tax agent T A, a peer needs to satisfy the following two criteria.

(a)

(3)

## 3. How to Evaluate a Tax Agent's Potential Compliance Risk

We evaluate an actual tax agent T A's potential compliance risk by comparing T A against its notional peers.

### 3.1 The normal distribution

Since T A's peers are created by random sampling with replacement and with stratification according to T A's rental properties' postcodes, all the peers are equal-size random samples from the same population.

### 3.3 The risk score

The risk score combines both the risk of underreporting rental gross income (z-score(income)) and the risk of overclaiming rental gross expense (z-score(expense)). Because a z-score is a standardised value that calculates how many counts of standard deviations the actual value of a tax agent falls away from the average value of its peers, z-score(income) and z-score(expense) are commensurate and hence we can apply mathematical operations on them to calculate the risk score. For T A we can calculate its z-score of rental gross income, z-score(income), as well as its z-score of rental expense, z-score(expense). The lower the value of z-score(income), the less the rental gross income declared by T A than its peers, and hence the higher the possible

- Peers' maximum $ value per property: the biggest mean rental gross income or expense value among all the peers.
- Peers' standard deviation: the standard deviation of the peers' mean rental gross income or expense values.
- z-score: the standardised difference between the tax agent's actual rental value and its expected value drawn from its peers.
- Risk score = z-score(gross expense) – z-score(gross income). It is used to rank actual tax agents in terms of compliance risk. The higher the risk score, the higher the potential compliance risk.
- Risk rank: this tax agent's rank among all actual tax agents in terms of

(a) Rental gross income

(b) Rental gross income

FIGURE 3: Compare Tax Agent X's mean rental gross income and mean rental gross expense respectively against its peers'. X underreports its rental income but overclaims its rental expense.

Thus, Tax Agent X underreports its rental income but overclaims its rental expense. Overall it incurs a risk score of 22.99 ( = 21.21 – (- 1.78)), which is the highest among

FIGURE 4: The risk score distribution of over 15,000 actual tax agents operating in a tax return year.

FIGURE 5: Individual tax agents' risk scores for a tax return year.

### 4.3 Effciency

Our proposed stratified random sampling algorithm is very efficient. Given the rental possesses potential compliance risk. But there can be many reasons behind such a symptom. Possibly Tax Agent X correctly reports gross income but significantly overclaims gross expense; or possibly it correctly claims gross expense but significantly underreports gross income; or possibly it both underreports gross income and overclaims gross expense. However, analysis of net income alone would not reveal these useful details.

FIGURE 7: Compare Tax Agent X's mean rental net income against its peers'.

Alternatively one can use behaviours more detailed than gross income and gross expense. For instance, gross expense can be further divided into expenses of bank loan interest, capital works and other expenses.

Risk score =
– z(gross income) (3)

Note that $(gross expense) = $(bank loan interest) + $(capital works) + $(other expenses). However, z(gross expense) • z(bank loan interest) + z(captial works) + z(other expenses) because a z-score is a standardised value. Instead 3xz(gross expense)§z(rental interest) + z(captial works) + z(other expenses).

### 5.2 The central limit theorem

According to Moore [5], the central limit theorem says that the distribution of a sum or average of many small random quantities is

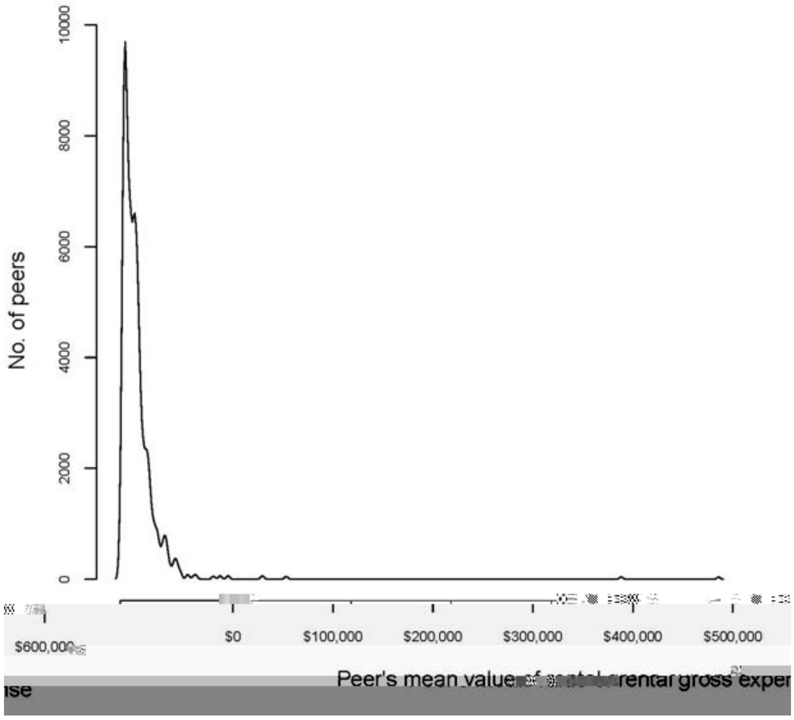


FIGURE 8: A small tax agent has only one rental property. Its peer means does not follow a normal distribution.

5.3 Median vs. mean

Sometimes people are interested in a tax agent's median rental value instead of its mean rental value.[6] Extra cautions are required when applying our stratified random sampling approach to compare a tax agent's median value against its peers'. Although it applies to the mean statistic, the central limit theorem does not necessarily apply to the median statistic. That is, the peers' median rental values do not necessarily follow a normal distribution. For instance, as illustrated in Figure 9(a) the median rental gross income values of Tax Agent Y's peers assume a bimodal distribution instead. As a result, a z-score is not always applicable and we cannot use Formula (2) to calculate the risk score. Nonetheless, it happens in this particular case that the median rental net income values of Tax Agent Y's peers still follow a normal distribution as depicted in Figure 9(b). Thus it is acceptable for one to

(a) For Tax Agent Y, the peers' median values of rental gross income follow a bimodal distribution instead of a normal distribution. Hence a z-score is not applicable.

(b) For Tax Agent Y, the peers' median values of rental net income do follow a normal distribution. Hence a z-score is applicable.

FIGURE 9: The central limit theorem does not cover the median statistic. If using median instead of mean to measure tax agent behaviour, one should always check whether peer median values follows a normal distribution before adopting the z-score to quantify a tax agent's potential compliance risk.

## 5.4 Ratio

In general, we discourage using ratio values as behaviour, such as . It is because a small denominator value will blow up the ratio and distort the behaviour. The extreme is when denominator is 0 and the ratio becomes infinitely big. Even if we replace 0 with some positive value to solve the infinity problem, the distortion problem still exists. Table 3 shows a true story. Tax Agent Z has 18 rental properties, whose rental gross income and gross expense are listed in Table 3. 10 out of the 18 properties have $0 gross income. In order to

- Risk rank = 1.

Thus Tax Agent Z incurs a very high risk score of 979.81 and is ranked as top risk, whereas the second highest risk score among tax agents is only 33.33. We suggest that Tax Agent Z's risk is largely exaggerated and ratio is the reason to the distortion. Hence one needs to be very cautious when using ratio.

## 6. RELATED WORK

Our concept of "notional peers" is inspired by Bloomquist, Albert and Edgerton's bootstrap approach to evaluating preparation accuracy of tax agents [1]. In Bloomquist etc.'s study the tax agent behaviour is AUR discrepancy rate, which equals to the number of tax returns lodged by a tax agent with potential misreported values divided by the total number of tax returns lodged by that tax agent. The misreported errors of tax returns are identified by the Automated Underreporter (AUR) program of the US Internal Revenue Service. Assume a tax agent T A lodges 12 tax returns of Postcode 20134 and 45 tax returns of Postcode 20143. The bootstrap approach creates T A's notional peers and evaluate T A's compliance risk by the following steps.

Step 1: Randomly pick 12 and 45 tax returns from all the tax returns of Postcode 20134 and Postcode 20143 respectively. The resulting 57 (= 12 + 45) picked tax returns will contribute to create a notional peer Peer1 for T A as in Step 2.

Step 2: For each of the above 57 tax returns, a uniform random number u (0, 1) is generated. If the value of u is less than or equal to the AUR discrepancy rate of the tax return's corresponding Postcode, a value 1 is added into Peer1's base; otherwise, a value 0 is added into Peer1's base.

Step 3: Compute Peer1's AUR discrepancy rate as = where {0, 1}.

Step 4: Repeat Steps 1-3 for 1000 times, creating 1000 notional peers for T A. The expected AUR discrepancy rate for T A equals to the average value of the 1000 notional peers' AUR discrepancy rates: = .

Step 5: Obtain the one-tailed 95% confidence interval by sorting the 1000 peer AUR discrepancy rates in ascending order and selecting the cutoff as the 950th value.

Step 6: If T A's AUR discrepancy rate exceeds the 95% confidence interval (the 950th value), it is identified as being a potential risk.

We respectfully suggest that the bootstrap approach does not quantify tax agent compliance risk. Consequently, it does not compare risk degrees across different tax agents to offer a risk ranking among multiple tax agents. However a proper risk ranking is highly desired in tax administration organisations such as the Australian Taxation Office because it enhances the effectiveness and efficiency of tax audit under resource constraints. Hence we have instead proposed a stratified random sampling approach where we have proved via the central limit theorem that one can use the z-score to quantify potential tax agent risk regarding a behaviour. Meanwhile, since z-scores are commensurate across different behaviours, we can apply mathematical operations on them to calculate a collective risk score for each tax agent. Multiple agents can be ranked according to their risk scores. These scores together with our proposed descriptive illustrations can provide important insight into the integrity an compliance level of a single tax agent as well as of the whole tax agent industry. Hsu etc. reported to use supervised learning to improve the audit selection procedure at the Minnesota Department of Revenue [3]. In the machine learning and data mining fields of computer science, there exist supervised learning versus unsupervised learning approaches [4, 6]. Supervised learning needs training data, that is, an unbiased and representative sample of the whole population where each of the sample returns has a known outcome (compliance or noncompliance). From the training data supervised learning infers a classifier to differentiate between compliance and non-compliance tax returns. This classifier is then used to classify other unlabelled tax returns. In their particular work, Hsu etc. had access to tax returns with auditing results and trained a naive Bayes classifier therefrom. In contrast, we lack the luxury of having good training data of agent compliance risk due to the fact that tax agent client bases are immensely diversified. Thus our proposed approach is unsupervised learning that does not demand a supply of labelled agents. As a result, our approach is of very low cost and can be easily made operational. A traditional risk identification approach in the Australian Taxation Offce is to use business expert rules. A rule system often first specifies non-compliance patterns according to domain experts' previous experience, normal distribution. Therefore one can use th

## REFERENCES

[1] K. M. Bloomquist, M. F. Albert, and R. L. Edgerton. Evaluating preparation accuracy of tax practitioners: A bootstrap approach. In Proceedings of the 2007 IRS Research Conference, pages 77-90, 2007.

[2] B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence errors, and other measures of statistical accuracy. Statistical Science, 1(1):54-75, 1986.

[3] K.-W. Hsu, N. Pathak, J. Srivastava, G. Tschida, and E. Bjorklund. Data mining based tax audit selection: A case study from Minnesota Department of Revenue. In Proceedings of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2009.

[4] T. Mitchell. Machine Learning. McGraw Hill, 1997.

[5] D. S. Moore. The Basic Practice of Statistics, 2nd Edition. W. H. Freeman and Company, Third printing 2000.

[6] I. H. Witten and E. Frank. Data Mining: Practical Machine Learning Tools and Techniques, Second Edition. Morgan Kaufmann, 2005.