Cavagnoli, International Journal of Applied Economics, 12(1), March 2015, 98-107

Does students' behaviour reflect a good proxy for the behaviour of the population? Evidence from New Zealand

Donatella Cavagnoli*

University of Ballarat, 3053 Australia

Abstract

The aim of this paper is to help clarifying the usefulness of students' samples in the tax and other disciplines studies. Data on tax behavioral responses of students' samples are obtained from a game simulation tax experiment in New Zealand. We test for the difference in tax behavioral responses of students in a two-period discrete choice economic framework, and we find that these differences are significant. The implications for the reliability of forecasted tax revenues and for measuring the efficacy of policies to curb tax evasion behaviour are also discussed.

Keywords: experimental students' samples, tax behavioural responses, difference in proportions, tax evasions behaviour, preferences, effective marginal tax rate, tax gap

JEL Classification: A23, H3, I20, K10, Z10

1. Introduction

Student samples used for experimental research are becoming very popular. The tax discipline, for example, appears to accept the use of student subjects in experimental research. Findings are particularly useful as there are difficulties in obtaining accurate behavioural data. Experimental techniques provide a relatively low cost environment for collecting data and test hypothesis on factors that it would be very difficult to obtain from observational or survey methods (Libby, Bloomfield, Nelson, 2002). The use of students' samples in experimental research, however, has been debated in the management and accounting literature, but also engineering, organisational psychology and economics.

Recent studies argued that research based on students' samples lacks both internal and external validity, and research findings based on students' samples cannot be generalized. The reasons for believing that research based on students' samples is 'weak' evolve around issues of measurements and heterogeneity.

The behavioral responses of individuals are, in general, difficult to measure. They are sensitive to the influence of economic, social and even political backgrounds. To measure people's tax behavioral responses become even more difficult in experimental research, where the conditions under which individuals behave can be manipulated (Sears, 1986; Wintre, North, and Sugar, 2001); hence, it becomes even more difficult to generalize the results to real-life situations (Cunningham, Anderson and Murphy, 1974). In addition, students are young and inexperienced, and all research based exclusively on samples with very specific age and life-stage characteristics are also argued to be biased (Bean and D'Aquila, 2003), and hence, results are arguably generalizable across diverse age populations.

Experimental research on students' samples, however, can be justified if there is homogeneity of sampling requirements. For example, if the research is focused on specific questions and within a particular theoretical framework, then there might be scope for the use of students' samples. Results may still be questionable for its 'narrowness' (Dobbins, Lane, and Steiner, 1988; Wintre et al., 2001), but not for its validity.

To help clarifying the usefulness of students' samples in the tax and other disciplines studies we ask the question: do students' tax behavioural responses reflect a good proxy for the tax behavioural responses of the population?

We provide an alternative approach to measure behavioural responses. The novelty lies in the use of a two-period discrete-choice economic framework to specify the theoretical assumptions about human behaviour and preferences. We compare the tax behavioural responses that are theoretically 'expected', and which are revealed at the initial stage of the experiment, with the 'actual' responses, which are revealed at the end of the experiment. We expect no significant differences.

By testing for the significance of differences in proportions, we can explain whether these differences arise from preferences for tax paying compliance, or not. We propose that if differences arise from preferences, then the results can be generalised. However, if differences arise from other factors, the results cannot be generalised.

We also address issues of homogeneity and sampling requirements. We extend the works by Cavagnoli (2008 and 2012) on measuring preferences for work, in labour supply studies. The method is new and provides the basis for a more systematic approach to measuring behavioral responses of individuals in the tax and other disciplines. The data, methodology and statistics are presented in section 2, the calculations are in sections 3 and 4. Section 5 is a discussion of the findings, and section 6 is the conclusion.

2. Data, Method and Statistics

The data utilised for the purpose of this paper is obtained from a 2010 tax behavioural simulation experiment study (Marriott, Randall and Holmes, 2010). Briefly summarised, a tax experiment includes the fundamental features of a basic tax system, which is found in most OECD countries: income tax, self-reporting, possibility of audit and sanctions for detected non-compliance. The experiment was designed to elicit individuals' preferences in tax compliant behaviour in response to a number of variables, such as audit and sanctions for non-compliance (Marriott, Randal, Holmes, 2010, p. 376-377).

The experiment was based on the responses of 483 students; of these, 222 were females and 261 were males. In order to obtain how much taxes students were to pay, tax rates were applied, and they varied between 20, 33 and 45 per cent. These tax rates are the actual tax rates applied in New Zealand to low-, medium- and high-income earners, respectively.

A fictitious figure reflecting the students' gross income was also given, and it varied between $ 3,000.00 and $ 4,500.00 (Marriott et al., 2010, p. 378 and 385-6). To measure tax evasion behaviour the above mentioned study proposed to utilise 'the proportion of income not reported' (Marriott et al., 2010, p. 379). The method of investigation is described in the next section.

2.1 Method

The investigation proceeds as per the following steps. For homogeneity of responses, we group individuals according to their income level to derive preferences for tax-payment compliance (alternatively, for tax-evasion behaviour). Within a *ceteris paribus* condition, and hence by holding the conditions of the experiment constant throughout the experiment, individuals respond to changes in the price of the tax (as the cost of buying a service) according to their preferences (see

for example the works of Nobel-Prize Economist Gary S. Becker, 1996). In the context of a simulation tax experiment, a tax is imposed; hence, their responses vary according to their preferences for tax-payment compliance.

Individuals that have preferences for tax-payment compliance declare their income and pay taxes according to the tax rate they belong to, and they reveal their preferences with what they pay in the initial round of the experiment. Individuals that have preferences for non-compliance declare less income, and hence, pay fewer than expected taxes, and they reveal their preferences with what they pay in the initial round. By grouping individuals according to their preferences for tax compliant behaviour and income, we overcome heterogeneity issues stemming from age differences, for example, or from education, or ethnical background.

A two-period discrete-choice framework allows us to capture variations in behavioural responses. We can compare the initial responses, which become the theoretically 'expected' responses, with the 'actual' responses, which are given at the end of the experiment. We test for significance in the difference. The next sub-sections present the data and our calculations in section 3, and the discussion of the findings in section 4.

2.2 Statistics

The responses of individuals in the initial stage, or round 1 (i.e. sample in period 1) are reported in table 1. The responses reveal that 116 (out of 222) females evade no taxes. The initial round shows that 52 per cent of females have preferences for tax compliance behaviour. The number of males who evade no taxes, is 99 out of 261; alternatively, 38 per cent of them have preferences for tax compliance behaviour (these proportions are derived from Marriott et al., 2010, p. 386). We compare the responses between two periods (the initial and the final, all rounds, samples).

When the final stage of the experiment is considered (i.e. sample in period 2), 16 out of 222 females students evade no taxes. The final round shows that, in period 2, 7.20 per cent of females have preferences for tax compliance behaviour. Males who evade no taxes in period 2 are 17 out of 261; alternatively, 6.51 per cent of them have preferences for tax compliance behaviour.

The proportion of taxes paid (or not evaded) is shown in table 2. In the initial round (in period 1), females do not evade 53.20 per cent of taxes, while males do not evade 41 per cent. The final round, in period 2, shows that females do not evade 48.10 per cent of taxes, while males do not evade 45.4 per cent (proportions derived from Marriott et. al., 2010, p. 386).

Table 3 shows the average proportion of taxes paid, by gender, in the initial and final rounds. We calculate an average gross income of about $ 3,750.00 based on figures of gross income given in the experimental study (Marriott et al., 2010, p. 378), and we obtain the same proportions of average gross income.

Table 4 shows the average proportion of individuals who do not evade taxes, by gender. The table reports the proportions in the initial and final round.

The above tables show immediately that the proportion of students not evading taxes decreases by repeating the experiment (table 1). This is true for both, males and females students. It seems that dishonest behaviour increases by repeating the experiment, and under the same experimental conditions. This means that there are variations in preferences for tax compliance behaviour.

The statistics also show that while the proportion of taxes paid by females decreases, the proportion of taxes paid by males increases (table 2). Moreover, while on average, the proportion of students evading taxes increases (table 4), the average proportion of taxes paid increases for females but decreases for males (table 3).

Females seem to pay more taxes than males at the initial stage of the experiment, but fewer of them do so by the end of the experiment. Males seem to pay fewer taxes than females at the initial stage of the experiment, and more of them do so, by the end of the experiment. We compute a 95% confidence interval on the difference in proportions in the next section.

3. Calculations for difference in the proportions: number of individuals paying taxes

The confidence interval on the difference between proportions is based on the same general formula as are other confidence intervals, and is computed in the following situation: there are two populations and the members of each population can be classified as falling into one of two categories, such as paying taxes and not paying taxes. In this section we compute the difference in the proportion of individuals in two samples (round 1 and final round, or all rounds) that pay taxes, and compute a 95% confidence interval on the difference. These students have preferences for tax compliance behaviour.

The average proportion of individuals paying taxes, in 'round 1' sample, is 0.45 (p_1), and 0.068 is the proportion of same type of individuals in the final or 'round 2' sample (p_2). The statistic $p_1 - p_2$ is used as an estimate of $\pi_1 - \pi_2$. The estimated value of $\pi_1 - \pi_2$ is: 0.45 – 0.0685 = 0.3815. As mentioned, to compute the confidence interval, the standard error of $p_1 - p_2$ is needed.

The standard error is:

$$\sigma_{p_1-p_2} = \sqrt{[\pi_1(1-\pi_1)/n_1] + [\pi_2(1-\pi_2)/n_2]} \quad (1)$$

The estimated standard error is:

$$s_{p_1-p_2} = \sqrt{[p_1(1-p_1)/n_1] + [p_2(1-p_2)/n_2]} \quad (2)$$

By substituting the proportions into the formula, we obtain the following value: $s_{p_1-p_2} = \sqrt{0.45(1-0.45)/222] + [0.0685(1-0.0685)/261]}$ = 0.0368. Finally, the value of z for the 95% confidence interval is computed using the z table: 1.96. The lower limit of the confidence interval is simply: $p_1 - p_2$ - (z) (estimated $\sigma_{p_1-p_2}$) = 0.45 - 0.0685 - (1.96)(0.0368) = 0.0275.

The upper limit is: = 0.45 - 0.0685 + (1.96)(0.0368) = 0.4536. The 95% confidence interval is therefore: $0.0275 \leq \pi_1 - \pi_2 \leq 0.4536$. This indicates that the proportion of individuals that have preferences for tax compliance behaviour, in the 'round 1' sample, is from 2.75 to 45.36 per cent higher than the proportion of individuals in 'round 2' sample. It is likely that p_1 is not equal to p_2.

In the hypothesis testing we test H_0: $p_1 = p_2$ against H_a: $p_1 \neq p_2$. We use the pooled proportion p* in the standard error and we calculate the test statistic. If H_0: $p_1 = p_2$, then the test statistic is:

$$Z = [p_1 - p_2 - 0]/\sqrt{[(p*)(1-p*)(1/n_1 + 1/n_2)]} \quad (3)$$

where p* = (215 +33) / (483+483) = 0.257

$$Z = (0.45 - 0.068 - 0)/\sqrt{[(0.257)(1-0.257)(1/483 + 1/483)]} = 11.90$$

Since 11.90 is more than 1.96 we reject the null hypothesis. The difference in the proportion of individuals with preferences for paying taxes, between the initial and the final rounds' samples, is relevant at 95% confidence interval. Given these results, it seems that, in terms of preferences for tax compliance behaviour, the tax behavioural responses of individuals in the students' samples cannot be generalised to the population. In the next section we test for the difference in the proportion of taxes not evaded.

4. Calculations for difference in the proportions: the amount of tax paid

In this section we compute the difference in the proportion of taxes not evaded (or taxes that are paid), in the two samples, and compute a 95% confidence interval on the difference. The proportion of taxes paid by individuals in the initial round of the experiment, sample 1, is 0.47 (p_1); and 0.46 is the proportion of taxes paid by individuals in the final round, sample 2 (p_2).

The statistic $p_1 - p_2$ is used as an estimate of $\pi_1 - \pi_2$. These reflect the proportion of taxes paid by individuals in the population (π_1), in round 1, and the proportion of taxes paid by individuals in the population (π_2), in the final round. The estimated value of $\pi_1 - \pi_2$ is 0.47 – 0.46 = 0.01. As per before, to compute the confidence interval, we require the standard error of $p_1 - p_2$. The estimated standard error is:

$$s_{p_1-p_2} = \sqrt{[p_1(1-p_1)/n_1] + [p_2(1-p_2)/n_2]} \quad (4)$$

Therefore, $s_{p_1-p_2} = \sqrt{0.47(1-0.47)/851287.5] + [0.46(1-0.46)/845250]}$=0.0007648. Finally, the value of z for the 95% confidence interval is computed using the z table: 1.96. The lower limit of the confidence interval is: $p_1 - p_2$ - (z) (estimated $\sigma_{p_1-p_2}$) = 0.47 - 0.46 - (1.96)(0.0007648) = 0.0051.

The upper limit is: = 0.47 - 0.46 + (1.96)(0.0007648) = 0.0114. The 95% confidence interval is therefore: $-0.0051 \leq \pi_1 - \pi_2 \leq 0.0114$ per cent. This indicates that the proportion of taxes paid by individuals in round 1, sample 1, is from 0.51 to 1.14 per cent higher than the proportion of taxes paid by the same individuals in the final round of the experiment, sample 2. It is likely that p_1 is not equal to p_2.

In the hypothesis testing we test H_0: $p_1 = p_2$ against H_a: $p_1 \neq p_2$. As per before, we use the pooled proportion p* in the standard error and we calculate the test statistic. If H_0: $p_1 = p_2$, then the test statistic is:

$$Z = \sqrt{[p_1 - p_2 - 0]/[(p*)(1-p*)(1/n_1 + 1/n_2)]} \quad (5)$$

where p* = (851287.50+845250.00) / (1811250.00+1811250.00) = 0.468

$$Z = (0.47 - 0.46 - 0)/\sqrt{[(0.468)(1-0.468)(1/1811250.00 + 1/1811250.00)]} = 19.071$$

The difference in the proportion of taxes paid between the two samples is relevant at 95% confidence interval. We find that the difference in the proportions of taxes paid by individuals in the initial and the final round of the experiment is relevant at 95% confidence interval. We conclude that the findings from students' samples cannot be generalised to the population.

5. Discussion

The results presented above reveal that the tax behavioural responses derived from students' samples differ from the population in terms of both the proportion of students paying taxes and the proportion of taxes paid. These findings suggest that students' behavioural responses cannot be a good proxy for the population. However, the proposed method also provides the means to test for the 'ceteris paribus' condition. It is the 'condition' under which standard economic models measure preferences and 'rational maximising behaviour'. The findings reveal that non-compliance cannot be explained by preferences, but also that the *ceteris paribus* condition, within which the responses are investigated, does not hold.

In a *ceteris paribus* condition we would expect that only changes in the price of the tax affect the responses of individuals to pay taxes (and hence, to comply or not with tax payments). Their preferences for tax compliance behaviour do not change, over time. What we found is precisely that individuals with preferences for tax compliance behaviour have changed their preferences for doing so. We found that both the proportion of individuals and the amount of taxes paid varied because the 'conditions' under which they responded to tax payments alter the 'real' cost of taxes.

The findings have important implications, in particular in terms of challenging conventional measures of tax evasion behaviour and of the tax gap. The insights are important also in terms of policy recommendations to curb dishonest behaviour, and in forecasting governments' tax revenues.

In this paper we tested for the usefulness of students' samples in experimental research. This is important not only because they are becoming popular in most fields of research (as a convenient and cheap alternative to the use of other samples), but also precisely because public policy recommendations arise from these type of research findings. Testing for the usefulness of students' samples becomes even more relevant as any study that attempts to investigate variations in human behaviour also has to deal with the ethical issues related to policy recommendations to 'punish' dishonest behaviour.

We find that preferences for tax compliance behaviour seem to be affected by learning from 'repeating' the task of paying taxes. This factor is excluded from the experimental model. The effect creates a gap between the theoretically 'expected' (as revealed by students in the initial round of the experiment) and the 'actual' responses, across the life-span of the experiment and across individuals. We find that 'learning from repeating' the experiment has altered the 'expected' preferences for honest behaviour, as well as worsened preferences for dishonest behaviour. This is a factor that is yet to be recognised.

From the theory of labour supply, for example, the leaning-by-doing would explain how well individuals learn to perform their tasks (Arrow, 1971). Learning-by-doing makes peoples more efficient in what they can do best; but, again it does not explain changes in preferences for what individuals choose to do. Hence, we found a type of learning that is not equivalent to the leaning-by-doing example, which is at the foundation of how learning is accounted for in measuring human behaviour. While the latter makes peoples more efficient in maximising their 'revenues', the former, or the 'learning by repeating' the experiment, decreases the probability of doing so. This factor, which arises from the learning environment, increases the risk of penalties and punishment for criminal behaviour. The type of learning that we uncovered changes the theoretical expectations of how individuals respond to the price (or cost) of the tax.

The literature on tax evasion behaviour, for example, shows that tax evasion (or non-compliance) behaviour can be measured in terms of a gap between the 'actual' and the 'nominal' cost of the tax. The 'real' cost of the tax is generally measured with the effective marginal tax rate (EMTR). The EMTR is one of the variables that is consistently reported in studies on tax evasion behaviour as highly likely to affect tax compliance behaviour (Andreoni, Erard, and Feinstein, 1998; Franzoni, 1999; Slemrod, 2007; Robinson and Slemrod, 2012). This measure was not included in the experiment, for example.

The EMTR is generally higher than the 'nominal' tax rate because it includes additional costs of taxes to the standard tax rate, with respect to changes in net earnings. For example, it includes the effects of changes in welfare payments on net earnings, as well as the cumulative effect of a progressive tax rate (of which NZ is representative). It includes the costs of paying taxes associated with changes in unemployment rates (i.e. the costs are perceived greater than the 'nominal' if unemployment rate is high), in per capita GDP (the adopted measure of income), in the Tax Act,

and even in the ability to shift the cost of the tax. Hence, the EMTR measures would explain non-compliance behaviour in terms of the 'effective cost' of paying taxes, which is higher than the 'nominal' tax rate.

However, the literature also shows that non-compliance is dominant amongst low-income and low-skilled individuals and in periods where economic conditions are generally bleak. Theoretically it is expected that the preferences of this group of individuals (for low income and low skills) are correlated with dishonest, or non-compliant, tax-payment behaviour. What is not theoretically expected is that dishonest behaviour is also found among individuals with preferences for high income and high skills. What is also theoretically unexpected are variations in responses within a group of individuals with similar preferences. This inconsistency is not accounted for in EMTR measures. The literature cannot explain a gap between the theoretically 'expected' and the 'actual' responses, and this is precisely what we found. The gap is yet to be recognised in the literature.

We applied an alternative approach for measuring the gap and for capturing variations in tax-behavioural responses of individuals with similar preferences. The insights from this investigation reveal that not only the validity of the findings from students' samples is undermined by the inability of researches to capture an 'endogenous' effect on individuals' preferences for tax compliance, but also that this effect is not accounted for in other studies on tax evasion behaviour.

We find that a gap arises from learning about the tax environment. It increases the probability for non-compliance behaviour amongst both, people that were once compliant and people that were not. It has implications for measuring, in its entirety, the 'real' cost of taxes, which in turn affects the reliability of forecasted tax revenue estimates, and the efficacy of policies to curb tax evasion behaviour. The next section presents our conclusion.

6. Conclusion

The aim of this paper was to clarify the usefulness of findings from experimental research using students' samples. Data from a game simulation tax experiment in New Zealand were investigated as an example. We proposed an alternative two-period discrete-choice economic framework to capture variations in students' behavioural responses, and tested for differences in proportions of students paying taxes and of taxes paid.

The main conclusion of this study is that students' samples cannot be used as a proxy for the population because the process of 'repeating the experiment' lead to more students paying less taxes (tables 1 and 4). We found an 'endogenous' (or a factor arising from within the condition of the experiment) change in the expected behavioural responses of students. It decreases the students' responsiveness to paying taxes.

The difference creates an additional (to the expected by the literature) gap between the expected marginal tax rate (EMTR) and the 'nominal' tax rate, which is yet to be captured and measured. The gap affects reliability measures of forecasted tax revenues, and efficacy measures of policies to curb tax evasion behaviour, including learning ethical behaviour.

In this paper we hope to have provided further evidence to argue against the use of research findings based on experimental students' samples. We proposed an alternative method to measure preferences for tax evasion behaviour to better capture the effectiveness of tax policies in future studies.

Endnote

*Dr. Donatella Cavagnoli, Adjunct Research Fellow, School of Business, University of Ballarat, 3353 Ballarat, Victoria, Australia. Email: d.cavagnoli@gmail.com

Acknowledgements

The initial version of the paper was presented at a departmental meeting in June 2012, in the School of Accounting, Victoria University of Wellington (New Zealand). The author is very grateful for the comments of the attendees.

References

Andreoni, J. Erard, B. Feinstein, J. 1998. "Tax Compliance'' *Journal of Economic Literature. American Economic Association*, 36, 2, 818-860.

Baldry, J.C. 1987. "Income Tax Evasions and the Tax Schedule: Some experimental results'' *Public Finance*, 42, 3, 357.

Bean, D.F. D'Aquila, J.M. 2003. ''Accounting Students as Surrogates for Accounting Professionals when Studying Ethical Dilemmas: a Cautionary Note'' *Teaching Business Ethic*., 7, 3, 187-204.

Becker, S. G. 1996. *Accounting for Tastes*, Harvard University Press.

Cavagnoli, D. 2008. ''Addiction to Work: An Inelastic Wage Elasticity of Labour Supply equals Long Hours of Work'' *Australian Journal of Labour Economics*, 11, 2, 129-147.

Cavagnoli, D. 2012. "The Labour Supply Curve: A Pluralist Approach to Investigate its Measurements'', *Journal of Economics and Labour Relations Review*, Vol. 23, N. 3, pp. 71-87.

Cunningham, W.H. Anderson, W. Thomas, Jr. Murphy, J.H. 1974. "Are Students Real People?'' *The Journal of Business*, 47, 3, 399-409.

Dobbins, G.H. Lane, I.M. Steiner, D.D. 1988. "A note on the role of laboratory methodologies in applied behavioural research: Don't throw out the baby with the bath water'' *Journal of Organizational Behaviour Behavior*, 9, 3, 281–286.

Franzoni, L.A. 1999. "Tax evasion and tax compliance''. *Encyclopedia of Law and Economics* eds. B. Bouckaert and G. De Geest. Edward Elgar and the University of Ghent.

Arrow, K. 1971. "The Economic Implications of Learning by Doing", *Review of Economic Studies*, 29, 155–173.

Libby, R. Bloomfield, R. Nelson, M.W. 2002. "Experimental Research in Financial Accounting'' *Accounting, Organizations and Society*, 27, 775-810.

Marriott, L. Randal, J. Holmes, K. 2010. "Influences on Tax Evasion Behaviour: Insights from a Behavioural Simulation Experiment'' *New Zealand Journal of Taxation Law and Policy*, 16, 4, 376-377.

Robinson, L. and Slemrod, J. 2012. "Understanding multidimensional tax systems'' *Journal of International Tax and Public Finance*,. 1-31.

Sears, D.O. 1986. "College Sophomores in the Laboratory: influences of a narrow data base on social psychology's view of human nature'' *Journal of Personality and Social Psychology*, 51, 3, 515-530.

Slemrod, J. 2007. "Cheating Ourselves, the Economics of Tax Evasion'' *Journal of Economic Perspectives*, 21, 1, 25-48.

Wintre, M.G. North, C. Sugar, L.A. 2001. ''Psychologists' response to criticisms about research based on undergraduate participants: A developmental perspective'' *Canadian Psychology*, 42, 3, 216–225.

Table 1. Proportion of individuals not evading taxes (i.e. paying taxes)

Rounds	Females	Males
Initial round	0.52	0.38
All rounds	0.072	0.065
Change	-0.448	-0.315

Source: table constructed by the author

Table 2. Proportion of Taxes Paid, by Gender

Rounds	Females	Males
Initial Round	0.532	0.41
All rounds	0.481	0.454
Change	-0.051	0.044

Source: table constructed by the author

Table 3. Average proportion of taxes paid

Rounds	Total	Females	Males
Initial Round	0.467	0.481	0.454
All rounds	0.471	0.532	0.41
Change	0.004	0.51	-0.044

Source: table constructed by the author

Table 4. Average proportion of individuals not evading taxes (i.e. paying taxes)

Rounds	Total	Females	Males
Initial Round	0.45	0.52	0.38
All rounds	0.0685	0.072	0.065
Change	-0.3815	-0.448	-0.315

Source: table constructed by the author