

**Modelling the Tax Compliance Profiles
of New Zealand Firms: Evidence From Audit Records**

David E. A. Giles*

Department of Economics, University of Victoria,

P.O. Box 1700, STN CSC, Victoria, B.C., CANADA, V8W 2Y2

Ph.: (250) 721 8540; Fax: (250) 721 6214

e-mail: dgiles@uvic.ca

Revised : May, 1998

Abstract

This paper analyzes a large sample of unit records relating to tax audit cases for individual New Zealand firms, undertaken by Inland Revenue New Zealand over the period 1993 to 1995. The data are used to reveal the key characteristics that are associated with firms that comply, or fail to comply, with their tax liabilities. Multinomial Logit modelling is then used to determine the relative significance of these characteristics, and to estimate the probabilities of compliance, evasion and avoidance that are associated with firms which have different profiles. The paper illustrates the usefulness of this analysis as an input into the process of selecting firms for tax audits.

1. Introduction:

This paper considers the use of information in the New Zealand Inland Revenue Department (IRD) "ORACLE" database to develop risk-profiling models of taxation compliance and non-compliance on the part of New Zealand firms. This database comprises information that is gathered in relation to tax-audits that are undertaken by the Department, including a formal assessment as to whether an audit case reveals compliance, avoidance or evasion. The study reported here is somewhat preliminary, as are the associated results. The latter should also be interpreted in the context of other related work on the risk-profiling of firms undertaken by the IRD. In particular, Heiler (1997) presents evidence relating to the probability of "risk" in terms of zero/low tax-paying activity on the part of New Zealand firms, using data from the IRD's "FIRST" database. She also examines the factors which determine the "gap" between observed tax-paying behaviour and expected effective tax rates (or "normal commercial behaviour") with respect to this same database. Law (1997) uses independent information from the New Zealand Corporate Database to analyze this "gap" between actual and expected tax-paying behaviour.

The purposes of this paper are two-fold. First, we appraise the information in the ORACLE database in order to appreciate the *actual* characteristics of compliant and non-compliant firms. The second, and major, objective is to estimate econometric models of the probability of compliance, avoidance and evasion. These models provide information about the relative *significance* of the different firm characteristics as far as compliance probability is concerned, and they also facilitate the *prediction* of the likelihood of (non-) compliance for any given set of firm characteristics. The latter, of course, has considerable potential in the context of IRD auditing activities, as well as providing important policy information.

In Section 2 we discuss the data sources, definitions, and the construction of various measures from the primary-source data. The data are analyzed, essentially in a descriptive manner, in Section 3. Section 4 describes the estimation of some multinomial Logit models for the probabilities of compliance, avoidance and evasion, and interprets the associated results¹. Some simple probability predictions, using the various Logit models, are provided in Section 5, in order to illustrate the potential use of these models for risk-profiling in an auditing context. Section 6 notes the nature of some ongoing and planned research to extend the work that is discussed here; and some concluding remarks appear in Section 7.

2. Data Sources, Definitions and Construction:

The data used in this analysis come from the IRD's ORACLE and FIRST databases, for the period 1993 to 1995. The ORACLE database, in which the data relate to individual audit cases associated with various firms, forms the primary basis for our analysis, as it provides information relating to compliance, avoidance and evasion. Typically, an audited firm has several audit cases "open" during a given year. For example, a firm can be audited in relation to income tax, its GST returns, *etc.*, separately but concurrently. Some firms are audited in more than one of the years in question, in relation to either the same, or different, taxation issues. Equally, it is possible for firms to be found

to be compliant with respect to one audit case, but non-compliant in some way with respect to another audit case (even in the same year). Much of the data are drawn from the FIRST database, which contains information based on their annual "IR4" tax returns, which is not recorded in ORACLE. In such cases, the firm-specific information in FIRST is cross-matched with the audit case-specific information on ORACLE via the firms' confidential IRD numbers.

Large random samples of observations have been drawn from the ORACLE and FIRST databases for 1993, 1994 and 1995. Although considerable information is available about each of the audit cases, only a limited number of variables are used in this study. These variables, and their sources, are described in Table 1. The useable samples comprise 6,520, 32,205 and 12,631 observations respectively (*i.e.*, 51,356 observations in total), once the data are filtered to eliminate incomplete or contaminated records, and to ensure consistency across the years. For the statistical analysis and modelling which follow, the samples were typically reduced further - *e.g.*, to eliminate situations where ratio variables could not be constructed due to zero values.

Several variable have been constructed from the raw data in Table 1. For example, we have considered two measures of firm efficiency : "Return on Net Assets", $RONA = (NP/NA)$; and "Activity Ratio", $AR = (SALES/NA)$. Various potential tax-reduction instruments which firms can use, such as Interest Expenses, Bad Debts, Depreciation and Losses Claimed in the Current Period have all been expressed as proportions of Total Assets prior to use, resulting in the variables (IE/TA) , (BD/TA) , $(DEPN/TA)$ and $(LOSS/TA)$. Finally, we define the "Effective Tax Rate" in two ways - first as tax paid (after auditing) relative to total assets, *i.e.*, $ETRA = (TAX/TA)$; and second as tax declared (before auditing) relative to total assets, *i.e.*, $ETRB = [(TAX - DISC)/TA]$. The change in tax liability as a result of the auditing process is also of interest, and we scale this in an analogous manner prior to use, working with $(DISC/TA)$ in the construction of the models which are reported below. Further modelling work which needs to be undertaken to consider the implications of defining these tax rates with Total Revenue (TR) or Net Profits (NP) as the base.

The 4-digit SIC codes were converted to dummy variables (S0 to S9) taking the value unity if the firm was in a particular sector (at the one-digit level), and zero otherwise. The details of this industry classification of the firms are given in Table 2. The YEAR variable is used to construct dummy variables (D94 and D95) which are unity only if the data are for 1994 or 1995 respectively.

We use the "Close Reason Code" (CLOSE) to construct an artificial variable $Y = (5 - CLOSE)$, to be used in the subsequent multinomial Logit analysis, discussed in Section 4 below. This variable classifies firms as compliant or non-compliant, and the latter category is in turn divided into avoiders and evaders. Official definitions of avoiders and evaders have been used, based on the coding in the IRD's *Codes and Indicators* (IR293) handbook. Strictly, "Y" takes one of five possible values (as is shown in Table 3), once an allowance is made for cases which are closed with the conclusion that an "error" occurred, and for those cases for which objections were filed. In our modelling work, cases associated with the last two categories have been discarded, and so "Y" takes one of three possible values : Y = 2, 3, 4 denote "compliance", "avoidance" and "evasion", respectively².

3. Some Basic Data Analysis:

Before undertaking any formal modelling, we have conducted some basic statistical analysis of the data. Tables 4, 5, 6 and 7 summarize some of the key data characteristics for each of our annual cross-section samples, first with respect to industry classification; second with respect to their tax liability positions; third in terms of other "firm profile" variables; and finally with respect to potential tax-minimization instrument variables. In each case, audit cases are divided according to compliance, avoidance and evasion, and the (reduced) sample sizes noted in the tables are determined by the number of observations available for the construction of the models discussed in Section 4. This takes into account, for example, the need for non-zero values for certain variables (such as total assets) which are used as divisors in the construction of ratio variables.

3.1 General Data Patterns

a. Sectoral Issues

Table 4 shows the distribution of firms by industrial classification and compliance/non-compliance category. We see that the bulk of audit cases relate to firms in sectors 2, 4, 5, and 7 (Manufacturing; Construction and Wholesale Trade; Retail Trade, Accommodation, Cafes and Restaurants; and Communication Services, Finance and Insurance, and Property and Business Services). By comparing the "All" percentage (for a particular sector) with its counterparts for compliant and non-compliant cases, we can determine which sectors show atypical non-compliant behaviour, for example. Firms in sectors 2, 4, and 5 (defined above) exhibit "excessive" avoidance in most years. The incidence of tax evasion is "above-average" (in the present sense) in Sectors 2, 5, 8 (Government Administration and Defence, Education, Health and Community Services), and 9 (Cultural and Recreational Services, Personal and Other Services), depending on the year in question.

b. Effective Tax Rates

The first interesting feature of the data in Table 5 is the fact that prior to auditing, firms typically have *negative* (or near-zero) effective tax rates. Comparing the data for ETRB with that for ETRA, we then see that generally the effect of auditing is to increase the sizeable negative effective tax rates for non-compliant firms to around a zero value. The "taxation discrepancy" data are analyzed further in Tables 9 and 10. A relatively high pre-audit effective tax rate tends to signal that the firm is compliant. In comparing this finding with Heiler's (1997) results in relation to effective tax rates and "risk", it should be borne in mind that her data relates only to incorporated firms, which cannot obtain a tax refund, but can carry losses forward. In our data-set, unincorporated firms can obtain net tax refunds. Finally we see that, except in 1993, the *value* of non-compliance (as measured by the audited "discrepancy" relative to total assets) is greater for avoiders than for evaders. This last result is based on *all* of the audit cases, and not just those which are audited with a positive discrepancy.

c. Other Firm Profile Variables

Table 6 shows that firms which are large (in terms of total revenue); efficient (in terms of their "Activity Ratio" or "Return on Net Assets"; and non-resident, tend to be relatively more compliant than are firms which are small, inefficient, or (self-declared as) resident for tax purposes. Law (1997) also finds that efficient firms (as measured by RONA) tend to be relatively "low risk", in the sense described in Section 1. As far as scale is concerned, our data-set differs from his and that of Heiler (1997), in that ours includes information for non-incorporated (and hence smaller) firms. Thus, our clear finding that evasion is associated with small firms, avoidance with medium-sized firms, and compliance with large firms, is consistent with the results of Law (1997) and Heiler (1997).

d. Tax-Minimization Instruments

Finally, the classification of the data in Table 7 shows that, in general, an aggressive use of tax-minimization instruments tends to be associated with compliance. This is true irrespective of the instrument in question. It seems that firms' behaviour was somewhat different in 1995 - then, interest expenses (and to a lesser degree depreciation expenses) were used relatively aggressively by evading firms, and bad debts were used relatively aggressively by firms subsequently found to be tax-avoiders. This feature of the 1995 data is taken up again in the conjunction with Table 11 below, in the context of the estimated Logit models.

3.2 Further Analysis

As can be seen from the sample proportions in Table 4, there appears to be a pattern between 1993 and 1995 involving a greater tendency for firms to comply, and a decreased tendency for them to avoid or evade, over time. Is this pattern statistically significant? In Table 8(a) we show the results of testing the null hypothesis that the population proportion of audit cases falling into any one of the three compliance/non-compliance categories is the same in each of two different years. To conduct this test we assume independence of the observations, both within and across years. If p_1 and p_2 are the sample proportions (say, associated with compliance) in two different years, and n_1 and n_2 are the associated sample sizes, then as long as all of n_1p_1 , n_2p_2 , $n_1(1-p_1)$ and $n_2(1-p_2)$ exceed 5 (which is the case here), the statistic, $z = \{(p_1 - p_2) / [(p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2)]^{1/2}\}$, is approximately standard Normally distributed if the null hypothesis of equal proportions is true.

As can be seen from Table 8(a), we strongly *reject* the null hypothesis in all cases. That is, there is strong evidence on the basis of these sampled audit records that firms became increasingly compliant, and less prone to avoid or evade over the period 1993 to 1995. Of course, some caveats are in order. For example, as we are dealing with audit cases, rather than with firms, it is not clear that the intra-sample independence assumption underlying the above distributional result is fully justified, though the results in Table 12 below offer some support in this respect. In addition, there is a form of selectivity bias, in the sense that IRD audit personnel select firms on the basis of various criteria, including the outcomes of previous audits. None the less, given our large samples and the clear outcomes of the tests, these results are quite convincing, and corroborative evidence is provided

in Section 4.2 below. Of course, none of this necessarily conflicts with the results of Giles (1995), who found that the size of the hidden economy in New Zealand *grew* between 1992 and 1994. His aggregate results relate to *all* agents, not just audited firms, and a recent update of his analysis suggests a small *downturn* in the hidden economy in 1995.

Table 9 provides information relating to the taxation "discrepancies" uncovered by IRD audits of firms, aggregated over the period 1993-1995. Column (1) shows the actual amount of *extra* tax revenue that is gathered as a result of audits, first just for those firms which under-reported their liability, and second in overall net terms (allowing for tax refunds arising from audits). The information in Table 10(a) shows that over this full period, 3.3% of audit cases resulted in a positive discrepancy; 96.6% in no discrepancy; and 0.1% in a negative discrepancy. There was also a steady decline in the proportion of cases yielding a positive discrepancy, between 1993 and 1995. The test results in Table 8(b) show that this decline was statistically significant (although this conclusion is, of course, subject to the same caveats that were noted in relation to the results in Table 8(a)). Also, as can be seen in Table 9, the extra tax recovered from (detected) avoiders was approximately three times that recovered from (detected) evaders. In Table 10(b) we see that, except in 1993, (detected) avoidance activity typically involved much larger dollar amounts than did typical (detected) evasion activity. Again, there is a steady decline in both the evasion and the avoidance amounts over the sample period. Whether or not this is true for tax-payers other than firms (or for undetected non-compliance in general) is an open question, of course.

Columns (2) and (3) in Table 9 show the total value of all *positive* discrepancies as a percentage of total sales, and of total revenue (respectively), where sales and revenue are totalled just for those cases recording a positive discrepancy. Columns (4) and (5) provide the same information when the base is total sales or revenue for *all* audit cases, regardless of the sign of their tax discrepancy. Considering just the non-compliant cases, we see from these figures that audit-recovered tax revenue amounts to a little over 1% of turnover in the case of avoidance; a little over 2% of turnover in the case of evasion; and about 1.2% of turnover for all non-compliant cases. Scaling these figures by a factor of 2.5 provides approximate corresponding ratios relative to GDP, rather than to sales, yielding an evasion-related Hidden Economy ratio of 5% of GDP, over 1993 to 1995.

This figure provides a (very rough) but interesting cross-check on the Hidden Economy series estimate by Giles (1995). The latter series showed that the size of the Hidden Economy ranged between 6.8% and 11.3% of measured GDP over the period 1968 to 1994, with an average value of 8.8%. This series has been analyzed by Giles and Caragata (1996) and Caragata and Giles (1996) to determine the responsiveness of underground activity in New Zealand to changes in the aggregate tax burden and/or the tax-mix. Giles and Caragata (1996; p.19) conduct simulations which indicate that evasion in the "illicit activity sector", which is driven by motives other than tax minimization, amounts to approximately 4.4% of measured GDP, on average. Caragata and Giles (1996; p.15) provide further simulation evidence which indicates that when the cyclical effects of changes in real GDP are taken into account, this "illicit activity sector" evasion ratio can be estimated at 4.95% of measured GDP in 1993 and 5.10% in 1994. These numbers accord remarkably well with the above audit-related figure of 5% on average over 1993-1995.

Two other figures, based on the positive discrepancy data in Column (1) of Table 9, are also of interest. If the total figure of \$38.25Million for all positive tax discrepancies is expressed as a ratio of total tax (after-audit) revenue, we get a crude measure of the Tax-Gap (for firms only, and only in terms of recovered revenue) of 10.8% when the base refers just to cases involving a positive discrepancy; or 25.6% when the base refers to all audit cases. The relative magnitudes of these two figures reflects the large proportion of situations where no tax or negative tax is being paid, as can be seen in Table 10(c). The apparent decline in this proportion over time is again statistically significant (subject to the earlier caveats), as can be seen in Table 8(c). Moreover, the above figure of 10.8% for the Tax-Gap for firms compares very favourably with the corresponding *aggregate* calculations based on Giles (1995), which suggest a Tax-Gap of the order of 10.2% in 1994.

4. Risk Profile Models:

4.1 Model Formulation and Estimation

Next, we discuss some preliminary analysis that has been undertaken with the above data, to construct, estimate and test some discrete-choice models of the factors determining firms' propensity to comply with tax legislation. The data analysis that has been described already provides a useful indication of basic patterns and associations. However, it fails to provide two crucially important pieces of information. The first is a measure of the (relative) *significance* of the different factors that affect the propensity to comply. The second relates to *predicting* this propensity for firms with a pre-specified profile. A formal model framework is needed to meet both of these requirements.

The outcomes of audit cases are categorized as "compliance", "avoidance" or "evasion" when cases are closed. So, we have used *multinomial* Logit analysis to model the probability that a case will fall into any one of these three categories, given a set of determining variables. The latter include various firm-profile variables, as described in Tables 4 (in the form of the dummy variables, S0 to S9) and 6, and the last column of Table 5; as well various instruments that firms can use to minimize their assessed tax (as in Table 8). It will be recalled that the qualitative dependent variable (Y, as defined in Section 2) takes the values 2, 3, and 4 for situations of compliance, avoidance and evasion respectively, and that the analysis and results are independent of these particular values. The parameter estimates, which have been obtained with the TSP (1996) package, are normalized (without loss of generality, of course) on the "compliance" category.

In the case of the individual years' samples, as well as for the "pooled" data-set for 1993-1995, a wide range of model specifications has been examined. The general strategy in specifying these models has been a "top-down" one, beginning with very general specifications, and then simplifying these on the basis of the significance of the explanatory variables. In general, variables have been retained in the models if they have an (absolute) asymptotic "t-value" in excess of unity (for at least one of the two categories). This is conservative, in the sense that it reduces the risk of wrongly omitting regressors which should be retained, which would result in parameter estimates which are inconsistent. Possibly over-specifying the models is less "costly" as the estimates would still be consistent, though they would be asymptotically inefficient. Some preliminary results appear in Table

11, and the implied partial derivatives are given in Table 12. The latter show the impact of a change in each explanatory variable on the *probabilities* of compliance, avoidance, and evasion.

4.2 Results

In terms of the specifications that have been investigated to date, no models have been estimated in which the RONA variable was statistically significant. The alternative measure of firm efficiency, the Activity Ratio (AR), was a significant determinant of compliance/non-compliance in 1993 and in 1995. No sensible results have been obtained with the "Non-Resident" dummy variable as an explanatory variable (either additively, or multiplicatively in conjunction with the other major profile variables). With this dummy in the model, convergence of the Logit estimation algorithm was not achieved (regardless of the number of iterations allowed). This is an issue that will receive further attention in future work, but it appears that this may be related to the particular formulation of the question to which firms respond in completing their tax returns. All of the other variables (though not every individual SIC dummy) discussed in Section 2 are represented in one or more of the estimated models in Table 11. The coefficient estimates (especially with regard to their signs) are most sensibly interpreted and appraised via the implied derivatives in Table 12; taken in conjunction with our knowledge of the general characteristics of the data, as in Tables 4 to 7. From these, a number of interesting results emerge.

a. Sectoral Effects

Generally, and not surprisingly, a firm's sectoral location has a significant influence on the probabilities that it will comply, avoid or evade. At this stage of the research, only a very limited amount of investigation has been undertaken with multiplicative SIC dummy specifications. Only purely additive sectoral effects are represented in Tables 11 and 12. As can be seen from Table 12, *ceteris paribus*, there is a *significantly* lower probability of compliance (and higher probabilities of avoidance and evasion) for firms in the Manufacturing and the Construction and Wholesale Trade sectors in all years, for example. Taking the pooled sample by way of illustration, we see that the probability of compliance for a firm in the Manufacturing sector will be 0.023 lower than that for an otherwise identical firm in (say) the Agricultural sector or the Mining sector. These derivatives can be used in conjunction with actual probability calculations of the type that are illustrated in Section 5 below, for the purposes of risk-profiling. Another interesting strategy that has not yet been explored in this context, is to partition the data by sector, and estimate separate model for different SIC classifications.

b. Temporal Effects

The dummy variables which allow for the different years in the models which are estimated from the "pooled" data-set are always highly significant, as is illustrated in Table 11. We see that there were significantly higher propensities to comply (and lower propensities to avoid or evade) in 1994 and 1995, than in 1993. This differential effect was roughly twice as large in 1995 as in 1994. This result provides confirmatory evidence in relation to the inter-temporal compliance trends that

were noted above in relation to Tables 4 to 7, and which were tested for in Table 8.

c. Firm Scale

The scale of a firm's operations is allowed for in the models via Total Revenue (TR). This is a significant factor in determining the probability of compliance/non-compliance in all years except 1995. More specifically, we find that the larger the firm, the greater is the probability of compliance, and the lower is the probability of avoidance or evasion. Work in progress investigates this result in more detail by modelling these effects using the (scale) quartils of the data, and by disaggregating according to sector. To illustrate the implications that can be drawn from our present results, as far as firm scale is concerned, consider the 1993 results in Table 12 and the corresponding data characteristics in Table 6. Suppose we compare two firms, one of median scale and a second of (larger) average scale. This would have entailed an *increase* of \$10.2Million in revenue in the case of the second firm. Taking the coefficient estimate of 4.167E-09 from Table 12, we see that for such firms, with otherwise typical characteristics, this would lead to a predicted *increase* in the probability of compliance of 0.043 on the part of the second firm. In the same way, the predicted *reductions* in the avoidance and evasion probabilities would have been 0.003 and 0.040 respectively, for the second firm, relative to the first firm.

d. Firm Efficiency

The efficiency of firms is represented in the various model specification via the "Activity Ratio" (AR) variable, measured as Sales/Net Assets. In Table 11, we see that this firm characteristic was a significant explanatory variable only in 1993 and 1995. Again, a more detailed analysis of the way in which firm efficiency might best enter the models' specifications is in progress. However, on the basis of the results in Tables 11 and 12, it appears that the *greater* is a firm's efficiency, the *greater* is the probability that it will be compliant, and the *lower* is the probability that it will avoid or evade. Again, by way of illustration, consider the results relating to 1993 in Table 6 and 12. We see that if we compare firms with median and mean Activity Ratios, the latter firm (with the *higher* Activity Ratio of 1475.1) is predicted to have had a compliance probability that was 0.386 *higher* than that of the former firm; and avoidance and evasion probabilities that were 0.340 and 0.046 *lower*, respectively, *ceteris paribus*.

e. Effective Tax Rate

It will be recalled that in Table 5 we distinguished between pre-audit and post-audit effective tax rates, each being measured relative to a total assets base (at this stage of our work). The former rate (ETRB) appears as a significant profile variable in all of our models. Some experimentation was also undertaken with the latter rate (ETRA), on the grounds that a firm might be expected to determine its compliance position according to its *desired* tax burden, and on the presumption that this target would be achieved because there would be no audit-based adjustment. No significant estimates were obtained.

In terms of constructing models with the potential for compliance-risk prediction, to assist in firm-selection for audit purposes, the pre-audit effective tax rate is obviously the relevant variable to use. We see from Table 12 that there is a *positive* relationship between this rate and the probability of compliance, and a *negative* relationship between ETRB and the probabilities of both avoidance and evasion. These results are consistent with the data characteristics displayed in Table 5. From the latter, it will be recalled that firms typically paid zero or negative pre-audit tax over the period in question. To illustrate the results in Table 12, consider two (otherwise similar) firms in 1995, one with a mean pre-audit effective tax rate (of 2.3%) and the other with a median value for ETRB (of 0.0%). Our estimated 1995 model (in which the ETRB variable is measured as a *proportion*, not as a percentage) predicts that former firm will have a probability of compliance that is 0.003 *higher* than the latter firm; and that its probabilities of avoidance and evasion will be 0.002 and 0.001 *lower*.

f. Tax-Minimization Instruments

All of the tax-minimization instruments that we have considered have effects which in general are "signed" in the anticipated manner (*i.e.*, positively in the case of compliance probability, and negatively for non-compliance probabilities). There are some exceptions in the 1995 model which require further exploration, but which are consistent with the data patterns described in Table 6 for that year. In addition, the "Bad Debts" variable was a *significant* explanatory variable only in 1993; and the "Depreciation" variable was *not* significant in the 1994 model. Generally, however, our results suggest that firms which use such tax-minimization instruments, do so legitimately, and not to illegally avoid or evade tax payments. To illustrate the interpretation of these results in Table 11, consider the associated data analysis for "Losses Claimed in Current Period" in 1993-1995 in Table 7. Our estimated model predicts that a firm with the average (LOSS/TA) ratio will have a compliance probability that is 0.064 *greater* than that for a firm with the (lower) median such ratio, *ceteris paribus*. Its avoidance and evasion probabilities will be 0.063 and 0.001 *lower*, respectively.

5. Some Probability Predictions:

When interpreting our estimated Logit models in Section 4.2, examples were given of the *differential* impacts of various firm attributes on compliance and non-compliance probabilities. In this Section we provide a few illustrative results relating to the actual *levels* of such probabilities, based on some hypothetical firm profiles. The purpose of this is to demonstrate some of the ways in which models such as those developed in the previous Section can be used for risk-profiling in an audit context.

The predicted probabilities for compliance, avoidance and evasion in Table 15 have been obtained from the multinomial Logit estimates based on the 1993-1995 pooled sample. The (non-dummy) explanatory variables were assigned their sample median values. As will be apparent, corresponding results can be obtained for any particular firm profile that may be of interest to IRD Audit personnel. The results in Table 15 are largely self-explanatory, and accord with the broad conclusions drawn from the data patterns in Table 4.

6. Future Research:

Although the research that is reported in this paper provides a good deal of information regarding the compliance behaviour of audited firms in New Zealand, there are many ways in which the analysis can be extended. Part of the work in progress deals with exploring alternative ways of organizing the data for analysis - for example, separate models are being constructed using samples defined in terms of the quartils of the data, by scale and by efficiency. In a similar vein, separate models for individual sectors are being explored, and there is the potential to undertake the sectoral analysis at a higher-digit level.

The majority of the work that is ongoing with this project addresses model specification issues of various sorts. In particular, we are refining the variable-selection procedures associated with the construction of the models, and considering multiplicative effects in the cases of the various dummy variables that are used as regressors. Formal tests for heteroskedasticity and variable omission for the multinomial Logit model are being explored; additional tests for IIA are being applied; and some experimentation is underway with nested Logit specifications.

In addition, various hypotheses are being tested within the context of the multinomial Logit models. For example, one can readily answer such questions as "is the probability of evasion in the Manufacturing sector the same as the probability of avoidance in the Construction and Wholesale sector?", *etc.*

7. Conclusions:

This paper provides a preliminary analysis of an important micro data-set relating to audited firms in New Zealand, and their propensity to comply with taxation laws. In addition to exploring some of the major features of the data, we have constructed some basic Logit models that enable us to predict the probabilities of compliance, avoidance, and evasion for firms which have specific profiles, in terms of industrial classification, scale, efficiency, residency, effective tax rate, and their use of various tax-minimization instruments. Our preliminary analysis leads to the following (*ceteris paribus*) results:

- (i) Larger firms tend to be significantly more compliant than smaller ones
- (ii) Firms which are relatively efficient (in terms of their "activity ratio") tend to be significantly more compliant than do relatively inefficient firms.
- (iii) There is a significant positive relationship between a firm's pre-audit effective tax rate and its propensity to comply with tax legislation.
- (iv) Firms which use standard tax-minimization instruments tend to do so in a legitimate manner, and there is a significant relationship between the active use of such instruments and taxation compliance.
- (v) There is a significant variation in the probabilities of compliance, avoidance and evasion across different industrial sectors of the economy, the details of which require further investigation before definitive judgements can be made.

In summary, the audit-record and tax-return databases of Inland Revenue New Zealand provide an extremely powerful foundation for predicting the likely non-compliance risk of New Zealand firms. The modelling framework outlined in this paper, once refined, should provide a useful basis for risk-profiling, for both audit and policy purposes.

Footnotes:

- * Earlier versions of this paper were presented at the Fourth Workshop on Monitoring the Health of the Tax System, Wellington, in May 1997, and at an Econometrics Workshop at the University of Victoria. I am most grateful to Patrick Caragata for initiating and supporting this research, and to him and Richard Bird, Johannah Branson, Erwin Diewert, Paul Dunmore, Michael Dunn, Judith Giles, Anna Heiler, Shee Boon Law, Knox Lovell, Jack Mintz, David Petterson, Gerald Scully and John Small for their many suggestions and constructive comments. My thanks, also, to Chris Gillion for his assistance with the data.
- 1. Some technical econometric issues associated with testing the specification of the models are discussed in Appendix 1.
- 2. The actual values taken by Y are irrelevant to the Logit analysis below - any three values can be used to classify the data into three groups.

References:

- Brooks, R. D., T. R. L. Fry, and M. N. Harris (1995), "The Size and Power Properties of Combining Choice Set Partition Tests for the IIA Property in the LOGIT Model", Working Paper 2/95, Department of Econometrics, Monash University.
- Caragata, P. J. and D. E. A. Giles (1996), "Simulating the Relationship Between the Hidden Economy and the Tax Mix in New Zealand", Working Paper No. 22, Working Papers on Monitoring the Health of the Tax System, Inland Revenue Department, Wellington.
- Cragg, J. G. and R. S. Uhler (1970), "The Demand for Automobiles", *Canadian Journal of Economics*, 3, 386-406.
- Cramer, J. S. and G. Ridder (1991), "Pooling States in the Multinomial Logit Model", *Journal of Econometrics*, 47, 267-272.
- Davidson, R. and J. MacKinnon (1984), "Convenient Specification Tests for Logit and Probit Models", *Journal of Econometrics*, 25, 241-262.
- Giles, D. E. A. (1995), "Measuring the Size of the Hidden Economy and the Tax-Gap in New Zealand: An Econometric Analysis", Working Paper No. 5a, Working Papers on Monitoring the Health of the Tax System, Inland Revenue Department, Wellington.
- Giles, D. E. A. and P. J. Caragata (1996), "The Learning Path of the Hidden Economy: Tax and Growth Effects in New Zealand", Working Paper No. 21, Working Papers on Monitoring the Health of the Tax System, Inland Revenue Department, Wellington.
- Greene, W. H. (1997), *Econometric Analysis* (3rd. ed.), Prentice-Hall, Upper Saddle River, N.J.
- Hall, B. H. (1996), *Time Series processor, Version 4.3: User's Guide*, TSP International, Palo Alto, CA.
- Hausman, J. and D. McFadden (1984), "A Specification Test for the Multinomial Logit Model", *Econometrica*, 52, 1219-1240.
- Heiler, A. L. (1997), "Establishing Risk Profiles of Corporate Tax Behaviour: The FIRST System Database", Working Paper on Monitoring the Health of the Tax System, Inland Revenue Department, Wellington.
- Law, S. B. (1997), "Establishing Risk Profiles of Corporate Tax Behaviour: New Zealand Corporate Database", Working Paper on Monitoring the Health of the Tax System, Inland Revenue Department, Wellington.
- Lechner, M. (1991), "Testing Logit Models in Practice", *Empirical Economics*, 16, 177-198.
- McFadden, D. (1974), "The Measurement of Urban Travel Demand", *Journal of Public Economics*, 3, 303-328.
- Maddala, G. S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.
- Orme, C. (1988), "The Calculation of the Information Matrix Test for Binary Data Models", *The Manchester School*, 56, 370-376.
- SHAZAM (1997), *SHAZAM Econometrics Computer Program: User's Reference Manual, Version 8.0*, McGraw-Hill, New York.

Table 1
Data Definitions

Symbol	Description	Units	Source
CASE	Case Key	None	ORACLE
CLOSE	Close Reason Code : see Table 4	None	ORACLE
DISC	Tax Discrepancy : difference between pre-audit and post-audit tax liability	\$	ORACLE
SIC	SIC code : ANZSIC Basis, as in Table 3	None	ORACLE
YEAR	Year for sample : 1993, 1994, or 1995	None	ORACLE
TA	Total Assets	\$	FIRST
TE	Total Expenses	\$	FIRST
SALES	Sales	\$	FIRST
DEPN	Depreciation	\$	FIRST
IE	Interest Expense	\$	FIRST
BD	Bad Debts	\$	FIRST
TR	Total Revenue : (sales + interest received + dividends received + rental income + royalties)	\$	FIRST
NP	Net Profit Before Interest and Tax	\$	FIRST
NA	Net Assets : total assets - accounts payable - bank overdraft	\$	FIRST
LOSS	Losses Claimed in Current Period	\$	FIRST
TAX	Tax Paid : after any auditing adjustment	\$	FIRST
NONRES	Non-Residency Dummy :	None	FIRST

1 for non-residents; 0 for residents

Table 2
SIC Dummy Variable Descriptions
(ANZSIC Basis : 1-Digit Level)

SIC Code Range	SIC Dummy	Industry Sector
0000 - 0999	S0	Agriculture; Forestry; Fishing
1000 - 1999	S1	Mining
2000 - 2999	S2	Manufacturing
3000 - 3999	S3	Electricity, Gas and Water Supply
4000 - 4999	S4	Construction and Wholesale Trade
5000 - 5999	S5	Retail Trade; Accommodation; Cafes and Restaurants
6000 - 6999	S6	Transport and Storage
7000 - 7999	S7	Communication Services; Finance and Insurance; Property and Business Services
8000 - 8999	S8	Government Administration and Defence; Education; Health and Community Services
9000 - 9999	S9	Cultural and Recreational Services; Personal and Other Services

Note: These codes are stored in six-digit format in the ORACLE database. The fifth digit is 0 or 5 to indicate an exact

or imperfect match between the new ANZSIC codes and the old NZSIC codes; and the sixth digit is always 0 in the case of ANZSIC codes. All SIC codes in ORACLE were updated (retrospectively) to an ANZSIC basis in 1995. A recorded code of "-1" indicates that the SIC code is unavailable.

Table 3
Classification of "Close Reason Codes"

Close Reason Code	Y	Category	Description (IR293 Code)
1	4	Evasion	Fraud/Omission (DFR) Prosecution/Penal Tax (DPP) Penal Tax Assessed (DPT)
2	3	Avoidance	Avoidance (DAV)
3	2	Compliance	All "No Discrepancy" codes All "Opened in Error codes Closed Special Conditions (CSC) No Net Change (CSO)
4	1	Errors	Accountant Error (DCF) Taxpayer Error (DTE) Technical Adjustment (DTA)

5

0

Objections

Case Stated Allowed in Full(DCF)

Case Stated Allowed in Part (DCO)

Objection Allowed in Full (DOF)

Objection Other (DOO)

Table 4
Data Analysis : Sectoral Distribution (%)

__SIC Group	0	1	2	3	4	5	6	7	89	
1993 (n=2,583)										
Compliance (94.3%)	6.4	0.5	15.6	0.6	21.8	23.9	4.8	22.0	1.3	3.1
Avoidance (3.7%)	4.1	0.0	17.5	0.0	27.8	21.6	7.2	17.5	2.1	2.1
Evasion (2.0%)	7.8	0.0	13.7	0.0	19.6	21.6	3.9	15.7	15.7	2.0
All (100.0%)	6.3	0.5	15.6	0.5	22.0	23.8	4.8	21.7	1.6	3.0
1994 (n=16,626)										
Compliance (97.0%)	9.2	0.5	16.6	0.1	22.1	23.6	4.7	18.8	1.2	3.1
Avoidance (2.2%)	3.0	0.0	22.2	0.0	23.0	24.9	3.8	14.8	1.6	6.8
Evasion (0.8%)	3.7	0.0	29.4	0.0	14.7	26.5	3.7	15.4	0.7	5.9
All (100.0%)	9.0	0.4	16.8	0.1	22.1	23.7	4.6	18.7	1.3	3.2
1995 (n=6,576)										
Compliance (98.4%)	5.5	0.2	17.4	0.0	22.4	23.5	5.7	20.1	1.7	3.4
Avoidance (1.2%)	3.9	0.0	26.0	0.0	20.8	27.3	3.9	14.3	0.0	3.9
Evasion (0.4%)	3.7	0.0	40.7	0.0	3.7	22.2	14.8	7.4	0.0	7.4
All (100.0%)	5.4	0.2	17.6	0.0	22.3	23.6	5.8	20.0	1.7	3.4
1993-1995 (n=25,785)										
Compliance (97.1%)	8.0	0.4	16.7	0.1	22.1	23.6	5.0	19.5	1.4	3.1
Avoidance (2.1%)	3.3	0.0	21.9	0.0	23.5	24.6	4.4	15.3	1.4	5.5
Evasion (0.8%)	4.7	0.0	27.1	0.0	14.5	24.8	5.1	14.5	4.2	5.1
All (100.0%)	7.8	0.4	16.9	0.1	22.1	23.7	4.9	19.4	1.4	3.2

__Note: The definitions of the abbreviated SIC codes are given in Table 2.

Table 5
Data Analysis : Taxation Variable Means (*Medians*)

	(DISC/TA) (%)	ETRA (%)	ETRB (%)
1993 (n=2,583)			
Compliance (94.3%)	0.0 (0.0)	-2.2 (0.0)	-2.2 (0.0)
Avoidance (3.7%)	9.1 (1.9)	-0.3 (0.0)	-9.4 (-1.6)
Evasion (2.0%)	23.7 (3.5)	0.0 (0.0)	-23.6 (-3.3)
All (100.0%)	0.8 (0.0)	-2.1 (0.0)	-2.9 (0.0)
1994 (n=16,626)			
Compliance (97.0%)	0.0 (0.0)	0.1 (0.0)	0.1 (0.0)
Avoidance (2.2%)	30.7 (1.5)	0.2 (0.0)	-30.5 (-1.5)
Evasion (0.8%)	17.3 (1.6)	0.5 (0.0)	-16.8 (-1.4)
All (100.0%)	0.8 (0.0)	0.1 (0.0)	-0.8 (0.0)
1995 (n=6,576)			
Compliance (98.4%)	0.0 (0.0)	2.4 (0.0)	2.4 (0.0)
Avoidance (1.2%)	11.8 (1.5)	6.6 (0.0)	-5.2 (-0.8)
Evasion (0.4%)	2.1 (1.2)	0.6 (0.0)	-1.4 (-1.2)
All (100.0%)	0.1 (0.0)	2.4 (0.0)	2.3 (0.0)
1993-1995 (n=25,785)			
Compliance (97.1%)	0.0 (0.0)	0.4 (0.0)	0.4 (0.0)
Avoidance (2.1%)	24.2 (1.5)	1.0 (0.0)	-23.2 (-1.4)
Evasion (0.8%)	16.9 (1.7)	0.4 (0.0)	-16.5 (-1.6)
All (100.0%)	0.7 (0.0)	0.5 (0.0)	-0.2 (0.0)

Table 6
Data Analysis : Other Profile Variable Means (*Medians*)

	TR (SM)	AR (ratio)	RONA (ratio)	NONRES (%)
1993 (n=2,583)				
Compliance (94.3%)	11.0 (0.3)	1564.3 (2.4)	78.5 (0.0)	5.3
Avoidance (3.7%)	2.6 (0.3)	8.2 (3.2)	0.1 (0.0)	2.1
Evasion (2.0%)	0.5 (0.3)	7.8 (3.6)	0.8 (0.0)	0.0
All (100.0%)	10.5 (0.3)	1475.1 (2.4)	74.0 (0.0)	5.1
1994 (n=16,626)				
Compliance (97.0%)	4.2 (0.5)	23.3 (2.4)	0.2 (0.1)	3.0
Avoidance (2.2%)	5.0 (0.5)	13.9 (3.6)	-0.1 (0.0)	1.9
Evasion (0.8%)	1.9 (0.5)	23.1 (3.6)	-2.4 (0.1)	1.5
All (100.0%)	4.2 (0.5)	23.1 (2.4)	0.2 (0.0)	3.0
1995 (n=6,576)				
Compliance (98.4%)	2.9 (0.5)	13.2 (2.6)	3.2 (0.1)	2.0
Avoidance (1.2%)	1.4 (0.7)	19.3 (4.0)	1.7 (0.1)	3.9
Evasion (0.4%)	1.1 (0.5)	4.3 (2.7)	1.2 (0.1)	3.7
All (100.0%)	2.9 (0.5)	13.3 (2.7)	3.2 (0.1)	2.0
1993-1995 (n=25,785)				
Compliance (97.1%)	4.5 (0.5)	170.6 (2.5)	8.6 (0.0)	3.0
Avoidance (2.1%)	4.0 (0.5)	13.7 (3.6)	0.2 (0.1)	2.2
Evasion (0.8%)	1.5 (0.4)	17.1 (3.5)	-1.2 (0.1)	1.4
All (100.0%)	4.5 (0.5)	166.0 (2.5)	8.4 (0.0)	3.0

Note: "NONRES" is a dummy variable, so its mean represents the fraction of non-resident firms in the sample.

Table 7
Data Analysis : Instrument Variable Means (*Medians*)

	(IE/TA)	(BD/TA)	(DEPN/TA)	(LOSS/TA)
	(%)	(%)	(%)	(%)
1993 (n=2,583)				
Compliance (94.3%)	824.7 (2.0)	1404.1 (0.0)	1562.4 (2.6)	58154.7 (0.0)
Avoidance (3.7%)	7.1 (3.1)	0.8 (0.0)	5.9 (4.2)	22.7 (0.0)
Evasion (2.0%)	6.4 (4.1)	0.0 (0.0)	4.1 (2.9)	283.3 (6.4)
All (100.0%)	777.8 (2.1)	1323.7(0.0)	1473.2 (2.6)	54829.0 (0.0)
1994 (n=16,626)				
Compliance (97.0%)	7.6 (1.8)	0.7 (0.0)	18.1 (2.6)	104.3 (0.0)
Avoidance (2.2%)	3.3 (2.5)	0.2 (0.0)	5.4 (3.7)	68.7 (0.0)
Evasion (0.8%)	5.2 (3.4)	0.4 (0.0)	5.3 (3.6)	176.2 (0.0)
All (100.0%)	7.5 (1.8)	0.7 (0.0)	17.7 (2.6)	104.1 (0.0)
1995 (n=6,576)				
Compliance (98.4%)	9.4 (2.1)	6.0 (0.0)	5.6 (3.3)	15054.4 (0.0)
Avoidance (1.2%)	11.0 (2.4)	62.6 (0.0)	5.8 (3.5)	90.6 (0.0)
Evasion (0.4%)	18.7 (5.7)	0.4 (0.0)	7.3 (3.8)	64.3 (0.0)
All (100.0%)	9.5 (2.1)	6.7 (0.0)	5.6 (3.3)	14817.7 (0.0)
1993-1995 (n=25,785)				
Compliance (97.1%)	87.6 (1.9)	138.6 (0.0)	165.1 (2.7)	9618.4 (0.0)
Avoidance (2.1%)	5.1 (2.6)	9.2 (0.0)	5.6 (3.7)	63.6 (0.0)
Evasion (0.8%)	7.2 (3.7)	0.3 (0.0)	5.3 (3.5)	187.6 (0.5)
All (100.0%)	85.2 (1.9)	134.7 (0.0)	160.4 (2.8)	9338.6 (0.0)

Table 8
Compliance Trends : Significance Tests
Asymptotic z-statistics

	1993 vs. 1994	1994 vs. 1995	1993 vs. 1995
(a) Compliance/Non-Compliance Ratios¹			
Compliance	5.68	6.88	8.51
Avoidance	3.86	5.68	6.33
Evasion	4.23	3.84	5.59
(b) Zero/Positive Discrepancy Ratios²			
Zero Discrepancy	7.31	8.29	10.79
Positive Discrepancy	7.31	8.29	10.79
(c) Non-Positive Tax Ratios³			
Post-Audit	11.38	26.55	29.45
Pre-Audit	11.65	26.60	29.77

Notes:

1. Based on Table 4, and the associated sample sizes.
2. Based on Table 10(a), and the associated sample sizes.
3. Based on Table 10(c), and the associated sample sizes.

Table 9
Relative Evasion and Avoidance : 1993 - 1995

	(1)	(2)	(3)	(4)	(5)
	DISC (Total +ve) [Total net] (\$Mill.)	DISC/SALES (+ve/Corresp.) (%)	DISC/TR (+ve/Corresp.) (%)	DISC/SALES (+ve/All) (%)	DISC/TR (+ve/All) (%)
Compliance	0.23 [0.19]	1.8	1.8	0.0	0.0
Avoidance	28.48 [28.36]	1.1	1.1	1.1	1.1
Evasion	9.54 [9.54]	2.1	2.1	2.1	2.1
Avoidance + Evasion	38.02 [37.90]	1.3	1.2	1.2	1.2
All	38.25 [38.09]	1.3	1.2	1.2	1.2

Table 10
Tax Discrepancy Analysis¹

	1993	1994	1995	1993-1995
(a) Audited Cases²				
Sample Size	3876	23751	9542	37169
Negative Discrepancy (%)	0.0	0.0	0.0	0.1
Zero Discrepancy (%)	93.6	96.6	98.1	96.6
Positive Discrepancy (%)	6.4	3.4	1.9	3.3
(b) Audit-Recovered Additional Tax³				
Avoiders				
Sample Size	165	561	132	858
Mean (\$'000)	37.1	35.5	18.5	33.2
Median (\$'000)	3.8	2.9	2.9	3.0
c.v. (%)	424.9	1012.0	515.9	906.0
Evaders				
Sample Size	82	228	41	351
Mean (\$'000)	62.0	17.9	9.1	27.2
Median (\$'000)	4.1	2.8	3.0	3.3
c.v. (%)	518.2	273.1	135.2	591.5
(c) Cases With Non-Positive Tax⁴				
Sample Size	3876	23751	9542	37169
Post-Audit (%)	86.6	79.7	65.0	76.6
Pre-Audit (%)	86.9	79.9	65.2	76.9

- Notes:**
1. "c.v." denotes "coefficient of variation."
 2. Based on all audit cases.
 3. Based only on audit cases yielding a positive tax discrepancy.
 4. Based on all audit cases.

Table 11
Estimated Multinomial Logit Models

Regressor ("t-value")	1993	1994	1995	1993-1995
Avoidance Equation				
Const	-3.206 (-21.52)	-4.221 (-36.37)	-4.931 (-17.14)	-3.616 (-25.86)
D94				-0.537 (-4.51)
D95				-1.141 (-7.27)
S2	0.330 (1.348)	0.744 (4.63)	0.926 (2.53)	0.751 (5.42)
S4	0.445 (1.781)	0.493 (3.09)	0.446 (1.17)	0.515 (3.78)
S5	0.491 (3.13)		0.641 (1.80)	0.464 (3.43)
S6			0.102 (0.16)	0.406 (1.75)
S8	0.601 (0.805)	0.710 (1.65)		0.555 (1.49)
S9		1.275 (5.43)	0.622 (0.96)	1.118 (5.25)
ETRB	-4.751 (-6.48)	-3.196 (-7.08)	-6.710	-3.434 (-3.55)
TR	-0.189E-07 (-1.68)	0.706E-09 (0.50)		-0.174E-08 (-0.93)
AR	-0.007 (-1.24)		-0.001 (-0.25)	
(IE/TA)	-0.547 (-1.03)	-0.726 -1.19	0.706 (2.01)	-0.842 (-7.98)

(BD/TA)	0.394 (0.34)				
(DEPN/TA)	-1.964 (-5.65)		0.470 (0.48)	-1.332 (-7.34)	
(LOSS/TA)	-0.194 (-1.61)	-0.041 (-1.70)	-0.014	-0.033 (-1.82)	(-1.95)

Table 11 (Continued)

Regressor ("t-value")	1993	1994	1995	1993-1995	
Evasion Equation					
Const	-3.907 (-16.81)	-5.162 (-28.44)	-6.744 (-10.10)	-4.181 (-20.71)	
D94				-0.901 (-5.34)	
D95				-1.590 (-6.55)	
S2	0.540 (1.22)	1.009 (4.17)	2.212 (3.04)	0.991 (4.78)	
S4	0.329 (0.84)	0.014 (0.05)	-1.068 (-0.82)	0.035 (0.15)	
S5		0.462 (1.85)	1.258 (1.63)	0.412 (1.95)	
S6			1.980 (2.40)	0.556 (1.62)	
S8	2.668 (5.56)	-0.132 (-0.13)		1.472 (3.82)	
S9		1.059 (2.65)	1.965 (2.04)	0.996 (2.91)	
ETRB	-5.113 (-6.82)	-3.063 (-6.75)	-7.651	-3.337 (-3.41)	(-9.29)
TR	-0.227E-06 (-1.42)	-0.143E-07 (-1.33)		-0.323E-07	(-2.09)
AR	-0.002 (-0.45)		-0.021 (-1.25)		
(IE/TA)	-0.020	-0.058	1.994	-0.743	

	(0.09)	(-0.280)	(4.43)	(-6.71)	
(BD/TA)	-227.387				
	(-1.11)				
(DEPN/TA)	-2.281		1.060	-1.329	
	(-3.94)			(1.16)	(-5.34)
(LOSS/TA)	-0.018	-0.004	-0.040	-0.003	
	(-0.58)		(-0.47)	(-4.04)	(-0.54)

Table 12
Partial Derivatives of Probabilities
With Respect to Regressors

Regressor	Compliance	Avoidance	Evasion
1993			
Const	0.169	-0.107	-0.062
S2	-0.022	0.013	0.009
S4	-0.020	0.015	0.005
S8	-0.062	0.017	0.044
ETRB	0.239	-0.159	-0.080
AR	2.619E-04	-2.310E-04	-3.096E-05
TR	4.167E-09	-3.327E-10	-3.835E-09
(IE/TA)	0.018	-0.019	0.001
(BD/TA)	3.517	0.345	-3.862
(DEPN/TA)	0.101	-0.066	-0.036
(LOSS/TA)	0.007	-0.007	-1.7152E-05
1994			
Const	0.129	-0.088	-0.041
S2	-0.023	0.016	0.008
S4	-0.010	0.011	-3.473E-05
S5	-0.014	0.010	0.004
S8	-0.014	0.015	-0.001
S9	-0.035	0.027	0.008
ETRB	0.091	-0.067	-0.024
TR	9.543E-11	1.927E-11	-1.147E-10
(IE/TA)	0.016	-0.015	-2.518E-04
(LOSS/TA)	8.853E-04	-8.688E-04	-1.650E-05

Table 12 (Continued)

Regressor	Compliance	Avoidance	Evasion
1995			
Const	0.081	-0.055	-0.026
S2	-0.019	0.010	0.009
S4	-8.220E-04	0.005	-0.004
S5	-0.012	0.007	0.005
S6	-0.009	9.199E-04	0.008
S9	-0.015	0.007	0.008
ETRB	0.105	-0.075	-0.030
AR	9.581E-05	-1.060E-05	-8.521E-05
(IE/TA)	-0.016	0.008	0.008
(DEPN/TA)	-0.009	0.005	0.005
(LOSS/TA)	3.139E-04	-1.566E-04	-1.572E-04
1993-1995			
Const	0.104	-0.072	-0.032
D94	0.018	-0.011	-0.007
D95	0.035	-0.022	-0.012
S2	-0.023	0.015	0.008
S4	-0.010	0.010	1.012E-04
S5	-0.012	0.009	0.003
S6	-0.012	0.008	0.004
S8	-0.022	0.011	0.012
S9	-0.030	0.022	0.008
ETRB	0.094	-0.068	-0.026
TR	2.836E-10	-2.376E-11	-2.598E-10
(IE/TA)	0.023	-0.017	-0.006
(DEPN/TA)	0.037	-0.026	-0.010

(LOSS/TA)

6.848E-04

-6.704E-04

-1.439E-05

Table 13
Goodness-of-Fit Statistics

Measure	1993	1994	1995	1993-1995
LRT	139.444	162.260	71.244	369.300
(d.o.f.)	(10)	(9)	(10)	(13)
[p-value]	[0.0]	[0.0]	[0.0]	[0.0]
Maddala (1983) R²	0.052	0.010	0.011	0.014
Cragg-Uhler (1970) R²	0.233	0.037	0.065	0.055
McFadden (1974) R²	0.105	0.032	0.060	0.060

Note: LRT = Likelihood Ratio Test for significance of non-intercept regressors. It is asymptotically Chi Square with (k-1) degrees of freedom (d.o.f.)

Table 14
Hausman Tests for Independence of Irrelevant Alternatives

	Delete Evasion	Delete Avoidance	Delete Compliance
		1993 (k=11)	
	0.217	6.891	7.576
	(1.00)	(0.81)	(0.75)
		1994 (k=10)	
	84.945	26.575	5.538
	(0.00)	(0.00)	(0.90)
		1995 (k=11)	
	0.492	0.031	0.749
	(1.00)	(1.00)	(1.00)

	1993-1995 (k=14)	
17.706	339.696	27.591
(0.12)	(0.00)	(0.02)

Note: The test statistics are asymptotically Chi Square with k degrees of freedom, and p-values appear in parentheses.

Table 15
Predicted Compliance/Non-Compliance
Probabilities Under Various Risk Profiles
(Based on the 1993-1995 Logit Model)

Year	Profile Sector	Pr(Compliance)	Pr(Avoidance)	Pr(Evasion)
1993	0, 1, 3, 7	0.96011	0.02613	0.01377
1993	2	0.91234	0.05245	0.03522
1993	4	0.94328	0.04272	0.01401
1993	5	0.93930	0.04037	0.02033
1993	6	0.93491	0.04171	0.02337
1993	8	0.88842	0.05608	0.05550
1993	9	0.88682	0.07877	0.03441
1994	0, 1, 3, 7	0.97952	0.01478	0.00570
1994	2	0.95462	0.03043	0.01495
1994	4	0.96981	0.02435	0.00584
1994	5	0.96842	0.02307	0.00851
1994	6	0.96629	0.02390	0.00980
1994	8	0.94309	0.03301	0.02391
1994	9	0.93898	0.04624	0.01478
1995	0, 1, 3, 7	0.98842	0.00870	0.00289
1995	2	0.97423	0.01811	0.00766

1995	4	0.98264	0.01439	0.00297
1995	5	0.98202	0.01365	0.00433
1995	6	0.98085	0.01415	0.00499
1995	8	0.96793	0.01976	0.01231
1995	9	0.96467	0.02771	0.00762

Note: Sectors are as defined in Table 2.

Appendix 1

Diagnostic Tests of the Logit Models

There is no clear way to measure "goodness-of-fit" in multinomial Logit models, though a variety of such measures have been suggested by different authors. Some of these are reported in Table 13. The values of such measures are typically very small when very large sample sizes are involved, as is the case here. The Likelihood Ratio Tests all imply rejection of the null hypothesis of no relationship between the dependent variable(s) and the full set(s) of regressors, so they support the model specifications.

Two major specification issues arise with the multinomial Logit model. The first involves the so-called "Independence of Irrelevant Alternatives" (IIA) problem. This is specific to the multinomial (as opposed to binomial) Logit model, and may be summarised as follows (*e.g.*, see Greene (1977, pp.920-921)). As a result of the standard assumptions that the model's errors are independent and homoskedastic, a property of the multinomial Logit model is that any one of the "odds ratios" (the ratio of the probability of one value of Y occurring, to the probability of another value of Y

occurring) does *not* depend on the other Y values. In terms of the implied behavioral characteristics of the agents, this can be a restrictive condition, although it makes the estimation of the model very straightforward. In our case, Y takes three possible values. Our use of the multinomial Logit model, and the associated IIA assumption, presumes that the relative likelihood of a firm choosing to comply as opposed to avoid, is unaffected by the fact that evasion is a "choice", *etc.*

Accordingly, it is very important to test the validity of the IIA assumption. A rejection of this hypothesis leaves us with some options with respect to the formulation of the model. The first is to consider the multinomial Probit model, but *computationally efficient* algorithms for the estimation of such models are not readily available. The second is to use the "nested" Logit model. The appropriate form of the nesting is determined by the outcomes of the tests for IIA, but in our case (were this model to be considered) a natural possibility would be to see firms as making an initial compliance/non-compliance decision, and then in the case of the latter choice, making an avoidance/evasion decision.

Testing the IIA assumption is of particular interest in the context of our data-sets for a very different reason. As noted already, the IIA assumption arises from the presumption that the model's disturbances are independent, and this was questioned in Section 3. Independence and Normality of the disturbances are assumed in the construction of the likelihood function that is maximized when the model is estimated. So, it is important to test for independence in order to examine the validity of the estimation procedure from a different standpoint. The IIA assumption also stems from the assumed homoskedasticity of the errors, which we discuss further below. In Table 14 we present the

results of the well-known Hausman specification test (Hausman and McFadden (1984)) for IIA. This test has the advantage that it is simple to apply, though other such tests are worthy of consideration here (*e.g.*, see Brooks *et al.* (1995)).

Although the results in Table 14 are somewhat mixed, they offer considerable support for the IIA assumption, other than in 1994. This is encouraging, given the preliminary model specifications presented here. Further, for all years, the Hausman tests indicate that IIA cannot be rejected (at the 1% level) if the compliance category is deleted from the model. So, if "nested Logit" analysis were pursued, then the primary "branch" in the model should be for the compliance/non-compliance decision, and the secondary one should be for the avoidance/evasion decision, in the latter case.

The second major specification issue is that of possible heteroskedasticity of the disturbances, as a result of working with cross-section data. If heteroskedasticity is present, then this is much more serious than in the usual linear regression model context, as here this will result in inconsistent parameter estimates. There is a very limited literature on testing for heteroskedasticity in the context of discrete-choice models. For example, Davidson and MacKinnon (1984) present such tests for the *binomial* situation, though these are better suited to the Probit model than to the Logit model. Tests for heteroskedasticity in the *multinomial* Logit context do not appear to be discussed in the existing literature, and the development of such a test remains as part of our work in progress. However, some comfort can be drawn from the Hausman test results in Table 14. Non-rejection of the IIA hypothesis provides support for both the independence and homoskedasticity assumptions.

Other specification issues (such as omitted variables), and the consequences of "pooling" the two non-compliance states into one, also could be considered, and these also would need extensions to existing such tests (*e.g.*, Orme (1988), Cramer and Ridder (1991) and Lechner (1991)) for the binomial Logit model.