Keeping it off the Books: An Empirical Investigation of Firms that Engage in Tax Evasion

Tedds, Lindsay

University of Victoria

July 2007

Keeping it off the Books:

An Empirical Investigation of Firms that Engage in Tax Evasion[*]

Lindsay M. Tedds

School of Public Administration, University of Victoria
PO Box 1700 STN CSC, Victoria, BC, Canada, V8W 2Y2
Phone: 250-721-8068, Email: ltedds@uvic.ca

This version: July 2007

Abstract

This paper uses a unique and recently available dataset that contains detailed information on firms from around the world to investigate factors that affect under-reporting behaviour by firms. The empirical strategy employed exploits the nature of the dependent variable, which is interval coded, and uses interval regression which provides an asymptotically more efficient estimator than the ordered probit, provided that the classical linear model assumptions hold. These assumptions are investigated using standard diagnostic tests that have been modified for the interval regression model. Evidence is presented that shows that firms in all regions around the world engage in under-reporting. Regression results indicate that government corruption has the single largest causal effect on under-reporting, resulting in the percentage of sales not reported to the tax authority being 53.4 percent higher. Taxes have the second single largest causal effect on under-reporting, resulting in the percentage of sales not reported to the tax authority being 20.2 percent higher. Access to financing, organized crime, political instability and the fairness of the legal system were found to have no effect on under-reporting. It is also found that there is a significant correlation between under-reporting and the legal organization of the business, size, age, ownership, competition and audit controls.

Keywords: Underground Economy, Tax Non-compliance, Firm Characteristics, Interval Regression

JEL Classification: C24, D21, O17

[*] I would like to thank Tom Crossley, Mike Veall and David Bjerk, all of McMaster University, Wayne Simpson, University of Manitoba, seminar participants at the University of Victoria and the 2005 Tax Policy: Perspectives from Law and Accounting conference for their helpful comments and invaluable guidance, Hassan Azziz for his assistance with the literature review, and Adian McFarlane for his diligent research support. The author would also like to gratefully acknowledge the financial support from the Social Sciences and Humanities Research Council (#752-2004-1096 and #501-2002-0107), the University of Manitoba, and the University of Victoria. I am solely responsible for any remaining errors and omissions.

Keeping it off the Books: An Empirical Investigation of Firms that Engage in Tax Evasion

1 Introduction

It is generally accepted that taxes and tax evasion are intrinsically linked; one cannot exist without the other. As a result of a great deal of theoretical, experimental, and empirical research conducted over the last twenty years, there exists an extensive knowledge base regarding tax evasion by individuals. However, research regarding tax evasion[1] by businesses is, by comparison, surprisingly modest. This is startling, given the importance of businesses and their decisions not only in economic models but also in tax systems and the economy as a whole.

Evidence does exist that supports the notion that there is cause for concern; that a substantial amount of income goes unreported to the tax authority. Schneider (2006) reports estimates of unreported activities for 145 countries around the world (including transition, developing and developed countries) over the period 1999/2000 to 2002/2003. He finds that the shadow economy averaged 42.3% of official GDP in developing African countries and 30% of official GDP in developing Asian countries, 39.6% in transitional countries in East and Central Europe and the former Soviet Republic, 16.5% in OECD countries, 21.7% in currently communist countries, and 33% of official GDP in countries in the South West Pacific Islands.[2] While estimates of the total size of the unreported activity, such as these, are valuable to

[1] The focus of this paper is on illegal tax evasion and not legal tax avoidance. Tax evasion or tax non-compliance refers to income tax that is legally owed but is not reported or paid whereas tax avoidance refers to legal actions taken to reduce tax liability. Tax avoidance includes such activities as "…purchasing tax-exempt bonds, which is certainly legal, not at all nefarious, but also certainly done for tax reasons." (Slemrod 2004, 4) There exists a voluminous literature on corporate tax avoidance, notably work by James Hines at the University of Michigan, and interested readers are encouraged to consult this literature.

[2] A number of detailed studies also exist that report estimates of the underground economy in specific countries. For example: Tedds (2005) and Giles and Tedds (2002) report estimates for Canada; Draeseke and Giles (2002), Giles and Johnson (2002) Giles and Caragata (2001) and Giles (1999) report estimate for New Zealand; Bajada (1999) reports estimates for Australia; and Dell'Anno, Gomez-Antonia, and Pardo (2007) report estimates for France, Greece and Spain.

understanding under-reporting behaviour in general, there is a great dealt that these estimates do not tell us. In particular, they do not distinguish between hidden legal activity (e.g. underreporting of profits or sales) and illegal or criminal activity (e.g. production and sale of illicit drugs, prostitution) and, more relevant to this paper, they are unable to distinguish between underreporting by individuals and firms. This detailed information is important for policy makers since different policy tools are required to curb criminal behaviour than those required to curb hidden legal activity, the causal factors that drive under-reporting differ between individuals and firms, and firms may respond differently to policy changes aimed at curbing hidden activity than individuals.

One of the only existing estimates on the distribution of share of legally-generated income that goes unreported by firms is produced by the United States Internal Revenue Service (IRS). The latest data from the IRS (2004) regarding the "tax gap" related to business activities are for the 2001 tax year. These estimates, denoted in U.S. dollars, indicate that: (1) the corporate tax gap amounted to $29.9 billion, of which corporations with over $10 million in assets contributed $25.0 billion; (2) the tax gap associated with business income earned by individuals amounted to $81.2 billion; (3) the self-employed evaded $61.2 billion in employment tax; and (4) corporations underpaid the amount of taxes due based on reported income by $2.3 billion. In total, businesses evaded $174.6 billion in taxes, which amounted to almost 10% of total taxes paid voluntarily. This is not an insignificant amount and yet should be considered a lower bound estimate because: it is based on twenty year old compliance rates; it does not include businesses that do not file tax returns (also known as non-filers, the hard-to-tax, ghosts or informal businesses); and/or it does not consider firms that are engaged in illegal activities.

Given that there is some evidence which supports the notion that businesses engage in tax non-compliance, several questions arise and require investigation. First, do businesses around the world engage in tax evasion or is it confined to a few countries or regions? Second, does the legal status of the business (e.g. sole proprietorship, partnership, corporation, etc.) affect the incidence and/or intensity of tax evasion? If so, then it may be possible to effect changes in the legal system in order to increase tax compliance. Third, do businesses that engage in tax non-compliance share common and observable characteristics, or is there too wide a variety of shapes and sizes to permit a useful generalization about them? If it were possible to define a typical evader, the tax authority could target their auditing activities more accurately. Finally, while there is considerable agreement internationally about the factors that likely trigger tax non-compliance (e.g. the tax burden, the degree of regulation, the level of enforcement, confidence in government, labour force characteristics, and morality), how do these features influence the intensity of non-compliance? With this information, policy makers could effect changes to increase the amount of tax revenues collected from businesses.

One of the main constraints to investigating and attempting to provide answers to these and related questions is the lack of data. To date, most of the existing empirical research on firm tax evasion uses data from tax audits (e.g. Rice 1992; Chan and Mo 2000; Giles 2000; Hanlon *et al.* 2005). However, these sources are only available for a very small number of select countries, the data is costly to collect, and access to the data is limited. In addition, tax audit data is subject to both selection and measurement error bias. The selection effect occurs because firms are generally not randomly audited. Rather, firms are usually selected for an audit because they have a particular characteristic (e.g. U.S. businesses with more than $10 million in assets are audited annually) or because their tax filing documents raise a "red-flag" with the tax authority.

The measurement error bias occurs because not only are audits unlikely to completely and accurately measure tax non-compliance but also the final non-compliance report is usually based on negotiations between the firm and the tax authority as well as various legal appeals.

More recently, however, an alternative data source has become available that is conducive to investigating issues related to business tax non-compliance namely, the World Business Environment Survey (WBES) which was hosted by the World Bank's Enterprise Analysis Unit. The survey was administered to more than 10,000 firms in eighty countries in late 1999 and early 2000, and provides information regarding a firm's characteristics and well as responses to multiple questions on the investment climate and business environment.[3] The survey also includes a question regarding the extent and intensity to which firms fail to report income to the tax authority. Rather than reporting the exact amount of income that went unreported to the tax authority, firms were provided with seven possible intervals of varying size that were grouped according to the percentage of sales that *are reported* for tax purposes. When a quantitative outcome is grouped into known intervals on a continuous scale the data is said to be interval-coded. Asymptotically consistent estimates can be obtained using interval regression. The advantage of interval regression is that, provided certain key assumptions are met, the coefficients on the exogenous variables can be interpreted as if ordinary least squares were applied to a continuous dependent variable.

Overall, and not surprisingly, high taxes and government corruption result in lower compliance and the magnitude of these effects are quite large. It is found that government corruption has the single largest effect, resulting in the percentage of sales not reported to the tax

[3] An updated data set recently became available after the writing of this paper. The Enterprise Survey data of the Enterprise Analysis unit has collected data, starting in 2002, on the investment climate in 97 countries for over 60,000 firms.

authority being a whopping 53.4 percent higher, and taxes have the second largest effect, causing the percentage of sales not reported to the tax authority to be 20.2 percent higher. Exchange rates have a positive effect on compliance, which is the only coefficient that fails to follow *a priori* expectations. Access to capital, political instability, organized crime, inflation, and the legal system are found to have no statistically significant effect on tax non-compliance.

It is also found that firms which are sole proprietorships and privately owned corporations report a smaller percentage of their sales to the tax authority. The previous literature has suggested that firms in the service and construction sector should be less compliant and this study finds that firms in the service sector are more compliant. It is likely this result stems from the fact that industry categories in the WBES are likely too broadly defined to obtain any significant relationship between industry sector and tax non-compliance. For example, the service sector will include restaurants and hair salons as well as law and accounting firms. There is no consensus in the previous literature about the relationship between firm size and under-reporting, but it is found, unambiguously, that small firms are less and large firms are more compliant. Government owned firms and firms that have audited financial statements are also more compliant, as was also found by other researchers, though foreign ownership and export status have no relationship with compliance.

The paper begins with a review of the relevant tax evasion literature. Section 3 describes the WBES data. Section 4 outlines the empirical technique used in this paper as well as a description of the key statistical assumption and associated tests. The results are then summarized and discussed in Section 5. The paper ends with some concluding comments.

2 Literature Review

As previously noted, extensive literature exists on tax evasion by individuals and the seminal contribution was provided by Allingham and Sandmo (1972).[4] Their model was sparse but surprisingly robust in modeling *an individual's* decision to evade taxes and, if so, how much to evade. However, the theoretical outcomes are muddied when the choice is not only the level of evasion but also the level of output, as with most businesses, and when there is no direct link between the owner's income, risk preference, audit probability, and evasion decisions. This is particularly true for businesses where evasion decisions are made by managers and not the owner(s). Hence, the reasons driving many businesses to evade taxes is likely different from that of individuals and, hence, should be modeled differently.

There is a much smaller pool of literature that addresses tax non-compliance by businesses.[5] The majority of this work utilizes the basic framework of the Allingham-Sandmo (1972) model, but the key variation in these models is that they explore how the tax rate, probability of detection, and penalty rate affect the two choices of evasion and output.[6] These models, however, continues to assume that *an individual* is at the centre of the tax decision. That is, that the firm *owner* makes the tax reporting decision. This assumption, however, likely only applies to the self-employed and other small, privately-held businesses. This suggests that the outcomes predicted by the Allingham-Sandmo (1972) model may not apply to other types of businesses, notably those businesses where financial decisions, including those related to taxes, are not made by the owner/shareholder but, rather, by their agents. With this in mind, Chen and

[4] The Allingham-Sandmo (1972) model has been extended in a number of dimensions over the last thirty years (e.g. Watson 1985, Trandel and Snow 1999, Mookherjee and P'ng 1989, Border and Sobel 1987, Scotchmer 1987, Cremer, Marchand and Pestieau 1990, and Sanchez and Sobel (1993)) and this literature is nicely surveyed by Andreoni, Erard, and Feinstein (1998) and Slemrod and Yitzhaki (2002).
[5] Cowell (2004) provides an excellent review of this literature.
[6] There is no agreement in this literature if the firm in these models should be modeled as risk-averse or risk neutral.

Chu (2002) extend the standard model to include a firm that hires a risk-averse manager who is offered some form of compensation (e.g. stock options, bonuses, etc.) as an incentive to engage in tax evasion on behalf of the firm (a principle-agent type model). Under this model, a firm will evade tax only when the expected profit from evasion is *significantly* greater. This is because tax evasion by a business actually involves the interaction of many agents and is much more complicated than individual income tax evasion. Crocker and Slemrod (2005) extend this further to consider the role of penalties in the tax evasion decision. Not unsurprisingly, the model implies that corporate tax evasion is reduced when penalties are imposed on the CFO directly, as opposed to the shareholder, and that tax evasion increases if the CFO's compensation contract optimally adjusts to offset the penalties imposed when evasion is detected.

The theoretical findings clearly suggest that the type of business, whether it be owner managed or otherwise, along with other factors such as tax rates, audit rates, and penalties greatly affects the tax evasion decision of firms. Whether or not these theoretical results hold remains unclear however due to the fact that empirical analysis of business tax evasion is not extensive and this is mainly due to a lack of data. There have, however, been a small number of empirical studies that will be summarized here.

A number of studies have been conducted on tax non-compliance by the self-employed. The self-employed are commonly believed to have lower compliance rates than wage and salary earners, primarily due to a lack of third-party reporting and withholding. One approach, popularized by Pissarides and Weber (1989), uses household income and expenditure data to estimate the degree of income under-reporting. This approach, along with various modifications, has been applied by: Apel (1994) and Engstom and Holmulnd (2006) for Sweden; Schuetze (2002) and Tedds (2007) for Canada; Lyssiotou *et al.* (2004) for Great Britain; and Feldman and

Slemrod (2006) for the United States. U.S. tax audit data has also been used to explore tax evasion by the self-employed and these studies include Christian (1994), Erard (1992), and Joulfaian and Rider (1998).

Several studies have investigated business tax non-compliance using data from tax audits. Probably the most comprehensive tax audit dataset in the world is the U.S. IRS Tax Compliance Measurement Program (TCMP).[7] Rice (1992) used TCMP data from 1980 to investigate tax compliance by small corporations (defined as corporations with assets between $1 and $10 million). His findings suggest four key results. First, compliance is higher among publicly traded corporations, which he attributes to the requirement that publicly traded corporations must disclose more information to the public about their operations.[8] Second, high profit companies are more likely to under-report their income, and corporations whose profits are below the industry mean tend to resort to non-compliance, perhaps as a means of limiting costs. Third, the marginal tax rate is negatively associated with compliance. Finally, firm size and tax non-compliance are positively related, and corporations engaging in tax non-compliance appear to be geographically bundled. The latter finding is interesting and may be evidence that the compliance behaviour of others directly affects ones own compliance behaviour. Joulfaian (2000), using TCMP data from 1987, found that non-compliant corporations are three time more likely to be managed by executives who have evaded personal taxes.

Hanlon *et al.* (2005) are the first to the use operational data from the Voluntary Compliance Baseline Measurement (VCBLM) program compiled by the Large and Mid-Sized

[7] The TCMP features data from a random sample of individual and small corporate income tax returns filed in a given year that were subject to intensive audits by experienced examiners. For certain groups, such as non-filers and proprietors who tend to not report a significant amount of their income, the results from special research studies are used to supplement TCMP data. Finally, data for large corporations are obtained from routine operational audits.
[8] Tannenbaum (1993), however, disagrees with this statement and instead argues that higher compliance could be the result of managers in these corporations having greater independence from the owners.

Business (LMSB) Research Division of the IRS to examine corporate tax non-compliance, as measured by deficiencies identified during audit. This data is more up to date, containing audit files up to and including 2002, than the TCMP, which ended in 1988. They find that business tax non-compliance relative to scale is a convex function, with medium-sized businesses (within the population of firms with assets exceeding $10 million) having the lowest rate of non-compliance. Private firms, multinationals and firms with incentivized executive compensation schemes are associated with higher non-compliance (providing empirical evidence for the predictions from the model proposed by Crocker and Slemrod (2005)) while foreign controlled firms are more compliant. In general, firms in the manufacturing industry, trade, transportation and warehousing industry, and education, healthcare and warehousing industry are less compliant. They also find no association between compliance and effective tax rates.

Giles (2000) discusses some of the factors that determine the probability of non-compliance among a very large population of New Zealand businesses that were audited by the Inland Revenue Department in that country between 1993 and 1995. Contrary to Rice (1992) and Hanlon *et al.* (2005), Giles finds that an increase in the scale of the business, regardless of how this is measured, unambiguously raises the probability of *compliance*, once other characteristics are controlled for. That is, businesses that are relatively small, in terms of sales revenue or before-tax profit, are more likely to evade taxes than are large corporations, all other things being equal. Businesses in the construction, wholesale trade, retail trade, accommodation, and cafes and restaurants sectors exhibited below-average compliance rates over the study period. He also considers several other characteristics that were found to be important in reducing the tax compliance rate among New Zealand businesses. Relatively inefficient businesses tended to be less compliant than more efficient ones, where efficiency was defined as either “return on net

assets", or "activity ratio" (sales as a percentage of net assets). In addition, businesses which were registered off-shore were generally *more* compliant than their on-shore counterparts, a finding that is shared by Hanlon *et al.* (2005). Finally, it was found that in general, an aggressive use of legitimate tax-minimization instruments (i.e. tax avoidance behaviour such as the deduction of interest and depreciation costs, and the writing-off of bad debts) tended to be associated with increased compliance.

Many countries use tax holidays to attract foreign investment by providing a limited period of tax exemptions and reductions for qualified investors. Chan and Mo (2000) examine the effect of tax holidays on foreign investors' tax non-compliance behaviour in China. They analyzed 583 tax audit cases, made available by the Chinese tax authorities, on corporate tax non-compliance by foreign investors. Their results indicate that the corporate taxpayers tax holiday position significantly affects non-compliance, notably: (1) companies in the pre-holiday position are least compliant; (2) companies are most compliant in the tax exemption period that has a zero tax rate and a heavy penalty for evasion; (3) domestic market-oriented companies have a higher rate of non-compliance than their export-oriented counterparts; and (4) wholly foreign-owned and manufacturing-oriented companies have higher compliance than joint ventures and service-oriented companies.

Only two studies appear to exist which use data on firm-reported tax non-compliance rather than audit data. Using firm level data from a 1997 survey of *private manufacturing* firms in Poland, Romania, Russia, Slovakia, and the Ukraine, Johnson *et al.* (2000) investigate the relationship between government corruption, criminal activities, and firm tax compliance. The dependent variable (percentage of sales that are unreported to the tax authority) is similar to that which is used in this paper except that it is a continuous variable (rather than grouped data as in

the WBES).[9] The authors found a positive and significant relationship between under-reporting of sales and bribing of corrupt officials, but no relationship between under-reporting of sales and protection payments to the mafia, tax payments, or efficiency of the legal system. Finally, firm tax non-compliance was greater in Russia and the Ukraine than in the other countries included in the study.

Batra *et al.* (2003) were the first to use the WBES to investigate the determinants of under-reporting by firms. The WBES asks each firm to provide an estimate of the percentage of sales revenues that firms like their own report to the tax authority. Rather than the answer being recorded on a continuous scale, seven answers were possible: 100%, 90-99%, 80-89%, 70-79%, 60-69%, 50-59%, and less than 50%. The authors define seven binary variables based on these seven intervals and each binary variable takes a value of one if the firm selected the particular interval of interest and zero otherwise. They then estimate seven different OLS regressions which include variables for firm characteristics, rule of law, business constraints and country fixed effects. Overall they find that: (1) "…small or medium-size firms that produce for the domestic market (non-exporters), lack foreign investment, and are located in large cities (but not necessarily in the capital) tend to engage more in unofficial activity" (p. 76); (2) the prevalence, though not the unpredictability, of corruption also significantly affects non-compliance; and (3) "…a firm's age, sector or mode of ownership do not influence [a firm's] under-reporting of revenue." (p. 78).

There are several reasons to be concerned about these reported results. First, the number of observations in each of the seven regressions ranged from a low of 3,802 to a high of 4,781

[9] The question regarding the percentage of sales not reported used by Johnson *et al*. was prefaced by "It is thought that many firms in your industry, in order to survive and grow, may need to misreport their operational and financial results. Please estimate the degree of under-reporting by firms in your area of activity."

from a total number of observations of over 10,000 available in the WBES. The authors do not discuss the reasons for the attrition but it is likely the result of missing observations and non-response for both the dependent and independent variable. As attrition of this magnitude can lead to biased results, it is important that it be investigated further. Second, the authors do not link their choice of explanatory variables to the existing theoretical or empirical literature regarding firm tax non-compliance. As a result, some potential controls are overlooked and some controls may be mis-specified. Third, the equations across categories are not all identically specified and it is not clear if the results should be compared across regressions, given the different specifications. Finally, Batra *et al.* (2003) do not exploit the nature of the dependent variable. The firm's response regarding under-reporting behaviour is grouped into categories. When a quantitative outcome is grouped into known intervals on a continuous scale, the data are said to be "interval-coded".[10] However, Batra *et al.* (2003) define the dependent variable as a binary outcome for each category and estimate the resulting equations by Ordinary Least Squares (OLS). The considerable statistical limitations of such a linear probability model are well known.[11] There is an estimation technique that has been developed specifically for interval-coded data. This estimation procedure is known as "interval regression" and is undertaken using maximum likelihood techniques.

This paper addresses these shortcomings. First, it uses the interval regression technique and the extent to which various covariates affect the estimation results is also explored. Second, information from existing theoretical and empirical literature of firm tax non-compliance is used

[10] When the interval thresholds are unknown and have to estimated along with the other parameters, an ordered logit/probit is the preferred estimation framework.

[11] In the WBES, over 40% of the firms surveyed indicate that 100% of their sales are reported to the tax authority and between 6-14% of firms appear in one of the other categories. This means that the linear probability models estimated by Batra *et al.* will be either zero or one inflated. This results in a large number of predicted probabilities falling outside the unit interval.

to build the empirical model explored in this paper. Third, the issue of missing observations or non-response is explored. This paper also builds on the findings of Johnson *et al.* by extending the number of countries included in the sample and examines firms in a variety of different industries.

3 Data

In 1998, the World Bank Group launched its World Business Environment Survey (WBES). The WBES used many of the same questions from the enterprise survey conducted for the *1997 World Development Report*[12] (World Bank 1997) but expanded the number of businesses and countries surveyed and the questions/issues covered. The survey is a sample from the universe of registered businesses.[13] The WBES uses a form of stratified random sampling along with oversampling and nonrandom methods to correct for biases in the representation across firms with respect to industry characteristics that were not common.[14] To ensure adequate representation of firms by industry, size, ownership, export orientation, and location, sampling targets were agreed on across all regions.

The survey was completed by over 10,000 firms across eighty countries, including the West Bank and Gaza over a 20 month period between the latter half of 1998 and mid-2000. Table 2 lists the countries available in the data set by region along with the associated sample size. Approximately, one third of these countries were low income, one half were middle income and a little less than one fifth were high income. Private contractors, rather than

[12] Unfortunately, this dataset contains neither detailed information on characteristics of the firms nor the key variable of interest contained in the WBES dataset. As a result, it cannot be appended to the WBES dataset.

[13] It is very difficult to obtain information on unregistered or informal suppliers. Smith and Adams (1987) endeavored to do so by examining the extent of tax non-compliance by informal suppliers in six "at risk" sectors using results from a survey commissioned by the U.S. Internal Revenue Service (IRS) on the expenditures made by consumers on goods and services provided by these suppliers. For the 1992 tax year, the authors find that unreported income by these informal suppliers amounted to approximately US$59.6 billion and that informal suppliers tend to report only about 20% of their net business income.

[14] Unfortunately, the survey does not contain survey weights. The lack of survey weights means that individual indicators should not be used for precise country rankings in any particular dimension measured by the survey.

government entities, conduct the survey on behalf of the World Bank to ensure confidentiality and the greatest degree of participation. The survey is completed by managing directors, accountants, human resource managers and other company staff. The purpose of the survey was to assess and compare the business environment in a large number of countries. To achieve this goal, the survey gathered information regarding the firm's characteristics, such as size and ownership structure, as well as responses to multiple questions on the investment climate and the local business environment as shaped by domestic economic policy, governance, regulatory, infrastructural, and financial impediments, as well as assessments of public service quality. A more detailed description of the survey can be found in Batra *et al.* (2003).

Of interest to this study, the survey included a question regarding the extent and intensity to which firms fail to report income to the tax authority which permits exploring the links between tax non-compliance and various firm, business, governance, and political characteristics.[15] The main advantages of using the WBES to explore tax evasion over audit data include the fact that: (1) audit data are not widely available, unlike the WBES, which covers eighty countries; (2) audit data only include firms that are selected (for diverse reasons) or caught by the tax authorities, while the WBES is a broad sample of firms; and (3) the WBES includes additional information that is not included in tax audits. In particular, factors that firms perceive as business obstacles, such as taxes and regulations, which may effect a firm's decision to under-report, are included.

The WBES does not come without disadvantages. One disadvantage shared by both the WBES and audit data is that neither data source includes "ghosts" (Erard and Ho 2001); firms that operate solely in cash and avoid normal business obstacles and regulations. The effect of

[15] Similar to the data used by Johnson *et al.*, the exact phrasing of the question was "*What percentage of total sales would you estimate the typical firm in your area of activity reports for tax purposes?*"

omitting these firms from the estimates will bias estimates of tax non-compliance downwards and this bias will be larger for those countries that have a larger percentage of unregistered firms. Further, the percentage of sales that the firm fails to report to the tax authority is not expressed in actual quantities, as with Johnson *et al.* (2000), but rather it is expressed in intervals. As discussed above, there were seven possible answers: 100%, 90-99%, 80-89%, 70-79%, 60-69%, 50-59%, and less than 50%. This represents something of a constraint but its key advantage is that the interval-coded nature of the responses is likely to reduce a respondent's reporting bias. The large number of intervals, most of them of equal size, is beneficial as it provides information on the scale of under-reporting rather than simply its incidence. Given that the incidence of under-reporting is relatively high, as will be shown below, information on the scale of under-reporting is extremely useful.[16]

The distribution of answers to the sales reporting question over the entire sample is given in Table 1a. The first column of the table indicates that just over 18% of the firms in the sample did not respond to the question of interest. It is quite common in survey data to have missing data of this nature. If the data is missing for unknown reasons and the missing data is unrelated to the completeness of the other observations then it is reasonable to exclude the missing observations from the analysis, though efficiency is sacrificed. If, however, the missing data are systematically related to the phenomenon being modeled then there will be a sample selection problem. For example, if firms whose response is missing for the sales reporting question are more likely to be firms who under-report their income, then the missing observations represent more than just missing information. Unfortunately, there is insufficient evidence to definitively

[16] Further, the interval coding of the data along with the lack of data on the dollar amount of sales for the firms surveyed in the WBES means that the tax gap cannot be calculated. The tax gap is the difference between the taxes paid and the taxes that should have been paid. Such a calculation is useful in determining the amount of tax revenues lost to non-compliance.

ascertain the reasons for the missing observations in this particular case. Rather, the relationship between these missing observations and firm characteristics that will be used as explanatory variables in our model were explored. If the missing observations are randomly distributed amongst these explanatory variables then this provides some evidence that the missing information may not cause systematic bias in the estimates. There appeared to be no significant systematic relationship between nonresponse to the sales reporting question and the explanatory variables, with one exception noted below. Since no obvious explanation is available for the missing observation on the sales reporting questions, these observations (1,716) were dropped from the data set leaving 8,153 observations.

Of the firms that responded to the sales reporting question, Table 1b shows that 60% of firms worldwide indicate that the typical firm fails to report their sales in full to the tax authority and, of those firms, over 19% of them fail to report more than half their sales. This shows that business tax compliance is a significant issue. Not surprisingly, there appears to be some differences in perceived tax non-compliance across regions, as is shown in Figure 1. In particular, in OECD countries only approximately 40% of firms' under-report their sales to the tax authority and of those, approximately 50% fail to report only up to 10% of their sales. Further, compared to other regions, significantly more firms in Latin America and Asia fail to report more than 50% of their sales. In addition, there are significant differences in perceived firm tax compliance across countries.

There are, however, some differences in the distribution of the missing observations discussed above with respect to region and country. Only 11% of firms located in the former Soviet Union, approximately 15% of firms in the OECD and Latin America, approximately 20% in Asian and Africa and the Middle East, and 27% in Transition Europe failed to respond to the

sales reporting question. The missing observations appear to be randomly allocated amongst the countries in Latin America and Africa and the Middle East. In Transition Europe, no firm located in Albania responded to the question, indicating that survey was not carried out in a similar fashion to that in other countries rather than a response bias. As a result, Albania is excluded from the analysis (163 observations). Higher non-response also occurs in Malaysia, India and Thailand in Asia and Russia, Georgia, Kazakhstan, and the Ukraine in the Former Soviet Union but in all cases the non-response rate is not dramatically higher from that experienced in other countries in the regions. In the OECD, France, Canada and the United States all have a higher response rate than the other countries in the group but again the difference is not alarming.

4 Empirical Framework

4.1 Interval Regression Model

As discussed above, the survey question which forms the basis for the dependent variable refers to categories. As a result, a firm's reporting behaviour is not directly observed. Rather, firms are categorized on the basis of the percentage of sales that go unreported. When a quantitative outcome is grouped into known intervals on a continuous scale, the data is said to be "interval-coded". An ordered probit (or logit) is ideal when the dependent variable is discrete, ordinal in nature, and when the categories or thresholds are unknown.[17] In this case, the thresholds are estimated along with the model's coefficients and the variance of the error term is normalized to be one. It is possible, however, to modify the ordered probit model so that the thresholds are fixed at their known values and only the coefficients and the error variance are

[17] That is, for an ordered logit/probit model, while it can be said that two is greater than one, it cannot be determined if the difference between two and one is somehow twice as important as the difference between one and zero. The latter is not true of interval coded data.

estimated. This estimation procedure is known as "interval regression" and is undertaken using maximum likelihood techniques. The key advantage of interval regression over the ordered probit is that it provides an asymptotically more efficient estimator as it uses the known threshold information and involves estimating fewer parameters. It is also preferred to OLS, as OLS on the grouped dependent variable model is inconsistent.[18]

The following is the general setup of the model and is based on the discussion contained in Stewart (1983). The responses for the dependent variable are coded 1, 2, 3, 4, 5, 6, and 7 to capture seven distinct sales under-reporting categories. Let y_i denote the observable ordinal variable coded in this way for the ith firm and let y_i^* denote the underlying variable that captures the sales under-reporting of the ith firm. This can be expressed as a linear function of a vector of explanatory variables x_i using the following relationship:

$$y_i^* = x_i'\beta + u_i, \quad u_i \sim N(0, \sigma^2). \tag{1}$$

It is assumed that y_i^* is related to the observable ordinal variable y_i as follows:

$$\begin{aligned} y_i &= 1 \text{ if } y_i^* \leq 49\% \\ y_i &= 2 \text{ if } 49\% < y_i^* \leq 59\% \\ y_i &= 3 \text{ if } 59\% < y_i^* \leq 69\% \\ y_i &= 4 \text{ if } 69\% < y_i^* \leq 79\% \\ y_i &= 5 \text{ if } 79\% < y_i^* \leq 89\% \\ y_i &= 6 \text{ if } 89\% < y_i^* \leq 99\% \end{aligned} \tag{2}$$

[18] Another popular alternative is to define an artificial dependent variable that takes the value of the midpoint of the reported interval and then estimate the model by OLS. The resulting estimates are inconsistent but can be transformed using a minimum χ^2 method and the transformed estimates are consistent. However, the transformation requires equally strong distributional assumptions, usually that of normality and homoskedasticity, to that required of the interval regression model.

$$y_i = 7 \text{ if } y_i^* > 99\%$$

Based on the above, the seven possible components of the general log likelihood function for the i[th] individual is expressed as:

$$\begin{aligned} L_i = {} & I[y_i = 1] \times \log_e \{\Phi(\frac{49 - x_i'\beta}{\sigma}) - \Phi(\frac{a_0 - x_i'\beta}{\sigma})\} + I[y_i = 2] \times \log_e \{\Phi(\frac{59 - x_i'\beta}{\sigma}) - \\ & \Phi(\frac{49 - x_i'\beta}{\sigma})\} + I[y_i = 3] \times \log_e \{\Phi(\frac{69 - x_i'\beta}{\sigma}) - \Phi(\frac{59 - x_i'\beta}{\sigma})\} + I[y_i = 4] \times \\ & \log_e \{\Phi(\frac{79 - x_i'\beta}{\sigma}) - \Phi(\frac{69 - x_i'\beta}{\sigma})\} + I[y_i = 5] \times \log_e \{\Phi(\frac{89 - x_i'\beta}{\sigma}) - \\ & \Phi(\frac{79 - x_i'\beta}{\sigma})\} + I[y_i = 6] \times \log_e \{\Phi(\frac{99 - x_i'\beta}{\sigma}) - \Phi(\frac{89 - x_i'\beta}{\sigma})\} + \\ & I[y_i = 7] \times \log_e \{\Phi(\frac{a_K - x_i'\beta}{\sigma}) - \Phi(\frac{99 - x_i'\beta}{\sigma})\} \end{aligned} \tag{3}$$

where Φ denotes the cumulative distribution function of the standard normal, $\Phi(-\infty) = 0$, $\Phi(\infty) = 1$, $\log_e(\cdot)$ denotes the natural logarithmic operator, and I[·] is an indicator function that takes the value of one when the statement in the square brackets is true and zero when it is false. The relevant part of the log-likelihood is then triggered by the indicator function for whether the individual falls within one of the seven intervals in question. The maximum likelihood procedure now involves the estimation of the β parameter vector and the ancillary standard error parameter σ.

4.2 Diagnostic Tests

Unlike the situation with the ordered probit estimation, the estimated coefficients from an interval regression are interpretable as if y_i^* is observed for each i and estimated $E(y^* \mid x) = x\beta$ by OLS. That is, the estimated coefficients can be interpreted as the marginal effects (i.e. the change in percentage of sales reported given a change in the independent variable, holding all

else constant). It should be noted that the estimates contained in the β parameter vector are only interpretable in this way due to the assumption that y* given x satisfies the classical linear model assumptions. If these assumptions do not hold then the interval regression estimator of β would be inconsistent. As a result, it is important to test the key assumptions of functional form, homoskedasticity, and normality, all of which are used to derive the log likelihood function noted in equation (3). Unfortunately, these assumptions, despite their importance, are seldom tested in applied research. Equally, these assumptions, particularly homoskedasticity and normality, are rarely met when using micro data.

Chesher and Irish (1987) outline diagnostic tests for (pseudo) functional form, normality and homoskedasticity for the ordered probit model that are easily modified for the interval regression model.[19] These tests are all score (or Lagrange Multiplier) tests for which the test statistics take the form

$$\xi = 1' F(F'F)^{-1} F'1 \tag{4}$$

where 1 is an n-dimensional vector of ones, and F is a matrix with row order n where each row contains the score contributions for all the parameters of the model. ξ can be easily calculated as n times the non-centered R^2 from a regression of 1 on the columns of F.

The construction of the F matrices for these tests, which are described below, are based on computations of the pseudo-residuals. Usually, residuals are defined as the difference between the observed and estimated values of the dependent variable. However, the estimated values of the dependent variable obtained in the interval regression have no counterpart in the

[19] While it is worthwhile considering the results of the aforementioned tests, care should be exercised when interpreting the associated results. Orme (1990) has questioned the use of such score tests in the context of a simple binary probit and demonstrated their poor finite sample properties in this setting. In particular, he notes that there is upward size-distortion. Assuming that these findings extend to the interval regression model, the diagnostic tests may indicate that the model does not satisfy the classical linear model assumptions when in fact it does. However, the sample size for the data used in this paper is large enough that this concern likely does not apply.

data. Rather, pseudo-errors need to be computed. An expression for the pseudo-errors is obtained by differentiating equation (3) with respect to σ and denoted for the i^{th} individual as:

$$u_i = \frac{\phi(\frac{a_{(j-1)i} - x_i'\beta}{\sigma}) - \phi(\frac{a_{ji} - x_i'\beta}{\sigma})}{\sigma[\Phi(\frac{a_{ji} - x_i'\beta}{\sigma}) - \Phi(\frac{a_{(j-1)i} - x_i'\beta}{\sigma})]} \tag{5}$$

where $\phi(\cdot)$ denotes the probability density function for the standard normal, and a_{j-1} and a_j denote the known interval parameters for individual i (e.g. if y_i=2 then a_{j-1}=49 and a_j=59). The pseudo-residuals, e_i, are obtained by replacing the unknown parameters in (5) with their maximum likelihood estimates.

For the homoskedasticity and non-normality tests, higher-order moment residuals are required, specified as:

$$M_{\tau i} = \frac{(\frac{a_{(j-1)i} - x_i'\beta}{\sigma})^\tau \phi(\frac{a_{(j-1)i} - x_i'\beta}{\sigma}) - (\frac{a_{ji} - x_i'\beta}{\sigma})^\tau \phi(\frac{a_{ji} - x_i'\beta}{\sigma})}{\sigma[\Phi(\frac{a_{ji} - x_i'\beta}{\sigma}) - \Phi(\frac{a_{(j-1)i} - x_i'\beta}{\sigma})]} \tag{6}$$

The higher-order moment residuals are obtained by replacing the unknown parameters in (6) with their maximum likelihood estimates. The first four moment residuals are required for the desired tests and are defined as follows:

$$\begin{aligned} e_i^1 &= e_i \\ e_i^2 &= \hat{M}_{1i} \\ e_i^3 &= 2e_i^1 + \hat{M}_{2i} \\ e_i^4 &= 3e_i^2 + \hat{M}_{3i} \end{aligned} \tag{7}$$

The F matrix, or score contributions, is obtained by multiplying the pseudo-residuals by the various auxiliary variables in question.

4.2.1 Pseudo-Functional Form Test

The (pseudo) functional form test is a modified version of the RESET test (Ramsey 1969). F is given as

$$F = (e^1 x, e^1 \hat{y}^{*2}, ..., e^1 \hat{y}^{*K}, e^2) \tag{8}$$

where x includes a column of 1's if the grouped model contains an intercept and $\hat{y}^{*K}$ is the Kth power of model's predicted standardized index, $\hat{y}^* = \frac{x'\hat{\beta}}{\sigma}$. That test statistic ξ is distributed as $\chi^2(K-1)$. If the null hypothesis is rejected, this is evidence of model misspecification that may be rectified by considering additional variables and/or alternative functional forms (such as a semi-log model).

4.2.2 Test for Homoskedasticity

For the test of homoskedasticity, F is given as

$$F = (e^1 x, e^2 d_i)^{20} \tag{9}$$

The test statistic ξ is distributed as $\chi^2(q)$ where q is equal to the number of variables that are interacted with e^2. If the form of the heteroskedasticity is unknown, then $d_i = xx'$. This is simply White's test for heteroskedasticity of unknown form, modified for the interval regression model. The test is sensitive to model misspecification so if the test for (pseudo) functional form is rejected, it is also likely that the test for homoskedasticity will also be rejected.

As indicated previously, the estimated coefficients are inconsistent in the presence of uncorrected heteroskedasticity but it is possible to address this inconsistency by correcting for heteroskedasticity. Typically, the Huber (1967) "sandwich" estimator of the variance is used in

[20] When an intercept is estimated so that x always contains a unit element, e^2 is redundant in the test for homoskedasticity.

place of the conventional Maximum Likelihood variance estimator. This estimator is expressed as:

$$Var(\hat{\beta}) = [I(\hat{\beta})]^{-1}(x_i' e_i^2 x_i)[I(\hat{\beta})]^{-1} \quad (5)$$

where $I(\hat{\beta})$ is the information matrix for the $\hat{\beta}$ vector, computed at the maximum likelihood estimates. However, unlike with linear regression, in an interval regression model is it not possible to correctly specify $E(y^* \mid x)$ but misspecify Var(y*|x) which implies that the validity of this correction is highly questionable in the interval regression model and, hence, is not used in this paper. In this situation, the preferred choice is to correct for heteroskedasticity using a variation of Harvey's (1976) multiplicative heteroskedasticity, denoted as $Var(u_i)=\exp(d_i\gamma)^2$ where d_i is a matrix of independent variables, including a column of one's, that are the source of the heteroskedasticity and γ is a coefficient vector. The expression $\exp(d_i\gamma)$ replaces the σ noted in equation (3).[21]

4.2.3 Non-Normality Test

Finally, F in the usual $\chi^2(2)$ test for zero skewness and/or excess kurtosis is given by

$$F = (e^1 x, e^2, e^3, e^4) \quad (10)$$

4.3 Explanatory & Control Variables

As was outlined above, previous empirical work using audit data found relationships between tax non-compliance and various economic, political, and firm characteristics and these relationships will also be explored in this paper. Generally anything that is perceived by the firm as affecting their ability conduct business provides incentive for the firm to engage in tax non-compliance, in part to reduce costs and be more competitive, but it is generally acknowledge that

[21] In STATA, the heteroskedastic corrected interval regression model is estimated using the 'het' option on the 'intreg' command.

factors such as high tax rates, corruption, extortion by organized crime, and the legal environment are primary drivers of tax evasion. It is widely acknowledged high taxes, whether they be corporate, payroll, and/or sales taxes, provide incentives for firms to hide output to reduce their tax burden. If corruption is common, then among other effects, it will increase the cost of business, reduce morality, and reduce a firm's confidence in government; all of which are likely to have a negative relationship with tax compliance. The relationship between tax compliance and inconsistency in the application of laws and regulations, however, is more ambiguous. If a firm can individually garner the favour of the government(s) and/or courts in the interpretation and application of laws and regulations, then this may reduce tax non-compliance. On the other hand, if the firm does not benefit from this inconsistency, then it may resort to tax non-compliance.

Others drivers of firm tax non-compliance include: access to capital, inflation, exchange rate, political instability, and regulatory burden. Andreoni (1992) argues that "…individuals facing binding borrowing constraints may use tax evasion to transfer resources from the future to the present. Even if a person finds tax evasion undesirable in the absence of borrowing constraints, it could become desirable if a borrowing constraint is binding. Tax evasion, therefore, may be a high-risk substitute for a loan." (p. 35-36). Price inflation may influence under-reporting in two ways. First, if tax rates are not indexed then the tax burden rises simply due to "bracket creep". Second, inflation generates uncertainty and businesses may hedge against this uncertainty by engaging in (more) underground activity. Exchange rates can be an obstacle to doing business if the exchange rate is unpredictable and can increase the input costs faced by businesses. Various governmental factors can also influence under-reporting. Political instability constrains a countries ability to collect taxes and also reduces the efficiency of

collection, thereby increasing non-compliance. The degree of regulation is often cited as a factor that influences people to engage in underground activity as regulations reduce a firms' freedom of choice (e.g. Deregulation Commission 1991 and Schneider and Enste 2000).

Various characteristics of the firm will also be included as control variables. The previous literature makes it clear that tax non-compliance likely varies across governance models. The WBES provides an opportunity to explore the relationship between the legal structure of the business and tax non-compliance. Based on the literature, it is expected that publicly traded corporations will be the most compliant. In most countries, publicly traded corporations have a greater probability of being audited and are subject to public disclosure requirements and independent financial auditing, which tends to expose any under-reporting behaviour to the authorities. As audit levels and detection probability are greater, compliance should be higher. Industry sector, firm size, number of competitors, age, ownership, export status, and whether the firm subjects its financial statements to audits will also be included. Table 3 provides summary statistics for the key variables described in this section.

5 Results

The first task is to specify the values of the two end intervals, denoted above as a_0 and a_K in Equation (3). The values of $-\infty$, 0, and 1 are considered for the values of a_0. It is difficult to argue that a firm would report less than 0% of their sales. Instead, it could be argued that the first interval should begin at either 1 or 0%: 1% since it can be argued that firms that report 0% of their sales would likely be operating completely as "ghosts" and would not have been selected for an interview since there would be no formal record of the firm. The results are quite

insensitive to the choice of a_0 hence the value of a_0 is set to 1%, which seems economically reasonable given the sampling design which excluded unregistered firms.

The values of ∞ and 100 are considered for the values of a_K. When the last interval is treated as open ended ($a_K=\infty$), the maximum predicted value is in excess of 140% and over 25% of the predicted values exceed 100%. These numbers seem economically implausible, particularly that over 25% of firms report in excess of 100% of their sales to the tax authority. Hence, the last interval is set to 100. With the bounds of the intervals set at 1 and 100, the maximum predicted value is approximately 105% and approximately 1% of the predicted values exceed 100%. These figures seem to be economically plausible and accord with the findings of Rice (1992) who found that 6% firms do over-report to some extent. In particular, firms may over-report due to a misinterpretation of tax laws, to avoid a tax audit, to secure financing, or particularly for public corporations, to appear more competitive. In the WBES, firms that over-report will likely be included in the full compliance category. The effect of choosing different values for a_0 and a_K on the resulting estimates will be explored. It should be noted that the sensitivity of the results and predicted values to the treatment of the end intervals is often overlooked in the literature yet can be of clear importance to the economic validity of the resulting estimates.

Prior to estimation, it is important to consider the necessary assumptions for consistent estimates and in particular, the assumption of normality. It is anticipated that the assumption of normality is likely violated in this model. Table 1b clearly shows that the distribution of sales reporting is negatively skewed. Normality may be more closely approximated by taking the natural logarithm of 100 minus the dependent variable, which is simply the natural logarithm of the percentage of sales *not* reported to the tax authority. Such a transformation requires that the

dependent variable take only strictly positive values and care exercised in interpreting the resulting estimates. In particular, the coefficients in such a model are no longer marginal effects but 100 multiplied by the estimated coefficient represents the percentage change.[22] Preliminary investigation supports this transformation and is the specification used to obtain the results reported here.

The results are presented in Table 4. The basic model is presented in the first column (Model 1a). Various goodness of fit measures and the results from the various diagnostic tests are reported near the end of the Table. Larger (less negative) log-likelihood values are indicative of a better fit. The log-likelihood values can be compared across the models only if they have the same samples and dependent variable. The R-square, for technical reasons, cannot be computed in the same way in interval regressions as it is in OLS regression. Various pseudo R-square measures, however, have been proposed, but there is no generally accepted measure. Veall and Zimmermann (1996) recommend the measure of McKelvey and Zavoina (1975), which is reported in Table 4.[23]

For model 1a, the null of correct functional form is not rejected at all the usual significance levels, but the tests for homoskedasticity and normality are both rejected at the usual significance levels. Plots of the pseudo residuals from this models supports the finding of non-normality and it is not clear if any further transformations of the dependent variable, beyond those already pursued, would result in normality. In addition, further transformations would

[22] Kennedy (1981), however, notes that in a semi-log model the coefficient on a dummy variable must be appropriately transformed. That is, if b_i is the estimated coefficient on a dummy variable and $V(b_i)$ is the estimated variance of b_i then $g_i = 100\ (\exp(\ b_i - \frac{V(b_i)}{2}) - 1)$ gives an estimate of the percentage impact of the dummy variable on the dependent variable. The effect of this transformation on the resulting coefficients was explored. As the resulting change to the coefficients reported in Table 6 were marginal, the results are not reported but are available from the author upon request.

[23] The McKelvey and Zavoina (1975) R-square is computed in STATA with the 'fitstat' command.

make it difficult to interpret the estimated coefficients. It is, however, possible to correct for the heteroskedasticity, as noted above, if a form for it is specified. Correcting for the heteroskedasticity can often achieve normality in the (pseudo) residuals. It is not unreasonable to assume that much of the heteroskedasticity results from countries and business type. Applications of the test for homoskedasticity of known form across subsets of the variables confirm this suspicion.[24] Model 1a is re-estimated to account for this known form of the heteroskedasticity and these results are presented in Table 4 under the column heading Model 1b. These results will now be discussed.

With respect to the causal variables, firms that perceive high taxes and regulations, and government corruption as obstacles to doing businesses report significantly less of their sales to the tax authority. The results presented in Table 4 indicate that government corruption has the single largest effect, resulting in the percentage of sales not reported to the tax authority being a whopping 53.4 percent higher and taxes at 20.2 percent higher. In comparison, Johnson *et al.* (2000) also found a positive relationship between non-compliance and government corruption but failed to find a relationship between compliance and tax payments. Exchange rates have a positive effect on compliance, which is the only coefficient that fails to follow *a priori* expectations. All other variables are statistically insignificant.

The control variables also provide interesting results and these results also accord with those obtained using tax audit data. Sole proprietorships and private corporations are the least compliant form business of all the business models included in this study. This result accords with those found by Rice (1992), that found that public corporations were more compliant than other types of firms, and Hanlon *et al.* (2005) who found that private corporations were less

[24] The results are not reported but are available from the author upon request.

compliant. Small firms, and firms with more than three competitors are also found to be the least compliant while large firms and firms with no competitors are more compliant. Giles (2000) reports a similar result with respect to size, whereas Rice (2002) found firm size and tax non-compliance were positively related and Hanlon *et al.* (2005) found that compliance was a convex function over firm size. Finally, being foreign, government owned, and having financial statements audited all lead to significantly higher compliance. Both Giles (2000) and Chan and Mo (2000) found that foreign owned firms are more compliant, and the results in Table 6 provide further support for this result. The relationship between internal audit controls of the firm and tax compliance has not been investigated in previous empirical work.

Numerically, the percentage of sales *not* reported to the tax authority is: 34 percent higher for sole proprietorships and 13.2 percent higher for private corporations; 8.6 percent lower for firms in the service sector; 20.1 percent higher for small firms and 19.3 percent lower for large firms; 5.7 percent greater for firms less than five years of age; 11.9 percent lower for firms with no competitors; 15.4 percent and 15 percent lower for foreign owned firms and firms with audited financial statements, respectively. Chan and Mo (2000) report that export oriented firms are more compliant but this characteristic is found to be insignificant. Previous empirical studies all found a significant relationship between industry and tax non-compliance whereas only firms in the service industry are found to have a statistically different reporting pattern. The direction of the relationship, however, is a bit surprising however it must be kept in mind that the industry the categories included in the WBES data set are likely too broadly defined to yield any informative results. For example, the service sector will include restaurants and hair salons as well as law an accounting firms.

6 Conclusion

Very little is actually known about firm tax compliance due to a lack of detailed and readily available data. The purpose of this paper was to use a unique and recently available dataset that contained information on firms from around the world to investigate some of the factors that effect business tax compliance. This is one of the first studies to examine firm tax compliance using worldwide data. The majority of previous empirical studies were confined to examining firms within a particular country using tax audit data.

The empirical strategy employed in this paper exploits the nature of the dependent variable, which is interval coded, and uses interval regression. The estimated coefficients from an interval regression are interpretable as marginal effects provided that the model satisfies the assumptions of correct functional form, homoskedasticity, and normality. If these assumptions do not hold then the interval regression estimator of β is inconsistent. These assumptions are investigated using standard diagnostic tests that have been modified for the interval regression model, a step frequently ignored in applied research. The test results support the use of the semi-log model that is corrected for heteroskedasticity, the origin of which is found to result at the country level and from the business type.

Overall, evidence is presented that shows that firms in all regions around the world engage in tax non-compliance, but that there is substantial variation within regions. The detailed regression results indicate that, not surprisingly, taxes and government corruption are positively related to tax non-compliance and the magnitude is quite large in both cases. Access to capital, political instability, organized crime, inflation, and the legal system are found to have no statistically significant effect on tax non-compliance. There is also a large correlation between the legal organization of a business and under-reporting, with sole proprietorships and private

corporations being the least compliant business type. Firm size is also correlated with tax compliance, with small firms reporting less and large firms more of their sales to the tax authority. Firms with no competitors and government owned firms are significantly more compliant.

The findings do suggest a role for public policy, as well as actions to be considered by the tax authority and items that require further study. First, the findings suggest that administrations interested in reducing business tax non-compliance should consider reducing business taxes, minimizing the number of regulations, and reducing, if not eliminating, government corruption. Admittedly, taking action of these issues is complex and involves more than just the tax authority. Second, tax authorities should consider auditing sole proprietorships, private corporations, small firms at a higher rate and requiring all firms to have their financial statement audited by a third party. Finally, based on this study, it is not entirely clear why large firms and firms that are less than five years of age are more compliant and how precisely exchange rates lead to higher tax compliance. Further exploration into these relationships appears to be a worthwhile venture. In addition, recently firm level data has been made available that was collected following 2000. This data includes similar variables to those included in this study, including the dependent variable. It would be worthwhile to pool these new data sources together with the WBES to investigate firm under-reporting in more detail.

TABLES

Table 1a: Univariate Frequencies of Percentage of Sales Reported to Tax Authorities, including missing observations

	Missing	<50%	50-59%	60-69%	70-79%	80-89%	90-99%	100%
Frequency	1,879	936	694	501	703	916	1,096	3,307
Percent	18.73%	9.33%	6.92%	4.99%	7.01%	9.31%	10.93%	32.96%
Observations	10,032							

Table 1b: Univariate Frequencies of Percentage of Sales Reported to Tax Authorities, excluding missing observations

	<50%	50-59%	60-69%	70-79%	80-89%	90-99%	100%
Frequency	936	694	501	703	916	1,096	3,307
Percent	11.48%	8.51%	6.14%	8.62%	11.24%	13.44%	40.56%
Observations	8,153						

Table 2: Countries Surveyed, Categorized by Region, and Number of Observations in Each Country

Country	Observations
Africa and Middle East	
Botswana	87
Cameroon	45
Côte d'Ivoire	66
Egypt	102
Ethiopia	77
Ghana	85
Kenya	89
Madagascar	84
Malawi	42
Namibia	66
Nigeria	73
Senegal	85
South Africa	99
Tanzania	66
Uganda	106
West Bank and Gaza	68
Zambia	64
Zimbabwe	102
Total	1,456
Asia	
Bangladesh	39
Cambodia	239
China	85
India	160
Indonesia	72
Malaysia	50
Pakistan	80
Philippines	90
Singapore	88
Thailand	369
Total	1,272
OECD	
Canada	95
France	89
Germany	77
Italy	81
Portugal	84
Spain	89
Sweden	82
United Kingdom	79
United States[1]	94
Total	770
Transition Europe	
Bosnia and Herzegovina	104
Bulgaria	97
Croatia	112
Czech Republic	106
Estonia	126
Hungary	121
Lithuania	29
Poland	208
Romania	125
Slovak Republic	24
Slovenia	123
Turkey	124
Total	1,299
Former Soviet Union	
Armenia	111
Azerbaijan	114
Belarus	116
Georgia	99
Kazakhstan	102
Kyrgyzstan	124
Moldova	116
Russia	481
Ukraine	188
Uzbekistan	118
Total	1,569
Latin American and Caribbean	
Argentina	83
Belize	33
Bolivia	87
Brazil	172
Chile	94
Colombia	92
Costa Rica	80
Dominican Republic	90
Ecuador	85
El Salvador	95
Guatemala	86
Haiti	95
Honduras	85
Mexico	80
Nicaragua	86
Panama	80
Peru	102
Trinidad and Tobago	93
Uruguay	87
Venezuela	82
Total	1,787

Notes: [1]Denotes the omitted category in estimation.

Table 3: Data Summary

Variable	Obs.	Mean	Standard Deviation	Minimum	Maximum
% of Sales Reported	8153	2.978	2.166	1	7
Causes of Under-reporting					
Financing	7526	0.811	0.391	0	1
Political Instability	7376	0.842	0.365	0	1
Organized Crime	7196	0.631	0.482	0	1
High Taxes & Regulations	7660	0.892	0.311	0	1
Corruption	7485	0.586	0.497	0	1
Inflation	7425	0.846	0.361	0	1
Exchange Rate	7259	0.741	0.438	0	1
Laws & Regs In consistent	8029	0.478	0.493	0	1
Legal Organization of Company					
Sole Prop.	8153	0.180	0.384	0	1
Partnerships	8153	0.173	0.378	0	1
Private Corp	8153	0.280	0.449	0	1
Public Corp. [1]	8153	0.130	0.336	0	1
Other Business	8153	0.154	0.361	0	1
Industry Sector					
Manufacturing[1]	7418	0.370	0.483	0	1
Service	7418	0.424	0.494	0	1
Other	7418	0.035	0.185		
Agriculture	7418	0.075	0.264	0	1
Construction	7418	0.096	0.294	0	1
Firm Size					
Small	8139	0.393	0.488	0	1
Medium[1]	8139	0.412	0.492	0	1
Large	8139	0.194	0.396	0	1
Firm Age					
<5	8153	0.257	0.437	0	1
5-15[1]	8153	0.354	0.478	0	1
>15	8153	0.389	0.487	0	1
Number of Competitors					
No Competitors	7907	0.111	0.314	0	1
1-3[1]	7907	0.399	0.490	0	1
>3	7907	0.490	0.500	0	1
Other					
Foreign Owned	8153	0.183	0.387	0	1
Gov. Owned	8153	0.120	0.325	0	1
Exporter	8153	0.339	0.473	0	1
Fin. Statements Audited	8153	0.586	0.493	0	1

Notes: [1]Denotes the omitted category in estimation.

Table 4: Estimation Results for Semi-log Model

Dependent Variable: Natural Log of Percentage of Sales Not Reported to Tax Authority

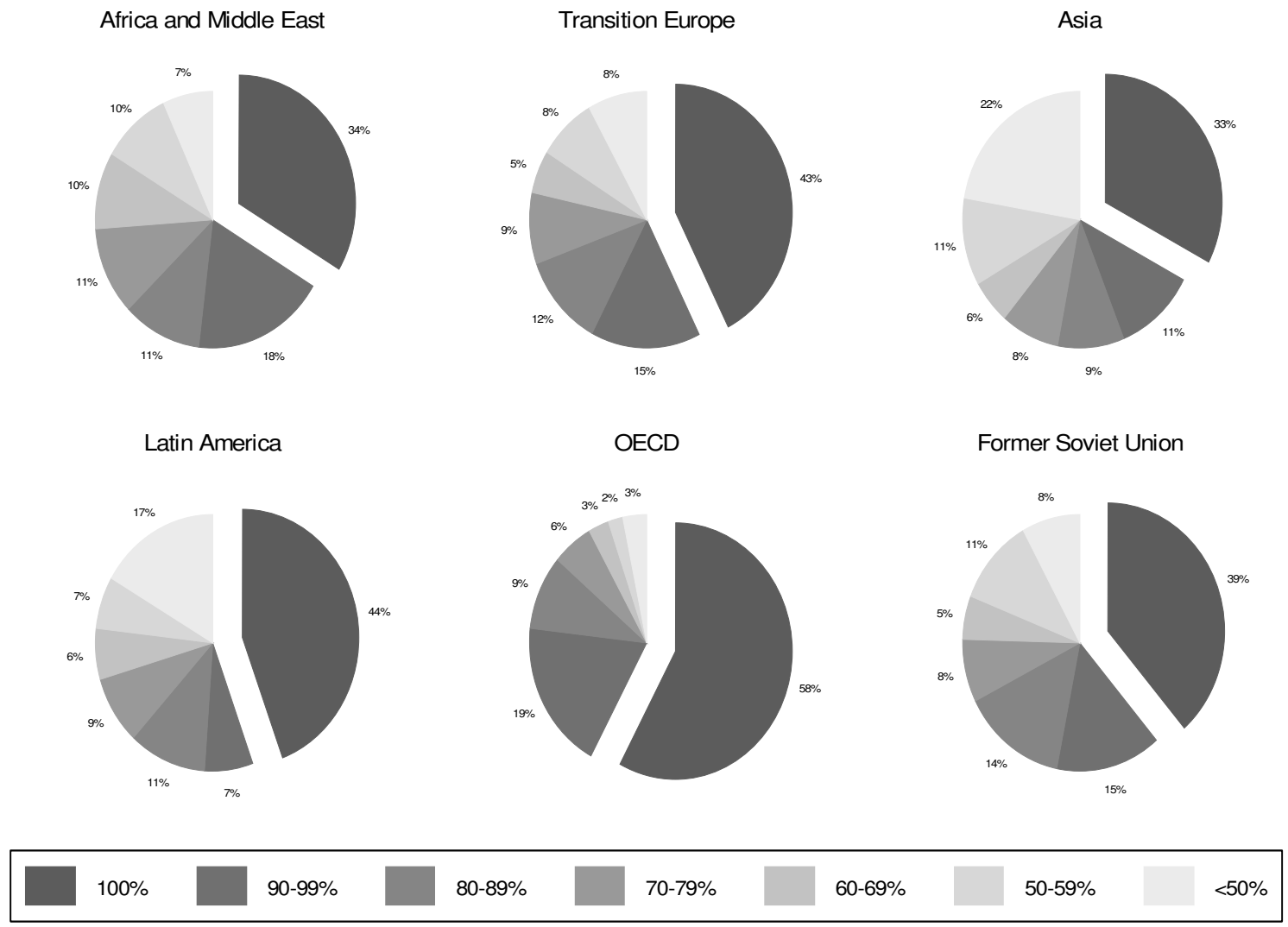
Variable	Model 1a	Model 1b Het. Corr.
Constant	**1.185**	**0.934**
	(0.182)*	**(0.155)***
Causes of Under-reporting		
Obstacle: Financing	**0.096**	0.045
	(0.04)**	(0.035)
Obstacle: Political Instability	-0.012	-0.0154
	(0.048)	(0.036)
Obstacle: Organized Crime	**0.092**	(0.065)
	(0.045)**	(0.042)
Obstacle: Taxes & Regs.	**0.212**	**0.202**
	(0.049)***	**(0.046)*****
Obstacle: Corruption	**0.527**	**0.534**
	(0.045)***	**(0.046)*****
Obstacle: Inflation	-0.013	0.028
	(0.051)	(0.044)
Obstacle: Exchange Rate	0.068	**-0.057**
	(0.047)	**(0.027)****
Obstacle: Laws & Regs. Inconsistent	**-0.067**	-0.046
	(0.04)*	(0.039)
Control Variables‡		
Sole Prop.	**0.211**	**0.340**
	(0.082)***	**(0.070)*****
Partnerships	0.031	0.060
	(0.077)	(0.065)
Other	0.040	0.118
	(0.086)	(0.083)
Private Corporation	**0.131**	**0.132**
	(0.066)**	**(0.059)****
Other Industry	-0.069	0.030
	(0.136)	(0.125)
Service	-0.070	**-0.086**
	(0.046)	**(0.043)****
Agriculture	-0.096	-0.087
	(0.083)	(0.080)
Construction	0.029	0.041
	(0.072)	(0.069)
Small: <50 employees	**0.174**	**0.201**
	(0.050)***	**(0.048)*****
Large: >500 employees	**-0.163**	**-0.193**
	(0.055)***	**(0.051)*****
<5 years of age	-0.041	**-0.057**
	(0.050)	**(0.027)****
>15 years of age	-0.063)	-0.059
	(0.050)	(0.045)
No Competitors	-0.064	**-0.119**
	(0.065)	**(0.062)****

>3 Competitors	**0.148**	0.084
	(0.060)**	(0.053)
Foreign Owned	**-0.124**	-0.011
	(0.053)**	(0.036)
Gov. Owned	**-0.150**	**-0.154**
	(0.071)**	**(0.065)****
Export	-0.049	-0.024
	(0.046)	(0.042)
Audits	**-0.239**	**-0.150**
	(0.048)***	**(0.042)*****
Country Dummies	Yes	Yes
σ	1.347	-
(s.e.)	(0.013)	
Pseudo Log-Likelihood Value	-11788.379	-11614.318
McKelvey & Zavoina Pseudo R^2	0.228	-
LRT-OS	1357.86	16652.51
[d.o.f.; P-Value]	[102;0.000]	[102; 0.000]
Pseudo Functional Form Test	0.001	-
[P-Value]	[0.979]	
Homoskedasticity Test of Unknown Form	1111.411	-
[P-Value]	[0.000]	
Normality Test	1699.6542	-
[P-Value]	[0.000]	
Observations	5393	5393

NOTES: ***, **, and * denote statistical significance at the 1%, 5% and 10% level respectively. d.o.f. denotes degrees of freedom. Standard errors (s.e.) are noted in parenthesis.

‡ Omitted categories are Public Corporations, Manufacturing, Medium, Between 5 and 15 years of age, Between 1 and 3 competitors, and "United States".

Figure 1: Percentage of Sales Reported to Tax Authorties



Source: World Business Environment Survey

REFERENCES

Allingham, M. and A. Sandmo. 1972. Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics* 1:323-338.

Andreoni, James. 1992. IRS as Loan Shark: Tax Compliance with Borrowing Constraints. *Journal of Public Economics* 49:35-46.

_____. Brian Erard, and Jonathan Feinstein. 1998. Tax Compliance. *Journal of Economic Literature* 36:818-860.

Apel, Mikael. 1994. An Expenditure-Based Estimate of Tax Evasion in Sweden. Working Paper, Department of Economics, Uppsala University.

Bajada, Christopher. 1999. Estimates of the Underground Economy in Australia. *The Economic Record* 75: 369-84

Batra, Geeta, Daniel Kaufmann and Andrew H.W. Stone. 2003. *Investment Climate Around the World: Voices of the Firms from the World Business Environment Survey*. Washington, D.C.: The World Bank.

Border, Kim C. and Joel Sobel. 1987. Samurai Accountant: A Theory of Audit and Plunder. *Review of Economic Studies* 54:525-540.

Chan, K. Hung and Phyllis Lai Lan Mo. 2000. Tax Holidays and Tax Non-compliance: An Empirical Study of Corporate Tax Audits in China's Developing Economy. *The Accounting Review* 75:469-484.

Chen, Kong-Pin and C. Y. Cyrus Chu. 2005. Internal Control and External Manipulation: A Model of Corporate Income Tax Evasion. *RAND Journal of Economics* 26:151-164.

Chesher, A. and M. Irish. 1987. Residual Analysis in the Grouped and Censored Normal Regression Model. *Journal of Econometrics* 34:33–61.

Christian, Charles W. 1994. Voluntary Compliance with Individual Income Tax: Results from the 1988 TCMP study. *The IRS Research Bulletin Publication 1500* 9:35-42.

Cowell, Frank. 2004. Sticks and Carrots. In *The Crises in Tax Administration* , eds. Henry Aaron and Joel Slemrod, 230-258. Washington, D.C.: The Brookings Institution Press.

Cremer, Helmuth, Maurice Marchand and Pierre Pestieau. 1990. Evading, Auditing and Taxing: the Equity-Compliance Tradeoff. *Journal of Public Economics* 43:67-92.

Crocker, Joel and Joel Slemrod. 2005. Corporate Tax Evasion with Agency Costs. *Journal of Public Economics* 89: 1593-1610.

Dell'Anno, Roberto, Miguel Gómez-Antonio, and Angel Pardo. 2007. The Shadow Economy in Three Mediterranean Countries: France, Spain and Greece. A MIMIC approach. *Empirical Economics* 33: 51-84.

Deregulation Commission (1991) Opening of Markets and Competition, Report Presented to German Federal Government, Bonn.

Draeseke, Robert and David .E.A. Giles. 2002. Modelling the New Zealand Underground Economy Using Fuzzy Logic Techniques. *Mathematics and Computers in Simulation* 59: 115-123.

Engström, Per and Bertil Holmlund. 2006. Tax Evasion and Self-Employment in a High-Tax Country: Evidence from Sweden. Working Paper, Department of Economics, Uppsala University.

Erard, Brian. 1992. The Influence of Tax Auditing on Reporting Behaviour. In Why People Pay Taxes, ed. Joel Slemrod. Ann Arbor: University of Michigan.

_____ and Chih-Chin Ho. 2001. Searching for Ghosts: Who are the Nonfilers and How Much Tax do they Owe? *Journal of Public Economics* 81:25-50.

Feldman, Naomi and Joel Slemrod. 2007. Estimating Tax Noncompliance with Evidence from Unaudited Tax Returns. *Economic Journal* 117: 327-352.

Giles, David E.A.. 1999. The Hidden Economy and the Tax-Gap in New Zealand: A Latent Variable Analysis. *Empirical Economics* 24: 621-640.

_____. 2000. Modeling the Tax Compliance Profiles of New Zealand Firms. In Taxation and the Limits of Government, eds. Gerald W. Scully and Patrick J. Caragata, 243-270. Boston: Kluwer Academic Publishers.

_____ and P.J. Caragata. 2001. The Learning Path of the Hidden Economy: Tax and Growth Effects in New Zealand. *Applied Economics* 33: 1857-1867.

Giles, David E.A. and B.J. Johnson. 2002. Taxes, Risk-Aversion, and the Size of the Underground Economy: A Nonparametric Analysis With New Zealand Data. Pacific *Economic Review* 7: 97-113.

Giles, David E.A. and Lindsay M. Tedds. 2002. Taxes and the Canadian Underground Economy. Toronto: The Canadian Tax Foundation.

Hanlon, Michelle, Lillian Mills, and Joel Slemrod. 2005. An Empirical Examination of Corporate Tax Non-compliance. Working Paper, University of Michigan.

Harvey, Andrew. 1976. Estimating Regression Models with Multiplicative Heterocedasticity. *Econometrica* 44: 461-65.

Huber, P.J. 1967. The Behaviour of Maximum Likelihood Estimates under Non-standard Conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221-223. Berkeley: University of California Press.

Internal Revenue Service. 2004. Interactive Tax Gap Map. National Headquarters Office of Research.

Johnson, Simon, Daniel Kaufmann, John McMillan, and Christopher Woodruff. 2000. Why do firms hide? Bribes and Unofficial Activity After Communism. *Journal of Public Economics* 76:495-520.

Joulfaian, David. 2000. Corporate Tax Evasion and Managerial Preferences. *Review of Economics and Statistics* 82:698-701.

_____ and Mark Rider. 1998. Differential Taxation and Tax Evasion by Small Business. *National Tax Journal* 51:676-87.

Kennedy, P. 1981. Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations. *American Economic Review* 71: 801.

Lyssiotou, Panayiota, Panos Pashardes and Thanasis Stengos. 2004. Estimates of the Black Economy Based on Consumer Demand Approaches. *Economic Journal* 114:622-640.

McElvey, W. and Zavoina, R. 1975. A Statistical Model for the Analysis of Ordinal Level Dependent Variables. *Journal of Mathematical Sociology* 4:103-120.

Mookherjee, Dilip and Ivan P.L. P'ng. 1989. Optimal Auditing, Insurance and Redistribution. *Quarterly Journal of Economics* 104:399-415.

Orme, C. 1990. The Small-Sample Performance of the Information Matrix Test. *Journal of Econometrics* 46:309–331.

Pissarides, Christopher A., and Guglielmo Weber. 1989. An Expenditure-based Estimate of Britain's Black Economy. *Journal of Public Economics* 39:17-32.

Ramsey, J.B. 1969. Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society* 31:350 – 71.

Rice, Eric M. 1992. The Corporate Tax Gap: Evidence on Tax Compliance by Small Corporations. In *Why People Pay Taxes*, ed. Joel Slemrod, 125-166. Ann Arbor: University of Michigan Press: Ann Arbor.

Sanchez, Isabel and Joel Sobel. 1993. Hierarchical Design and Enforcement of Income Tax Policies. *Journal of Public Economics* 50:345-369.

Schuetze, Herb J. 2002. Profiles of Tax Non-compliance Among the Self-Employed in Canada, 1969-1992. *Canadian Public Policy* XXVIII: 219-237.

Schnieder, Friedrich. 2006. Shadow Economies of 145 Countries all over the World: What do we Know? Working Paper, University of Linz, Austria.

_____ and Dominik Enste (2000) "Increasing Shadow Economies All Over the World – Fiction or Reality? A Survey of the Global Evidence of Their Size and Their Impact From 1970 to 1995." 38 *Journal of Economic Literature* 77-114.

Scotchmer, Suzanne. 1987. Audit Classes and Tax Enforcement Policy. *American Economic Review* 77:229-233.

_____ and Shlomo Yitzhaki. 1996. The Cost of Taxation and the Marginal Efficiency Cost of Funds. *International Monetary Fund Staff Papers* 43:25-34.

Smith, J. and T. Adams. 1987. The Measurement of Selected Income Flows in Informal Markets, 1981 and 1985-1986. Working Paper, University of Michigan Survey Research Centre.

Stewart, M. 1983. On Least Squares Estimation when the Dependent Variable is Grouped. *Review of Economic Studies* 50:737-753.

Tannenbaum, Robert. 1993. Corporate Tax Disclosure: Good or Bad for the Commonwealth? Working Paper, Massachusetts Special Commission on Business Tax Policy.

Tedds, Lindsay M., 2005, "The Underground Economy in Canada," in Chris Bajada and Friedrich Schneider (eds.) *Size, Causes and Consequences of the Underground Economy* (Ashgate Publishing, UK).

Tedds, Lindsay M. 2007. Estimating the Income Report Function for the Self-Employed. Working Paper, School of Public Administration, University of Victoria.

Trandel, Greg and Arthur Snow. 1999. Progressive Income Taxation and the Underground Economy. *Economics Letters* 62:217-222.

Veall, Michael R. and Kurt F. Zimmermann. 1996. Pseudo-R^2 Measures for Some Common Limited Dependent Variable Models. *Journal of Economic Surveys* 10:241-259.

Watson, Harry. 1985. Tax Evasion and Labor Markets. *Journal of Public Economics* 27:231-246.

World Bank. 1997. World Development Report 1997: The State in a Changing World. Washington, D.C.: World Bank Group.