Paulus, Alari

Working Paper

Tax evasion and measurement error: An econometric analysis of survey data linked with tax records

ISER Working Paper Series, No. 2015-10

Provided in Cooperation with:
Institute for Social and Economic Research (ISER), University of Essex

Suggested Citation: Paulus, Alari (2015) : Tax evasion and measurement error: An econometric analysis of survey data linked with tax records, ISER Working Paper Series, No. 2015-10, University of Essex, Institute for Social and Economic Research (ISER), Colchester

This Version is available at:
https://hdl.handle.net/10419/126489

ISER Working Paper Series

Tax evasion and measurement error: An econometric analysis of survey data linked with tax records

Alari Paulus
Institute for Social and Economic Research
University of Essex

No. 2015-10
May 2015

INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

www.iser.essex.ac.uk

Non-technical summary

Given its concealed nature, tax non-compliance raises non-trivial challenges for its measurement and it has proved very difficult to provide hard evidence on its scale and incidence. The paper estimates the extent and determinants of individual tax compliance behaviour by utilising a dataset, where income survey data have been linked with tax records for Estonia at the individual level. The linkage is unique in this context for not requiring consent by the survey respondents and is achieved for nearly all respondents, hence, avoiding potential issues with a biased sample.

We focus on wages and salaries and propose a novel econometric method to model income reporting to the tax authority and in the household survey jointly, in both cases allowing observed income values to differ from their true values. Our paper connects the empirical literature on tax evasion with another strand of literature, where linked datasets of a similar nature have been used to study survey measurement error but assuming register incomes to be accurate.

Our identification strategy is based on the assumption that public sector employees have no opportunities to hide their (public) employment income but are comparable to private sector employees in other aspects. This allows us to determine true employment income for some employees and estimate an econometric model, where the pair of observed income measures is related to individual characteristics and (partly unobserved) true income.

We find significant effects of various socio-demographic characteristics (e.g. gender, age, education, nationality, marital status, region) and labour-market characteristics (occupation, industry, firm size) on tax compliance. Our results indicate substantial tax non-compliance for wages and salaries overall: about 12% of total employment income is not reported to the tax authority, mostly due to partial underreporting of incomes which concerns more than 20% of employees. The share of employees who fully evade taxes is estimated to be marginal (2-3%). Across the estimated true earnings distribution, compliance is estimated to be lower for the bottom and the top decile group. There are also substantial measurement errors in survey income, revealing a pattern where low levels of true income tend to be overreported in the survey while it is the opposite for medium and high values of true income.

Sizeable underreporting of earnings highlights the limitations of third-party reporting and tax withholding, which these incomes are subject to. It suggests scope for employees and employers to collude for tax evasion, raising questions about the effectiveness of such tax enforcement mechanisms. The results also challenge the common view in the literature that only a marginal share of taxes on wages and salaries is evaded and suggests that more attention to employment income could be warranted. Finally, tax compliance patterns have implications for the progressivity and redistributive aspects of the tax system.

Tax evasion and measurement error

An econometric analysis of survey data linked with tax records *

Alari Paulus†

Institute for Social & Economic Research (ISER)
University of Essex

April 30, 2015

Abstract

We use income survey data linked with tax records at the individual level for Estonia to estimate the determinants and extent of income tax compliance in a novel way. Unlike earlier studies attributing income discrepancies between such data sources either to tax evasion or survey measurement error, we model these processes jointly. Focussing on employment income, the key identifying assumption made is that people working in public sector cannot evade taxes. The results indicate a number of socio-demographic and labour market characteristics, which are associated with non-compliance. Overall, people in the bottom and the top part of earnings distribution evade much more and about 12% of wages and salaries in total are underreported, which is very substantial for a major income source subject to third party reporting and tax withholding.

Keywords: tax compliance, measurement error, linked data, Estonia
JEL codes: D31, H26, H31

*Acknowledgements: The paper uses Estonian Social Survey 2008 linked with administrative tax records, made available by Statistics Estonia. I am very grateful to Steve Pudney for all the valuable discussions and Mari Toomse-Smith for her help with the data access and related queries. The paper has also benefitted from comments and suggestions received from Chris Bollinger, Francesco Figari, Carlo Fiorio, Henry Ohlsson, Karsten Staehr and the participants of the 3rd General Conference of the International Microsimulation Association, the Shadow-2011 Conference, the 2nd Microsimulation Research Workshop, the 69th IIPF Annual Congress, the 2014 HMRC/ESRC International Tax Analysis Conference and seminars at the University of Tartu, University of Essex, Tallinn University of Technology and Bank of Estonia. This work was supported by the Economic and Social Research Council (ESRC) through the Research Centre on Micro-Social Change (MiSoC), award no. RES-518-28-001. Any errors and the views expressed in the paper are the author's sole responsibility.

†Contact: apaulus@essex.ac.uk; ISER, University of Essex, Colchester, CO4 3SQ, UK.

1 Introduction

Income tax evasion, i.e. a deliberate act of non-compliance with legal requirements to disclose income (obtained by legal means) to tax authorities in order to reduce tax liability, undermines the intended effects of a tax by eroding the tax base and altering the distribution of tax burden among individuals. It also affects labour supply (and demand) behaviour by introducing an additional choice margin in the form of undeclared work as opposed to declared work and, hence, can distort the allocation of economic resources. Furthermore, it increases the costs for the society to enforce tax rules.[1] On the other hand, tax evasion may have not only negative consequences, e.g. (partly) undeclared work could provide the only employment option for the most vulnerable. For this reason and because enforcement is costly, it is neither optimal nor feasible to eliminate tax evasion completely. However, to design optimal tax and enforcement rules one needs to know who evades taxes, their reasons for doing so and the extent of non-compliance.

The main constraint for empirical research on tax evasion is, unsurprisingly, the lack of suitable data, this being especially pronounced for developing countries. To study and explain income tax evasion at the micro-level, one would essentially need a measure of undeclared income for individuals. This kind of data are usually unreliable and very difficult and/or expensive to obtain. There are two main sources: audited tax reports and surveys from which the incidence and the degree of tax evasion can be inferred either directly or indirectly. An alternative to the actual income data is to rely on laboratory experiments. Each of these has its own advantages and disadvantages, which will be explained in more detail in the next section.

As various data sources can complement each other, a combination of them has potential to provide more exhaustive information about non-compliance. In particular, combining survey income data with tax records at the *individual* level offers new possibilities to study tax evasion. Matching and linking such information is usually very restricted due to privacy concerns and indeed, to the author's knowledge, the only study so far using such data to estimate tax evasion is by Baldini et al. (2009). They compare the two income measures and assuming that people report their true income in the survey, obtain a measure of non-reporting. However, the survey data can often contain notable measurement errors (which they acknowledge but do not deal with in their analysis). In fact, there have been several studies in the measurement error literature[2], which assume that administrative data is error-

[1] As Shaw et al. (2010) emphasise, enforcement is a true resource cost to a society and it does not produce any resource gains because any resulting increase in tax revenues is a transfer from private citizens.

[2] See Bound et al. (2001) and Chen et al. (2011) for surveys of this literature (in economics).

free and hence differences between survey income and income from the tax records are due to survey measurement error alone, analysing its determinants, e.g. Bound and Krueger (1991) and Bollinger (1998).[3] Later studies have started to relax this assumption by considering matching errors (Kapteyn and Ypma, 2007; Meijer et al., 2012) or errors in register data (Abowd and Stinson, 2013), but have not attempted to assess the scale and nature of error in administrative data, let alone tax evasion as a possible source.

This paper provides estimates of the pattern and determinants of tax evasion based on a unique dataset combining a household income survey and tax records for Estonia. The main research questions are: (1) Which individual characteristics contribute to evading taxes on wages and salaries? (2) What is the extent and distribution of undeclared income? Unlike earlier studies attributing income discrepancies between different data sources either to tax evasion or survey measurement error, here both reporting processes are modelled in a joint framework. Focusing on employment income, the key assumption is that measurement error is unrelated to the sector where the individual works while the same does not hold for tax compliance. Specifically, it is assumed that taxes cannot be evaded in the public sector. This assumption provides some parallels with the methodology pioneered by Pissarides and Weber (1989), where underreported income for one population group (like self-employed) is inferred from a comparison with a reference group (e.g. employed), which is assumed to have negligible non-compliance but to be similar in other respects. In addition to different data strategy and econometric model, the assumption used in this study is, arguably, less restrictive as it considers the possibility that (private sector) employees engage in tax evasion as well. Furthermore, Pissarides and Weber (1989) type of studies have assumed implicitly that the underreporting of income in a survey corresponds to the underreporting of income to the tax authority, which is not required here.

We use the Estonian Social Survey, which is the basis for the Estonian component of the European Union Statistics on Income and Living Conditions (EU-SILC) survey, linked to tax records for 2007. As the underlying data linkage has been carried out (legitimately) without the requirement for consent by the survey respondents, this allows us to retain all relevant sample and, more importantly, avoid potential selection biases, which can arise from the consent decision (see e.g. Sakshaug and Kreuter, 2012). This is the main problem for data linkages as they often require

[3]The linked administrative data (referred to as validation data in this literature) could also originate from other sources. For example, Duncan and Hill (1985) and Bound et al. (1994) are based on linked survey and employer reports. Apart from limited representability due to a small single firm sample characterising these two examples, there is also an important conceptual difference for studying tax compliance – the information what employers have reported in the validation study is not necessarily identical with that reported to the tax authority.

respondents agreement beforehand. For example, previous evidence suggests that consenting can be correlated with income (Jenkins et al., 2006) and as it is conceivable that the consent decision for linking tax records could be influenced by the tax compliance behaviour as well, it is crucial to avoid such sample restrictions.

The paper extends the empirical tax evasion literature in several ways. First, it proposes a novel econometric model to analyse tax evasion, taking into account potential survey measurement error. As far as the author is aware of, this is the first such attempt. Second, it provides new evidence on non-compliance in a post-socialist country, which extends the rather limited empirical literature on countries other than the US. Third, it studies specifically tax non-compliance related to wages and salaries which has received less attention in the literature, for example, compared to self-employment income.

The estimates show that compliance is associated with a number of socio-demographic and labour market characteristics. Overall, people in the bottom and the top part of earnings distribution are found much less compliant. The results indicate substantial non-reporting of wages and salaries, mainly in the form of partial rather than full evasion. This highlights that third party reporting and tax withholding, which this income source is subject to, have limitations and suggests that tax audits might be less effective in revealing true wages and salaries than previously thought.

The paper is structured as follows. The next section gives an overview of the relevant tax compliance literature, focusing on previous theoretical and empirical findings on the individual characteristics associated with tax evasion. Section 3 provides information on the main elements of the Estonian income taxes and their administration. Section 4 presents the econometric model used to estimate jointly tax compliance and survey measurement error. Section 5 gives an overview of the data sources, their linkage and summarises earnings information. Section 6 presents and discusses findings, in terms of who is more likely to evade income taxes as well as the extent and pattern of non-compliance, and tests the robustness of results through sensitivity checks. The last section concludes with some policy implications and suggestions for further extensions.

2 Related literature

We first review previous work which has provided insights into the factors influencing income tax compliance, both in the form of theoretical predictions and empirical evidence. The focus here is on individuals rather than firms or the tax authority. For more comprehensive recent reviews, see Andreoni et al. (1998), Alm (1999), Slemrod and Yitzhaki (2002), Sandmo (2005), Shaw et al. (2010), Alm (2012), Hashimzade

et al. (2013), Pickhardt and Prinz (2014).

2.1 Theoretical work on tax evasion

The economic theory of tax evasion has evolved over the past 40 years starting with the seminal paper by Allingham and Sandmo (1972), who provided a relatively simple framework for analysis, but demonstrated the complexity of the subject as they could provide clear predictions only in certain dimensions. Theoretical models have advanced significantly since then, however, as Alm (2012) points out, more complex approaches tend to yield more ambiguous results. For this reason, we start from the standard model.

In the Allingham-Sandmo (A-S) paper, a risk-averse individual maximises expected utility by choosing how much income to report to the tax authority. While non-compliance reduces tax liability (levied at the proportional rate), the individual would have to pay a fine (proportional to the non-reported income) if this was detected. This so-called deterrence model predicts that evasion is decreasing in the probability of detection and the penalty rate but gives ambiguous results in other aspects. The effect of an increase in total income on the fraction of income reported depends on relative risk aversion: the effect is positive (constant or negative) if relative risk aversion is increasing (constant or decreasing). Assuming decreasing absolute risk aversion, which has been generally accepted since then, it can be further shown that the *level* of underreported income increases with total income and that more risk-averse individuals would evade less (Cowell, 1990). An increase in the tax rate has an ambiguous effect on evasion.[4] In a similar model, Srinivasan (1973) analysed generic tax and penalty schedules with a risk-neutral individual and showed that evasion decreases as the probability of detection increases. The effect of an increase in total income on the proportion of income reported depends now on the nature of the tax schedule and the probability of detection: it decreases with a progressive tax if the probability of detection is independent of income, while it increases with a proportional tax if the probability of detection is an increasing function of (total) income.

While the A-S model has been criticised for various reasons, it has remained central in economic analysis with much of the theoretical work maintaining a focus on the rational agent making his decision on the basis of a cost-benefit analysis. The main weakness of the original model is that it seems to predict much lower

[4]Yitzhaki (1974) pointed out that if instead the penalty is proportional to the evaded *tax* then, surprisingly, a tax increase has a positive effect on compliance (if the individual has decreasing absolute risk aversion).

compliance than the empirical evidence suggests[5] and various additional factors have been proposed to explain this, for example, the differences between actual and perceived probabilities of auditing, third party reporting and non-pecuniary costs. The standard economic analysis of tax compliance has been also criticised in other disciplines for overlooking legal issues, e.g. Graetz and Wilde (1985), and for taking taxpayer's motivation as given, e.g. Weigel et al. (1987). Indeed, its focus is mainly on enforcement activities – as Alm (1999) stressed, a person would *only* pay taxes because of the fear of detection and punishment with this approach.

Further theoretical work starting with Andersen (1977) and Pencavel (1979) extended the A-S framework with endogenous income where the individual decides jointly with compliance his labour supply. The relationship between the key parameters and evasion, however, becomes even less straightforward in this case. Nevertheless, one relevant insight for our purposes is from Cowell (1985) who points out that one form of cheating involves taking additional jobs. One strand of the subsequent literature focused on the interactions with the tax authority[6], which in general is outside the scope of interest here as they offer little insights on individual characteristics relevant for compliance. Among a few exceptions is a study by Erard and Feinstein (1994) who confirm with a game-theoretic model that evasion (in general) increases with total income. There is also a useful hint on firm characteristics: Kleven et al. (2009) show that in the presence of third-party reporting, it is optimal for large firms to comply fully.

A relatively recent part of the literature considers more realistic behavioural elements like various forms of non-expected utility and social interactions, though the focus often remains on enforcement parameters. See Hashimzade et al. (2013) for a detailed discussion. This branch has considered additional factors such as different subjective costs (feeling guilty or ashamed, stigma, damage to reputation) which can explain why there seem to be fewer non-compliant people than the standard model predicts. While the extent of evasion depends on the utility function in the A-S model, the condition for compliance is determined solely by the audit risk, tax and penalty rate. Adding nonpecuniary costs to the utility function makes this condition more restrictive, as pointed out by Allingham and Sandmo (1972) themselves[7] and later by Gordon (1989) and Sandmo (2005). The decision to comply is then affected by the extent of disutility from cheating, which naturally varies between individuals. However, these unobservable parameters are difficult to test empirically.

[5]See Alm (1999) and Slemrod and Yitzhaki (2002) for numeric illustrations.

[6]Two main approaches rely on principal-agent and game-theoretic models. See Andreoni et al. (1998) for a detailed discussion.

[7]This together with other extensions in their paper – endogenous probability of detection and a dynamic case – seem surprisingly often overlooked in the later literature.

In this paper, we focus on the association between tax evasion and total income. As explained above, the theory tends to suggest that evasion (in absolute amount) increases with income, while it is inconclusive about the proportion of income evaded. This has great political importance as, for example, if people with higher income were more likely to evade taxes on larger proportions of their income, this would raise important questions about the fairness of tax system. Given the nature of the dataset used (more in Section 5), we will not be able to test the effect of risk preferences and enforcement parameters on compliance as these are not observed directly. The probability of auditing/detection is likely to vary, for example, with industry (and occupation) and this we can control for but we have no detailed information about the actual auditing strategy. Furthermore, what is likely to be more important is the *perceived* probability of getting caught.

Due to the flat income tax in Estonia (more in Section 3), there is also very little variation in the marginal effective tax rates in the cross-sectional data which does not allow studying their effect on compliance. On the other hand, this can be also a useful feature as it allows us to set aside a component which is generally difficult to identify due to endogeneity.

The broad set of socio-demographic information available in our dataset allows us to identify which personal characteristics are associated with tax compliance. While economic theory remains rather vague in this context, one useful framework has been suggested in the psychology literature by Weigel et al. (1987) where tax evasion behaviour is influenced by social and psychological (or personal) factors. In both cases, they further distinguish between two groups of factors: those instigating tax evasion behaviour and those that constrain it. Social norms are given as an example of social instigations, while financial difficulties and perceived unfairness of tax laws and authorities are part of personal instigations; access to cash receipts for a given occupation and tax enforcement, among else, operate as social constraints, while the perceived risk of punishment and attitudes towards evasion represent personal constraints. This provides some useful guidelines for selecting specific variables in the econometric model.

2.2 Empirical work on tax evasion

We now turn to the empirical literature on tax evasion, retaining the focus on individual. We limit our attention further to the studies utilising individual-level income data, grouping these by the type of data source used: audits, surveys and experiments.

2.2.1 Audited tax records

Audited tax returns are considered to offer the most reliable information on tax compliance (Andreoni et al., 1998; Feldman and Slemrod, 2007). On the grounds of cost-efficiency, audits are typically non-random as the cases are already chosen based on some predictions of which individuals are more susceptible of evasion, making it difficult if not impossible to generalise findings to the wider population. To overcome this problem, there have been also randomised audits carried out in some countries. These have been most extensive and regular in the US in the form of the Taxpayer Compliance Measurement Program (TCMP) in 1965-88 and the National Research Program (NRP) since 2001.[8] The individual-level data from these audits have been used in several papers, typically regressing the difference between reported income and actual income as established on the basis of an audit against variables such as the marginal effective tax rate, total true income, presence and proportion of particular income sources and the limited socio-demographic information that is available from tax reports (e.g. age group, marital status, region). The first study was by Clotfelter (1983) whose primary interest was the effect of marginal tax rates on evasion. This has been followed with extensions including partial detection (Feinstein, 1991), the role of tax practitioners (Erard, 1993, 1997), non-filers (Erard and Ho, 2001) and multi-mode evasion (Martinez-Vazquez and Rider, 2005).

Despite similar sets of regressors, the findings have been surprisingly varied. For the marginal tax rate, Clotfelter (1983) and Martinez-Vazquez and Rider (2005) find a positive effect on non-compliance, while Erard (1997) finds a negative effect (for reports where tax practitioners were used). Feinstein (1991) provides mixed results with a positive effect for each of two years analysed separately (i.e. as in other studies) but a negative effect for the pooled model. In Erard (1993), the effect of the marginal tax rate is also significant and goes in either direction depending on a particular tax preparation mode.

Findings on the relationship between (true) income and evasion are also mixed. Clotfelter (1983) found that underreporting increases with income[9], which Feinstein (1991) confirmed with single-year audits, while showing an insignificant (and opposite) effect with the pooled model. Martinez-Vazquez and Rider (2005) found a negative effect with the whole sample yet a positive link emerged when the sample was split into three audit classes. Evidence in Erard (1993, 1997) suggest an inverted U-shape for some paid-prepared returns (and non-significant or a negative effect for others). It is even less clear how evasion, measured as the proportion of underreported income, varies across the income distribution. This has been shown in Johns

[8] For an overview of US studies, see Slemrod (2007).

[9] He used after-tax income, while later papers have relied on (adjusted) gross income.

and Slemrod (2010) who analysed the distributional impact of non-compliance. They find that the proportion of *total (true) income* not reported is larger for higher income groups (although peaking between the 90th and 95th percentile), while the proportion of underreporting in wages and salaries declines over the same income groups, and amounts to only about 1% overall.

In terms of other personal characteristics there seems to be evidence that evasion is higher among married people and lower for the elderly (Clotfelter, 1983; Feinstein, 1991; Martinez-Vazquez and Rider, 2005). The latter also find that the number of dependents is positively related to non-compliance.

There are several shortcomings commonly acknowledged in the literature about audited tax information: even thorough audits are unlikely to detect all income and models accounting for imperfect detection have been suggested in Alexander and Feinstein (1987) and Feinstein (1991); TCMP/NPR data typically exclude non-filers who have been studied by Erard and Ho (2001) and non-compliance can also include unintentional reporting errors which have been considered by Alexander and Feinstein (1987) and Erard (1997). One critical aspect from our point of view is the lack of socio-demographic variables, though some studies have addressed this by matching audit data with information from other source, see e.g. Witte and Woodbury (1985) and Dubin and Wilde (1988), though using aggregated rather than individual-level data. Furthermore, analyses based on audited returns typically consider all taxable income together which come from very different sources characterised by different opportunities for evasion and potentially different factors influencing compliance decisions. Evasion can also take place in the form of underreporting income or overreporting deductions which have been distinguished only in a few of studies (Feinstein, 1991; Martinez-Vazquez and Rider, 2005). Overall, US audits have suggested very low evasion of incomes from wages/salaries, although this might be underestimated as any undeclared payments could be concealed both by the individual and the employer and, hence, very difficult to detect.

A study by Kleven et al. (2011) for Denmark is a rare one based on random audits outside the US. They find that tax evasion has a statistically significant positive association with being male, a homeowner, working in a small firm and working in sectors like agriculture, construction and real estate. The strongest predictors are, however, variables reflecting the presence and size of self-reported income, and once these are controlled for only gender (and marital status, after changing sign) are statistically significant.

2.2.2 Surveys

Surveys can provide wide-ranging information. On the one hand, they can ask respondents directly whether they have engaged in tax evasion activities of various forms, see e.g. Kinsey (1992), Sheffrin and Triest (1992), Forest and Sheffrin (2002). There are also two studies for Estonia which rely on such data to estimate individual determinants for tax evasion/undeclared work (or its proxies). Kriz et al. (2008) use a survey by the Estonian Institute of Economic Research (*Eesti Konjuktuuriinstituut*, EKI) on the self-reported receipt of undeclared earnings (i.e. the so-called envelope wages) together with two other data sources: non-random tax audits and the Estonian Labour Force Survey (LFS). While the first two sources provided explicit information on whether a person had evaded taxes, the LFS could only provide a proxy in the form of self-reported work under a verbal contract. Using logit models they find higher propensities for being a tax evader for those working in small firms, in construction and agricultural sector; for part-time employees, non-Estonians, men, young and elderly; for those with less education as well as regional differences. Meriküll and Staehr (2010) reach similar conclusions with their estimations for all three Baltic States on the basis of the Working Life Barometer survey for 1998 and 2002. Using a logit model where the dependent variable indicates self-reported receipt of envelope wages, they confirm earlier findings for Estonia by showing a higher likelihood of tax evasion for people with more than one job, a lower skilled job, working in a smaller firm or expanding firm; and in the construction, trade and agricultural sector. Both studies have limitations due to a small number of cases of tax evasion and/or limited sets of explanatory variables.

The main problem with self-reported data is that it is unclear how truthful respondents are, given the sensitivity of the subject (Weigel et al., 1987; Elffers et al., 1991), even more so when asked about the magnitude of evasion. Such measurement problems with survey data prompt Slemrod and Weber (2012) to even conclude that the empirical research in tax compliance is (largely) yet to experience a 'credibility revolution', and call for more creativity and attention to appropriate econometric techniques. Studies employing methods determining the extent of tax non-compliance indirectly from survey data are, however, a step in that direction. For example, several studies have followed the Pissarides and Weber (1989) approach deriving such estimates from the comparison of income and (food) expenditure by contrasting self-employed with employees as the prevalence of tax evasion is usually lower for the latter, see e.g. Schuetze (2002), Lyssiotou et al. (2004), Engström and Holmlund (2009), Kukk and Staehr (2014) and Hurst et al. (2014). Feldman and Slemrod (2007) take a similar approach but compare claimed tax deductions for different population sub-groups (using unaudited tax returns). However, these

studies offer little insights to the determinants of tax evasion.

Combining survey data with administrative sources may offer the most promising route, though there are very few previous studies on tax compliance using survey data linked with tax reports and even less with income information from both sources at the individual level. Mork (1975) provides an early example for Norwegian men where respondents were asked about their income (in intervals). He compared income interval midpoints in the survey with the average declared income for the same persons and found that the proportion of register income was lower at higher income levels. Elffers et al. (1987) analysed a sample of Dutch taxpayers whose tax returns had been carefully audited (without their knowledge) and then asked to participate in a survey, relying on a complex procedure to link two data sources while preserving people's anonymity. Participants were asked in the survey whether they had underreported income or overreported deductions, but not about the magnitude. Their most important finding is essentially zero correlation between assessed and admitted non-compliance.

Baldini et al. (2009) is apparently the only other study on tax compliance using individual income from linked survey and administrative data.[10] They do acknowledge the presence of measurement errors (potentially in either source) but do not attempt to account for these and attribute all differences between two income measures to tax evasion, assuming that survey income represents true income. Their findings suggest that evasion is higher (both in absolute and relative terms) for higher income groups, people with more education and self-employed. However, the analysis includes only a few explanatory variables and the data have clear limitations in terms of a relatively small sample (about 1,000 observations), representativeness (as it refers to the residents of Modena in Italy) and accuracy (a period mismatch between two sources). Most importantly, their finding of (average) register income exceeding (average) survey income at lower survey income levels points to substantial measurement errors in the survey. Hence, an analysis based on raw differences between two income measures can give a rather misleading picture of evasion.

As discussed in the introduction, linked survey and administrative data are more common in the survey measurement error literature where, in turn, potential misreporting of earnings in tax records due to non-compliance is ignored.

2.2.3 Experiments

Another method is generating data through laboratory experiments, see Alm (1991) and Alm and Jacobson (2007) for relevant reviews. Experiments have confirmed

[10] See Fiorio and D'Amuri (2005) and Benedek and Lelkes (2011) for examples of studies comparing survey income with administrative records at aggregate levels without involving matching.

the role of auditing and penalties (though evidence on the effect of marginal tax rates remains mixed), provided useful guidance on various auditing strategies as well as highlighted additional factors influencing compliance decisions. Similar to audited tax returns, only a small number of socio-demographic variables have been examined: older people have been found more compliant (Friedland et al., 1978; Baldry, 1987; Pudney et al., 2000) and males less compliant (Spicer and Becker, 1980; Baldry, 1987; Pudney et al., 2000). There is also evidence that the propensity to evade (Becker et al., 1987; Pudney et al., 2000) increases with true income, but the results for the extent of underreporting are less clear with Baldry (1987) showing a positive effect and Pudney et al. (2000) a negative effect (conditional on evasion).

While experiments can provide unique insights into the behaviour underpinning tax evasion and avoid usual problems with measurement error, the main challenge is its ability to represent individuals' behaviour in the real world and at the population level. Several studies have found notable framing effects (Baldry, 1986; Webley and Halstead, 1986; Schepanski and Kelsey, 1990), meaning that results can be sensitive to how the nature or purpose of the experiment is perceived by the participants. Furthermore, experiments are naturally limited as not all determinants can be (easily) tested in a laboratory setting. For example, all job-related characteristics (e.g. occupation, industry, firm size) are difficult if not possible to relate to the income generated in a lab session. The income distribution arising from a lab experiment is also hardly representative of the actual income distribution and the same applies to the estimates of the level of non-compliance.

3 The institutional setting

Estonia is one of the three Baltic States in the northeastern part of Europe and one of the smallest EU member states with a population of 1.3 million. The Estonian tax system is fairly simple and linear; it was the first country in Europe to (re)introduce flat income tax in 1994. The five largest tax instruments – personal and corporate income taxes, social security contributions, VAT and excises – are all levied at the national level and account for about 96-98% of total tax receipts in 1995-2009 (European Commission, 2011). Property taxes are marginal and there are no wealth taxes. The structure of taxes has remained largely unchanged since 2000.

Personal income tax is applied on comprehensive income, pooling all sources of income including realised capital gains. The main deductions from taxable income are personal allowance, child allowance, pension allowance, mortgage interest payments and education related expenses.[11] This leaves rather limited possibilities

[11] As of 2007, the personal and the child allowance (per child starting from the second) were both

for overreporting tax deductions and, hence, non-compliance can mainly take place in the form of underreporting income to the tax authority. A single marginal tax rate (22% in 2007) is applied on the final tax base.[12] Nearly all social insurance contributions (SIC) are paid by employers and consist of the social tax (33% of gross earnings since its introduction in 1994), which funds pension and health care systems, and unemployment insurance contribution (0.3% in 2007). Employees pay only contributions to the funded pension scheme (2% in 2007), which is voluntary for older generations, and unemployment insurance contributions (at twice the rate of employers). This means that the effective marginal tax rate varies very little between employees and cannot be an important determinant of non-compliance at the individual level.

The fiscal year is the calendar year and tax reports must be submitted by the end of March next year. Individual declarations are pre-populated with the information received from employers as well as social insurance funds who administer taxable benefits (public pensions, unemployment insurance benefit, parental benefit, sickness pay etc). Married couples can choose to file a joint report, in which case all the income and allowances are considered together. While this would be beneficial only for couples where one spouse has unused allowances, for other couples the joint liability would be the same as the sum of individual liabilities (but never higher). For employment income and taxable benefits, income tax and SIC are withheld at source. As only the personal allowance and the pension allowance can be applied on a monthly basis, individuals entitled to other allowances and deductions need to file a report to benefit from them. The same applies to those who have been employed only part of the year. Otherwise, as of 2007, residents whose taxable income does not exceed the personal allowance[13] or who have no additional tax liability, i.e. final tax liability corresponds to the withholding tax, do not have to file a tax report. A relatively simple personal income tax system places low compliance burden on individuals and little professional assistance is required and used. As the tax authority also offers free phone and email support service, the overall compliance costs for individuals ought to be low.

Due to employers' obligation to report salaries and wages (on a monthly basis), evading taxes on employment income cannot take place without their knowledge

1,534 EUR per year (24,000 EEK); the pension allowance was 2,301 EUR per year (36,000 EEK) and the upper limit on deductible expenses was 3,196 EUR (50,000 EEK). All applied on individual basis, except the child allowance which can be claimed by one of the parents. For comparison, average gross annual salary was 8,694 EUR in 2007.

[12]Companies only pay corporate income tax on distributed earnings, while retained earnings are not taxed. Dividends are only taxed once and not considered as taxable income for individuals.

[13]Also the pension allowance and the allowance applicable to the compensation for accident at work or occupational disease, if applicable.

and consent. Furthermore, given how the (statutory) tax burden is shared between employees and employers, this provides significant incentives for both sides to evade taxes. The employer would gain from cost reductions, providing an advantage over law-abiding competitors, though it is important to note that such incentives are unlikely to hold for the *public sector* in Estonia. This is supported by the evidence from the Working Life Barometer survey in the Baltic countries, according to which only 2% of the public sector employees in 2002 admit having received (sometimes) undeclared payments in cash, see Antila and Ylöstalo (2003)[p. 128]. (The estimate covers wages and salaries from second jobs as well and, hence, does not appear to refer strictly to income from the public sector employment.) Along with potential gains from non-compliance, the employer must consider risks – there is always the possibility that any current or previous employee might tip off the tax authorities, which in Estonia is likely to result in the employer being fined and not the employee. In this respect the risk of being exposed is significantly lower for self-employed and, arguably, for smaller companies.

The employee in turn might benefit from higher net earnings or having employment at all. There are also significant disadvantages built into the system for those undertaking fully undeclared work as they would not have health insurance coverage, their (expected) future pension would be lower, especially when it comes to the funded scheme (the so-called second pillar), and they would have difficulties getting a mortgage or a loan.[14] Hence, a common practice for tax evasion is believed to entail declaring part of the earnings, e.g. at the level of the legal minimum wage or slightly higher to raise less suspicion. A similar practice is mentioned in Besim and Jenkins (2005) for North Cyprus. They also suggest that by employing people through contracts with smaller firms, larger firms can benefit from tax evasion without increasing the risk of exposure for themselves. They also point out that as firms need to make unrecorded cash sales to pay their employees undeclared income, the evasion of payroll taxes also results in part of value added taxes and, possibly, corporate income tax being evaded.

Overall, it is not obvious whether it is the employee or the employer who has the decisive role in evading income and payroll taxes. Unless one side has a much stronger bargaining position, for example, if the employee has few or no job alternatives and the employer is well aware of that, it is effectively a joint decision.

[14]Given the real estate boom in mid-2000s and a very large increase in mortgage loans to households, this must have become a rather important incentive.

4 Model

The general model structure is the following. Let y_i^T denote the true value of earnings of individual i.[15] Employed persons have positive earnings ($y_i^T > 0$) and non-employed people have zero earnings ($y_i^T = 0$). Generally, true earnings are not directly observable and instead each person states her earnings in the survey, y_i^s, which can differ from the actual earnings due to intentional or unintentional misreporting (e.g. recall errors). Hence, there could be individuals with zero true earnings among those reporting positive survey income and such misreporting may have occurred, for example, because of confusing reference time periods or not wanting to reveal the non-employment status. People also choose how much of their actual earnings to declare to the tax authority, which we refer to as register income and denote with y_i^r. We can rule out negative taxable earnings and assume that people do not declare more income to the tax authority than they actually received.[16] Employed individuals have then three choices: full compliance ($y_i^r = y_i^T$), partial evasion ($0 < y_i^r < y_i^T$) or full evasion ($y_i^r = 0$), while non-employed persons always declare zero earnings ($y_i^r = 0$).

Our main interest is an estimate of income not reported to the tax authority, which is the difference between true earnings and declared earnings, $e_i = y_i^T - y_i^r$, and non-negative by assumption. This in turn requires a measure of true earnings and we seek to obtain this from observed survey and register income, assuming both relate to true earnings (and other personal characteristics), in a latent class framework. More specifically, our modelling strategy involves specifying a structural model for true earnings, survey earnings and declared earnings, and estimating it with a parametric method. As the econometric model consists of three separate equations estimated simultaneously while only two dependent variables are observed (y_i^r, y_i^s), we need further restrictions to identify all model parameters. Given the discussion about incentives to evade in Section 3, our key *identifying assumption* is that people working in the public sector are constrained in their choice and cannot evade taxes, i.e. $y_i^r = y_i^T$, while there are no systematic differences between the public and private sector employees with respect to (true) earnings formation and measurement error in the survey data. (In the empirical analysis, we are actually able to relax the latter assumption by allowing some key parameters to differ between the two sectors.)

[15] We focus throughout on wages and salaries and use terms *earnings* and *income* interchangeably.

[16] It is possible to report negative self-employment income in Estonia (similar to many other countries) as related expenses can be deducted from gross self-employment income, but the same does not apply to wages and salaries. Over-reporting of earnings could happen in practice, although one might expect this to be not very common. For example, Clotfelter (1983) shows evidence for the US that the proportion of people understating their taxable income greatly exceeds the proportion of people overstating their income.

This means that for a part of the sample, i.e. public sector employees, we observe true earnings as well and can therefore identify parameters for all three earnings equations.

Focussing on the sample of people with reported (full-time) employment and hence positive earnings in the survey ($y_i^s > 0$), we proceed by specifying the exact structure for each earnings function.[17] With probability p, an individual i in our sample is truly employed and has log-normally distributed true earnings:

$$\ln y_i^T = x_i \beta^T + \varepsilon_i^T \tag{1}$$

where x_i are her characteristics determining the log income and $\varepsilon_i^T \sim N(0, \sigma_T^2)$ is a random term. With probability $1 - p$, the employment status is misreported in the survey and the person has actually no earnings ($y_i^T = 0$) – assuming this could happen equally among those claiming to be working in the public sector and those in the private sector. We constrain the probability to be fixed, though this could be relaxed by allowing the probability to vary according to personal characteristics. We have chosen not to complicate the model structure with this as it seems to concern relatively few cases. The probability density of true earnings, conditional on having positive earnings, is:

$$f(y_i^T | x_i, y_i^T > 0) = \frac{1}{\sigma_T y_i^T} \phi \left(\frac{\ln y_i^T - x_i \beta^T}{\sigma_T} \right) \tag{2}$$

where $1/y_i^T$ is the Jacobian term and $\phi(.)$ is the probability density function of the standard normal distribution.

To reflect multiple choices of compliance, we model declared earnings (y_i^r) as a *fraction* of true income, using a two-limit Tobit model and a latent variable r_i^* ('the propensity to comply'):

$$y_i^r = \begin{cases} 0 & \text{if } y_i^T = 0 & \text{(no earnings)} \\ 0 & \text{if } y_i^T > 0 \text{ and } r_i^* \leq 0 & \text{(full evasion)} \\ r_i^* \cdot y_i^T & \text{if } y_i^T > 0 \text{ and } 0 < r_i^* < 1 & \text{(partial evasion)} \\ y_i^T & \text{if } y_i^T > 0 \text{ and } r_i^* \geq 1 & \text{(no evasion)} \end{cases} \tag{3}$$

where

$$r_i^* = \theta^r y_i^T + x_i \beta^r + \varepsilon_i^r \quad \text{if } y_i^T > 0 \tag{4}$$

and $\varepsilon_i^r \sim N(0, \sigma_r^2)$. Assuming ε_i^T and ε_i^r to be independent, the probability density

[17] Note that we maintain a wider scope compared with several previous studies on measurement error using linked data as their focus is typically limited to cases where positive earnings are reported in *both* sources, for example, Bound and Krueger (1991) and Kapteyn and Ypma (2007).

of declared earnings, conditional on true earnings, is the following:

$$f(y_i^r|x_i,y_i^T)=\begin{cases}\Pr(y_i^r=0|y_i^T=0) &=& 1 & \\ \Pr(y_i^r=0|x_i,y_i^T) &=& \Phi\left(-\frac{\theta^r y_i^T+x_i\beta^r}{\sigma_r}\right) & \forall y_i^T>0 \\ f(y_i^r|x_i,y_i^T) &=& \frac{1}{\sigma_r y_i^T}\phi\left(\frac{y_i^r/y_i^T-\theta^r y_i^T-x_i\beta^r}{\sigma_r}\right) & \forall y_i^T>0 \\ \Pr(y_i^r=y_i^T|x_i,y_i^T) &=& 1-\Phi\left(\frac{1-\theta^r y_i^T-x_i\beta^r}{\sigma_r}\right) & \forall y_i^T>0\end{cases} \tag{5}$$

We refer to this as the *multiplicative* model and additionally consider declared earnings in an *additive* form, where θ^r and β^r-s are interpreted in levels rather than the ratio of declared earnings.[18] The probability density function of declared earnings is very similar in the two cases. As a characteristic of the Tobit model, both specifications combine the extensive and intensive margin of decision making – whether to underreport incomes to the tax authority at all and, if so, to what extent. Modelling each choice margin explicitly would provide more flexibility but also further complicate the model structure and its identification. We have therefore opted for testing these two alternative Tobit specifications instead.

The multiplicative model combines the overall compliance decision (i.e. extensive margin) with underreporting in relative terms and part of its structure is akin to the model of fractional detection of income tax evasion in Feinstein (1991). The additive model combines the compliance decision with underreporting in absolute terms. While both types of model allow studying how compliance in *relative* terms varies across the income distribution (i.e. one of our main research questions), a slight advantage of the multiplicative model is that its parameter θ^r provides (some) direct insights into that. More specifically, θ^r provides a clear indication of the effect of true earnings on the latent variable. (The link with the censored variable is non-linear and depends on the values of other covariates as well.) With the additive

[18] Specifically:

$$y_i^r=\begin{cases}0 & \text{if } y_i^T=0 & \text{(no earnings)}\\ 0 & \text{if } y_i^T>0 \text{ and } y_i^{*r}\le 0 & \text{(full evasion)}\\ y_i^{*r} & \text{if } y_i^T>0 \text{ and } 0<y_i^{*r}<y_i^T & \text{(partial evasion)}\\ y_i^T & \text{if } y_i^T>0 \text{ and } y_i^{*r}\ge y_i^T & \text{(compliance)}\end{cases}$$

where

$$y_i^{*r}=\theta^r y_i^T+x_i\beta^r+\varepsilon_i^r \quad \text{if } y_i^T>0$$

and $\varepsilon_i^r\sim N(0,\sigma_r^2)$. The probability density of declared earnings, conditional on true earnings:

$$f(y_i^r|x_i,y_i^T)=\begin{cases}\Pr(y_i^r=0|y_i^T=0) &=& 1 & \\ \Pr(y_i^r=0|x_i,y_i^T) &=& \Phi\left(-\frac{\theta^r y_i^T+x_i\beta^r}{\sigma_r}\right) & \forall y_i^T>0 \\ f(y_i^r|x_i,y_i^T) &=& \frac{1}{\sigma_r}\phi\left(\frac{y_i^r-\theta^r y_i^T-x_i\beta^r}{\sigma_r}\right) & \forall y_i^T>0 \\ \Pr(y_i^r=y_i^T|x_i,y_i^T) &=& 1-\Phi\left(\frac{(1-\theta^r)y_i^T-x_i\beta^r}{\sigma_r}\right) & \forall y_i^T>0\end{cases}$$

model, θ^r reflects both the level of true resources and their effect on compliance, though it may capture more adequately the existence of a tax-free threshold.[19] The additive model could also reflect better the nature of compliance decisions if there are fixed costs involved and non-compliance is not worthwhile unless the amount of evaded taxes is substantial enough. On the other hand, the cost of compliance could be correlated with true earnings (for example, potential damage to reputation may increase with true earnings) for which the multiplicative model would be then more appropriate. Overall, it is difficult to establish a priori which specification is more relevant and people's actual behaviour could be more complex and involve elements of each. We therefore estimate both models to see which one fits the data better.

Finally, conditional on $y_i^s > 0$, log survey income y_i^s is modelled as a function of log true earnings and individual characteristics x_i, assuming ε_i^T and ε_i^s to be independent and including a separate dummy in the case true earnings are zero:

$$\ln y_i^s = \theta^s \ln y_i^T \cdot 1(y_i^T > 0) + \theta_0^s \cdot 1(y_i^T = 0) + x_i\beta^s + \varepsilon_i^s \tag{6}$$

where $1(\cdot)$ is an indicator function and $\varepsilon_i^s \sim N(0, \sigma_s^2)$.[20] The probability density of survey income, conditional on reporting employment in the survey, is

$$f(y_i^s|x_i, y_i^T, y_i^s > 0) = \frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s \ln y_i^T \cdot 1(y_i^T > 0) - \theta_0^s \cdot 1(y_i^T = 0) - x_i\beta^s}{\sigma_s}\right) \tag{7}$$

where $1/y_i^s$ is another Jacobian term. (Given our sample of interest, we omit the condition $y_i^s > 0$ below.)

The overall probability density function (PDF) for a pair of observed earnings measures (y_i^r, y_i^s) for individual i can be written conditional on true earnings, with the latter integrated out over its plausible range, i.e. any amount equal to or larger than declared earnings:

$$f(y_i^r, y_i^s|x_i) = f(y_i^T = y_i^r|x_i) f(y_i^r, y_i^s|x_i, y_i^T = y_i^r) + \int_{y_i^r}^{\infty} f(y^T|x_i) f(y_i^r, y_i^s|x_i, y^T)\, \mathrm{d}y^T \tag{8}$$

Assuming that, conditional on true earnings and other covariates, the statements of register and survey income are independent of each other, i.e. the error terms $(\varepsilon_i^r,$

[19] However, the threshold applies only to the personal income tax while employer social contributions are paid on all gross earnings (see Section 3). Furthermore, as we are focusing on full-time employees and the legal minimum wage exceeds substantially the tax-free threshold, we have decided not to model the threshold explicitly.

[20] We also experimented with survey earnings in levels but the model fit to the data was much poorer. The log form of earnings has been also commonly used in the measurement error literature, where the focus is typically on the sample of people with positive earnings in both data sources.

ε_i^s) are uncorrelated, this can be simplified further as

$$\begin{aligned} f(y_i^r, y_i^s|x_i) =& f(y_i^T = y_i^r|x_i)\Pr(y_i^r = y_i^T|x_i, y_i^T) f(y_i^s|x_i, y_i^T = y_i^r) \\ &+ \int_{y_i^r}^{\infty} f(y^T|x_i) f(y_i^r|x_i, y^T) f(y_i^s|x_i, y^T)\,\mathrm{d}y^T \end{aligned} \tag{9}$$

Among those with positive survey income, we can distinguish between two sets of observational outcomes, depending on whether register income is zero (A_{0s}) or positive (A_{rs}).[21] In the case of observations in set A_{0s}, the PDF combines the possibility of true earnings being zero and true earnings being positive and entirely undeclared. For observations in set A_{rs}, the PDF combines the possibility of all or part of earnings being declared, as positive register income implies that true earnings are also positive given our assumption of $y_i^r \leq y_i^T$:

$$f(y_i^r, y_i^s|x_i) = \begin{cases} f(\text{no earnings}) + f(\text{full evasion}) & \text{if } y_i^r = 0 \\ f(\text{compliance}) + f(\text{partial evasion}) & \text{if } y_i^r > 0 \end{cases} \tag{10}$$

The log likelihood function of the sample is

$$\ln L = \sum_{i \in A_{0s}} \ln f_{0s}(y_i^r, y_i^s|x_i) + \sum_{i \in A_{rs}} \ln f_{rs}(y_i^r, y_i^s|x_i) \tag{11}$$

We estimate the parameters p, β-s, θ-s and σ-s with the maximum likelihood method and use the Gauss-Hermite quadrature to evaluate the integrals numerically. Detailed components of the likelihood function for the multiplicative and the additive model are provided in Appendix A.

Model identification is based on the assumption that the public employees are constrained in their choice to be compliant, hence determining a priori some of those who are fully compliant (or actually non-employed). As true earnings are then directly observed for the public employees in the tax records, their sample drives the identification of parameters in the true earnings equation and also in the survey earnings equation. The sample of private sector employees, in turn, identifies parameters in the declared earnings equation. Survey earnings are instrumental in establishing to what extent observed income disparities between the constrained and

[21] There is also a small group of people who reported zero survey earnings and positive register income (see Section 5). These cases point to a specific type of survey measurement error and appear to be associated with very marginal employment, therefore, having less relevance for our purposes as we shall be focusing on full-time employees. Furthermore, as employment characteristics on which we draw in the analysis are only available in the survey data and cannot be established for this group, we have excluded such observations from the analysis. This is common in survey-based empirical literature on labour market behaviour in general, though typically the same choice is made implicitly there.

unconstrained employees are due to non-compliance rather than differences in their true earnings. Hence, a partial model omitting survey earnings and covering only true earnings (y_i^T) and register income (y_i^r), is likely to result in downward biased estimates of the scale of non-compliance. Intuitively, on the basis of register income alone, there would be weaker evidence to suggest that the actual level of earnings among unconstrained employees may be above what is recorded in the tax records and comparable to that for the public employees or, possibly, even higher. As long as part of private sector employees are fully compliant, some (indirect) evidence is still present. At extreme, if *all* private sector employees unreport the same amount of income or the same proportion of their true income, then it would not be possible to separate it from differences in true earnings compared with the public sector employees, using a single observed measure of income. Estimating a system of equations with two income measures, ensures that parameter estimates agree with both sets of observations.[22] We illustrate the importance of having two income measures by estimating also a partial model as part of the sensitivity analysis.

In principle, the model can be estimated with an identical set of covariates (x_i) for all three income equations (as shown later in the sensitivity analysis), but in order to improve the identification we have made some exclusion restrictions. For example, interview characteristics are only included in the survey earnings equation, while it excludes job characteristics present in other two equations. The full list is given in Section 6. In terms of identification, there are no substantial differences between the multiplicative and the additive model.

5 Data

5.1 Data sources and linkage

The analysis is based on the Estonian Social Survey 2008 (*Eesti Sotsiaaluuring*, ESU) linked to administrative tax records. ESU is a household income survey, carried out annually since 2004 by Statistics Estonia. It is based on a rotating panel where each household is surveyed for four waves and one fourth of the sample is replaced in every wave. (Only cross-sectional information is used in this paper.) Basic demographic information is collected for all household members, while detailed person interviews are conducted with those aged 16 or over. ESU is also used as the basis for the Estonian component in the European Union Statistics on Income and Living Conditions (EU-SILC) database.

[22] This is of course useful only if survey earnings are indeed strongly correlated with true earnings, otherwise they would give misleading information. The latter would have wider implications as it would then call into a question the reliability of income surveys in general.

Information from administrative tax records is based on individual tax declarations (FID), if available, or (employer) tax withholding reports (TSD), hence covering all residents.[23] Although individual and employer reports differ in their structure, this has little importance in our case, not least because individual reports are pre-populated with the information from employers. Both provide detailed income information, with the main (yet minor) difference that the TSD forms exclude income earned abroad as reporting is limited to resident firms. Where only information from TSD is available this means that neither the joint reporting for married couples was used nor additional tax allowances claimed (even if applicable). For each individual, income is provided separately by type and provider, e.g. employer or government institution administrating a given benefit. This is also the case for joint reporting affecting certain aggregates like total income, total income tax, total allowances and total deductions, which are then summed for the couple (and not needed in the analysis).

Individual records in the two data sources have been linked using a unique personal identification number (PIN). This is officially assigned to each person and included in the Population Registry which provides the sample frame. PIN is therefore known for all sampled individuals, while asked for other household members during the interview in return for excluding them from the sample frame while participating in the ESU panel, so that they would not have to take part in other surveys conducted by Statistics Estonia at the same time. Those who did not provide a PIN were matched with the Population Register using their address and individual characteristics (as the Population Register does not provide information about household composition).[24] This resulted eventually in only a very few people without a match and, hence, without an identified PIN. It is also possible that the matching involved some error with incorrect PINs being assigned, although it is likely to be negligible. All data linkage was carried out by Statistics Estonia without a requirement to inform sample members or obtain their consent on the basis of the legislation governing its activities.[25] This characteristic is very important as consenting could be systematically affected by factors which are of key interest in this context: for example, income in general, or tax compliance behaviour in particular. The final dataset used here is anonymised with people's names, addresses etc removed.

The initial sample for ESU 2008 included 14,942 individuals of whom only 71 could not be identified in the tax register (see Table 1). Omitting people younger

[23]This is different from studies on the US where non-filers are usually missing from administrative data (hence referred to as 'ghosts'). Erard and Ho (2001) is one of the very few exceptions.

[24]Seven out of 11 digits of the PIN are determined by person's gender and the date of birth.

[25]In comparison, 24-89% consent rate was achieved in studies summarised in Sakshaug and Kreuter (2012)[Table 1], where respondents' agreement for linkage was required.

than 16, who are not subject to a person interview, reduces the sample size to 12,699 persons. Of those, 1,910 did not respond to the survey (12.8% of the initial sample)[26] and another 87 people had no person interview carried out. A further 465 cases are omitted due to missing earnings information in ESU (mainly those who reported their earnings on an interval scale), which leaves 10,237 people with known survey earnings (including zero values).

[TABLE 1 HERE]

Essentially, we are interested in all individuals with (paid) employment in the income reference period but focus on those with more substantial employment experience to achieve greater sample homogeneity. For that purpose, we first exclude those who have *never* had a regular job, that is any full- or part-time work which lasted for at least 6 months. We then limit our sample to those employed, i.e. with positive survey earnings (5,500 people).[27] Besides *current* labour market characteristics (at the time of the interview), ESU also collects information about the *main* activity in any month of the income reference period, which is the previous calendar year before the survey interview, i.e. 2007. On this basis, we further select those who reported part- or full-time employment as the main activity at least for one month in the income reference period (5,327 cases).

In the final step, we limit our sample to 4,121 individuals who worked full time for the whole income reference period as a way to increase robustness with respect to potential measurement error in the number of months worked information. (This will be relaxed as part of sensitivity testing in Section 6.4, adjusting earnings with the number of months in receipt.) We also distinguish between people working in the constrained and in the unconstrained sector reflecting people's opportunities to engage in tax evasion. Following our key assumption, the *constrained sector* refers to people working in the public sector, but excluding those with a second job or who have changed jobs to take a more conservative approach. They account for about 29% of the final sample and are primarily located in set A_{rs}. Everyone else is assigned to the *unconstrained sector*, including those with missing employer status. As part of robustness checks, we will also test alternative definitions.

[26] For newly-added sample members, the number of non-respondents refers to sampled people only without other household members (as they remain unknown).

[27] There are also 343 cases where people (with regular job experience) have positive earnings in the tax records but zero earnings in ESU. These appear to represent very marginal employment, as reflected in the much lower average earnings compared to the main sample – see Table 2. Nearly 60% of these are old age or disability pensioners according to their labour market status.

5.2 Earnings information

The version of ESU used here includes all the income variables from the standard release as well as variables with *original values* before imputations by Statistics Estonia, i.e. incomes as they were reported (either net or gross, monthly or annual), including missing values. This allows us to avoid relying on the imputations in the standard release.

Among 5,500 individuals who reported positive earnings in ESU (Table 1), 95% stated earnings in monthly terms (rather than annual) and in 91% of cases net of (withheld) employee social contributions and income tax.[28] As derivation of gross values from net (or vice versa) is also affected by tax evasion, we keep the extent of such imputations for ESU data to a minimum by using the original net values in the subsequent analysis. Imputations are then only needed to obtain net values for cases where gross values were initially reported in the survey (about 10% of the sample). We carry out our own imputations drawing on the self-reported information about whether the employer (withheld and) paid social insurance contributions and income tax and whether a person participates in the funded pension scheme. Given the sensitivity of the question about withheld taxes, this is likely to overestimate compliance but provides nevertheless a better approximation compared to assuming full (or no) compliance. Among those who reported a gross income figure, nearly 97% said that income tax was fully paid and under 3% that taxes were not paid.[29] As part of sensitivity analysis in Section 6.4, the model is also estimated on a sample excluding all observations with imputed values.

The tax records indicate gross annual earnings together with withheld income tax and contributions, therefore, it is possible to construct an equivalent measure of net earnings. While there is only a single individual-level variable for wages and salaries in ESU (separate from self-employment income), earnings in the tax records are known in great detail, distinguishing payments by employer and type (as well as tax treatment).[30] On the other hand, unlike in ESU the number of months paid

[28] When asked about non-regular payments and bonuses, about 20% of people reported additional (net) remuneration, which they had omitted from their earnings reported initially.

[29] The proportion of those reporting that employee SIC (i.e. unemployment insurance contribution) had been fully paid was somewhat lower, about 90%. This is because less people are liable to pay this (e.g. it excludes those who have reached the legal retirement age or are receiving an early retirement pension) but also likely due to less awareness of that particular contribution (as it was introduced only in 2002). The same proportions are slightly lower for those who reported a net income figure, 92% for income tax and 87% for employee SIC, mainly due to higher prevalence of individuals who said they did not know or did not answer the question.

[30] The following type of payments have been included in the constructed earnings measure to match the content of the ESU earnings variable as closely as possible: salaries and wages, board member fees, compensation for termination of employment or service, remuneration or service fees paid on the basis of a contract for services (*töövõtuleping*). Payments to compensate loss of earnings due to health-related absence from work (by the Health Insurance Fund) or unemployment (by the

is not available in our dataset and we rely on corresponding information from ESU.

Table 2 shows mean log earnings in ESU and in the tax records (the annual net figure is divided by 12), distinguishing between non-respondents and respondents in ESU and in which of the two sources positive earnings were reported. There are several important features. First, a comparison of the mean value of log earnings in the tax records for ESU (unit) non-respondents (8.46) and respondents (8.33), see panel (a), shows that non-respondents' earnings are somewhat higher on average (the difference is non-zero with $p = 0.035$) and suggests that those with higher (register) income may be less likely to participate in the survey.[31] This is not necessarily a concern for our model estimates, as long as non-response patterns are the same for the public and private sector employees. Though we are unable to investigate non-response in much detail (due to the lack of information on non-respondents), the distribution of register earnings – not shown here – appears very similar for non-respondents and respondents.

[TABLE 2 HERE]

Second, there is a small group of people who reported zero earnings in ESU but had positive earnings in the tax records. The average value of their log register income (6.28) is much lower, see panel (b), which implies very marginal (formal) employment with a particular recall error. We therefore conclude that this group is rather specific and its omission (see previous sub-section) should not be problematic from the viewpoint of tax compliance. In contrast, mean log survey earnings are much more similar among those with no earnings in the tax records (8.54) and those with earnings in both sources (8.72).

Third, for those with positive earnings in both sources (A_{rs}), the difference in the mean log value of survey and register income is a modest 0.1. However, when distinguishing between those in the *constrained sector* and those in the *unconstrained sector*, a very clear pattern emerges. Mean log earnings in the tax records (8.84) exceed mean log earnings in ESU (8.77) in the constrained sector, which by our assumption means their true earnings are on average underreported in the survey. But it is the opposite in the unconstrained sector, where mean log earnings in the tax records (8.55) are lower than mean log survey earnings (8.70). Assuming that

Unemployment Insurance Fund) have been excluded.

[31] Toomse (2010) uses an earlier wave of the same data (ESU 2007), extended with additional information from the sample frame, to analyse non-response in depth. She finds that, conditional on making a contact, those living in the capital region and urban settlements, younger people and males were less likely to take part in the survey, while income (salary quintile) was not relevant for the probability to co-operate. However, income was significant for some particular modes of refusal and co-operation as high salary earners were more likely to firmly refuse at the first contact and more likely to be respondents requiring larger number of calls after the first contact.

survey earnings are similarly underreported by this group, this indicates substantial underreporting of earnings to the tax authority. Note also that the difference in mean log survey earnings between the two sectors is statistically significant only at the 10% level ($p = 0.068$). The same pattern holds for the final estimation sample, see panel (c). Similarly, Kapteyn and Ypma (2007)[p. 524] report that the mean difference between survey and administrative earnings (for A_{rs}) in their data is positive, while survey earnings are smaller than administrative earnings in most cases. More generally, measurement error studies (based on linked data) have commonly found very similar mean earnings in survey and administrative sources but significant differences at the individual level, in either direction (see Bound et al., 2001). There appear to be no distinction made between the private and the public sector in this literature though.

Figure 1 provides further details by showing the full distribution of each earnings variable for the final sample (excluding zero register incomes and some very high incomes for a better overview). While the overall shape of the distribution is similar for the two earnings measures, earnings reported in the survey have a number of pikes at round income levels (e.g. 5, 6, 7, 8, 10, 15 thousand EEK), which is a sign of a particular type of measurement error called heaping: a tendency to report rounded-off values. Earnings reported in the tax records show a much smoother distribution. It has been shown that heaping can cause notable problems in some applications, for example, for modelling the dynamics of (self-reported) household consumption (Pudney, 2008). Pischke (1995) noted the same feature in the US income survey (PSID 1983 and 1987) linked with employer reports. He imposed similar rounding pattern to register incomes and found only little correlation with the actual measurement error (defined as the difference between earnings in the survey and the employer records), suggesting that this is perhaps not a critical issue in our context. As our econometric approach is already quite complicated, we therefore chose not to model this feature explicitly.

[FIGURE 1 HERE]

As the final sample contains only people who (according to ESU) worked full time during the whole income reference period, in principle, there should not be anyone below the minimum wage level (denoted by the vertical lines in Figure 1). This does not hold strictly, especially for register income. It could mean either that survey information on work duration is not completely accurate and/or part of earnings have been unreported to the tax authority. As the distribution of log earnings (not shown here) is close to a normal distribution and there is no obvious spike at the minimum wage level as, for example, demonstrated for Hungary by

Elek et al. (2012), we do not model possible censoring of true earnings at the level of minimum wage. This also means that (despite of anecdotal evidence) there is little trace of a particular form of non-compliance, where only a part of earnings *equal* to the minimum wage is reported to the tax authority and taxes evaded on the rest of income. We therefore choose not to model this particular case of non-compliance explicitly, preferring instead a more generic model as set out in Section 4.

Figure 2 gives an overview of the correspondence between two earnings measure at the individual level, separately for the constrained and the unconstrained sector. (Again for the final sample excluding those with zero register income and some very high incomes.) The two groups of individuals reveal a similar pattern with most of the observations appearing around the 45-degree line, though survey earnings tend to exceed earnings in the tax records in cases where the latter have low values, and the opposite when the latter have high values. This is also reflected by the slope of a linearly fitted line which is about 0.65 for both sectors. The same pattern has been also found in the studies on survey measurement error (e.g. Bound and Krueger, 1991; Bound et al., 1994; Bollinger, 1998), where this has been interpreted as a negative correlation between the measurement error in the survey data and the true value of earnings – recall that these studies have commonly assumed earnings in the administrative data to reflect true values – though Kapteyn and Ypma (2007) show that this pattern can also occur without 'true' mean reversion. Additionally, there is visibly more variation in the unconstrained sector compared to the constrained sector and a greater mass of observations in the upper left region as one would expect in the presence of tax evasion (if earnings in the survey are disclosed more truthfully). This is also illustrated by a locally weighted regression line which has a U-shape at the low values of register earnings.

[FIGURE 2 HERE]

Finally, Table 3 shows (unweighted) sample means by sector for all the explanatory variables used in subsequent regression models. These are mostly dummy and categorical variables and provide information about socio-demographic and work characteristics as well as interview related aspects. Note that some labour market variables contain a few missing values and these observations are omitted at the estimation stage. The age variable has been centered around its mean (and re-scaled) to avoid linear correlation between the age and the age-squared variable. Furthermore, in several cases, the categories have been joined to avoid having very few observations in any subgroup.[32] There are some differences in the composition of

[32]Various groupings for the industry variable were tested and the final version chosen on the basis of similar tax compliance behaviour based on the modelling results.

people working in two sectors. In comparison with the unconstrained sector, there are less men in the constrained sector, they tend to be more educated and work primarily in the field of education, health and public administration; there is also a larger proportion of professionals but fewer craft workers and machine operators.

[TABLE 3 HERE]

6 Findings

6.1 Model estimates

The model is estimated both in the multiplicative and the additive form on the sample described in Table 1. (The effective sample has about 120 observations less due to item non-response.) The semi-infinite integrals (for true earnings) were solved numerically using Gauss-Hermite quadrature with the nodes and the weights as calculated in Steen et al. (1969). The log-likelihood functions (see Appendix A.2) were programmed in Stata 12 and estimated using 15 quadrature points. In addition to the main results discussed below, Section 6.4 provides an overview of results from a sensitivity analysis.

The following explanatory variables are included in all three earnings equations: age, age squared, gender, nationality and education. Further demographic characteristics (marital status, region, rural area, dummy for studying) and job characteristics (industry, occupation, number of employees, hours in the main job, dummy for the second job, hours in the second job) are included in the true earnings equation and in the declared earnings equation but not in the survey earnings equation as they are expected to have a negligible effect on the latter. Each equation also includes certain covariates which are excluded from the other two equations to improve identification: health status in the true earnings equation, a mortgage and a lease dummy in the declared earnings equation and interview characteristics (month, people present, rating, response mode, wave) in the survey earnings equation. Having a mortgage and/or a lease loan is assumed to be associated with higher compliance (other things equal) as in order to successfully apply for either of these, one needs to have earnings (in sufficient amount) deposited directly to a bank account on a regular basis. As such this creates an incentive to have a higher proportion of declared earnings if access to credit is desired (see also section 3). Finally, our baseline model specification allows certain parameters to differ between the unconstrained and the constrained sector: the intercept and variance for the true earnings and survey earnings equations as well as θ^s.

The results for all three equations (with robust standard errors) are presented in Table 4 for the multiplicative model and in Table 5 for the additive model. Most covariates for log true earnings ($\ln y^T$) are statistically significant at the 1% level and with expected signs. Earnings are higher for males, Estonian nationals and more educated people; they are higher in the northern (capital) region and notably lower in the north-east region.[33] Age has an inverted U-shape effect on the size of earnings, peaking at 40 years where the age premium is about 17% compared to people aged 20 and 60. There is also a statistically significant positive relationship with health status, job skill level (i.e. occupation), the size of firm and hours worked. Compared to employees in education, health and public administration – reflecting largely public sector employment – earnings are higher in construction, wholesale trade, transportation, professional services and finance. It is somewhat surprising that the sector premium is highest in construction, though the data refer to 2007 which marked the height of the boom in the real estate and construction sector. Finally, while the dummy for the constrained sector is very close to zero (and statistically non-significant), variance (σ_T^2) estimates are clearly higher for the unconstrained sector. Results with the additive model for the true earnings equation are very similar except for slightly larger coefficients for nationality, education, firm size and occupation.

[TABLE 4 AND 5 HERE]

In the case of declared earnings (y^r), the raw coefficients show the effect of independent variables on the latent dependent variable, while our key interest is the effect on the censored dependent variable. For that purpose, raw estimates are useful only to the extent of showing which covariates are statistically significant and the sign of the effect on the censored variable. Marginal effects on the (censored) declared earnings are shown in the next subsection.

Conditional on true earnings, declared earnings have a statistically significant positive association with age, Estonian nationality, education, studying, the size of the firm and whether the household has a mortgage or a lease loan. Having a mortgage has lower statistical significance and one explanation for this is that people interested in mortgage could be less constrained by lower declared earnings if they can compensate this by using (accumulated) undeclared earnings to make a larger downpayment. People requiring a lease loan are presumably less likely to have substantial savings of any form and, hence, the size of declared earnings is more important.

[33] The gender earnings gap is very large at 39%, calculated as $\exp(\hat{\beta}_{\text{male}}) - 1$. Estonia has the highest (unadjusted) gender earnings gap among the EU countries, see Eurostat indicator *tsdsc340*.

Declared earnings are lower for men and for non-married, in particular those who are separated, divorced or widowed. The north-east region, which has several specific characteristics, also stands out for a negative coefficient. First, it has suffered from the highest unemployment rate compared to other regions since the beginning of the 1990s (following the collapse of heavy industry which was central to the local labour market), at times even up to twice higher than in others. Second, with the highest share of non-Estonians, the region is ethnically much less homogenous and this may affect the overall level of trust in public institutions and tax morale. Across sectors, declared earnings are lower in construction, transportation (combined with storage and courier services), hotels and restaurants, and finance (combined with real estate and administrative support) in comparison with education, health and public administration as well as manufacturing, mining and utilities. Occupations associated with higher declared earnings are clerks as well as service and sales workers, while skilled agricultural workers and blue-collar workers have lower earnings. The results for declared earnings are well in line with findings in Kriz et al. (2008) and Merikϋll and Staehr (2010) based on self-reported compliance for Estonia, and also with the (few) general patterns found in the literature (e.g. gender and age – see Section 2.2). The main exception concerns marital status as being married has been found associated with *more* evasion in the previous audit-based US studies, though Kleven et al. (2011) also find link with less evasion like we do.

Again, in terms of statistical significance and the sign of coefficients, results for the additive model are very similar. The values and units of coefficients naturally differ given how declared earnings are specified, most notably for parameter θ^r, i.e. the coefficient of true earnings in the declared earnings equation, which is negative with the multiplicative model and positive with the additive model. But the interpretation of θ^r differs between the two models: unlike for the multiplicative model, it combines the effect of true earnings on declared earnings in levels and relative terms in the additive model.

Finally, conditional on true earnings, survey earnings are higher for males, Estonian nationals, those more educated. The dummy for working in the constrained sector is not statistically significant. There is also a positive link with the timing of interview[34] and its rating, while the number of waves has a negative effect on earnings reported in the survey. Survey earnings are higher when the interview was responded by another household member, however, there is no statistically significant relationship with who was present at the interview. The coefficient of true earnings (θ^s) is highly significant and in the range of 0.6-0.7, being slightly higher

[34]The interviews usually take place around the time when annual tax reports are due (i.e. the end of March) to reduce recall errors.

for the unconstrained sector.

For the model as a whole, both the AIC and the BIC statistic favour the multiplicative form.

6.2 Marginal effects on declared earnings

To give a quantitative interpretation for the effects of the independent variables in the declared earnings equation (y^r), we estimate their marginal effects on the probability of compliance and on the size of declared earnings, conditional on true earnings, as well as the elasticity of declared earnings with respect to true earnings. The underlying formulae are derived in Appendix B.

Figure 3 shows marginal effects of age, gender, education, region, industry and firm size on the probability of compliance, conditional on being truly employed.[35] It focuses on covariates for which estimated coefficients were statistically significant and relatively large in absolute size. Marginal effects are estimated at the sample means and modes of, respectively, continuous and discrete variables for a wide range of values of true earnings: from near 0 up to 25 thousand EEK per month, roughly 3 times the average value of earnings in the sample for the unconstrained sector.

[FIGURE 3 HERE]

Figure 3 shows that, based on the multiplicative model, the estimated probability of full compliance is up to 5 percentage points higher for an additional 10 years of age, increasing in the observed range of true earnings. Depending on the level of true earnings, the probability of compliance is up to 10-11 percentage points (pp) higher for females and people with tertiary education relative to basic education (or less). Similarly, the probability is up to 10 pp lower for the north-east region relative to the north, and as much as 24 pp lower for construction, relative to the pooled sectors of manufacturing, mining and utilities, and 28 pp lower for firms with 1-10 employees relative to firms with 50 or more employees.

In comparison, the additive model shows effects of similar magnitude with the exception of effects for region and firm size which are smaller. The plotted curves for the additive model also exhibit more curvature, reflecting greater sensitivity to the level of true earnings. Among else, the effects for industry and firm size are clearly not monotonically increasing in the covered range of true earnings – the highest effect is shown around the level of 20 thousand EEK (per month).

[35] That is $\partial \Pr(y_i^r = y_i^T | x_i, y_i^T)/\partial x_k$ in case x_k is a continuous variable ($\forall y_i^T > 0$). This equals $\frac{\beta_k^r}{\sigma_r}\phi\left(\frac{\theta^r y_i^T + x_i\beta^r - 1}{\sigma_r}\right)$ with the multiplicative model and $\frac{\beta_k^r}{\sigma_r}\phi\left(\frac{x_i\beta^r - (1-\theta^r)y_i^T}{\sigma_r}\right)$ with the additive model.

The marginal effects on the probability of full and partial evasion are not shown as the estimated probability of full evasion is low and varies rather little with true earnings. Therefore, the effect on the probability of partial evasion basically mirrors that on the probability of full compliance. The marginal effect on *full* evasion is most notable in the case of construction and small firms where the probability is up to 6-7 pp higher.

Figure 4 shows the marginal effect on the expected value of declared earnings for the same characteristics, conditional on true earnings. Overall, this gives a very similar picture in terms of direction and relative magnitude of effects. The key difference is that results for the multiplicative and additive model are now very similar, meaning that the marginal effects on the expected value of declared earnings are much more robust to the model specification than the marginal effects on the probabilities of full compliance.

[FIGURE 4 HERE]

Finally, to understand how the level of true earnings itself affects compliance (holding other characteristics constant), we consider the elasticity of the expected value of declared earnings with respect to true earnings. The mean elasticity across all employees in the unconstrained sector, calculated at predicted individual true earnings (conditional on being truly employed)[36], is 0.91-0.92 depending on the type of the model. This means that on average a 1% increase in (predicted) true earnings would result in a 0.9% increase in the expected value of declared earnings.

Figure 5 shows elasticity estimates for a person with sample mean/mode characteristics, varying one characteristic at a time and across the same range of true earnings. In all cases, elasticity estimates are below 1. Furthermore, elasticity estimates are lower at higher levels of true earnings, indicating that there is a negative association between compliance and true earnings (other things being equal).

[FIGURE 5 HERE]

Elasticity estimates for a person with sample mean/mode characteristics and true earnings at average declared (net) earnings in the sample (8,000 EEK), is 0.97. At this level of true earnings, estimates for the multiplicative and the additive model are basically the same and remain in a narrow range of 0.96-0.98 when varying key characteristics like age, gender, education and region. The estimates are slightly smaller (0.92-0.93) for construction sector and small firms.

Elasticity estimates for true (net) earnings at their mean estimated value in the unconstrained sector (10,000 EEK), are in the range of 0.94-0.97 for most cases in

[36] That is $\mathrm{E}(y_i^T | x_i, y_i^T > 0) = \mathrm{E}[\exp(x_i\beta^T + \varepsilon_i^T) | x_i, y_i^T > 0] = \exp(x_i\beta^T)\exp(\sigma_T^2/2)$

Figure 5. At higher levels of true earnings, the gap between two model estimates increases, exceeding 10 percentage points at 25,000 EEK in the case of construction and small firms.

6.3 Extent of tax evasion

As a last indicator, we provide (aggregate) estimates for the extent of tax evasion. Each individual is characterised by one of the four activities: $S \in \{$no income, partial evasion, full evasion, compliance$\}$. Applying Bayes's law on equation (10), the probability of being engaged in activity s for an individual i (observed in set k) can be expressed as

$$\Pr(s_i|y_i^r, y_i^s, x_i) = f_k(s_i)/f_k \qquad \text{where } s_i \in S \tag{12}$$

The proportion of the sample with outcome s can be estimated as

$$\frac{1}{N}\left[\sum_{i\in A_{0s}} \Pr(s_i|y_i^r, y_i^s, x_i) + \sum_{i\in A_{rs}} \Pr(s_i|y_i^r, y_i^s, x_i)\right] \qquad \text{where } s_i \in S \tag{13}$$

where N is the number of individuals in the sample. Additionally, we can estimate the amount of undeclared earnings and their share in total earnings. The expected value of undeclared earnings e_i for individual i is

$$\mathrm{E}[e_i|y_i^r, y_i^s, x_i] = \begin{cases} \mathrm{E}[y_i^T - y_i^r|y_i^r, y_i^s, x_i, y_i^T > y_i^r] \cdot \Pr(\text{full evasion}) & \text{if } y_i^r = 0 \\ \mathrm{E}[y_i^T - y_i^r|y_i^r, y_i^s, x_i, y_i^T > y_i^r] \cdot \Pr(\text{partial evasion}) & \text{if } y_i^r > 0 \end{cases} \tag{14}$$

which can be rewritten as

$$\frac{1}{f_k}\int_{y_i^r}^{\infty} (y^T - y_i^r) f(y^T|x_i, y_i^T > 0) f(y_i^r|x_i, y^T) f(y_i^s|x_i, y^T)\,\mathrm{d}y^T \qquad \forall i \in A_k \tag{15}$$

The aggregate share of undeclared earnings in total earnings is then[37]

$$\frac{1}{N}\frac{\sum_i E[e_i|y_i^r, y_i^s, x_i]}{\sum_i (y_i^r + E[e_i|y_i^r, y_i^s, x_i])} \tag{16}$$

Estimated proportions are given in Table 6 (panel a). With both types of the model, the estimated share of people in the unconstrained sector with no income is less than 1% and the share of people not reporting any earnings about 3%. The estimated share of people declaring only part of their true earnings exceeds 20% and differs

[37]More specifically, this is in terms of total *net* earnings. To obtain estimates in terms of total *gross* earnings, taxes paid (as they appear in the tax records) have been added to the denominator.

more between the models (28% with the multiplicative and 23% with the additive model), leaving about 70% of private sector employees estimated to be fully compliant (73% with the multiplicative and 68% with the additive model). On the other hand, the difference between the two models is only marginal when comparing the estimates of non-compliance in monetary terms: the aggregate share of undeclared earnings in total (gross) earnings is 15-16% in both cases (panel b). The table also provides estimates for the whole sample as the extent of non-compliance would be typically considered at the population level. Because employees in the constrained sector cannot evade taxes by assumption, the share of compliant individuals in the whole sample is naturally higher than for the unconstrained sample alone (75-80%), while the aggregate share of undeclared earnings is about 12%.

In comparison, a recent audit-based study by Johns and Slemrod (2010) for the US estimated that only 1% of wages and salaries are unreported. Similarly, Kleven et al. (2011) find from audited reports for Denmark that 1% of personal income (comprising labor income, transfers and pensions) is unreported and attribute this to third-party reporting. It is important to note though that unlike most other countries, Denmark has very high effective income tax rates in combination with very low social contribution rates for the employer, hence, the financial incentives implied by the statutory tax burden are very different from that in Estonia.

[TABLE 6 HERE]

Finally, we consider the extent of non-compliance over the (true) income distribution. Table 6 (panel b) shows undeclared earnings as a share of total (gross) earnings by decile groups and the pattern which emerges is similar for both types of model. The share is higher for the bottom and the top decile group: 17-23% of total earnings in the unconstrained sector and 13-18% for the whole sample are estimated to be undeclared, yielding a gently sloping U-shape profile. For the multiplicative model, the share of undeclared earnings for the bottom decile group exceeds that of the top decile group, while the opposite is the case for the additive model. This is further illustrated in Figure 6, which also shows the scale of measurement error by decile group.

[FIGURE 6 HERE]

The pattern of measurement error is clearly different from that of non-compliance showing a very substantial overreporting of survey earnings for the bottom decile group (20-40% of true earnings), a small overreporting for the second decile group and increasing underreporting for higher decile groups, reaching 15-20% of true earnings in the top decile group. Estimated misreporting of survey earnings in

the unconstrained sector follows closely what is found for the constrained sector (by assumption), with the main exception of the bottom decile group where misreporting for the constrained sector is notably larger. Our findings therefore support previous evidence on mean reverting survey measurement error, which stemmed from studies assuming administrative data to be error-free. Largely opposite patterns of non-compliance and measurement error may also explain why differences in mean values of survey and administrative earnings have been found to be rather muted in the measurement error literature.

6.4 Sensitivity analysis

The sensitivity of the main estimates presented above has been tested by estimating the multiplicative and the additive model (i) on alternative samples (models 1 to 3), (ii) with alternative definitions for the constrained sector (models 4 and 5), (iii) with different sets of covariates or parameter constraints (models 6 to 12), (iv) with modifications to the model specifications (models 13 to 15), and (v) taking into account survey design elements, i.e. weights and clustering (model 16).

Table 7 and Table 8 summarise estimates of the key parameters and overall model fit as measured by the AIC and BIC statistics. These show that results are fairly robust to extending the sample with part-time employees (model 1), which was discussed in Section 5.[38] Increased sample heterogeneity mainly affects parameter estimates for the constrained sector, resulting in a higher estimate of the variance of true earnings ($\hat{\sigma}_T^2$) and a smaller coefficient of true earnings in the survey earnings equation ($\hat{\theta}^s$). Estimates are also similar when the sample includes everyone who reported survey earnings for 12 months, i.e. also those whose main activity was not paid employment (model 2), or when employing more conservative sample restrictions, i.e. excluding those with self-employment income or who reported earnings in ESU in gross terms (model 3). The latter finding helps to confirm that the gross-to-net imputations, which were needed for a small sub-sample (see Section 5), have no substantial impact on estimates.

More relaxed definitions for the constrained sector, such as assuming that everyone working in large firms (model 4) or utilities, public administration, education and health (model 5) are also constrained, result in poorer model fit, especially for model 4 where the constrained sector becomes much more heterogeneous as a result (cf. $\hat{\sigma}_T^2$ and $\hat{\sigma}_s^2$). Unfortunately, the categorical variable for firm size makes it impossible to test the relevance of any other criteria for a large firm.

[TABLE 7 AND 8 HERE]

[38] In this case, the earnings variables are adjusted with the number of months paid.

Next, we test alternative sets of covariates and parametric restrictions. Model fit and the estimates of the key parameters are quite robust to omitting covariates for the declared earnings (y^r) equation (model 6) or the survey earnings (y^s) equation (model 7), i.e. imposing all $\beta^r = 0$ and $\beta^s = 0$ (apart from the intercept), respectively. The same applies to restricting the intercept β_0^T and σ_T^2 (model 8) or β_0^s, θ^s and σ_s^2 (model 9) to be the same for the constrained and the unconstrained sector, i.e. the parameters that are allowed to differ between the two sectors in the baseline model. As the main difference between the sectors concerns $\hat{\sigma}_T^2$, which is nearly two times larger for the unconstrained sector, the model fit is worse with model 8. Compared to the baseline, including additional covariates (model 10 and 11) improves the model fit according to AIC, though BIC indicates the opposite. Model 10 adds to the survey earnings equation (y^s) demographic and job characteristics, which were previously included only in the other two equations (marital status, region, industry, occupation etc), while model 11 includes all covariates in all three equations. In both cases, the key parameters change little.

Across models 1 to 11, the estimates of the coefficient of true earnings in the declared earnings equations ($\hat{\theta}^r$) are rather stable with the multiplicative type of models (ranging from -0.02 to -0.03) and always statistically highly significant. It varies more with the additive type of models (ranging from 0.05 to 0.6) and is not always statistically significant (cf. model 4). This implies that θ^r for the additive type of model is more sensitive and cannot be estimated so precisely. Estimates of another key parameter, the coefficient of true earnings in the survey earnings equation ($\hat{\theta}^s$), are similar for the two types of models ranging from 0.42 to 0.75 in these scenarios.

We also test alternative model specifications (besides the multiplicative and the additive form for the y^r equation). Most importantly, we assess the added value of having income also reported in the survey and not only in the tax records by estimating a partial model which contains the true earnings (y^T) equation and the declared earnings (y^r) equation and leaves the survey earnings (y^s) equation aside (model 13). This is equivalent to imposing $\theta^s = 0$ and $\theta_0^s = 0$ in the survey earnings equation (model 12) such that any direct link between the true earnings and survey earnings is ignored. The latter approach demonstrates how the overall model fit becomes much poorer with these restrictions and, hence, confirms the importance of combining two sets of income observations for estimating true earnings (see also discussion in Section 4). It is notable how much the estimates of θ^r for model 12 and 13 differ from other models. Second, we estimate a (sub)model using only the sample of employees with both positive survey and declared earnings (model 14) and here too we observe a sizeable effect on the estimate of θ^r. Third, assuming that everyone

has declared their earnings correctly to the tax authority (model 15), we estimate a model based only on simplified likelihood functions (see equations A.3 and A.4 in Appendix A.1 – in this case there is no difference between the multiplicative and the additive form). Much poorer model fit confirms that this is clearly an unrealistic assumption. Without the possibility of underreporting y^r, estimated variance of true earnings and survey earnings $(\hat{\sigma}_T^2, \hat{\sigma}_s^2)$ increase greatly and the link between true earnings and survey earnings becomes weaker (i.e. $\hat{\theta}^s$ decreases).

Finally, estimations with survey weights, which account for the sample design and non-response, and clustering at the household level (model 16), confirm their negligible effect on parameter estimates.

The second part of the sensitivity analysis focuses on the estimates of the extent of non-compliance under various scenarios. These are summarised in Table 9, both in terms of the proportion of sample and unreported earnings as a share of total earnings. The share of compliant people is between 72-82% with the multiplicative type of models (leaving aside model 15 where evasion is ruled out by assumption), while it is slightly more varying with the additive type of models (69-85%). It is notable that the estimated share of full evaders is highest when the sample includes part-time employees (model 1 and 2). Across models 1-11, the estimated share of undeclared earnings is quite stable ranging between 9-14% of the total for the multiplicative and the additive types, the latter often yielding marginally higher estimates. Among these models, the share of undeclared earnings is lowest when the constrained sector is extended to include employees in large firms (model 4) and highest with the extended sample under model 2. The proportion of undeclared earnings is only 6% with the partial model (12 and 13), where true earnings are estimated solely on the basis of declared earnings, ignoring survey earnings.

[TABLE 9 HERE]

Among models 1-11, non-compliance is higher in the bottom and the top decile group, and to some extent in the 2nd and the 9th decile group, hence, providing further support for the overall U-shape. The U-shape is especially pronounced for model 1 and model 2, which are estimated on extended samples including also individuals with lower work intensity (as employees). The estimates by decile groups are more robust for the multiplicative models.

The partial model (12 and 13), however, exhibits a different profile: the share of undeclared earnings is highest for the bottom decile group (25%), decreases smoothly across the estimated true income distribution and is only 2% for the top decile group. This illustrates how on the basis of declared earnings alone and without a secondary income measure, it is not possible to detect all undeclared earnings as estimates

of true earnings, especially at higher income levels, remain too conservative.[39] A declining ratio of unreported wages and salaries across the true income distribution is also shown in Johns and Slemrod (2010) based on audited reports and, in the light of evidence above, could therefore indicate limited success of audits to uncover non-compliance in earnings at higher levels. The structure of multiplicative model 13 is similar to Feinstein (1991) who modelled income underreporting and its partial detection by auditors using also audit data. Without means to identify absolute detection rates, he interpreted his estimates of non-compliance as if all auditors had the same detection rates as estimated for the best performers and our findings essentially confirm his intuition. Our empirical finding is also in line with recent work in the measurement error literature where Meijer et al. (2012), generalising the Kapteyn and Ypma (2007) model, demonstrate that the best predictors of true earnings are those combining survey and register income measures.[40]

Finally, as with the previous table, taking survey design into account (model 16) has only a limited effect on the estimates – the biggest change occurs in the top decile group where the estimated share of underreported earnings decreases by 2-4 percentage points.

7 Conclusions

The paper uses income survey data linked with tax records at the individual level for Estonia to estimate the determinants and extent of income tax compliance in a novel way. We propose and estimate an econometric model with three simultaneous equations for true income, register income and survey income. Unlike previous approaches in the tax compliance and survey measurement error literature, our model allows income to be misreported *both* in the survey and in the tax records. Focussing on employment income (i.e. wages and salaries), we model register and survey earnings conditional on true earnings and other personal characteristics. Our key identifying assumption is that people working in the public sector are constrained in their choice and cannot evade taxes, while there are essentially no systematic differences in true earnings and survey measurement error between the public and private sector employees (after controlling for individual characteristics). This enables us to observe true earnings for part of the sample.

Besides proposing a novel econometric model and identification strategy, the

[39] That is unless there are no earnings differences between the constrained and the unconstrained sector at any income level which would be a very strong assumption.

[40] In this case, the source of error in the administrative values is only due to mismatch in record linkage. Interestingly, Meijer et al. (2012) highlight unreported earnings in the register data when discussing potential reasons for the latter to perform relatively poorly.

paper extends the empirical tax evasion literature by providing new evidence of non-compliance in a post-socialist country. High-quality data sources for studying tax compliance are very rare, especially in other than major developed countries; the dataset used here is also unique for not requiring respondents' consent for linkage, which could result in a serious sample selection bias. A long-term characteristic of Estonia is its flat income tax due to which cross-sectional variation in effective marginal tax rates is very limited. Our study is therefore unable to shed light on the effect of marginal tax rates on compliance, but also avoids related endogeneity problems as progressive tax rates would be highly correlated with declared income.

The main findings are the following. First, our estimates show that, conditional on true earnings, earnings declared to the tax authority are positively associated with age, education levels, Estonian nationality, studying, the size of the firm and having a mortgage or a lease loan. Compliance is lower for men, non-married and for people living in the north-east region. There are also notable sectoral and occupational differences and, importantly, our results indicate a negative association between compliance and true earnings (other things being equal). In general, our estimates appear to be in line with findings in the previous literature. Second, we find overall substantial non-compliance with respect to wages and salaries. While the share of fully non-compliant employees is marginal (2-3%), our estimates show that more than 20% of employees underreport part of their earnings and about 12% of total employment income (and 15-16% of total income in the unconstrained sector) is not declared to the tax authority. Third, there are significant differences across the estimated true income distribution with much lower compliance among the people in the bottom and the top earnings decile group. Fourth, there are substantial measurement errors in survey income. These exhibit a mean-reverting pattern with large over-reporting at low values of true earnings and moderate under-reporting at medium and high values of true earnings.

In times when researchers are increasingly gaining access to linked survey and administrative data, our model represents a new improved method for studying prevalence and determinants of tax compliance as well as survey measurement error. Our analysis also highlights limitations for detecting non-compliance on the basis of audited tax reports alone, even with partial detection methods (commonly used by the US tax authority), as the resulting estimates are likely to be too conservative.

Our findings have also several important policy implications. Rather sizable underreporting of earnings, despite all such income being in principle subject to third-party reporting and tax withholding, highlights the limitations of such procedures to avoid non-compliance and confirms the (continuing) need for other measures as well to counter evasion. It also raises questions about the common view in the

literature that there is very little evasion of taxes on wages and salaries in the first place and about the ability of (randomised) audits, on which previous findings are mainly based, to capture non-declared earnings. This suggests that more attention to employment income by the tax authorities could be warranted. Finally, there are implications for the progressivity and redistributive aspects of the tax system. The overall pattern of non-compliance across the income distribution could induce more people to perceive that their effective tax burden is higher compared to those who are better off and subsequently weaken their motives to be compliant.

References

Abowd, J. M., and Stinson, M. H. (2013). "Estimating measurement error in annual job earnings: A comparison of survey and administrative data." *Review of Economics & Statistics*, *95*(5), 1451–1467.

Alexander, C., and Feinstein, J. (1987). "A microeconometric analysis of income tax evasion.", unpublished manuscript.

Allingham, M. G., and Sandmo, A. (1972). "Income tax evasion: A theoretical analysis." *Journal of Public Economics*, *1*(3-4), 323–338.

Alm, J. (1991). "A perspective on the experimental analysis of taxpayer reporting." *The Accounting Review*, *66*(3), 577–593.

Alm, J. (1999). "Tax compliance and administration." In W. B. Hildreth, and J. A. Richardson (Eds.), *Handbook on Taxation*, 741–768, New York: Marcel Dekker, Inc.

Alm, J. (2012). "Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies." *International Tax and Public Finance*, *19*, 54–77.

Alm, J., and Jacobson, S. (2007). "Using laboratory experiments in public economics." *National Tax Journal*, *60*(1), 129–152.

Andersen, P. (1977). "Tax evasion and labor supply." *Scandinavian Journal of Economics*, *79*(3), 375–383.

Andreoni, J., Erard, B., and Feinstein, J. (1998). "Tax compliance." *Journal of Economic Literature*, *36*(2), 818–860.

Antila, J., and Ylöstalo, P. (2003). "Working Life Barometer in the Baltic countries 2002." Labour Policy Studies 247, Ministry of Labour (Finland), Helsinki.

Baldini, M., Bosi, P., and Lalla, M. (2009). "Tax evasion and misreporting in income tax returns and household income surveys." *Politica Economica*, *XXV*(3), 333–348.

Baldry, J. C. (1986). "Tax evasion is not a gamble: A report on two experiments." *Economics Letters*, *22*(4), 333 – 335.

Baldry, J. C. (1987). "Income tax evasion and the tax schedules: some experimental results." *Public Finance / Finances Publiques*, *42*(3), 357–383.

Becker, W., Büchner, H.-J., and Sleeking, S. (1987). "The impact of public transfer expenditures on tax evasion: An experimental approach." *Journal of Public Economics*, *34*(2), 243–252.

Benedek, D., and Lelkes, O. (2011). "The distributional implications of income under-reporting in Hungary." *Fiscal Studies*, *32*(4), 539–560.

Besim, M., and Jenkins, G. P. (2005). "Tax compliance: when do employees behave like the self-employed?" *Applied Economics*, *37*(10), 1201–1208.

Bollinger, C. B. (1998). "Measurement error in the Current Population Survey: A nonparametric look." *Journal of Labor Economics*, *16*(3), 576–594.

Bound, J., Brown, C., Duncan, G. J., and Rodgers, W. L. (1994). "Evidence on the validity of cross-sectional and longitudinal labor market data." *Journal of Labor Economics*, *12*(3), 345–368.

Bound, J., Brown, C., and Mathiowetz, N. (2001). "Measurement error in survey data." In J. J. Heckman, and E. Leamer (Eds.), *Handbook of Econometrics*, vol. 5, chap. 59, 3705–3843, Amsterdam: Elsevier.

Bound, J., and Krueger, A. B. (1991). "The extent of measurement error in longitudinal earnings data: Do two wrongs make aright?" *Journal of Labor Economics*, *9*(1), 1–24.

Chen, X., Hong, H., and Nekipelov, D. (2011). "Nonlinear models of measurement errors." *Journal of Economic Literature*, *49*(4), 901–937.

Clotfelter, C. (1983). "Tax evasion and tax rates: An analysis of individual returns." *The Review of Economics and Statistics*, *65*(3), 363–373.

Cowell, F. A. (1985). "Tax evasion with labour income." *Journal of Public Economics*, *26*(1), 19–34.

Cowell, F. A. (1990). *Cheating the government: the economics of evasion*. Cambridge, Massachusetts: The MIT Press.

Dubin, J. A., and Wilde, L. L. (1988). "An empirical analysis of federal income tax auditing and compliance." *National Tax Journal*, *41*(1), 61–74.

Duncan, G. J., and Hill, D. H. (1985). "An investigation of the extent and consequences of measurement error in labor-economic survey data." *Journal of Labor Economics*, *3*(4), 508–532.

Elek, P., Köllo, J., Reizer, B., and Szabó, P. A. (2012). "Detecting wage underreporting using a double-hurdle model." In S. Polachek, and K. Tatsiramos (Eds.), *Informal Employment in Emerging and Transition Economies*, *Research in Labor Economics*, vol. 34, chap. 4, 135–166, Emerald Group Publishing Limited.

Elffers, H., Robben, H. S., and Hessing, D. J. (1991). "Under-reporting income: Who is the best judge – tax-payer or tax inspector?" *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *154*(1), 125–127.

Elffers, H., Weigel, R. H., and Hessing, D. J. (1987). "The consequences of different strategies for measuring tax evasion behavior." *Journal of Economic Psychology*, *8*(3), 311–337.

Engström, P., and Holmlund, B. (2009). "Tax evasion and self-employment in a high-tax country: evidence from Sweden." *Applied Economics*, *41*(19), 2419–2430.

Erard, B. (1993). "Taxation with representation: An analysis of the role of tax practitioners in tax compliance." *Journal of Public Economics*, *52*(2), 163–197.

Erard, B. (1997). "Self-selection with measurement errors. A microeconometric analysis of the decision to seek tax assistance and its implications for tax compliance." *Journal of Econometrics*, *81*(2), 319–356.

Erard, B., and Feinstein, J. S. (1994). "Honesty and evasion in the tax compliance game." *The RAND Journal of Economics*, *25*(1), 1–19.

Erard, B., and Ho, C.-C. (2001). "Searching for ghosts: who are the nonfilers and how much tax do they owe?" *Journal of Public Economics*, *81*(1), 25–50.

European Commission (2011). *Taxation trends in the European Union: Data for the EU Member States, Iceland and Norway*. Luxembourg: Publications Office of the European Union.

Feinstein, J. S. (1991). "An econometric analysis of income tax evasion and its detection." *The RAND Journal of Economics*, *22*(1), 14–35.

Feldman, N. E., and Slemrod, J. (2007). "Estimating tax noncompliance with evidence from unaudited tax returns." *The Economic Journal*, *117*, 327–352.

Fiorio, C. V., and D'Amuri, F. (2005). "Workers' tax evasion in Italy." *Giornale degli Economisti e Annali di Economia*, *64*(2/3), 241–264.

Forest, A., and Sheffrin, S. M. (2002). "Complexity and compliance: An empirical investigation." *National Tax Journal*, *55*(1), 75–88.

Friedland, N., Maital, S., and Rutenberg, A. (1978). "A simulation study of income tax evasion." *Journal of Public Economics*, *10*(1), 107–116.

Gordon, J. P. (1989). "Individual morality and reputation costs as deterrents to tax evasion." *European Economic Review*, *33*(4), 797–805.

Graetz, M. J., and Wilde, L. L. (1985). "The economics of tax compliance: fact and fantasy." *National Tax Journal*, *38*(3), 355–363.

Hashimzade, N., Myles, G. D., and Tran-Nam, B. (2013). "Applications of behavioural economics to tax evasion." *Journal of Economic Surveys*, *27*(5), 941–977.

Hurst, E., Li, G., and Pugsley, B. (2014). "Are household surveys like tax forms: evidence from income underreporting of the self-employed." *Review of Economics & Statistics*, *96*(1), 19–33.

Jenkins, S. P., Cappellari, L., Lynn, P., Jäckle, A., and Sala, E. (2006). "Patterns of consent: Evidence from a general household survey." *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *169*(4), 701–722.

Johns, A., and Slemrod, J. (2010). "The distribution of income tax noncompliance." *National Tax Journal*, *63*(3), 397–418.

Kapteyn, A., and Ypma, J. Y. (2007). "Measurement error and misclassification: A comparison of survey and administrative data." *Journal of Labor Economics*, *25*(3), 513–551.

Kinsey, K. A. (1992). "Deterrence and alienation effects of IRS enforcement: an analysis of survey data." In J. Slemrod (Ed.), *Why people pay taxes: tax compliance and enforcement*, 259–285, Ann Arbor: The University of Michigan Press.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., and Saez, E. (2011). "Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark." *Econometrica*, *79*(3), 651–692.

Kleven, H. J., Kreiner, C. T., and Saez, E. (2009). "Why can modern governments tax so much? An agency model of firms as fiscal intermediaries." NBER Working Paper 15218.

Kriz, K., Meriküll, J., Paulus, A., and Staehr, K. (2008). "Why do individuals evade payroll and income taxation in Estonia?" In M. Pickhardt, and E. Shinnick (Eds.), *Shadow Economy, Corruption and Governance*, 240–264, Cheltenham: Edward Elgar.

Kukk, M., and Staehr, K. (2014). "Income underreporting by households with business income: evidence from Estonia." *Post-Communist Economies*, *26*(2), 257–276.

Lyssiotou, P., Pashardes, P., and Stengos, T. (2004). "Estimates of the black economy based on consumer demand approaches." *The Economic Journal*, *114*(497), 622–640.

Martinez-Vazquez, J., and Rider, M. (2005). "Multiple modes of tax evasion: Theory and evidence." *National Tax Journal*, *58*(1), 51–76.

Meijer, E., Rohwedder, S., and Wansbeek, T. (2012). "Measurement error in earnings data: Using a mixture model approach to combine survey and register data." *Journal of Business & Economic Statistics*, *30*(2), 191–201.

Meriküll, J., and Staehr, K. (2010). "Unreported employment and envelope wages in mid-transition: Comparing developments and causes in the Baltic countries." *Comparative Economic Studies*, *52*(4), 637–670.

Mork, K. A. (1975). "Income tax evasion: some empirical evidence." *Public Finance / Finances Publiques*, *30*(1), 70–76.

Pencavel, J. H. (1979). "A note on income tax evasion, labor supply, and nonlinear tax schedules." *Journal of Public Economics*, *12*(1), 115–124.

Pickhardt, M., and Prinz, A. (2014). "Behavioral dynamics of tax evasion – a survey." *Journal of Economic Psychology*, *40*, 1–19.

Pischke, J.-S. (1995). "Measurement error and earnings dynamics: Some estimates from the PSID validation study." *Journal of Business & Economic Statistics*, *13*(3), 305–314.

Pissarides, C. A., and Weber, G. (1989). "An expenditure-based estimate of Britain's black economy." *Journal of Public Economics*, *39*, 17–32.

Pudney, S. (2008). "Heaping and leaping: Survey response behaviour and the dynamics of selfreported consumption expenditure." ISER Working Paper 2008-09, University of Essex, Colchester.

Pudney, S. E., Pyle, D. J., and Saruc, T. (2000). "Income tax evasion: an experimental approach." In Z. MacDonald, and D. J. Pyle (Eds.), *Illicit activity: the economics of crime, drugs and tax fraud*, 267–283, Dartmouth: Ashgate.

Sakshaug, J. W., and Kreuter, F. (2012). "Assessing the magnitude of non-consent biases in linked survey and administrative data." *Survey Research Methods*, *6*(2), 113–122.

Sandmo, A. (2005). "The theory of tax evasion: A retrospective view." *National Tax Journal*, *LVII*(4), 643–663.

Schepanski, A., and Kelsey, D. (1990). "Testing for framing effects in taxpayer compliance decisions." *Journal of the American Taxation Association*, *12*(1), 60–77.

Schuetze, H. J. (2002). "Profiles of tax non-compliance among the self-employed in Canada: 1969 to 1992." *Canadian Public Policy / Analyse de Politiques*, *28*(2), 219–238.

Shaw, J., Slemrod, J., and Whiting, J. (2010). "Administration and compliance." In J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, and J. Poterba (Eds.), *Dimensions of Tax Design: The Mirrlees Review*, chap. 12, 1100–1162, Oxford: Oxford University Press.

Sheffrin, S. M., and Triest, R. K. (1992). "Can brute deterrence backfire? Perceptions and attitudes in taxpayer compliance." In J. Slemrod (Ed.), *Why people pay taxes: tax compliance and enforcement*, 193–218, Ann Arbor: The University of Michigan Press.

Slemrod, J. (2007). "Cheating ourselves: The economics of tax evasion." *The Journal of Economic Perspectives*, *21*(1), 25–48.

Slemrod, J., and Weber, C. (2012). "Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy." *International Tax and Public Finance*, *19*, 25–53.

Slemrod, J., and Yitzhaki, S. (2002). "Tax avoidance, evasion, and administration." In A. J. Auerbach, and M. Feldstein (Eds.), *Handbook of Public Economics*, vol. 3, chap. 22, 1423–1470, Amsterdam: Elsevier.

Spicer, M. W., and Becker, L. A. (1980). "Fiscal inequity and tax evasion: An experimental approach." *National Tax Journal*, *33*(2), 171–175.

Srinivasan, T. (1973). "Tax evasion: A model." *Journal of Public Economics*, *2*(4), 339–346.

Steen, N. M., Byrne, G. D., and Gelbard, E. M. (1969). "Gaussian quadratures for the integrals $\int_0^\infty \exp(-x^2)f(x)dx$ and $\int_0^b \exp(-x^2)f(x)dx$." *Mathematics of Computation*, *23*(107), 661–671.

Toomse, M. (2010). "Looking for a middle class bias: Salary and co-operation in social surveys." ISER Working Paper 2010-03, University of Essex, Colchester.

Webley, P., and Halstead, S. (1986). "Tax evasion on the micro: significant simulations or expedient experiments?" *Journal of Interdisciplinary Economics*, *1*, 87–100.

Weigel, R. H., Hessing, D. J., and Elffers, H. (1987). "Tax evasion research: A critical appraisal and theoretical model." *Journal of Economic Psychology*, *8*(2), 215–235.

Witte, A. D., and Woodbury, D. F. (1985). "The effect of tax laws and tax administration on tax compliance: The case of the U.S. individual income tax." *National Tax Journal*, *38*(1), 1–13.

Yitzhaki, S. (1974). "A note on 'Income tax evasion: A theoretical analysis'." *Journal of Public Economics*, *3*(2), 201–202.

Table 1: Evolution of the sample

Sample	Number of persons			Omitted at
	Total	A_{0s}	A_{rs}	each step
Initial sample of ESU 2008	14,942	-	-	-
Linked with tax records	14,871	-	-	71
Aged 16 or older[a]	12,699	-	-	2,172
Respondent household[b]	10,789	-	-	1,910
Respondent individual	10,702	-	-	87
Complete earnings information	10,237	-	-	465
Ever had a regular job	8,587	-	-	1,650
Employed (positive survey earnings)	5,500	294	5,206	3,087
Employment main activity[c]	5,327	249	5,078	173
Full time employment[d]	4,121	138	3,983	1,206
- constrained sector[e]	921	12	909	-
- unconstrained sector	3,200	126	3,074	-

Notes: [(a)] subject to a personal interview in the survey; [(b)] for new sample members the number of non-respondents includes only sampled persons without other household members; [(c)] part- or full-time employment reported as the main activity at least for one month in the income reference period; [(d)] full-time employment reported as the main activity (and employment income received) for 12 months in the income reference period; [(e)] constrained sector sub-sample includes public sector employees, except those who changed jobs or have a second job.

Table 2: Mean log survey and register income

Sample	$\ln y^s$		$\ln y^r$		Difference		N
	b	se	b	se	b	se	
(a) All (adults) with positive earnings in the tax records							
ESU non-respondents	-	-	8.46	0.06	-	-	1,114
ESU respondents	-	-	8.33	0.03	-	-	6,698
(b) ESU respondents – intermediate sample							
Positive earnings in the tax records (A_{r0})	-	-	6.28	0.14	-	-	343
Positive earnings in ESU (A_{0s})	8.54	0.09	-	-	-	-	294
Positive earnings in both sources (A_{rs})	8.72	0.02	8.61	0.02	0.10	0.01	5,206
- constrained sector	8.77	0.03	8.84	0.03	-0.07	0.02	1,040
- unconstrained sector	8.70	0.02	8.55	0.02	0.16	0.02	4,166
(c) ESU respondents – final estimation sample							
Positive earnings in ESU (A_{0s})	8.99	0.07	-	-	-	-	138
Positive earnings in both sources (A_{rs})	8.92	0.01	8.84	0.02	0.08	0.01	3,983
- constrained sector	8.87	0.03	8.95	0.03	-0.08	0.02	909
- unconstrained sector	8.93	0.01	8.80	0.02	0.14	0.02	3,074

Notes: annual (net) earnings in EEK divided by 12, in log terms; estimates take into account design weights and clustering at the household level; intermediate sample contains respondent individuals with complete earnings information and who have had a regular job; final estimation sample contains full-time employed; constrained sector sub-sample includes public sector employees, except those who changed jobs or have a second job.

Figure 1: Distribution of survey and register income

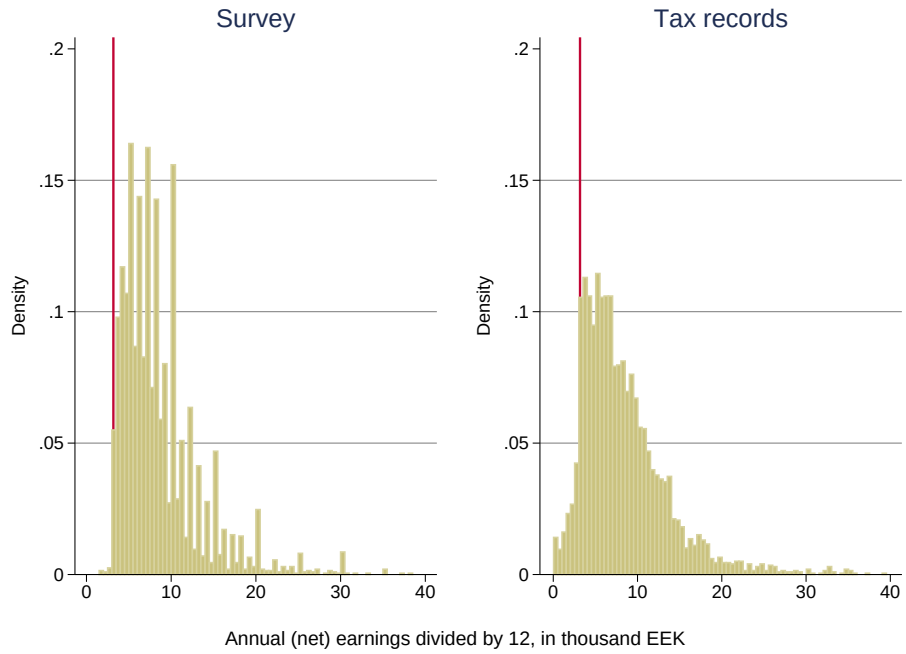

Figure 2: Survey and register income by sector

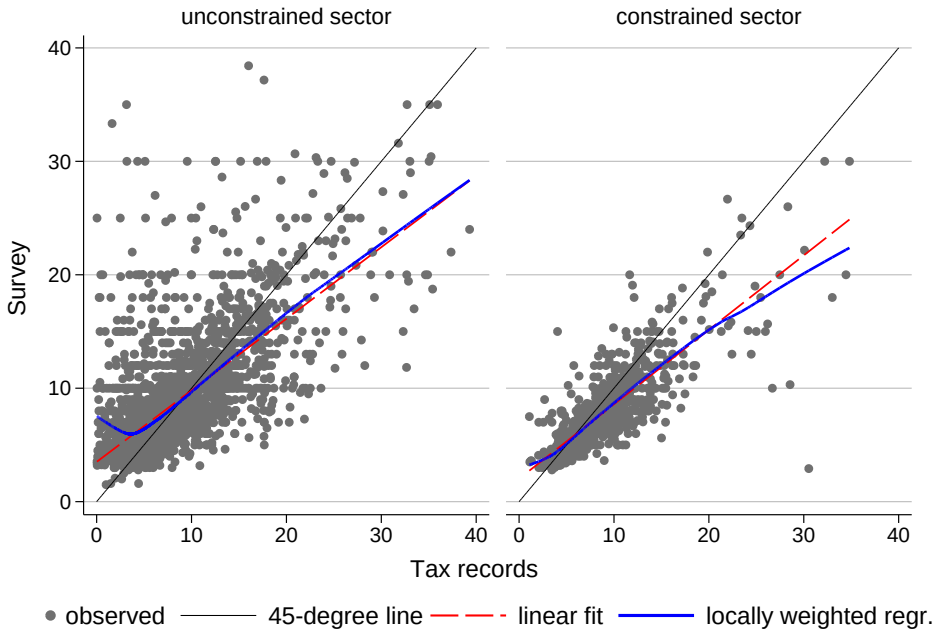

Table 3: Sample means

Variable	Uncon-strained	Con-strained	All	N
Monthly (net) earnings in tax report, thousand EEK	8.09	8.76	8.24	4,121
Monthly (net) earnings in ESU, thousand EEK	8.94	7.85	8.69	4,121
Age[a]	-0.10	0.33	-0.01	4,121
Age[a] squared	1.36	1.23	1.33	4,121
Gender=male	0.56	0.30	0.50	4,121
Nationality=Estonian	0.73	0.79	0.74	4,121
Education=basic or less	0.11	0.05	0.10	4,121
Education=secondary	0.63	0.45	0.59	4,121
Education=tertiary	0.25	0.50	0.31	4,121
Marital status=single	0.16	0.12	0.15	4,121
Marital status=married	0.54	0.56	0.54	4,121
Marital status=cohabiting	0.19	0.15	0.18	4,121
Marital status=divorced, widow or separated	0.11	0.17	0.12	4,121
Dummy for studying	0.03	0.05	0.04	4,121
Region=north	0.30	0.26	0.29	4,121
Region=central	0.14	0.12	0.14	4,121
Region=north-east	0.10	0.12	0.10	4,121
Region=west	0.17	0.17	0.17	4,121
Region=south	0.28	0.32	0.29	4,121
Area=rural	0.41	0.40	0.41	4,121
Occupation=senior managers, legislators	0.11	0.12	0.11	4,120
Occupation=professionals	0.09	0.36	0.15	4,120
Occupation=technicians, associate professionals	0.11	0.14	0.11	4,120
Occupation=clerks	0.05	0.06	0.05	4,120
Occupation=service and sales workers	0.10	0.13	0.11	4,120
Occupation=skilled agricultural workers	0.02	0.01	0.01	4,120
Occupation=craft and related trade workers	0.22	0.04	0.18	4,120
Occupation=plant and machine operators	0.22	0.05	0.19	4,120
Occupation=elementary occupations	0.08	0.10	0.09	4,120
Industry=agriculture, forestry	0.06	0.02	0.05	4,028
Industry=manufacturing, mining, utilities	0.32	0.05	0.26	4,028
Industry=construction	0.15	0.01	0.12	4,028
Industry=wholesale trade, motor vehicles	0.06	0.00	0.05	4,028
Industry=retail trade	0.09	0.00	0.07	4,028
Industry=transportation, storage, courier	0.09	0.07	0.08	4,028
Industry=hotels, restaurants	0.04	0.01	0.03	4,028
Industry=prof. services, information, communication	0.04	0.02	0.04	4,028
Industry=finance, real estate, admin/support	0.07	0.01	0.05	4,028
Industry=education, health, public admin.	0.08	0.80	0.25	4,028

Notes: unweighted means for the final estimation sample (i.e. full-time employed); [(a)] constructed as $(age - 43)/10$, where 43 is (unweighted) sample mean.

(Table continues on next page)

Table 3 continues

Variable	Uncon-strained	Con-strained	All	N
Dummy for constrained sector[b]	0.00	1.00	0.22	4,121
No of employees=1-10	0.20	0.14	0.19	4,019
No of employees=11-19	0.17	0.16	0.16	4,019
No of employees=20-49	0.22	0.26	0.23	4,019
No of employees=50 or more	0.37	0.42	0.38	4,019
No of employees=uncertain (more than 10)	0.04	0.03	0.04	4,019
Number of hours in main job (usual per week)	40.84	40.03	40.66	4,024
Dummy for second job	0.07	0.00	0.05	4,028
Number of hours in second job (usual per week)	0.86	0.00	0.67	4,028
Health=very good	0.08	0.07	0.08	4,120
Health=good	0.59	0.58	0.59	4,120
Health=neither good or bad	0.30	0.32	0.30	4,120
Health=poor or very poor	0.03	0.03	0.03	4,120
Dummy for health problems limiting work/study	0.15	0.14	0.15	4,121
Dummy for HH having a mortgage	0.25	0.20	0.24	4,110
Dummy for HH having a lease	0.25	0.23	0.25	4,121
Month of interview (since Feb)	1.65	1.53	1.62	4,121
Dummy for young child at interview	0.04	0.03	0.03	4,121
Dummy for older child at interview	0.09	0.12	0.09	4,121
Dummy for spouse at interview	0.29	0.28	0.29	4,121
Dummy for other relative at interview	0.10	0.07	0.09	4,121
Interview rating=very well	0.63	0.63	0.63	4,121
Interview rating=well	0.32	0.30	0.31	4,121
Interview rating=ok	0.06	0.06	0.06	4,121
Interview responded=alone	0.84	0.89	0.85	4,121
Interview responded=with someone's help	0.03	0.02	0.03	4,121
Interview responded=by other HH member	0.13	0.09	0.13	4,121
Number of waves	2.14	2.19	2.15	4,121

Notes: [(b)] constrained sector sub-sample includes public sector employees, except those who changed jobs or have a second job.

Table 4: Estimates for the multiplicative model

	Dependent variable					
	$\ln y^T$		y^r		$\ln y^s$	
	coef.	se	coef.	se	coef.	se
Age[a]	-0.025***	0.008	0.073***	0.021	-0.027***	0.004
Age[a] squared	-0.039***	0.005	0.021	0.015	-0.007**	0.003
Male	0.327***	0.018	-0.181***	0.055	0.089***	0.014
Estonian nationality	0.166***	0.024	0.230***	0.055	0.044***	0.011
Education (ref=basic or less)						
- secondary	0.066**	0.026	0.168***	0.056	0.051***	0.016
- tertiary	0.223***	0.030	0.331***	0.079	0.136***	0.019
Marital status (ref=married)						
- single	-0.042*	0.024	-0.128**	0.065		
- cohabiting	-0.011	0.020	-0.165***	0.051		
- divorced/widow/separated	-0.021	0.022	-0.267***	0.069		
Region (ref=north)						
- central	-0.141***	0.025	0.080	0.063		
- north-east	-0.228***	0.027	-0.146**	0.066		
- west	-0.146***	0.024	0.097	0.061		
- south	-0.172***	0.022	0.025	0.053		
Rural area	-0.020	0.016	-0.043	0.044		
Studying	0.006	0.036	0.418**	0.169		
Industry (ref=edu/health/pub.adm)						
- agriculture/forestry	0.008	0.043	-0.085	0.146		
- manufacturing/mining/utilities	0.054*	0.030	-0.006	0.116		
- construction	0.323***	0.039	-0.364***	0.116		
- wholesale trade	0.199***	0.044	0.002	0.131		
- retail trade	0.054	0.034	-0.223	0.137		
- transportation/storage/courier	0.235***	0.036	-0.334***	0.120		
- hotels/restaurants	0.046	0.044	-0.386***	0.139		
- prof. services/inform./commun.	0.160***	0.046	-0.104	0.139		
- finance/real estate/admin-support	0.128***	0.043	-0.437***	0.129		
Occupation (ref=clerks)						
- senior managers	0.409***	0.039	-0.127	0.134		
- professionals	0.345***	0.037	-0.207	0.148		
- technicians/associate prof.	0.227***	0.038	-0.163	0.134		
- service/sales workers	-0.065*	0.039	-0.104	0.156		
- skilled agricultural workers	0.139*	0.082	-0.617***	0.191		
- craft/trade workers	0.119***	0.041	-0.323**	0.129		
- plant/machine operators	0.039	0.037	-0.318**	0.128		
- elementary	-0.205***	0.038	-0.268*	0.142		

Notes: [(a)] constructed as $(age - 43)/10$, where 43 is (unweighted) sample mean. Robust standard errors shown. * p<0.1, ** p<0.05, *** p<0.01

(Table continues on next page)

Table 4 continues

	Dependent variable					
	$\ln y^T$		y^r		$\ln y^s$	
	coef.	se	coef.	se	coef.	se
Constrained sector[b]	0.003	0.024			0.062	0.055
No of employees (ref=1 to 10)						
- 11 to 19	0.107***	0.025	0.110**	0.052		
- 20 to 49	0.162***	0.023	0.339***	0.057		
- 50 or more	0.273***	0.022	0.416***	0.055		
- uncertain (more than 10)	0.246***	0.051	0.159*	0.086		
Hours in main job	0.013***	0.002	-0.005	0.003		
Second job	0.109*	0.057	-0.016	0.155		
Hours in second job	0.004	0.003	0.002	0.007		
Health status (ref=neutral)						
- very good	0.183***	0.031				
- good	0.077***	0.018				
- poor/very poor	-0.082*	0.046				
Health affected work/studying	-0.055***	0.021				
HH has a mortgage			0.077*	0.043		
HH has a lease			0.154***	0.041		
Number of waves					-0.020***	0.004
Month of interview (since Feb)					0.008**	0.003
Interview rating (ref=very well)						
- well					-0.014	0.010
- ok					-0.051**	0.023
Interview responded (ref=alone)						
- with someone's help					-0.045	0.030
- by other HH member					0.037**	0.016
At interview: young child					0.030	0.028
At interview: older child					-0.012	0.013
At interview: spouse					0.012	0.010
At interview: other relative					0.004	0.018
Intercept	0.934***	0.095	1.646***	0.240	0.479***	0.039
p	0.993***	0.002				
θ (unconstrained sector)			-0.024***	0.004	0.689***	0.018
θ (constrained sector)					0.642***	0.025
θ_0					1.113***	0.084
σ (unconstrained sector)	0.474***	0.015	0.583***	0.035	0.247***	0.008
σ (constrained sector)	0.354***	0.014			0.233***	0.012
Sample size	4,006					
AIC	39,017					
BIC	39,741					

Notes: [(b)] constrained sector sub-sample includes public sector employees, except those who changed jobs or have a second job. Robust standard errors shown. * p<0.1, ** p<0.05, *** p<0.01

Table 5: Estimates for the additive model

	Dependent variable					
	$\ln y^T$		y^r		$\ln y^s$	
	coef.	se	coef.	se	coef.	se
Age[a]	-0.024***	0.009	1.180***	0.319	-0.028***	0.004
Age[a] squared	-0.038***	0.005	0.356*	0.215	-0.008**	0.003
Male	0.329***	0.018	-2.911***	0.820	0.095***	0.014
Estonian nationality	0.182***	0.025	2.740***	0.898	0.041***	0.011
Education (ref=basic or less)						
- secondary	0.071***	0.027	1.961**	0.773	0.049***	0.016
- tertiary	0.232***	0.032	4.531***	1.311	0.133***	0.019
Marital status (ref=married)						
- single	-0.045*	0.025	-1.528	0.962		
- cohabiting	-0.013	0.021	-2.446***	0.845		
- divorced/widow/separated	-0.029	0.023	-3.525***	1.022		
Region (ref=north)						
- central	-0.145***	0.025	0.942	0.964		
- north-east	-0.233***	0.028	-1.428	0.975		
- west	-0.150***	0.025	1.158	0.940		
- south	-0.178***	0.023	0.546	0.833		
Rural area	-0.021	0.017	-0.367	0.661		
Studying	0.007	0.036	4.940*	2.769		
Industry (ref=edu/health/pub.adm)						
- agriculture/forestry	0.008	0.044	-2.094	2.309		
- manufacturing/mining/utilities	0.062**	0.031	-1.077	1.948		
- construction	0.340***	0.042	-6.985***	2.014		
- wholesale trade	0.203***	0.047	-0.587	2.239		
- retail trade	0.052	0.036	-4.713**	2.161		
- transportation/storage/courier	0.249***	0.038	-6.500***	2.080		
- hotels/restaurants	0.027	0.044	-6.009***	2.322		
- prof. services/inform./commun.	0.173***	0.048	-3.275	2.360		
- finance/real estate/admin-support	0.123***	0.045	-7.685***	2.206		
Occupation (ref=clerks)						
- senior managers	0.429***	0.041	-2.589	2.000		
- professionals	0.349***	0.038	-3.647	2.276		
- technicians/associate prof.	0.234***	0.039	-3.198	1.989		
- service/sales workers	-0.065	0.040	-1.655	2.193		
- skilled agricultural workers	0.110	0.083	-9.570***	2.919		
- craft/trade workers	0.108**	0.042	-5.282***	1.941		
- plant/machine operators	0.031	0.038	-4.815**	1.921		
- elementary	-0.208***	0.039	-3.886*	2.078		

Notes: [(a)] constructed as $(age - 43)/10$, where 43 is (unweighted) sample mean. Robust standard errors shown. * p<0.1, ** p<0.05, *** p<0.01

(Table continues on next page)

Table 5 continues

	Dependent variable					
	$\ln y^T$		y^r		$\ln y^s$	
	coef.	se	coef.	se	coef.	se
Constrained sector[b]	0.017	0.024			-0.021	0.057
No of employees (ref=1 to 10)						
- 11 to 19	0.119***	0.026	1.089	0.795		
- 20 to 49	0.181***	0.024	4.199***	0.993		
- 50 or more	0.295***	0.023	5.191***	1.054		
- uncertain (more than 10)	0.263***	0.053	0.843	1.355		
Hours in main job	0.013***	0.002	-0.124**	0.059		
Second job	0.118*	0.061	-0.400	2.408		
Hours in second job	0.005	0.004	-0.023	0.110		
Health status (ref=neutral)						
- very good	0.183***	0.032				
- good	0.077***	0.019				
- poor/very poor	-0.085*	0.046				
Health affected work/studying	-0.053**	0.022				
HH has a mortgage			0.750	0.650		
HH has a lease			2.099***	0.653		
Number of waves					-0.021***	0.004
Month of interview (since Feb)					0.008**	0.003
Interview rating (ref=very well)						
- well					-0.015	0.010
- ok					-0.046**	0.023
Interview responded (ref=alone)						
- with someone's help					-0.050*	0.030
- by other HH member					0.043***	0.016
At interview: young child					0.036	0.028
At interview: older child					-0.007	0.014
At interview: spouse					0.011	0.010
At interview: other relative					0.004	0.018
Intercept	0.860***	0.100	20.137***	4.301	0.567***	0.037
p	0.996***	0.001				
θ (unconstrained sector)			0.300**	0.108	0.653***	0.018
θ (constrained sector)					0.642***	0.026
θ_0					1.129***	0.099
σ (unconstrained sector)	0.478***	0.019	8.553***	0.944	0.254***	0.008
σ (constrained sector)	0.354***	0.014			0.233***	0.012
Sample size	4,006					
AIC	39,189					
BIC	39,913					

Notes: [(b)] constrained sector sub-sample includes public sector employees, except those who changed jobs or have a second job. Robust standard errors shown. * p<0.1, ** p<0.05, *** p<0.01

Figure 3: Marginal effect on the probability of full compliance, conditional on true earnings

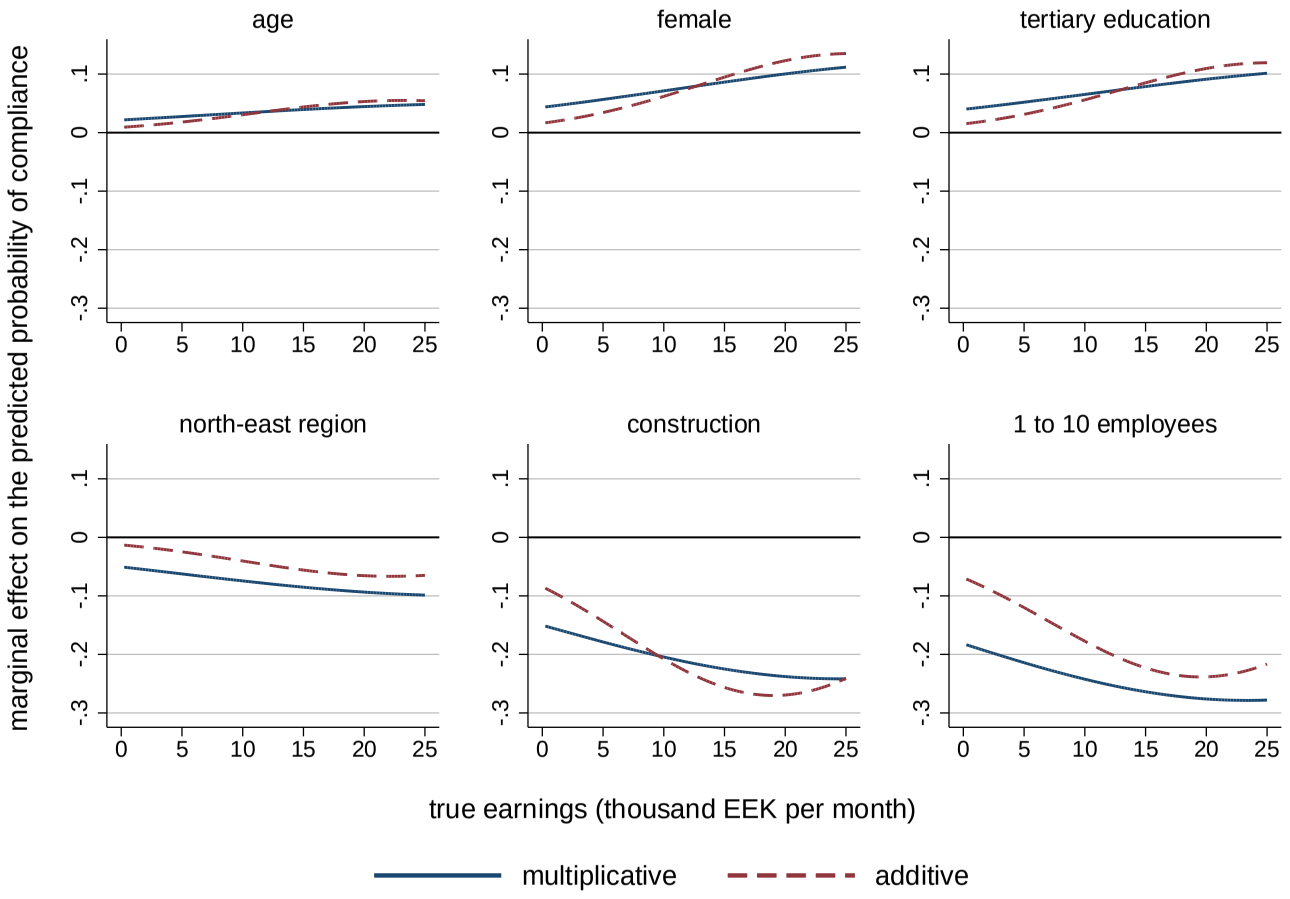

Figure 4: Marginal effect on the expected value of declared earnings ($\mathrm{E}[y^r|x, y^T]$)

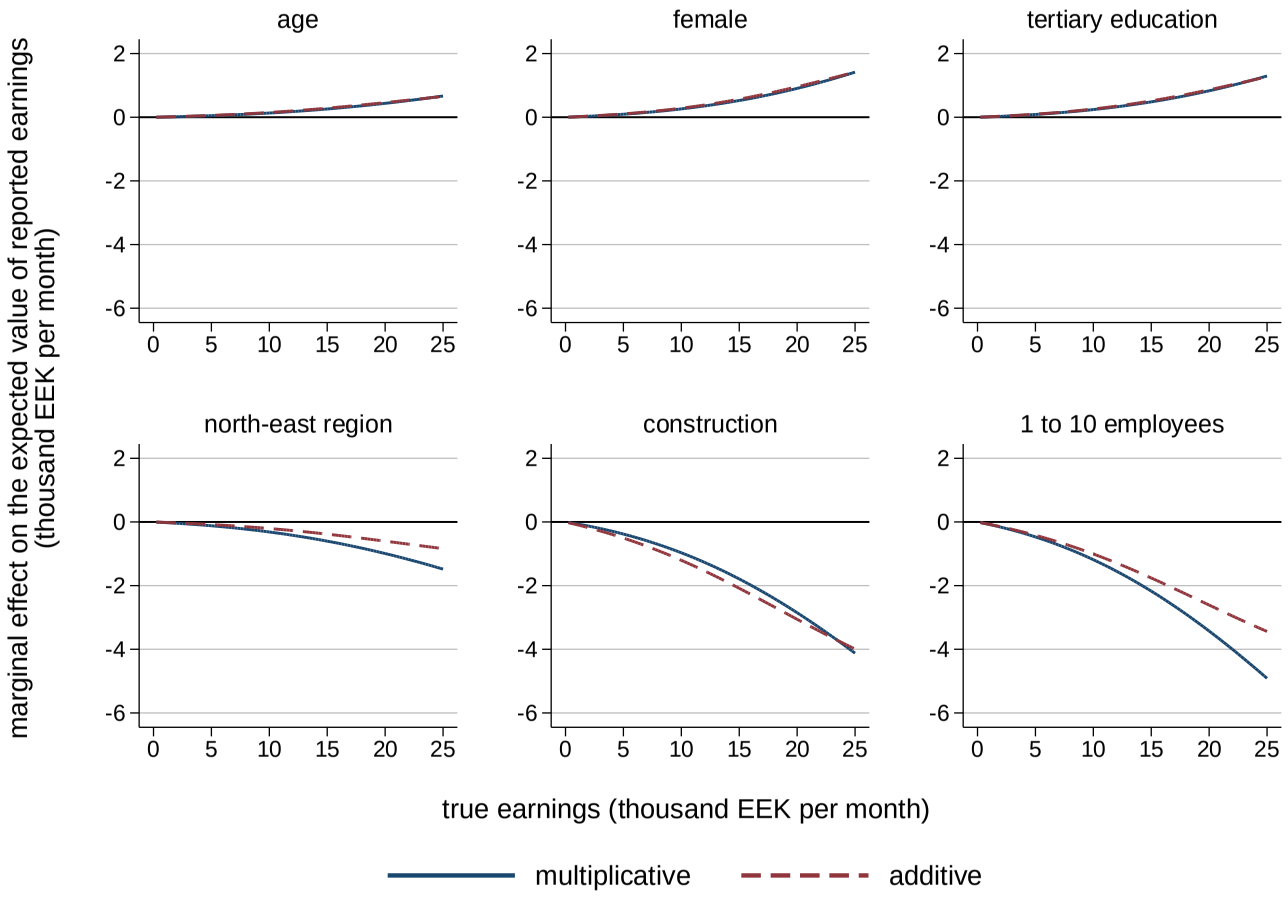

Figure 5: Elasticity of expected value of declared earnings ($\mathrm{E}[y^r|x, y^T]$) with respect to true earnings (y^T)

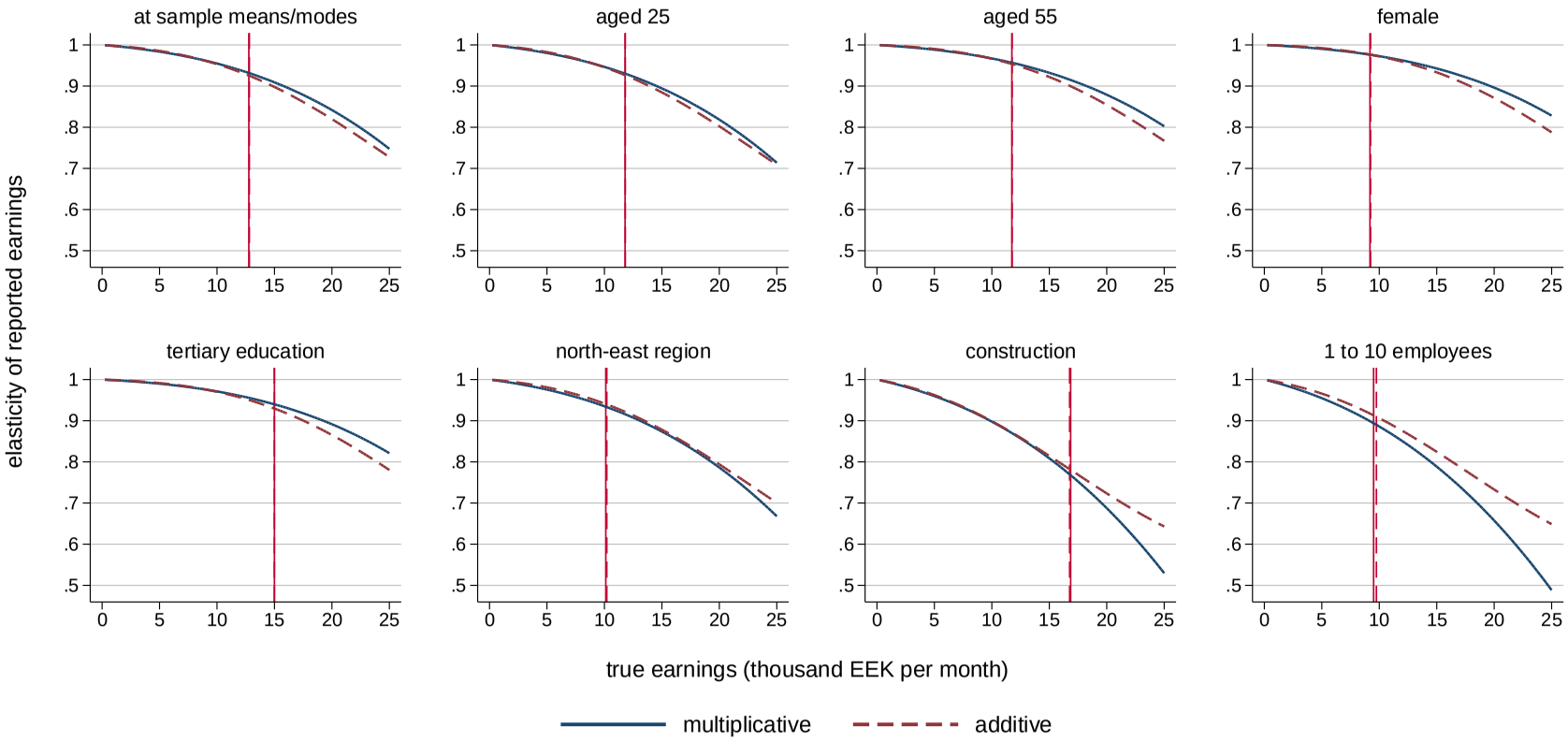

Table 6: Estimation of tax non-compliance

	Unconstrained sector		Whole sample	
	Multiplicative	Additive	Multiplicative	Additive
(a) Proportion of sample, %				
no income	0.5	0.2	0.7	0.4
full evaders	3.1	3.5	2.4	2.7
part evaders	28.2	22.9	21.8	17.7
compliant	68.2	73.4	75.1	79.2
(b) Undeclared earnings as a share of total gross true earnings, %				
All	15.4	15.8	12.1	12.5
Decile 1	23.8	17.2	17.4	12.9
Decile 2	12.2	12.7	9.7	10.2
Decile 3	13.7	11.7	11.1	9.4
Decile 4	12.8	10.6	10.2	8.3
Decile 5	11.4	11.1	8.9	8.8
Decile 6	14.0	16.0	10.1	11.6
Decile 7	12.4	11.3	8.8	7.8
Decile 8	13.1	15.1	9.4	10.8
Decile 9	15.6	16.7	12.5	13.4
Decile 10	19.4	20.6	16.9	18.0
	N = 3,093		N = 4,006	

Notes: deciles are constructed on the basis of estimated gross true earnings using the whole estimation sample.

Figure 6: Tax evasion and measurement error by decile groups

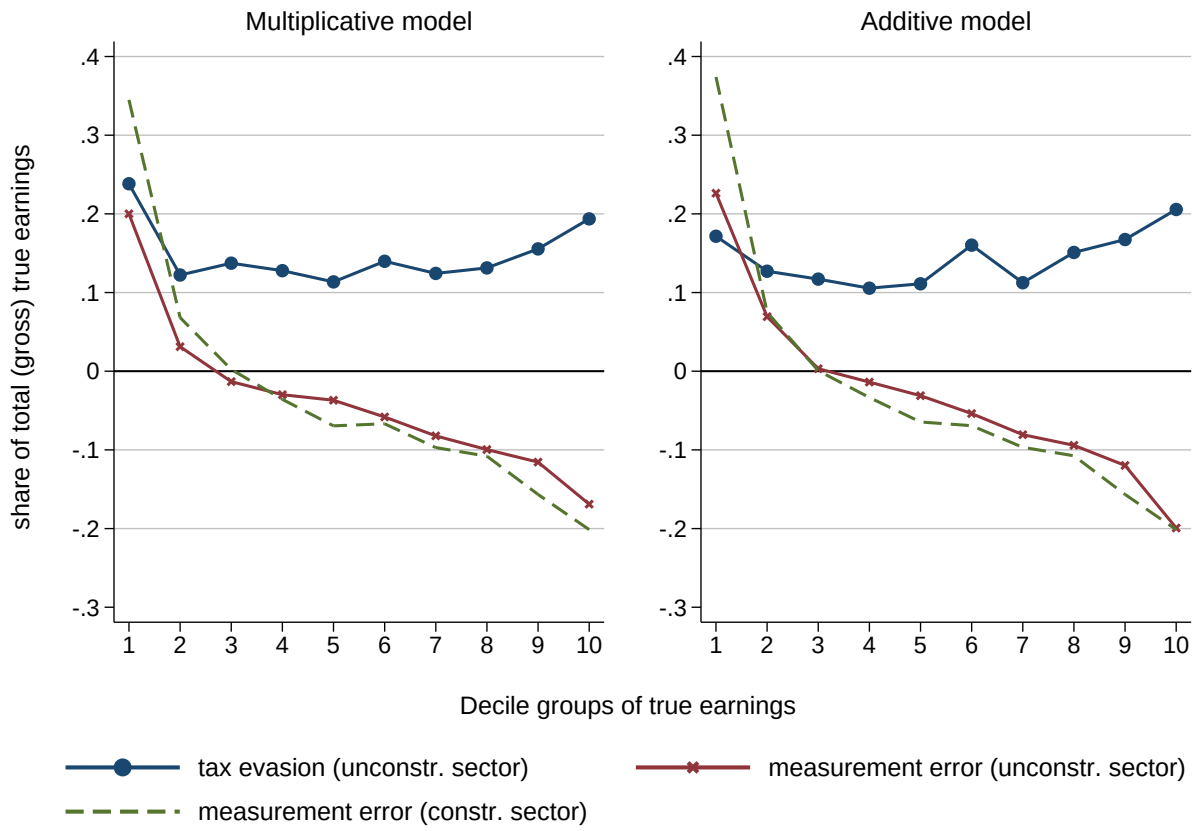

Table 7: Sensitivity analysis: selected parameter estimates for alternative multiplicative model specifications

		Multiplicative models															
	base	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
p	0.993	0.990	0.992	0.996	0.985	0.993	0.990	0.992	0.993	0.993	0.993	0.994	0.992	0.992		0.968	0.992
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)		(0.003)	(0.003)
θ^r	-0.024	-0.020	-0.020	-0.024	-0.030	-0.025	-0.023	-0.020	-0.021	-0.023	-0.028	-0.026	0.016	0.016	-0.010		-0.018
	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.012)	(0.012)	(0.002)		(0.004)
θ^s (unconstr.)	0.689	0.687	0.715	0.693	0.651	0.677	0.679	0.754	0.694	0.681	0.630	0.621	0.000		0.676		0.671
	(0.018)	(0.016)	(0.016)	(0.021)	(0.020)	(0.020)	(0.019)	(0.016)	(0.019)	(0.016)	(0.022)	(0.022)	(.)		(0.020)		(0.026)
θ^s (constrained)	0.642	0.584	0.657	0.644	0.426	0.613	0.637	0.721	0.644	0.681	0.571	0.556	0.000		0.646	0.353	0.594
	(0.025)	(0.034)	(0.024)	(0.029)	(0.036)	(0.036)	(0.025)	(0.024)	(0.025)	(0.016)	(0.028)	(0.029)	(.)		(0.026)	(0.021)	(0.044)
θ_0^s	1.113	1.039	1.159	1.304	0.927	1.050	1.060	1.231	1.116	1.150	0.970	0.941	0.000			0.778	1.150
	(0.084)	(0.078)	(0.081)	(0.084)	(0.089)	(0.092)	(0.080)	(0.094)	(0.087)	(0.082)	(0.084)	(0.101)	(.)			(0.070)	(0.109)
β_0^T x constrained	0.003	-0.008	0.012	-0.024	-0.110	0.026	-0.001	-0.011	0.000	-0.007	0.009	0.014	0.059	0.059	-0.006		0.024
	(0.024)	(0.025)	(0.024)	(0.025)	(0.027)	(0.027)	(0.024)	(0.024)	(.)	(0.023)	(0.024)	(0.025)	(0.028)	(0.028)	(0.024)		(0.031)
β_0^s x constrained	0.062	0.186	0.094	0.072	0.499	0.114	0.056	0.041	0.066	0.000	0.095	0.111	-0.064		0.034		0.127
	(0.055)	(0.066)	(0.053)	(0.063)	(0.069)	(0.075)	(0.055)	(0.060)	(0.056)	(.)	(0.059)	(0.060)	(0.015)		(0.060)		(0.094)
σ_T (unconstr.)	0.474	0.482	0.486	0.473	0.490	0.478	0.474	0.469	0.446	0.479	0.476	0.473	0.424	0.424	0.458		0.470
	(0.015)	(0.013)	(0.014)	(0.017)	(0.017)	(0.016)	(0.016)	(0.015)	(0.013)	(0.016)	(0.014)	(0.014)	(0.013)	(0.013)	(0.017)		(0.021)
σ_T (constrained)	0.354	0.427	0.375	0.355	0.502	0.386	0.353	0.356	0.446	0.354	0.353	0.351	0.353	0.353	0.354	0.554	0.389
	(0.014)	(0.023)	(0.015)	(0.016)	(0.027)	(0.017)	(0.014)	(0.014)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.016)	(0.029)
σ_r	0.583	0.630	0.623	0.591	0.530	0.593	0.720	0.538	0.600	0.551	0.623	0.617	0.725	0.725	0.397		0.608
	(0.035)	(0.037)	(0.033)	(0.036)	(0.036)	(0.038)	(0.051)	(0.031)	(0.036)	(0.033)	(0.040)	(0.039)	(0.095)	(0.095)	(0.017)		(0.054)
σ_s (unconstr.)	0.247	0.273	0.262	0.239	0.257	0.249	0.251	0.248	0.253	0.242	0.245	0.242	0.429		0.248		0.251
	(0.008)	(0.007)	(0.008)	(0.009)	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.006)		(0.008)		(0.011)
σ_s (constrained)	0.233	0.269	0.249	0.236	0.321	0.260	0.233	0.245	0.233	0.242	0.224	0.223	0.354		0.231	0.356	0.262
	(0.012)	(0.012)	(0.011)	(0.013)	(0.012)	(0.017)	(0.012)	(0.012)	(0.012)	(0.007)	(0.012)	(0.013)	(0.010)		(0.012)	(0.007)	(0.021)
AIC	39,017	47,594	44,521	34,220	40,715	39,285	39,485	39,213	39,112	39,032	38,855	38,734	41,917	21,369	37,299	42,258	.
BIC	39,741	48,340	45,259	34,929	41,433	40,009	39,951	39,836	39,824	39,737	39,787	39,868	42,622	21,948	38,007	42,674	.
N total	4,006	4,853	4,545	3,515	4,006	4,006	4,016	4,006	4,006	4,006	4,006	4,006	4,006	4,006	3,881	4,016	4,006
N constrained	3,093	3,807	3,558	2,742	1,958	2,816	3,100	3,093	3,093	3,093	3,093	3,093	3,093	3,093	2,980	0	3,093

Notes: robust standard errors are shown in parentheses under parameter point estimates. Alternative model specifications as follows. **Alternative samples:** (1) include those working part-time or working less than 12 months, (2) survey earnings reported for 12 months, (3) exclude those with survey self-employment income or survey earnings reported in gross terms. **Alternative definitions for the constrained sector:** (4) include private sector workers in large firms (50+ employees), (5) include private sector workers in utilities, public administration, education and health. **Alternative set of co-variates or constraints:** (6) no co-variates in the register income (y^r) equation, (7) no co-variates in the survey income (y^s) equation, (8) common parameters (intercept, σ_T^2) for the constr./unconstr. sector in the true income (y^T) equation, (9) common parameters (intercept, θ^s, σ_s^2) for the constr./unconstr. sector in the survey income (y^s) equation, (10) extended co-variates in the survey income (y^s) equation, (11) same co-variates in all earnings equations, (12) true income omitted among covariates in the survey income equation ($\theta^s = 0$). **Alternative model specifications:** (13) partial model without the survey income (y^s) equation, (14) limit to those with positive earnings in both sources, i.e. set A_{rs} only, (15) everyone assumed constrained, i.e. no evasion. **Other:** (16) survey design (weights, clustering) taken into account.

Table 8: Sensitivity analysis: selected parameter estimates for alternative additive model specifications

		Additive models															
	base	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
p	0.996	0.995	0.989	0.990	0.988	0.996	0.988	0.996	0.989	0.996	0.996	0.989	0.991	0.991		0.968	0.994
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)		(0.003)	(0.002)
θ^r	0.300	0.385	0.519	0.500	0.048	0.268	0.465	0.353	0.618	0.283	0.229	0.603	1.090	1.090	0.541		0.433
	(0.108)	(0.078)	(0.057)	(0.082)	(0.116)	(0.120)	(0.074)	(0.084)	(0.104)	(0.099)	(0.150)	(0.116)	(0.080)	(0.080)	(0.037)		(0.099)
θ^s (unconstr.)	0.653	0.642	0.636	0.624	0.610	0.640	0.597	0.712	0.617	0.652	0.598	0.545	0.000		0.618		0.624
	(0.018)	(0.016)	(0.019)	(0.024)	(0.018)	(0.019)	(0.023)	(0.018)	(0.022)	(0.017)	(0.021)	(0.023)	(.)		(0.024)		(0.029)
θ^s (constrained)	0.642	0.584	0.724	0.695	0.425	0.613	0.708	0.724	0.715	0.652	0.571	0.625	0.000		0.647	0.353	0.616
	(0.026)	(0.035)	(0.026)	(0.031)	(0.037)	(0.037)	(0.027)	(0.025)	(0.027)	(0.017)	(0.028)	(0.030)	(.)		(0.026)	(0.021)	(0.052)
θ_0^s	1.129	1.051	2.035	2.260	0.922	1.060	1.954	1.269	2.004	1.125	0.976	1.790	0.000			0.778	1.357
	(0.099)	(0.108)	(0.139)	(0.122)	(0.114)	(0.108)	(0.117)	(0.119)	(0.127)	(0.093)	(0.090)	(0.133)	(.)			(0.070)	(0.269)
β_0^T x constrained	0.017	0.002	0.016	-0.022	-0.094	0.039	-0.007	-0.002	0.000	0.005	0.031	0.041	0.060	0.060	-0.022		0.032
	(0.024)	(0.025)	(0.025)	(0.026)	(0.027)	(0.028)	(0.028)	(0.024)	(.)	(0.024)	(0.024)	(0.024)	(0.027)	(0.027)	(0.027)		(0.031)
β_0^s x constrained	-0.021	0.090	-0.209	-0.179	0.409	0.030	-0.261	-0.061	-0.249	0.000	0.016	-0.220	-0.064		-0.078		-0.022
	(0.057)	(0.070)	(0.060)	(0.074)	(0.072)	(0.078)	(0.065)	(0.063)	(0.064)	(.)	(0.060)	(0.066)	(0.015)		(0.062)		(0.121)
σ_T (unconstr.)	0.478	0.494	0.488	0.465	0.515	0.484	0.476	0.473	0.434	0.484	0.480	0.455	0.423	0.423	0.470		0.476
	(0.019)	(0.016)	(0.017)	(0.023)	(0.020)	(0.021)	(0.021)	(0.019)	(0.016)	(0.020)	(0.019)	(0.017)	(0.014)	(0.014)	(0.021)		(0.026)
σ_T (constrained)	0.354	0.428	0.376	0.357	0.502	0.386	0.354	0.357	0.434	0.355	0.353	0.350	0.353	0.353	0.356	0.554	0.391
	(0.014)	(0.023)	(0.015)	(0.016)	(0.027)	(0.016)	(0.014)	(0.014)	(0.016)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.016)	(0.029)
σ_r	8.553	7.498	5.629	5.713	8.395	8.701	6.346	7.636	5.294	8.218	9.697	5.882	3.670	3.670	3.895		7.541
	(0.944)	(0.658)	(0.493)	(0.500)	(0.762)	(1.098)	(0.681)	(0.782)	(0.437)	(0.936)	(1.105)	(0.552)	(0.563)	(0.563)	(0.379)		(0.969)
σ_s (unconstr.)	0.254	0.282	0.275	0.248	0.259	0.255	0.262	0.254	0.267	0.245	0.253	0.261	0.429		0.257		0.259
	(0.008)	(0.007)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.007)	(0.008)	(0.009)	(0.006)		(0.008)		(0.010)
σ_s (constrained)	0.233	0.269	0.266	0.248	0.321	0.260	0.249	0.245	0.249	0.245	0.224	0.239	0.354		0.231	0.356	0.263
	(0.012)	(0.012)	(0.013)	(0.015)	(0.012)	(0.017)	(0.013)	(0.012)	(0.014)	(0.007)	(0.012)	(0.014)	(0.010)		(0.012)	(0.007)	(0.021)
AIC	39,189	47,774	44,749	34,322	40,804	39,445	39,637	39,391	39,297	39,206	39,036	38,948	41,913	21,365	37,409	42,258	.
BIC	39,913	48,520	45,488	35,031	41,522	40,169	40,103	40,014	40,008	39,911	39,967	40,081	42,618	21,945	38,117	42,674	.
N total	4,006	4,853	4,545	3,515	4,006	4,006	4,016	4,006	4,006	4,006	4,006	4,006	4,006	4,006	3,881	4,016	4,006
N constrained	3,093	3,807	3,558	2,742	1,958	2,816	3,100	3,093	3,093	3,093	3,093	3,093	3,093	3,093	2,980	0	3,093

Notes: robust standard errors are shown in parentheses under parameter point estimates. Alternative model specifications as follows. **Alternative samples:** (1) include those working part-time or working less than 12 months, (2) survey earnings reported for 12 months, (3) exclude those with survey self-employment income or survey earnings reported in gross terms. **Alternative definitions for the constrained sector:** (4) include private sector workers in large firms (50+ employees), (5) include private sector workers in utilities, public administration, education and health. **Alternative set of co-variates or constraints:** (6) no co-variates in the register income (y^r) equation, (7) no co-variates in the survey income (y^s) equation, (8) common parameters (intercept, σ_T^2) for the constr./unconstr. sector in the true income (y^T) equation, (9) common parameters (intercept, θ^s, σ_s^2) for the constr./unconstr. sector in the survey income (y^s) equation, (10) extended co-variates in the survey income (y^s) equation, (11) same co-variates in all earnings equations, (12) true income omitted among covariates in the survey income equation ($\theta^s = 0$). **Alternative model specifications:** (13) partial model without the survey income (y^s) equation, (14) limit to those with positive earnings in both sources, i.e. set A_{rs} only, (15) everyone assumed constrained, i.e. no evasion. **Other:** (16) survey design (weights, clustering) taken into account.

Table 9: Sensitivity analysis: estimation of tax compliance (whole sample)

							Multiplicative models										
	base	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(a) Proportion of sample, %																	
no income	0.7	0.9	0.8	0.4	1.5	0.7	1.0	0.8	0.7	0.7	0.7	0.6	0.8	0.8	0.0	3.2	0.8
full evaders	2.4	3.1	3.2	2.7	1.6	2.4	2.2	2.4	2.4	2.4	2.4	2.5	2.3	2.3	0.0	0.0	2.5
part evaders	21.8	22.9	24.0	22.3	15.5	20.4	20.7	24.5	21.2	23.5	19.8	19.6	14.9	14.9	28.6	0.0	22.0
compliant	75.1	73.1	71.9	74.6	81.4	76.5	76.1	72.3	75.6	73.3	77.1	77.3	82.0	82.0	71.4	96.8	74.7
(b) Undeclared earnings as a share of total gross true earnings, %																	
All	12.1	12.6	13.6	13.0	8.9	11.8	11.5	13.1	11.4	13.0	11.3	11.3	6.3	6.3	10.1	0.0	11.7
Decile 1	17.4	26.7	25.2	16.2	11.9	16.5	19.9	17.2	17.4	17.6	16.4	17.2	14.5	14.5	11.3	0.0	17.0
Decile 2	9.7	14.0	15.3	11.6	6.5	9.7	10.6	11.3	10.4	10.4	12.0	11.3	16.2	16.2	9.8	0.0	12.4
Decile 3	11.1	14.8	13.5	11.2	8.1	10.5	10.0	9.7	11.1	11.5	9.4	9.7	17.6	17.6	10.0	0.0	11.2
Decile 4	10.2	11.2	12.5	10.3	6.9	9.3	9.1	10.7	10.4	9.6	11.2	9.9	12.4	12.4	8.8	0.0	10.6
Decile 5	8.9	10.8	10.1	9.3	6.2	9.1	7.6	10.6	9.3	9.8	8.1	9.5	12.4	12.4	10.4	0.0	11.2
Decile 6	10.1	9.3	10.6	10.3	7.0	9.4	9.8	10.3	10.0	9.5	9.9	9.6	7.6	7.6	8.8	0.0	9.5
Decile 7	8.8	11.0	11.2	10.2	6.9	8.4	8.7	10.4	8.8	10.1	8.9	8.8	6.0	6.0	8.5	0.0	9.6
Decile 8	9.4	9.7	10.0	8.8	7.1	9.2	8.4	10.5	9.2	10.0	8.0	7.7	3.8	3.8	8.7	0.0	11.0
Decile 9	12.5	11.8	13.4	13.7	9.7	12.2	11.9	14.4	12.0	13.5	10.9	11.3	1.8	1.8	11.4	0.0	12.8
Decile 10	16.9	15.4	17.4	18.8	12.5	16.5	15.9	17.7	14.0	18.7	15.3	15.2	0.8	0.8	11.4	0.0	12.9
									Additive models								
	base	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(a) Proportion of sample, %																	
no income	0.4	0.5	1.1	1.0	1.2	0.4	1.2	0.4	1.1	0.4	0.4	1.1	0.9	0.9	0.0	3.2	0.6
full evaders	2.7	3.5	3.0	2.1	1.9	2.7	2.0	2.7	2.0	2.7	2.7	2.0	2.2	2.2	0.0	0.0	2.7
part evaders	17.7	20.0	26.2	24.1	11.5	16.6	24.0	21.1	23.5	19.3	14.6	18.7	20.8	20.8	31.2	0.0	21.4
compliant	79.2	76.0	69.7	72.8	85.4	80.3	72.8	75.8	73.3	77.5	82.3	78.2	76.0	76.0	68.8	96.8	75.3
(b) Undeclared earnings as a share of total gross true earnings, %																	
All	12.5	13.8	14.5	13.1	9.4	12.2	12.3	14.2	11.0	13.7	10.8	8.9	6.2	6.2	13.1	0.0	13.2
Decile 1	12.9	19.2	21.9	12.4	11.4	13.4	16.8	13.0	14.5	13.5	12.9	13.7	24.9	24.9	8.5	0.0	12.6
Decile 2	10.2	13.3	14.2	11.1	5.7	9.0	13.1	10.6	12.9	9.9	10.0	13.0	18.1	18.1	8.7	0.0	10.9
Decile 3	9.4	14.0	15.8	11.7	5.7	9.3	13.7	8.6	12.1	9.6	8.3	11.4	15.1	15.1	8.5	0.0	11.9
Decile 4	8.3	12.1	12.9	10.7	6.4	8.0	10.7	10.3	9.7	9.0	10.3	9.7	10.1	10.1	9.4	0.0	10.7
Decile 5	8.8	9.2	13.8	11.7	5.3	8.6	11.3	10.2	12.9	8.9	6.8	11.2	8.3	8.3	11.2	0.0	11.8
Decile 6	11.6	11.3	13.6	13.0	7.3	10.4	8.5	10.8	11.4	11.2	10.1	9.7	6.5	6.5	10.2	0.0	11.7
Decile 7	7.8	11.9	10.2	11.0	6.2	8.0	9.0	11.2	8.6	9.2	7.5	7.1	4.6	4.6	10.8	0.0	11.2
Decile 8	10.8	10.6	11.7	11.1	7.4	10.4	10.7	12.0	9.7	12.2	9.7	8.7	3.9	3.9	12.5	0.0	10.8
Decile 9	13.4	13.8	14.6	12.9	8.9	12.9	10.7	16.0	10.7	14.0	10.7	6.3	2.6	2.6	13.9	0.0	14.7
Decile 10	18.0	18.0	17.3	16.7	15.8	18.0	16.2	20.3	11.6	20.9	14.8	8.4	2.1	2.1	18.5	0.0	16.3
Sample size	4,006	4,853	4,545	3,515	4,006	4,006	4,016	4,006	4,006	4,006	4,006	4,006	4,006	4,006	3,881	4,016	4,006

Notes: deciles are constructed on the basis of estimated gross true earnings using the whole estimation sample. Alternative model specifications as follows. **Alternative samples:** (1) include those working part-time or working less than 12 months, (2) survey earnings reported for 12 months, (3) exclude those with survey self-employment income or survey earnings reported in gross terms. **Alternative definitions for the constrained sector:** (4) include private sector workers in large firms (50+ employees), (5) include private sector workers in utilities, public administration, education and health. **Alternative set of co-variates or constraints:** (6) no co-variates in the register income (y^r) equation, (7) no co-variates in the survey income (y^s) equation, (8) common parameters (intercept, σ_T^2) for the constr./unconstr. sector in the true income (y^T) equation, (9) common parameters (intercept, θ^s, σ_s^2) for the constr./unconstr. sector in the survey income (y^s) equation, (10) extended co-variates in the survey income (y^s) equation, (11) same co-variates in all earnings equations, (12) true income omitted among covariates in the survey income equation ($\theta^s = 0$). **Alternative model specifications:** (13) partial model without the survey income (y^s) equation, (14) limit to those with positive earnings in both sources, i.e. set A_{rs} only, (15) everyone assumed constrained, i.e. no evasion. **Other:** (16) survey design (weights, clustering) taken into account.

Appendices

A Detailed presentation of the model

A.1 The likelihood function

A.1.1 The multiplicative model

Recall from Section 4 that $\Pr(y_i^T > 0) = p$ and $\Pr(y_i^r = 0 | y_i^T = 0) = 1$ by assumption. For the unconstrained employees (U), probability density functions are the following:

$$\begin{aligned}
f_{0s}^U &= f_{0s}^U(y_i^r, y_i^s | x_i) = f_{0s}^U(\text{no earnings}) + f_{0s}^U(\text{full evasion}) \\
&= \Pr(y_i^T = 0) \Pr(y_i^r = 0 | x_i, y_i^T = 0) f(y_i^s | x_i, y_i^T = 0) \\
&\quad + \Pr(y_i^T > 0) \int_0^\infty f(y^T | x_i, y_i^T > 0) \Pr(y_i^r = 0 | x_i, y^T) f(y_i^s | x_i, y^T) \, \mathrm{d}y^T \\
&= (1-p) 1 \frac{1}{\sigma_s y_i^s} \phi \left(\frac{\ln y_i^s - \theta_0^s - x_i \beta^s}{\sigma_s} \right) \\
&\quad + p \int_0^\infty \frac{1}{\sigma_T y^T} \phi \left(\frac{\ln y^T - x_i \beta^T}{\sigma_T} \right) \Phi \left(- \frac{\theta^r y^T + x_i \beta^r}{\sigma_r} \right) \\
&\qquad \cdot \frac{1}{\sigma_s y_i^s} \phi \left(\frac{\ln y_i^s - \theta^s \ln y^T - x_i \beta^s}{\sigma_s} \right) \mathrm{d}y^T
\end{aligned} \tag{A.1}$$

$$\begin{aligned}
f_{rs}^U &= f_{rs}^U(y_i^r, y_i^s | x_i) = f_{rs}^U(\text{partial evasion}) + f_{rs}^U(\text{full compliance}) \\
&= \Pr(y_i^T > 0) f(y_i^T = y_i^r | x_i, y_i^T > 0) \Pr(y_i^r = y_i^T | x_i, y_i^T) f(y_i^s | x_i, y_i^T = y_i^r) \\
&\quad + \Pr(y_i^T > 0) \int_{y_i^r}^\infty f(y^T | x_i, y_i^T > 0) f(y_i^r | x_i, y^T) f(y_i^s | x_i, y^T) \, \mathrm{d}y^T \\
&= p \frac{1}{\sigma_T y_i^r} \phi \left(\frac{\ln y_i^r - x_i \beta^T}{\sigma_T} \right) \left[1 - \Phi \left(\frac{1 - \theta^r y_i^r - x_i \beta^r}{\sigma_r} \right) \right] \frac{1}{\sigma_s y_i^s} \phi \left(\frac{\ln y_i^s - \theta^s \ln y_i^r - x_i \beta^s}{\sigma_s} \right) \\
&\quad + p \int_{y_i^r}^\infty \frac{1}{\sigma_T y^T} \phi \left(\frac{\ln y^T - x_i \beta^T}{\sigma_T} \right) \frac{1}{\sigma_r y^T} \phi \left(\frac{y_i^r / y^T - \theta^r y^T - x_i \beta^r}{\sigma_r} \right) \\
&\qquad \cdot \frac{1}{\sigma_s y_i^s} \phi \left(\frac{\ln y_i^s - \theta^s \ln y^T - x_i \beta^s}{\sigma_s} \right) \mathrm{d}y^T
\end{aligned} \tag{A.2}$$

In the case of constrained employees (C), $\Pr(y_i^r = y_i^T) = 1$, and their probability density functions simplify to:

$$f_{0s}^C = (1-p) \frac{1}{\sigma_s y_i^s} \phi \left(\frac{\ln y_i^s - \theta_0^s - x_i \beta^s}{\sigma_s} \right) \tag{A.3}$$

$$f_{rs}^C = p \frac{1}{\sigma_T y_i^r} \phi \left(\frac{\ln y_i^r - x_i \beta^T}{\sigma_T} \right) \frac{1}{\sigma_s y_i^s} \phi \left(\frac{\ln y_i^s - \theta^s \ln y_i^r - x_i \beta^s}{\sigma_s} \right) \tag{A.4}$$

A.1.2 The additive model

For the unconstrained employees (U), probability density functions are the following. First

$$
\begin{aligned}
f_{0s}^{U} = (1-p)1 & \frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta_0^s - x_i\beta^s}{\sigma_s}\right) \\
& + p\int_0^\infty \frac{1}{\sigma_T y^T}\phi\left(\frac{\ln y^T - x_i\beta^T}{\sigma_T}\right)\Phi\left(-\frac{\theta^r y^T + x_i\beta^r}{\sigma_r}\right) \\
& \cdot \frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s \ln y^T - x_i\beta^s}{\sigma_s}\right)\mathrm{d}y^T
\end{aligned} \tag{A.5}
$$

which is the same as for the multiplicative model (equation A.1), and then

$$
\begin{aligned}
f_{rs}^{U} = & p\frac{1}{\sigma_T y_i^r}\phi\left(\frac{\ln y_i^r - x_i\beta^T}{\sigma_T}\right)\left[1-\Phi\left(\frac{(1-\theta^r)y_i^r - x_i\beta^r}{\sigma_r}\right)\right]\frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s \ln y_i^r - x_i\beta^s}{\sigma_s}\right) \\
& + p\int_{y_i^r}^\infty \frac{1}{\sigma_T y^T}\phi\left(\frac{\ln y^T - x_i\beta^T}{\sigma_T}\right)\frac{1}{\sigma_r}\phi\left(\frac{y_i^r - \theta^r y^T - x_i\beta^r}{\sigma_r}\right) \\
& \cdot \frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s \ln y^T - x_i\beta^s}{\sigma_s}\right)\mathrm{d}y^T
\end{aligned} \tag{A.6}
$$

For the constrained employees (C), both probability density functions are the same as with the multiplicative model, see equation (A.3) and (A.4).

A.2 Log likelihood function with the application of Gauss-Hermite quadrature

A.2.1 The multiplicative model

First, rewrite the integral for f_{rs}^U in equation (A.1) by making the substitution $u = \frac{\ln y^T}{\sqrt{2}\sigma_T}$, implying $y^T = \exp(\sqrt{2}\sigma_T u)$ and $\mathrm{d}y^T = \sqrt{2}\sigma_T \exp(\sqrt{2}\sigma_T u)\,\mathrm{d}u$:

$$
\begin{aligned}
& \int_0^\infty \frac{1}{\sigma_T y^T}\phi\left(\frac{\ln y^T - x_i\beta^T}{\sigma_T}\right)\Phi\left(-\frac{\theta^r y^T + x_i\beta^r}{\sigma_r}\right)\frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s \ln y^T - x_i\beta^s}{\sigma_s}\right)\mathrm{d}y^T \\
& = \int_0^\infty \frac{1}{\sigma_T \exp(\sqrt{2}\sigma_T u)}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma_T^2}(\sqrt{2}\sigma_T u - x_i\beta^T)^2\right]\Phi\left(-\frac{\theta^r \exp(\sqrt{2}\sigma_T u) + x_i\beta^r}{\sigma_r}\right) \\
& \quad \cdot \frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s\sqrt{2}\sigma_T u - x_i\beta^s}{\sigma_s}\right)\sqrt{2}\sigma_T \exp(\sqrt{2}\sigma_T u)\,\mathrm{d}u \\
& = \int_0^\infty \frac{1}{\sigma_T}\frac{1}{\sqrt{2\pi}}\exp\left[-u^2 + \frac{\sqrt{2}u}{\sigma_T}x_i\beta^T - \frac{1}{2\sigma_T^2}(x_i\beta^T)^2\right]\Phi(\cdot)\frac{1}{\sigma_s y_i^s}\phi(\cdot)\sqrt{2}\sigma_T\,\mathrm{d}u \\
& = \frac{1}{\sigma_T}\phi\left(\frac{x_i\beta^T}{\sigma_T}\right)\int_0^\infty \exp(-u^2)\exp\left(\frac{\sqrt{2}u}{\sigma_T}x_i\beta^T\right)\Phi(\cdot)\frac{1}{\sigma_s y_i^s}\phi(\cdot)\sqrt{2}\sigma_T\,\mathrm{d}u
\end{aligned} \tag{A.7}
$$

This semi-infinite integral can be approximated using the Gauss-Hermite quadrature rule

$$\int_0^\infty \exp\left(-u^2\right) f(u)\,\mathrm{d}u \simeq \sum_{j=1}^{n} \omega_j f(\tau_j) \tag{A.8}$$

as follows

$$\begin{aligned}\sum_{j=1}^{n} \omega_j \exp\left(\frac{\sqrt{2}\tau_j}{\sigma^T} x_i\beta^T\right) \Phi\left(-\frac{\theta^r \exp(\sqrt{2}\sigma_T\tau_j) + x_i\beta^r}{\sigma_r}\right) \\ \cdot \frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s\sqrt{2}\sigma_T\tau_j - x_i\beta^s}{\sigma_s}\right)\sqrt{2}\sigma_T\end{aligned} \tag{A.9}$$

using the nodes τ_j and the weights ω_j as calculated in Steen et al. (1969). Finally, the log likelihood of observation i in set A_{0s} is:

$$\begin{aligned}\ln f_{0s}^U = \ln\Bigg\{(1-p)\frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta_0^s - x_i\beta^s}{\sigma_s}\right) + p\frac{1}{\sigma_T}\phi\left(\frac{x_i\beta^T}{\sigma_T}\right)\sum_{j=1}^{n}\omega_j \exp\left(\frac{\sqrt{2}\tau_j}{\sigma_T}x_i\beta^T\right) \\ \cdot\Phi\left(-\frac{\theta^r\exp(\sqrt{2}\sigma_T\tau_j) + x_i\beta^r}{\sigma_r}\right)\frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s\sqrt{2}\sigma_T\tau_j - x_i\beta^s}{\sigma_s}\right)\sqrt{2}\sigma_T\Bigg\}\end{aligned} \tag{A.10}$$

In analog the integral for f_{rs}^U in (A.2) is rewritten by making the substitution $u = \frac{\ln y^T - \ln y^r}{\sqrt{2}\sigma_T}$, implying $y^T = \exp(\sqrt{2}\sigma_T u + \ln y^r)$ and $\mathrm{d}y^T = \sqrt{2}\sigma_T \exp(\sqrt{2}\sigma_T u + \ln y^r)\,\mathrm{d}u$:

$$\begin{aligned}&\int_{y_i^r}^\infty \frac{1}{\sigma_T y^T}\phi\left(\frac{\ln y^T - x_i\beta^T}{\sigma_T}\right)\frac{1}{\sigma_r y^T}\phi\left(\frac{y_i^r/y^T - \theta^r y^T - x_i\beta^r}{\sigma_r}\right) \\ &\quad\cdot\frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s\ln y^T - x_i\beta^s}{\sigma_s}\right)\mathrm{d}y^T \\ &= \int_0^\infty \frac{1}{\sigma_T}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma_T^2}(\sqrt{2}\sigma_T u + \ln y^r - x_i\beta^T)^2\right]\frac{\phi(\cdot)}{\sigma_r\exp(\sqrt{2}\sigma_T u + \ln y^r)}\frac{\phi(\cdot)}{\sigma_s y_i^s}\sqrt{2}\sigma_T\,\mathrm{d}u \\ &= \frac{1}{\sigma_T}\phi\left(\frac{\ln y^r - x_i\beta^T}{\sigma_T}\right)\int_0^\infty \exp(-u^2)\exp\left(-\frac{\sqrt{2}u}{\sigma_T}(\ln y^r - x_i\beta^T)\right) \\ &\quad\cdot\frac{\phi(\cdot)}{\sigma_r\exp(\sqrt{2}\sigma_T u + \ln y^r)}\frac{\phi(\cdot)}{\sigma_s y_i^s}\sqrt{2}\sigma_T\,\mathrm{d}u \\ &\simeq \frac{1}{\sigma_T}\phi\left(\frac{\ln y^r - x_i\beta^T}{\sigma_T}\right)\sum_{j=1}^{n}\omega_j\exp\left(-\frac{\sqrt{2}\tau_j}{\sigma_T}(\ln y^r - x_i\beta^T)\right) \\ &\quad\cdot\frac{\phi(\cdot)}{\sigma_r\exp(\sqrt{2}\sigma_T\tau_j + \ln y^r)}\frac{\phi(\cdot)}{\sigma_s y_i^s}\sqrt{2}\sigma_T\end{aligned} \tag{A.11}$$

Unlike with $\ln f_{0s}^U$, taking the logarithm of f_{rs}^U allows us to separate several terms:

$$\begin{aligned}\ln f_{0s}^U = &\ln p - \ln \sigma_T - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\left(\frac{\ln y_i^r - x_i\beta^T}{\sigma_T}\right)^2 \\ &+ \ln\Bigg\{\frac{1}{y_i^r}\left[1-\Phi\left(\frac{1-\theta^r y_i^r - x_i\beta^r}{\sigma_r}\right)\right]\frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s \ln y_i^r - x_i\beta^s}{\sigma_s}\right) \\ &\quad + \sum_{j=1}^{n}\omega_j \exp\left[-\frac{\sqrt{2}\tau_j}{\sigma_T}(\ln y_i^r - x_i\beta^T)\right]\frac{1}{\sigma_r \exp(\sqrt{2}\sigma_T\tau_j + \ln y_i^r)} \\ &\quad \cdot \phi\left(\frac{y_i^r / \exp(\sqrt{2}\sigma_T\tau_j + \ln y_i^r) - \theta^r \exp(\sqrt{2}\sigma_T\tau_j + \ln y_i^r) - x_i\beta^r}{\sigma_r}\right) \\ &\quad \cdot \frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s(\sqrt{2}\sigma_T\tau_j + \ln y_i^r) - x_i\beta^s}{\sigma_s}\right)\sqrt{2}\sigma_T\Bigg\}\end{aligned} \tag{A.12}$$

A.2.2 The additive model

The log likelihood of an observation i in set A_{0s} is identical to (A.10):

$$\begin{aligned}\ln f_{0s}^U = \ln\Bigg\{&(1-p)\frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta_0^s - x_i\beta^s}{\sigma_s}\right) + p\frac{1}{\sigma_T}\phi\left(\frac{x_i\beta^T}{\sigma_T}\right)\sum_{j=1}^{n}\omega_j \exp\left(\frac{\sqrt{2}\tau_j}{\sigma_T}x_i\beta^T\right) \\ &\cdot\Phi\left(-\frac{\theta^r \exp(\sqrt{2}\sigma_T\tau_j) + x_i\beta^r}{\sigma_r}\right)\frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s\sqrt{2}\sigma_T\tau_j - x_i\beta^s}{\sigma_s}\right)\sqrt{2}\sigma_T\Bigg\}\end{aligned} \tag{A.13}$$

In analog the integral for f_{rs}^U in (A.6) is rewritten by making the substitution $u = \frac{\ln y^T - \ln y^r}{\sqrt{2}\sigma_T}$:

$$\begin{aligned}&\int_{y_i^r}^{\infty}\frac{1}{\sigma_T y^T}\phi\left(\frac{\ln y^T - x_i\beta^T}{\sigma_T}\right)\frac{1}{\sigma_r}\phi\left(\frac{y_i^r - \theta^r y^T - x_i\beta^r}{\sigma_r}\right)\frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s \ln y^T - x_i\beta^s}{\sigma_s}\right)\mathrm{d}y^T \\ &= \int_0^{\infty}\frac{1}{\sigma_T}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma_T^2}(\sqrt{2}\sigma_T u + \ln y^r - x_i\beta^T)^2\right]\frac{1}{\sigma_r}\phi(\cdot)\frac{1}{\sigma_s y_i^s}\phi(\cdot)\sqrt{2}\sigma_T\,\mathrm{d}u \\ &\simeq \frac{1}{\sigma_T}\phi\left(\frac{\ln y_i^r - x_i\beta^T}{\sigma_T}\right)\sum_{j=1}^{n}\omega_j \exp\left[-\frac{\sqrt{2}\tau_j}{\sigma_T}(\ln y_i^r - x_i\beta^T)\right]\frac{1}{\sigma_r}\phi(\cdot)\frac{1}{\sigma_s y_i^s}\phi(\cdot)\sqrt{2}\sigma_T\end{aligned} \tag{A.14}$$

The log likelihood of an observation i in set A_{rs} is:

$$\begin{aligned}\ln f_{rs}^{U} = &\ln p - \ln \sigma_T - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\left(\frac{\ln y_i^r - x_i\beta^T}{\sigma_T}\right)^2 \\ &+ \ln\Bigg\{\frac{1}{y_i^r}\left[1-\Phi\left(\frac{(1-\theta^r)y_i^r - x_i\beta^r}{\sigma_r}\right)\right]\frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s \ln y_i^r - x_i\beta^s}{\sigma_s}\right) \\ &+ \sum_{j=1}^{n}\omega_j \exp\left[-\frac{\sqrt{2}\tau_j}{\sigma_T}(\ln y_i^r - x_i\beta^T)\right]\frac{1}{\sigma_r}\phi\left(\frac{y_i^r - \theta^r \exp(\sqrt{2}\sigma_T\tau_j + \ln y_i^r) - x_i\beta^r}{\sigma_r}\right) \\ &\cdot \frac{1}{\sigma_s y_i^s}\phi\left(\frac{\ln y_i^s - \theta^s(\sqrt{2}\sigma_T\tau_j + \ln y_i^r) - x_i\beta^s}{\sigma_s}\right)\sqrt{2}\sigma_T\Bigg\} \qquad \text{(A.15)}\end{aligned}$$

B Model interpretation

B.1 The multiplicative model

B.1.1 Expected value of y^r, conditional on true employment

Let us define $a = -(\theta^r y^T + x\beta^r)/\sigma_r$ and $b = (1 - \theta^r y^T - x\beta^r)/\sigma_r$, omitting the subscript i. For any positive y^T, the probability of full evasion is $\Phi(a)$, the probability of full compliance $[1-\Phi(b)]$ and the probability of partial evasion $[\Phi(b)-\Phi(a)]$. The expected value of the *truncated* reported earnings is (for any $y^T > 0$):[41]

$$\begin{aligned}\mathrm{E}\left[y^r \middle| 0 < y^r < y^T, x, y^T\right] =& \mathrm{E}\left[r^* y^T \middle| 0 < r^* < 1, x, y^T\right] \\ =& y^T(\theta^r y^T + x\beta^r) + y^T\sigma_r \mathrm{E}\left[\frac{\varepsilon^r}{\sigma_r}\middle| a < \frac{\varepsilon^r}{\sigma_r} < b, x, y^T\right] \\ =& y^T\sigma_r(-a) + y^T\sigma_r\int_a^b \left(\frac{\varepsilon^r}{\sigma_r}\right)\frac{f\left(\frac{\varepsilon^r}{\sigma_r}\middle| x, y^T\right)}{\Pr\left(a < \frac{\varepsilon^r}{\sigma_r} < b \middle| x, y^T\right)}\,\mathrm{d}\frac{\varepsilon^r}{\sigma_r} \\ =& y^T\sigma_r(-a) + y^T\sigma_r\int_a^b \left(\frac{\varepsilon^r}{\sigma_r}\right)\frac{\phi\left(\frac{\varepsilon^r}{\sigma_r}\right)}{\Phi(b)-\Phi(a)}\,\mathrm{d}\frac{\varepsilon^r}{\sigma_r} \\ =& y^T\sigma_r(-a) + y^T\sigma_r\frac{\phi(a)-\phi(b)}{\Phi(b)-\Phi(a)} \qquad \text{(B.1)}\end{aligned}$$

[41] To solve the integral, note that $\mathrm{d}\phi(x) = -x\phi(x)\,\mathrm{d}x$.

The expected value of the *observed* reported earnings (for any $y^T > 0$):

$$\begin{aligned}
\mathrm{E}(y^r|x, y^T) =& 0 \cdot \Pr(y^r = 0|x, y^T) + y^T \cdot \Pr(y^r = y^T|x, y^T) \\
&+ \mathrm{E}\left[y^r \middle| 0 < y^r < y^T, x, y^T\right] \cdot \Pr(0 < y^r < y^T|x, y^T) \\
=& y^T[1 - \Phi(b)] + \left[y^T\sigma_r(-a) + y^T\sigma_r\frac{\phi(a) - \phi(b)}{\Phi(b) - \Phi(a)}\right][\Phi(b) - \Phi(a)] \\
=& y^T\Phi(-b) + y^T\sigma_r(-a)[\Phi(b) - \Phi(a)] + y^T\sigma_r[\phi(a) - \phi(b)] \qquad \text{(B.2)}
\end{aligned}$$

B.1.2 Partial effects for $\mathrm{E}(y^r)$

If x_k is a continuous variable then (for any $y^T > 0$):

$$\begin{aligned}
\frac{\partial \mathrm{E}(y^r|x, y^T)}{\partial x_k} &= y^T\phi(-b)\left(\frac{\beta_k^r}{\sigma_r}\right) + y^T\beta_k^r[\Phi(b) - \Phi(a)] + y^T\sigma_r(-a)\left[\phi(b) - \phi(a)\right]\left(-\frac{\beta_k^r}{\sigma_r}\right) \\
&\quad + y^T\sigma_r\left[\phi(a)(-a) - \phi(b)(-b)\right]\left(-\frac{\beta_k^r}{\sigma_r}\right) \\
&= y^T\beta_k^r[\Phi(b) - \Phi(a)] + y^T\phi(a)\left(-\frac{\beta_k^r}{\sigma_r}\right)[-\sigma_r(-a) + \sigma_r(-a)] \\
&\quad + y^T\phi(b)\left(-\frac{\beta_k^r}{\sigma_r}\right)[-1 + \sigma_r(-a) - \sigma_r(-b)] \\
&= y^T\beta_k^r[\Phi(b) - \Phi(a)] \qquad \text{(B.3)}
\end{aligned}$$

If x_k is a dichotomous variable (for any $y^T > 0$):

$$\frac{\Delta \mathrm{E}(y^r|x, y^T)}{\Delta x_k} = \mathrm{E}(y^r|x, y^T, x_k = 1) - \mathrm{E}(y^r|x, y^T, x_k = 0) \qquad \text{(B.4)}$$

Finally, differentiate with respect to y^T (for any $y^T > 0$):

$$\begin{aligned}
\frac{\partial \mathrm{E}(y^r|x, y^T)}{\partial y^T} &= \Phi(-b) + y^T\phi(-b)\left(\frac{\theta^r}{\sigma_r}\right) + [\sigma_r(-a) + y^T\theta^r][\Phi(b) - \Phi(a)] \\
&\quad + y^T\sigma_r(-a)\left[\phi(b) - \phi(a)\right]\left(-\frac{\theta^r}{\sigma_r}\right) + \sigma_r[\phi(a) - \phi(b)] \\
&\quad + y^T\sigma_r\left[\phi(a)(-a) - \phi(b)(-b)\right]\left(-\frac{\theta^r}{\sigma_r}\right) \\
&= \Phi(-b) + [\sigma_r(-a) + \theta^r y^T][\Phi(b) - \Phi(a)] + \sigma_r[\phi(a) - \phi(b)] \\
&\quad + \phi(a)\left(-\frac{\theta^r}{\sigma_r}\right)\left[-y^T\sigma_r(-a) + y^T\sigma_r(-a)\right] \\
&\quad + \phi(b)\left(-\frac{\theta^r}{\sigma_r}\right)\left[-y^T + y^T\sigma_r(-a) - y^T\sigma_r(-b)\right] \\
&= \Phi(-b) + [\sigma_r(-a) + \theta^r y^T][\Phi(b) - \Phi(a)] + \sigma_r[\phi(a) - \phi(b)] \qquad \text{(B.5)}
\end{aligned}$$

B.1.3 Elasticity of $\mathrm{E}(y^r)$

Combining equation (B.2) and (B.5), the elasticity of $\mathrm{E}(y^r)$ with respect to y^T (for any $y^T > 0$):

$$\begin{aligned}\frac{\partial \mathrm{E}(y^r|x,y^T)/\partial y^T}{\mathrm{E}(y^r|x,y^T)/y^T} &= \frac{\Phi(-b)+[\sigma_r(-a)+\theta^r y^T][\Phi(b)-\Phi(a)]+\sigma_r[\phi(a)-\phi(b)]}{y^T\Phi(-b)+y^T\sigma_r(-a)[\Phi(b)-\Phi(a)]+y^T\sigma_r[\phi(a)-\phi(b)]}y^T \\ &= 1+\frac{\theta^r y^T[\Phi(b)-\Phi(a)]}{\Phi(-b)+\sigma_r(-a)[\Phi(b)-\Phi(a)]+\sigma_r[\phi(a)-\phi(b)]} \end{aligned} \tag{B.6}$$

B.2 The additive model

B.2.1 Expected value of y^r, conditional on true employment

Define now $a = -(\theta^r y^T + x\beta^r)/\sigma_r$ and $b = (y^T - \theta^r y^T - x\beta^r)/\sigma_r$, omitting again the subscript i. The expected value of the *truncated* reported earnings (for any $y^T > 0$):

$$\begin{aligned}\mathrm{E}\left[y^r\middle|0<y^r<y^T,x,y^T\right] =& \mathrm{E}\left[y^{*r}\middle|0<y^{*r}<y^T,x,y^T\right] \\ =& \theta^r y^T + x\beta^r + \sigma_r \mathrm{E}\left[\frac{\varepsilon^r}{\sigma_r}\middle| a<\frac{\varepsilon^r}{\sigma_r}<b,x,y^T\right] \\ =& \sigma_r(-a)+\sigma_r\int_a^b\left(\frac{\varepsilon^r}{\sigma_r}\right)\frac{f\left(\frac{\varepsilon^r}{\sigma_r}\middle|x,y^T\right)}{\Pr\left(a<\frac{\varepsilon^r}{\sigma_r}<b\middle|x,y^T\right)}\mathrm{d}\frac{\varepsilon^r}{\sigma_r} \\ =& \sigma_r(-a)+\sigma_r\int_a^b\left(\frac{\varepsilon^r}{\sigma_r}\right)\frac{\phi\left(\frac{\varepsilon^r}{\sigma_r}\right)}{\Phi(b)-\Phi(a)}\mathrm{d}\frac{\varepsilon^r}{\sigma_r} \\ =& \sigma_r(-a)+\sigma_r\frac{\phi(a)-\phi(b)}{\Phi(b)-\Phi(a)} \end{aligned} \tag{B.7}$$

The expected value of the *observed* reported earnings (for any $y^T > 0$):

$$\begin{aligned}\mathrm{E}(y^r|x,y^T) =& 0\cdot\Pr(y^r=0|x,y^T)+y^T\cdot\Pr(y^r=y^T|x,y^T) \\ &+\mathrm{E}\left[y^r\middle|0<y^r<y^T,x,y^T\right]\cdot\Pr(0<y^r<y^T|x,y^T) \\ =& y^T[1-\Phi(b)]+\left[\sigma_r(-a)+\sigma_r\frac{\phi(a)-\phi(b)}{\Phi(b)-\Phi(a)}\right][\Phi(b)-\Phi(a)] \\ =& y^T\Phi(-b)+\sigma_r(-a)[\Phi(b)-\Phi(a)]+\sigma_r[\phi(a)-\phi(b)] \end{aligned} \tag{B.8}$$

B.2.2 Partial effects for $\mathrm{E}(y^r)$

If x_k is a continuous variable then (for any $y^T > 0$):

$$\begin{aligned}
\frac{\partial \mathrm{E}(y^r|x, y^T)}{\partial x_k} &= y^T \phi(-b) \left(\frac{\beta_k^r}{\sigma_r}\right) + \beta_k^r[\Phi(b) - \Phi(a)] + \sigma_r(-a)\left[\phi(b) - \phi(a)\right] \left(-\frac{\beta_k^r}{\sigma_r}\right) \\
&\quad + \sigma_r\left[\phi(a)(-a) - \phi(b)(-b)\right] \left(-\frac{\beta_k^r}{\sigma_r}\right) \\
&= \beta_k^r[\Phi(b) - \Phi(a)] + \phi(a) \left(-\frac{\beta_k^r}{\sigma_r}\right) \left[-\sigma_r(-a) + \sigma_r(-a)\right] \\
&\quad + \phi(b) \left(-\frac{\beta_k^r}{\sigma_r}\right) \left[-y^T + \sigma_r(-a) - \sigma_r(-b)\right] \\
&= \beta_k^r[\Phi(b) - \Phi(a)] \qquad \text{(B.9)}
\end{aligned}$$

If x_k is a dichotomous variable then (for any $y^T > 0$):

$$\frac{\Delta \mathrm{E}(y^r|x, y^T)}{\Delta x_k} = \mathrm{E}(y^r|x, y^T, x_k = 1) - \mathrm{E}(y^r|x, y^T, x_k = 0) \qquad \text{(B.10)}$$

Finally, differentiation with respect to y^T (for any $y^T > 0$) yields:

$$\begin{aligned}
\frac{\partial \mathrm{E}(y^r|x, y^T)}{\partial y^T} &= \Phi(-b) + y^T \phi(-b) \left(-\frac{1 - \theta^r}{\sigma_r}\right) + \theta^r[\Phi(b) - \Phi(a)] \\
&\quad + \sigma_r(-a) \left[\phi(b) \left(\frac{1 - \theta^r}{\sigma_r}\right) - \phi(a) \left(-\frac{\theta^r}{\sigma_r}\right)\right] \\
&\quad + \sigma_r \left[\phi(a)(-a) \left(-\frac{\theta^r}{\sigma_r}\right) - \phi(b)(-b) \left(\frac{1 - \theta^r}{\sigma_r}\right)\right] \\
&= \Phi(-b) + \theta^r[\Phi(b) - \Phi(a)] \\
&\quad + \phi(a) \left(-\frac{\theta^r}{\sigma_r}\right) \left[-\sigma_r(-a) + \sigma_r(-a)\right] + \phi(b) \left(\frac{1 - \theta^r}{\sigma_r}\right) \left[-y^T + \sigma_r(-a) - \sigma_r(-b)\right] \\
&= \Phi(-b) + \theta^r[\Phi(b) - \Phi(a)] \qquad \text{(B.11)}
\end{aligned}$$

B.2.3 Elasticity of $\mathrm{E}(y^r)$

Combining equation (B.8) and (B.11), we can express the elasticity of $\mathrm{E}(y^r)$ with respect to y^T (for any $y^T > 0$):

$$\begin{aligned}
\frac{\partial \mathrm{E}(y^r|x, y^T)/\partial y^T}{\mathrm{E}(y^r|x, y^T)/y^T} &= \frac{y^T \Phi(-b) + \theta^r y^T[\Phi(b) - \Phi(a)]}{y^T \Phi(-b) + \sigma_r(-a)[\Phi(b) - \Phi(a)] + \sigma_r[\phi(a) - \phi(b)]} \\
&= 1 - \frac{x\beta^r[\Phi(b) - \Phi(a)] + \sigma_r[\phi(a) - \phi(b)]}{y^T \Phi(-b) + \sigma_r(-a)[\Phi(b) - \Phi(a)] + \sigma_r[\phi(a) - \phi(b)]} \qquad \text{(B.12)}
\end{aligned}$$