Three Essays on Tax Compliance and the Estimation of Income-gaps

by

Ana Cinta González Cabral

Submitted to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Economics,
March 2017.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Author ....

March, 2017

"There is freedom waiting for you,
On the breezes of the sky,
And you ask 'What if I fall?'
Oh but my darling,
What if you fly?"

Erin Hanson

Three Essays on Tax Compliance and the Estimation of Income-gaps

by

Ana Cinta González Cabral

Submitted to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Economics,
March 2017.

Abstract

Quoting James Andreoni, 'the problem of tax compliance is as old as taxes themselves'. The sources of missing tax revenues have traditionally concerned tax administrations and particularly now in times when public finances are striving. In the quest for analysing the revenue that is foregone, tax administrations have started to produce a report of their tax gap—understood as the difference between the theoretical tax liability and the actual collection—to obtain a measure of the extent of non-compliance.

Due to the complexity of the non-compliance behaviour and the lack of visibility of certain types of income, different methods are usually put in place in order to offer a plausible range for the estimates. This dissertation dedicates its two first chapters to providing an alternative method for estimating the income-gap—defined to be one minus the proportion of reported to actual income—for two populations: the self-employed and the employees. The underlying data used for both cases is publicly available survey data on expenditures and income that is generated on a timely manner. This carries substantial advantages. First, relying on a general purpose survey dataset means that the estimation can be updated more frequently than if it was to rely solely on either the timing of administrative data or on survey data that is specifically targeted to measure non-compliance. Second, it provides an alternative estimation using an independent source of data which allows for the triangulation of the estimate obtained using administrative sources. Third, it allows tax administrations which do not have readily available administrative data to perform estimations using a type of survey widespread available in most countries.

More specifically, the first chapter estimates the income-gap for the self-employed in Great Britain. It emerges that self-employed report, on average, around 80.4% of their income to the tax authority, which translates into an income-gap of 19.6%. The link of the income-gap to non-compliance is critically assessed and differing preferences between the self-employed and employed, the measurement error of incomes and the use of savings to fund current consumption do not appear to be the drivers of the income-gap. A framework is posed to analyse the heterogeneity of the evasion response according to characteristics of the self-employed. It is found that the self-employed income-gap

varies significantly by sex, age, and region. In particular, male self-employed taxpayers underreport more than female ones, and they, in general, become more compliant as they age. The framework allows tax administrations the possibility of identifying the profile of evasion in order to inform compliance activities.

The second chapter, proposes a methodology to estimate the income-gap for the *moonlighters*—individuals who are on formal employment and their income is subject to withholding taxes but that have a second undeclared source of labour income. An expenditure approach based on industrial classification is proposed. Two methods are presented to identify occupations that concentrate high-risk of moonlighting practices: a factual and a data-driven approach. Using the latter, we find that individuals working in Construction display an income-gap of 19.6%, Real Estate and Professionals of 22.9% and those working on the Distribution sector display an income-gap of 19.8%. This translates into a 2.6% of GDP.

The third chapter of this thesis explores the role of the extrinsic and intrinsic incentives in explaining engagement in the hidden economy—defined as undeclared work practices. This chapter contributes firstly to the literature on shadow economy and to the debate of whether crowding effects are found between extrinsic and intrinsic motivations in a tax environment. Using micro-level European survey data, we demonstrate that both motivations matter. While intrinsically motivated individuals are less prone to participation in general, extrinsic incentives are key in deterring participation particularly for those that are not intrinsically motivated. We find no crowding-out effects between the intrinsic and the extrinsic motivation which entails interesting policy implications for the tax administration. Our results also suggest that undeclared work participation seems to be more widely spread among those who face financial strain.

Acknowledgments

Countless names come to my mind that I would like to thank for making this beautifully crazy adventure possible. First and foremost, I would like to express extreme gratitude to my supervisors Christos Kotsogiannis and Gareth D. Myles, for embarking on this journey with me. Their constant support both from an academic and personal perspective has been invaluable. Thank you for showing that you believed in me in every step, for your time and for always being available. I treasure all the advice that you have given me along these years and that will accompany me along my career.

This thesis is dedicated to my grandparents, Mercedes y Manuel for being my biggest inspiration. To my parents, Mar y Bartolomé for their constant sacrifice and endless love. To Mar, for being my most preferred sister and my confidante. To María José, for showing us that we are the fighting type. To Manuel, for his unconditional support and endless patience.

This thesis would not have been possible without the help, care and support of all my very best friends and colleagues, both from home and during my years in the UK. A special thank you goes to Polly, Ploy and Sara—specially for saying 'You are next', to Cliff, Janire, Hao and Poru. Thank you for making this happen while keeping me sane.

My deepest gratitude also goes to Prof. Juan José García del Hoyo, Prof. Ramón Jiménez Toribio, Prof. David Castilla Espino, y Prof. Félix García Ordaz from the University of Huelva who planted on me the curiosity for doing empirical research and guided me on my first steps. Thank you for introducing me to this world and letting me participate from your knowledge and projects.

I would like to also send a very special thank you to someone who is probably the ultimately responsible person for this PhD being on Economics, Gaspar León, whose passion for economics, enthusiasm and stimulating talks during my studies at Instituto La Rábida opened a new world for me.

Lastly, I would like to thank the University of Exeter, the members of the Business School and specially TARC for enabling me to pursue my dream. Lastly, I gratefully acknowledge the sources of funding that has made my Ph.D. possible. I was supported by a Formula Santander Scholarship from Santander Bank during my Masters programme and through TARC, I gratefully acknowledge the ESRC for supporting me through my PhD.

Contents

List of Figures

List of Tables

Chapter 1

Self-Employment Income-Gap in Great Britain: How Much and Who?

1.1 Introduction

Tax non-compliance—taken to be, in its broader form, illegal tax evasion of ones tax liabilities, either deliberate or from ignorance—undermines revenues, distorts competition (since it puts non-compliant taxpayers at an advantage) and increases inequality. Unsurprisingly, therefore, over the past few years enhancing tax compliance has been a central policy concern of governments in many countries across the world.[1] But tackling non-compliance necessitates an understanding of the scale of the phenomenon and, importantly, its determinants.[2] There are substantial challenges with measuring non-compliance, since, tax evasion, being illegal, it does not naturally lend itself to measurement. Indeed a consequence of this is that to date there is no systematic way of measuring the misreporting of income and consequently the corresponding 'tax gap' (defined as the difference between actual tax collected and the potential tax collection under full compliance) and diverse methodologies have been utilised generating various

[1] Amplified by the recent economic crisis as countries are struggling to restore public deficits.

[2] Something that international organisations (such as the IMF, and its 'gap analysis program', and the OECD, through the 'closing the gap action plan') have recently started paying particular attention to.

estimates.[3]

The purpose of this paper is to focus on a particular aspect of the income-gap, the income-gap of the self-employed and the heterogeneity of the evasion response, an issue for which there is limited recent empirical evidence, and there are good grounds to believe that it is significant. Since self-employment income is not subject to a withholding tax (pay-as-you-earn, PAYE) it might offer significant more opportunities for someone to understate their tax liabilities either directly or by overstating their business expenses.[4] This might have become more significant in the aftermath of the financial crisis (and in the midst of sovereign debt tensions) since all EU countries have experienced a rise in the number of self-employed individuals and this has been especially pronounced in the UK. Since the first quarter of 2008, the number of self-employed individuals in the UK has increased by 15% while the number of employees has increased only by 1%. The self-employed now represent 14% of the labour force in the UK and this share has seen steadily growing since the beginning of the economic downturn. Only since the first quarter of 2013 the number of self-employed grew at a rate of 8%, the highest in the EU-15 (only Slovenia, Cyprus, Bulgaria, and Lithuania exhibited higher growth rates).

To conduct the estimation use is made of a framework—that of Pissarides and Weber (1989), appropriately modified to incorporate the *profile* of the non-compliant self-employed—that has some very desirable properties: it relies on survey data which gives us access to a much wider range of observable characteristics[5] allowing us to take a closer look at the traits related to non-compliance. The framework, briefly considers two groups of taxpayers, employed and self-employed who, while having similar preferences

[3] And it is even also common in policy discussions for the tax-gap to be proxied by the shadow economy (the latter of which is much broader than the former). See Fiscalis TGPG (2016) for a discussion of the issues.

[4] Some contributions, to which we turn to below, do though bear on some aspects of the issue but none of them looks at the particular aspects addressed here.

[5] Than those available, for example, from tax returns. No data are, of course, problem free. There is a well-known problem with survey data: respondents are normally wary of answering questions that can be revealing of their non-compliance, even if the individuals are ensured confidentiality and despite assurances that the data will not be linked without their consent to administrative data. However, the use of a general scope survey shelters us from this shortcoming that affects specific surveys that rely on self-reported non-compliance. We recognise that despite this, survey data has its own caveats that we discuss further on Section 1.4.

for consumption (and record and state their expenditures correctly), they differ in the opportunity to underreport their income: employed (being subject to withholding taxes) have *more limited opportunities* for evading taxes than the self-employed and thus they report their income fully.[6] What this implies is that with a given level of expenditure, any discrepancy in the level of income reported by the self-employed is due to income not declared. The similar preference for consumption of employed and self-employed is demonstrated within the text as it stands as particularly meaningful in this approach. Expenditure is assumed to be correctly reported due to the nature of the expenditure categories used and the methods of collection of this information within the survey as will be discussed thoroughly in Section 1.4.

Taxpayers differ in terms of the characteristics potentially associated with their income-gap. Therefore, the framework allows for analysing the heterogeneity of the evasion response providing a tool for tax administrations to determine the profile of the non-compliant in order to effectively direct compliance activities. Central to the identification strategy is whether the discrepancy observed can indeed be related to underreporting or whether other alternative reasons more related to the heterogeneous behaviour of the occupational groups and differences in preferences, are the drivers of the gap observed. The availability of a wide range of variables allows us to critically assess the assumptions of the model and confirm that the observed discrepancy is due to income underreporting—and not driven by preference heterogeneity, savings, or measurement error.

The data set used is the Secure Access version of the Living Costs and Food Survey for the period 2010-2012[7] which provides us with a rich set of variables. We derive alternative estimates based on three different measures of expenditure and three different measures of self-employment income. The results from the three product groups are entirely consistent and the findings run in the expected direction. As is common with

[6]We do not rule out some minimal evasion behaviour of employed due to either collusion with their employers or moonlighting practices. However, on their main source of income, the third-party reporting of their income makes evasion a much harder task than for the self-employed. We acknowledge that some evasion will be picked up in the baseline.

[7] Office for National Statistics, Department for Environment, Food and Rural Affairs. (2014). Living Costs and Food Survey, 2006-2012: Secure Access. [data collection]. 4th Edition. UK Data Service. SN: 7047

models of the type considered here, there might be several other explanations that challenge the underlying assumptions of the model which could be potentially biasing the results, leading to a false attribution to underreporting. The main results show that, on average, the self-employed report only 80.4% of their actual income (or to put it differently, the self-employed income-gap is 19.6%) which translates into a lower-bound estimate of unreported taxable income of 1.6% of GDP. This result is not caused by different preferences between the self-employed and the employed, nor is it caused by savings funding current consumption, or by measurement error. It also emerges that using documents utilised to report to the revenue authority to fill in the income section of the survey leads to a higher level of underreporting being detected. Regarding the characteristics of the non-compliant taxpayers, the results suggest that males underreport more than females and that individuals become more compliant as they age. We also find suggestive evidence of the geographical heterogeneity of the evasion response.

The organisation of the paper is as follows. Section 1.2 contains a brief literature survey. Section 1.3 describes the model and Section 1.4 presents the data and outlines briefly the empirical methodology. Section 1.5 provides and discusses the results, and, finally, Section 1.6 summarises and concludes.

1.2 Brief literature review

One of the first contributions exploiting the discrepancy between income and expenditure to measure the size of the black economy in the UK is that of Dilnot and Morris (1981). They computed excess expenditures from survey data for 7200 households and characterised as 'black-economy households' all those where expenditure exceeded income by 20% and by at least £3. They reported an upper and a lower bound distinguishing whether the discrepancy could be explained by the circumstances of the household or whether they included pensioner and unemployed households. They found that the black economy was between £3.2 and £4.2 billion representing around 2.3-3% of 1977 GNP. Most of the black economy households were headed by an individual working in a skilled or semiskilled occupation and 22% of the black economy households were headed by a self-employed individual. Dilnot and Morris (1981) calculated participation ratios for different observed characteristics. The self-employed were much more likely than other groups to be part of the black economy sample than other employees. Individuals in skilled or semiskilled manual occupations were more likely to participate in black economy activities than unskilled.

More recently, Pissarides and Weber (1989), building on the contribution by Dilnot and Morris (1981), provided a more structural framework for the estimation of underreporting of the self-employed in Great Britain. Their approach consisted in obtaining a measure of income underreporting by the self-employed through a comparison of the relationship between food expenditure and income from this group to that of the employees who are assumed to be honest reporters. Pissarides and Weber (1989) recorded that true self-employment income is on average 1.55 times the income reported by the self-employed in Great Britain using the 1982 Family Expenditure Survey, with the uplift[8] factor being higher in blue-collar households than in white-collar households (1.65 versus 1.5). Using this estimate of underreporting they obtained that the size of the black economy was 5.5% of GDP. Lyssiotou *et al.* (2004), using the same approach for 1992, estimated the coefficient of self-employment underreporting to be on average 1.28,[9] and again higher for blue-collar households (1.39) than white-collar (1.18). In addition using a complete demand system, they found that self-employment income should be multiplied by a factor of 2.18 for the case of blue-collar households and 1.64 for white-collar households. They estimated the size of the black economy to be 10.6% of GDP in 1993.

Similar studies have also been conducted in other countries. In Canada, Schuetze (2002), using the equivalent to the Family Expenditure Survey for 1969-1992, found an average underreporting factor of 1.2 (1.12 lower bound and 1.23 upper bound). More recently, Hurst *et al.* (2014) for the US found that self-employment income is underreported by 30% which is equivalent to what was found by Engström and Holmlund (2009) for Sweden. Johansson (2005) found that the self-employed underreport their incomes in Finland by 16.5% for households of one self-employed (42% for households

[8] The term uplift factor, underreporting factor or coefficient of underreporting are going to be used interchangeably. They all refer to the factor by which reported self-employment income needs to be multiplied by to obtain true income.

[9] This result was obtained by averaging the values of blue-collar and white-collar workers.

of two self-employed) representing 1.3% (3.2%) of GDP.[10]

Like these contributions, the present paper focuses on estimating the degree of income underreporting by the self-employed (income-gap). Unlike them, however, the emphasis is also on testing the assumptions of the model to ensure the correct measurement of underreporting and the identification of the *profile* of the individuals associated with non-compliance which is an issue of significant policy relevance.

1.3 Description of the model

The framework is that of Pissarides and Weber (1989) extended to encompass the different (observed in the data) individual characteristics of the taxpayers. The model builds on a log-linear Engel curve that relates expenditure of household i on good j, denoted by C_{ij}, and income as,

$$\ln C_{ij} = \beta_j \ln Y_i^p + X_i \alpha_j + \varepsilon_{ij}, \tag{1.1}$$

where Y_i^p is permanent income, the measure of income that affects consumption, β_j is the elasticity of income with respect to consumption for good j, X_i is a vector that contains other determinants that affect consumption of household i, α_j is a vector of parameters, and ε_{ij} is a white noise error.

There are two groups of households in the model: employed and self-employed households. Expenditure on goods are reported correctly by all households.[11] Three measures of income enter the model: Permanent income Y_i^p, which is the measure of income that affects consumption decisions; reported income as recorded in the survey, Y_i^r; and actual income, Y_i^a, which is the true income the household holds. They are related as follows. The key assumption in the model is that unlike consumption, income may be

[10] Caution should be exercised when comparing estimates between different countries, or even for the same country for different years, as macroeconomic conditions and samples as well as definitions of the variables used for the analyses vary.

[11] Note that the assumption of expenditure being correctly recorded and the type of expenditure variables chosen are discussed thoroughly in Section 1.4.3. At this stage, we concentrate on the mechanics of the model itself and how the expenditure and income variables act to inform on income underreporting.

misreported. Self-employed and employed households in the model differ in that employed households (being taxed at source through pay-as-you-earn tax (PAYE[12])) have no opportunity to underreport their income, but self-employed may do so by a factor k_{in} (to be estimated), which, importantly, can vary with the individual characteristics of those self-employed. Each characteristic, denoted by N, has n different categories: N, for example, could be gender, with n reflecting the two categories of the variable, male and female.

For self-employed households it is the case that reported and actual income are related as,

$$\ln Y_i^a = \ln k_{in} + \ln Y_i^r, \tag{1.2}$$

where, as noted above, k_{in} is a random variable that captures the factor by which self-employed income in category n has to be scaled to arrive to their true income. For the employed, by assumption, for all n, $\ln Y_i^a = \ln Y_i^r$.

Actual income does not need to coincide with permanent income (unobserved) since there might be fluctuations (transitory component of income). Thus, actual and permanent income are related as,

$$\ln Y_i^a = \ln p_i + \ln Y_i^p, \tag{1.3}$$

where p_i is random variable with mean being the same for both type of households. However, income from self-employment is more variable than income from employment and thus the variance of the latter, denoted by $\sigma^2_{u_E}$, is lower than the variance of the former, denoted by $\sigma^2_{u_S}$, that is

$$\bar{p}_S = \bar{p}_E \quad ; \quad \sigma^2_{u_S} > \sigma^2_{u_E}. \tag{1.4}$$

It is assumed that both k_{in} and p_i are log-normally distributed which implies that they

[12] Employed households in the UK are subject to pay-as-you-earn tax (PAYE), a withholding tax on income payments to employees. Income withheld is treated as advance payments of income tax due and they are refundable to the extent they exceed tax as determined on tax returns.

can be written as deviations from their means

$$\ln p_i = \mu_p + u_i \quad ; \quad \ln k_{in} = \mu_{kn} + \nu_{in}, \tag{1.5}$$

where u_i and ν_{in} are random variables with zero mean and (constant) variances σ_u^2 and $\sigma_{\nu n}^2$, respectively.[13]

Following from the log-normality of p_i and k_i, the mean of p_i and k_i and the mean of their log, μ_p and μ_{kn} respectively, are related as follows

$$\ln \bar{p} = \mu_p + \frac{1}{2}\sigma_u^2 \quad ; \quad \ln \bar{k}_n = \mu_{kn} + \frac{1}{2}\sigma_{v_{Sn}}^2. \tag{1.6}$$

Given that the mean of p_i is the same for both groups by assumption,[14] it follows that

$$\mu_{p_S} - \mu_{p_E} = -\frac{1}{2}\left(\sigma_{u_S}^2 - \sigma_{u_E}^2\right) \tag{1.7}$$

One can then combine (1.2), (1.3) and (1.5) into (1.1) to obtain,

$$\ln C_{ij} = \beta_j \ln Y_i^r - \beta_j\left(\mu_p - \mu_{kn}\right) - \beta_j\left(u_i - v_{in}\right) + X_i\alpha_j + \epsilon_{ij}. \tag{1.8}$$

Equation (1.8) defines household i's expenditure on good j, the variation of which, across the self-employed with characteristic n and the employed, is captured by $\beta_j\left(\mu_p - \mu_{kn}\right)$. This can be straightforwardly captured empirically using the interaction of a self-employment dummy variable and the characteristic N of interest, that is

$$\ln C_{ij} = \beta_j \ln Y_i^r + \gamma_{jn} SE_i * N + X_i\alpha_j + \eta_i, \tag{1.9}$$

where SE_i is a dummy variable that takes the value 1 if the household is self-employed and zero otherwise, N is a taxpayer's characteristic and η is an error term defined as $\eta_i = \beta_j(u_i - v_{in}) + \epsilon_{ij}$.[15] Thus, the parameter γ_{jn} estimated from the Engel curve

[13] But the variance of v for the self-employed within different categories n of the same characteristic N is allowed to vary.

[14] The variance of u, σ_u^2, is the same for the different characteristics of the self-employed and so $\sigma_{u_n}^2 = \sigma_u^2$.

[15] A precise definition of a self-employed household is provided in Section 1.4.

regression (1.9) captures,

$$\gamma_{jn} = -\beta_j((\mu_{p_S} - \mu_{p_E}) - \mu_{kn}) \quad = \beta_j \left[\mu_{kn} + \tfrac{1}{2}(\sigma^2_{u_S} - \sigma^2_{u_E})\right] \tag{1.10}$$

where the term $\mu_{p_S} - \mu_{p_E}$ captures the correction in volatility of permanent income.[16] Substituting (1.10) for μ_{kn} in (1.6) one arrives at,

$$\ln \bar{k}_n = \frac{\gamma_{jn}}{\beta_j} + \frac{1}{2}\left(\sigma^2_{v_{Sn}} - \sigma^2_{u_S} + \sigma^2_{u_E}\right). \tag{1.11}$$

The term $\sigma^2_{v_{Sn}} - \sigma^2_{u_S} + \sigma^2_{u_E}$ can be identified off differences in income volatility. Given that in regression (1.9), the measure of income used is reported income in the survey as opposed to the appropriate permanent income measure (unobserved), income is measured with error and therefore instrumental variables can be used to mitigate this issue. We, therefore, estimate the first-stage regressions of income as,

$$\ln Y_i^r = X_i\delta_1 + Z_i\delta_2 + \varsigma_i, \tag{1.12}$$

where Z_i is a set of identifying instruments. This gives us an independent estimate of the variance of income for both occupational groups given the characteristic chosen. The residual ς_i in (1.12) is a composite of three errors: unexplained variations in permanent income, deviations from actual to permanent income, u_i, which is assumed to be the same for both groups, and deviations from actual to reported income, ν_{in}, which are allowed to differ across categories n. Assuming, as we shall, that the unexplained variations of permanent income have the same variance for both groups, the residual income variance for the self-employed with characteristic n, denoted by $var\varsigma_{Sn}$, would be larger than the residual income variance, denoted by $var\varsigma_E$, for the employed. This is because $var\varsigma_{Sn}$ contains the variance of v_i, the underreporting component, and the variance of u_i is larger for the self-employed. Defining $var\varsigma_{Sn} \equiv \sigma^2_{Y_{Sn}}$ and $var\varsigma_E \equiv \sigma^2_{Y_E}$,

[16] Note that the effect on the mean of transitory events is the same but allows for the volatility to differ from equation (1.4)

it is the case that

$$\begin{aligned} var\varsigma_{Sn} - var\varsigma_E &\equiv \sigma^2_{Y_{Sn}} - \sigma^2_{Y_E}, && (1.13) \\ &= var(u - v_n)_S - var(u)_E, && (1.14) \\ &= \sigma^2_{u_S} + \sigma^2_{v_{Sn}} - 2cov(uv)_{Sn} - \sigma^2_{u_E}, && (1.15) \end{aligned}$$

where the equality in (1.14) follows from (1.8) and the equality in (1.15) follows from calculating the variances in (1.14).

The expression in (1.15) can be combined with (1.11) to obtain an estimation of the average value of income underreporting.[17] A certain value for $\bar{k}_n$ cannot be provided, as it depends on $\sigma^2_{v_{Sn}}$ and $\sigma^2_{u_S}$ but bounds for the estimate can be attained. Clearly, following (1.15), $\sigma^2_{v_{Sn}}$ and $\sigma^2_{u_S}$ are negatively related and, therefore, $\bar{k}$ reaches a lower bound when $\sigma^2_{v_{Sn}}$ takes its lowest value of 0. Using (1.15) into (1.11) the lower bound, denoted by $\bar{k}_{ln}$, is obtained as

$$\ln \bar{k}_{ln} = \frac{\gamma_{jn}}{\beta_j} - \frac{1}{2}(\sigma^2_{Y_{Sn}} - \sigma^2_{Y_E}). \tag{1.16}$$

Similarly, there is an upper bound for $\bar{k}_n$. Since the variance of u for the self-employed is larger than for the employed, the minimum feasible value of $\sigma^2_{u_S}$ is $\sigma^2_{u_E}$ which will provide the upper bound for $\bar{k}_n$, denoted by $\bar{k}_{hn}$, and that can then be written as

$$\ln \bar{k}_{hn} = \frac{\gamma_{jn}}{\beta_j} + \frac{1}{2}(\sigma^2_{Y_{Sn}} - \sigma^2_{Y_E}). \tag{1.17}$$

Central to this paper is to estimate the average scaling factor, $\bar{k}_n$, and the income-gap, $\bar{\kappa}_n$ (defined, to be recalled, as one minus the proportion of reported to actual income).[18] These two quantities are related as

$$\bar{\kappa}_n \equiv 1 - \frac{Y^r_n}{Y^a_n} = 1 - \frac{1}{\bar{k}_n}. \tag{1.18}$$

[17] It is assumed for simplicity that underreporting and income volatility are uncorrelated for the self-employed.

[18] Hurst *et al.* (2014) obtain a point estimate of income under-reporting. Appendix A discusses the implications of the restrictions required to achieve a point estimate, and provides a justification for the choice of not doing so.

Section 1.5 reports the magnitudes of the income-gap, $\bar{\kappa}_n$.

1.4 Description of the data

1.4.1 Living Costs and Food Survey

The data comes from the Secure Access version of the Living Costs and Food Survey (LCFS) and covers the years 2010-12. The LCFS uses as the unit of survey the household and it captures expenditure decisions and income earned from all the individuals within a household. For consistent comparison across households, and to avoid concerns arising from differences in non-compliance due to the composition of the household, the sample is restricted to households of two adults, either cohabitees, married or civil partners who live in Great Britain and the household reference person (HRP) is either employed or self-employed.[19] The age of HRP is also restricted to be less than 60 in order to leave out the different expenditure behaviour that has been documented in the literature after retirement, Aguiar and Hurst (2005).

To obtain a reasonable sample size data for all three waves are pooled. They are also deflated using the consumer price index base of 2010. Income reported has also been adjusted in order to account for the fact that income reported by the self-employed dates back to the last available record which could well have been obtained a year before and consequently, not to the date of the interview. Failing to correct this time lag could lead to spurious results. Self-employment income reported is updated to the time of the interview using the monthly rate of inflation calculated from the amount of self-employment income per self-employed worker drawn from the Blue Book for the corresponding years.

A self-employed household is defined to be one that draws more than 25% of their income from self-employment sources. This threshold is imposed in order to avoid households that have a substantial amount of self-employment income (e.g. from a subsidiary

[19]The HRP is the householder who: either owns the household accommodation, or 1) is legally responsible for the rent of the accommodation, or 2) has the household accommodation as an emolument or perquisite, or 3) has the household accommodation by virtue of some relationship to the owner who is not a member of the household. If there are joint householders, HRP is the one with the higher income. If the income is the same, then the eldest householder is taken.

source) to classify themselves as employees. For robustness, alternative specifications of the self-employment dummy variable have also been considered as will be discussed in Section 1.5.

1.4.2 Measuring income

In the LCFS, self-employed individuals were asked about the profit from their activity and about how much their drawings amount was for both business and non-business purposes. In the event that the individuals were not able to respond to any of these questions they were asked for an estimate of how much their income was once expenses were deducted. To capture all income the self-employed take into consideration, it is assumed that their consumption decision is based on their total earnings which include both labour earnings and those reinvested back into the business (a form, perhaps, of precautionary savings). But it is conceivable, too, that income for the self-employed is better proxied by drawings from their business. To obtain a comprehensive picture (and also being agnostic of which form of income approximates true income better) use is made of three different measures of self-employment income: a *comprehensive measure*; *profits* and *drawings*.

The *comprehensive measure* of income takes into account as much data on income as possible, and, therefore, avoids missing values in the answers, and is computed as follows. The profit figure (transformed into a weekly amount) is taken if reported. If the individual reported a loss (or zero profit) or is not able to report a profit figure, then the estimate of the weekly drawings is taken. If none of the former are available, then the weekly equivalent of the estimation of income minus expenses is taken. The second and third measures consider, respectively, only *profits* and *drawings* as the measure of self-employment income.

1.4.3 Measuring Expenditure

The analysis considers three different measures of expenditure: *food*, *utilities* and a *basket of nondurable goods*. Survey data are known to suffer from well-known problems, but food expenditure is considered to be reported accurately for several reasons. Firstly, food is one of the items of expenditure that is better captured on the LCFS. Brewer and O'Dea (2012), through a comparison of the National Accounts with the LCFS, find that food together with household fuel and the running costs associated with motoring appear as the categories of expenditure where coverage is higher. Food had a coverage rate above 80% for each year in the period 1974-2009. This is consistent with the patterns found in the US by Meyer and Sullivan (2010) who, comparing the Personal Consumption Expenditure and the Consumption Expenditure, have found a good coverage for food eaten in of 85% while the coverage of food eaten away from home has been declining over time. Secondly, food, being a necessity,[20] is not an expenditure that can be altered by transitory shocks, its consumption cannot be postponed to future periods and zeros for infrequency of purchase are not present.[21] Thirdly, food does not represent an item of expenditure that can generally be claimed as a business expense.

The UK tax authority (HMRC) establishes that some of the costs the self-employed face can be claimed as allowable expenses. These include: office costs, travel costs, clothing, staff costs, resell of goods, financial costs, costs of your business premises and advertising or marketing. Food expenses do not feature in the list of allowable expenses, but there is an exception: food expenses can be claimed in the case of overnight business travel. Other food costs such as those derived from entertaining clients, suppliers and customers or those related to event hospitality cannot be claimed, HMRC (2015). In fact, none of the individuals in the sample mentions food as a business expense. Besides, food is expected to be correctly reported as there is no social stigma associated to it. It is well documented that other items of expenditure such as tobacco or alcohol are usually misreported on surveys, particularly in the LCFS the coverage ratio is 40% (Brewer and O'Dea, 2012). Food does not represent an item that can be suspected of showing

[20] For a discussion on this see Attanasio *et al.* (2004) who also present some evidence that for very poor households (defined as having their income or consumption below 60% of median consumption or income) this might not hold. The fraction of such households, however, in the general population is, in general, very small to significantly alter our results.

[21] This is one of the reasons why durable goods were not taken into account in the measure of expenditure. Ways to mitigate problems with infrequency of purchases have been proposed but there are limitations too (see Meghir and Robin, 1992). Lyssiotou *et al.* (2004) estimated underreporting using a demand system in which they included durable goods, but problems with infrequency of purchase still remain. Additionally, Brewer and O'Dea (2012) find that durable goods (household and personal services, vehicle purchases and other durable leisure goods) are an item of expenditure with very volatile coverage ratio in the LCFS (previously named as the Family Expenditure Survey) ranging between 55-80%.

a certain lifestyle as opposed to expenditure on holidays or newly bought cars, and therefore the interest in misreporting is minimal. Lastly, food expenditures are recorded using a diary during a two week period by all individuals over 16 years old and children between 7 and 15 are offered a simplified version. Other items of regular expenditure, such as the mortgage, are captured in the household questionnaire.[22] Expenditure on utilities was chosen as an item of expenditure that could be easily recalled due to the regular payments. However, it raises concerns as it represents an item typically claimed as a business expense by the self-employed. The higher level of consumption could be masquerading the differing nature of the self-employment activity and its particular fiscal treatment. Therefore, estimates although presented for comparison should, in principle, be taken with caution. A composite measure of expenditure is also presented and is formed of a basket of nondurable goods.[23]

Table 1.1 reports the summary statistics for the expenditure measures and disposable income for the two groups.

[22] The Living Costs and Food Survey uses diaries instead of interviews to collect food expenditure data. Recall data has been long recognised to suffer from different problems: quantities are difficult to remember (Gray, 1955), memory declines with the length of the recall period which Deaton (1997) refers to as 'progressive amnesia' (see Sudman and Bradburn, 1973; and Scott and Amenuvegbe, 1991) and telescoping errors (Neter and Waksberg, 1964). Diaries minimize these problems as respondents are asked to record their expenditure on the day it occurs. However, as Browning *et al.* (2003) and Battistin (2003) recognise diary methods are also recognised to suffer from problems of 'diary fatigue' due to the high burden they pose on the respondents. If they are only required to be kept for a short period of time then problems of infrequency of purchase arises. As Deaton (2005) puts it there is a trade-off between memory and match between consumption and purchases. Ahmed *et al.* (2006) exploiting a database with both interview (recall) and diary data find that none of the sources are free of errors but that recall measures of expenditure were substantially underrecorded. There is an inclination to mix both techniques in order to minimise their challenges (Gibson, 2002; Battistin, 2003). The Consumer Expenditure Survey in the US and the LCFS in the UK combine recall and diary data with the intention of capturing more infrequently purchased items of expenditure with the former. Browning et al. (2003) comparing recall and diary expenditure on food at home for the US recognise that individuals do a 'remarkably good job' when recording food at home as opposed to total expenditure. As food expenditure is a frequent item diaries seem to be the appropriate way of recording it.

[23] The basket of non-durable goods comprises expenditure on food, alcohol and tobacco, clothing, utilities, non-durable expenditure on recreation, non-durable expenditure on transportation and communication, health, education and other miscellaneous nondurable expenditures.

Table 1.1: Summary Statistics

	Employed			Self-Employed		
	N	Mean	S.D.	N	Mean	S.D.
Disposable Labour Income (ln)	4034	6.528	0.504	738	6.473	0.695
Food Expenditure (ln)	4026	4.35	0.558	738	4.432	0.579
Utilities (ln)	4021	3.222	0.565	736	3.314	0.585
Non-Durables (ln)	4036	5.734	0.524	739	5.81	0.552

Table 1.1 shows, interestingly, and across all expenditure measures, the self-employed exhibit higher levels of expenditure whereas their reported income is lower than that of the employed.[24]

1.5 Estimation and results

The estimation method is two-stage least squares based on equation (1.9) being the first-stage as in equation (1.12). Endogeneity is a pervasive and well-known problem in consumption function regressions of the kind explored here leading to biased and inconsistent estimators. To mitigate the endogeneity bias, income is instrumented. The instruments to be used are: educational attainment variables (whether the household reference person and spouse had higher education) and whether the household reference person of the household is in a white-collar or blue-collar occupation.[25] A summary of results is available within the section. Full results are available in Appendix C.

1.5.1 Self-employment income-gap

Estimation of the income-gap for the self-employed as evidenced in equations (1.16) and (1.17) requires four key parameters: the estimation of the residual income variance

[24] The difference in means is statistically significant. Summary statistics of the covariates are provided in Appendix B.

[25] Pissarides and Weber (1989) also instrument for the self-employment dummy variable. The reason behind this is that a self-employed household is defined from the proportion that self-employment income represents to total household income, and if the first is underreported then the proportion may vary and consequently whether the household is considered self-employed or not. However, instrumenting can bring its own problems. The bias that can result in the estimates due to weak instruments could outweigh the effect of failing to instrument and accepting the bias introduced by considering that some households that behave as self-employed are classified in our sample as employed. Instead of instrumenting, our approach is to use alternative definitions of a self-employed household in order to assess the impact of definition on the estimation of underreporting.

for the two employment groups (self-employed and employed) from the first-stage regressions (σ^2_Y); the elasticity of consumption (given by β_j), and the coefficient of the self-employment dummy variable (γ_{jn}) from the second-stage regressions.

Table 1.2 summarises the estimation of the income-gap calculated as the midpoint between the upper and lower bound as exposed in equations (1.16) and (1.17) on Section 1.3. For comparison purposes, it also summarises the effects on the estimate of failing to correct for the differing volatility of income for both groups, this is $\sigma^2_{Y_{Sn}} = \sigma^2_{Y_E}$.[26] For simplicity, the discussion in what follows will be centred around the midpoint estimations of the income-gap.[27]

Table 1.2 shows that, with *food expenditure*, the self-employed income-gap is 19.6% which translates into a midpoint multiplier of $\bar{k} = 1.24$, implying that income reported by the self-employed should be multiplied by a factor of 1.24 to obtain true income.[28] This value is similar to that obtained for countries such as Sweden or Canada (Engström and Holmlund, 2009; Schuetze, 2002). Table 1.2 also reports estimation using two alternative measures of expenditure: expenditure in *utilities* and a *nondurable basket of goods*. The concern with using expenditure in utilities, despite being an item of expenditure one may expect to be easily recalled, is that it can be claimed as a business expense. This may act as an incentive for the self-employed to carry some of their activity at home and claim the expense on utilities. Thereby we could observe a higher level of expenditure on utilities for a certain level of income that would not be due to underreporting but that it is only a feature of the self-employed activity and the fiscal treatment. The degree of underreporting revealed using utilities is only slightly larger: 22.3%. The basket of all non-durable goods as a proxy for expenditure gives a similar degree of underreporting as using food expenditure, 19.9%.[29]

[26] We assume away the correction for volatility to enable the comparison to Hurst et al. (2014).

[27] For the sake of brevity, we refer in the summary table in this section and in the discussion of the results to a midpoint income gap estimated by averaging the upper and lower bound of the income-gap reported in equation (1.17) and (1.16) respectively. Full results including the bounded estimate of the income-gap and the multiplier of underreporting is available in Appendix C. First-stage regressions are not shown but statistics of their quality are contained in the each regression table of Appendix C.

[28] Notice that the estimates presented are obtained assuming the covariance between u_i and v_i is zero. We cannot empirically assess the correlation between the two but assuming the highest correlation of 1, our estimates would need to be scaled up by 35%. On the other hand if the correlation was 0, the estimates would be correct, and would need to be adjusted upwards by between 0-35% if the correlation lies in between 0 and 1.

[29] Arguably, estimation should also consider housing as a measure of expenditure. The problem with

Using alternative measures of self-employment income—capturing the possibility that the self-employed make their consumption decision based on their available and retained earnings—the level of underreporting estimated using the comprehensive measure of self-employment income and profit is identical (compare last column of rows 4 and 6 of Table 1.2). Using only drawings from the business results in a substantial decrease on the sample size as not many individuals report taking drawings from their business or report missing answers. Table 1.2 shows more significant income underreporting with the use of drawings as the measure of self-employment income. This result may be overestimating the amount of underreporting as it might be failing to consider the influence of retained earnings in shaping consumption decisions.

The definition of a self-employed household (being those who derive more than 25% of their income from self-employment) might be of concern in the sense that it relies on reported income potentially leading to the misclassification of households. This definition is aimed at circumventing the use of self-reported employment status in order to avoid the selection of households into employment when their main source of income is self-employment. We test the robustness of the results to alternative definitions of a self-employed household. Table 1.2 shows that the income-gap estimated is robust to the definition of the self-employed household, be it based on income (25% rule) or defined from self-reports. Two alternative definitions are utilised: the first one uses *self-reported employment status* of the HRP in their main or subsidiary jobs (*Self-Employed Head* in Table 1.2), while the second relates to *opportunity* a self-employed household[30] has to underreport income (*Any Self-Employed* in Table 1.2).

this, however, is that it would require imputed rent for individuals who own their house something for which data are not available and, more importantly, not easily computed.

[30] A household is defined as self-employed if any self-employment income source is available to them. This definition neither accounts for the importance nor amount of self-employment income: only the opportunity to underreport income matters.

Table 1.2: Estimation of the Income-Gap.

	No Volatility Correction	Volatility Correction
	Point Estimate	Midpoint
Expenditure		
Food	0.187***	0.196***
	(0.043)	(0.042)
Utilities	0.214***	0.223***
	(0.045)	(0.045)
Non-Durables	0.190***	0.199***
	(0.031)	(0.03)
Income		
Profit	0.180***	0.196***
	(0.066)	(0.065)
Drawings	0.368***	0.391***
	(0.099)	(0.095)
Proxy	0.187***	0.196***
	(0.043)	(0.042)
Alternative SE definitions		
Self-Employed Head	0.215***	0.221***
	(0.038)	(0.037)
Any Self-Employed	0.215***	0.221***
	(0.037)	(0.036)
SE (25% rule)	0.187***	0.196***
	(0.043)	(0.042)

Notes: 2SLS regressions based on equation (1.9) assuming no heterogeneity of the income-gap is analysed $N = 1$. Robust standard errors in parentheses. The stars indicate significance at the following levels: $^{*}p < 0.10$,** $p < 0.05$,*** $p < 0.01$. Survey data pooled for waves 2010-2012. The dependent variable for the first three regressions estimations (first three rows of the table) are (1) Food Expenditure, (2) Utilities and (3) Non-Durables. The dependent variable is food expenditure for the rest of the regressions results unless specified otherwise. The independent variable of interest is household disposable labour income instrumented using the level of education and the type of occupation (white or blue collar, defining white-collar as those individuals who are employers in small or large organisations, or hold a higher managerial, higher professional, lower professional and higher technical, lower managerial and higher supervisory and intermediate position) and the self-employment dummy variable. All specifications include year and quarter fixed effects and controls for household and personal characteristics (age and number of children in a non-quadratic fashion, number of cars and rooms, type of tenure, availability of drier and central heating). All dependent and independent variables are in logs and deflated accordingly. The sample is restricted to individuals aged less than 60 years of age. All specifications are weighted using sample weights. The income-gap presented in this table are the IV estimated income-gaps. First stage F-statistics, tests on the quality of the instrument, sample sizes and the OLS specification and estimation are provided in Appendix C. The estimates provided in this table are obtained assuming the covariance between the evasion and reporting component is zero (v and u in the model).

1.5.2 Is it underreporting?

A concern with the analysis so far (and the results obtained) might be whether the estimates obtained can indeed be attributed to underreporting or whether they are a feature of the assumptions posed to obtain the estimate. It is precisely this issue that the analysis now turns to, paying particular attention to three aspects: heterogeneous preferences for food consumption, heterogeneous spending behaviour, and measurement error in the survey.

1.5.2.1 Preference heterogeneity

It has been assumed (see equation (1.8)) that both taxpayer groups (self-employed and employed) have the same preferences over consumption and, therefore, the same income elasticity of consumption. One, however, can certainly think of a more structural model where occupational choice is driven by sector-specific skills and different preferences over consumption. To test this assumption, the consumption functions of the two occupational groups have been estimated separately showing that the income elasticities for both occupational groups, using the IV specification, are statistically indistinguishable from each other.[31]

On another matter, data on food expenditure distinguishes between food eaten in the house and eaten out of the house. Since food eaten out is typically more expensive, a difference in the pattern of consumption between the two groups could explain the estimation results. We have, therefore, tested whether there is a different preference between food eaten in and out between the occupational groups. Estimating the share of food eaten in with respect to total food expenditure, we find that there is no significant difference between the two groups. Therefore, the higher level of expenditure observed for the self-employed cannot be justified by a higher expenditure on food outside from home.

[31] The results are available upon request.

1.5.2.2 Heterogeneous spending behaviour: Financial constraints

Another reason why we might observe a discrepancy between expenditure and income may be a different level of financial stability for the self-employed. What we know, and is controlled for in the model, is that income for the self-employed is more volatile than income for the employed. However, it could still be the case that in the years observed, 2010-12, the level of expenditure of the self-employed is not funded by current income only but also by past savings. If this was the case the model would be misleadingly attributing to underreporting the fact that some current consumption is not funded by current but by past income. Using variables from the Secure Access data, we are able to filter out those individuals who claim to be financially constrained: by this is meant individuals who find their current income not sufficient to fund their current consumption and need to recourse to other alternative sources such as loans, savings, or money from relatives. We have created a smaller data set of individuals not financially constrained and we rerun the analysis and we are still able to find a substantial income-gap of 22.9% as displayed in Table 1.3, which comes to show that the income-gap obtained cannot be explained because of the use of past incomes to fund current expenditure.

The main category of expenditure analysed in this paper is food expenditure for the reasons discussed previously on Section 1.4.3. However, we also present estimates using utilities as an expenditure category. One of the main concerns about using utilities is that it is one of the main categories claimed by the self-employed as a business expense. Therefore, a higher level of expenditure on utilities relative to income may not be a sign of underreporting but the result of a differential fiscal treatment for the self-employed. We have sieved out of the estimation all those self-employed that claim any utilities bill as a business expense. Using utilities as the expenditure category on those self-employed individuals who do not claim any business expenses, we still observe a very similar level of underreporting (21.9%)—see Table 1.3.

Therefore, there is evidence that neither the role of savings in funding current consumption nor the possibility of claiming of business expenses from some expenditure categories are the drivers of the income-gap obtained.

Table 1.3: Estimation of the Income-gap: Alternative explanations of the gap.

	No Volatility Correction	Volatility Correction
	Point Estimate	Midpoint
Income Measurement Error		
Any Documents	0.206**	0.218***
	(0.081)	(0.08)
HMRC Documents	0.310**	0.320**
	(0.127)	(0.125)
No Financial Constraints	0.218***	0.229***
	(0.083)	(0.082)
No business expenses claims	0.212***	0.219***
	(0.05)	(0.049)

Notes: 2SLS regressions based on equation (1.9). Robust standard errors in parentheses. The stars indicate significance at the following levels: $^*p < 0.10$,** $p < 0.05$,*** $p < 0.01$. The dependent variable for the last row regression results is expenditure on utilities. The dependent variable is food expenditure for the rest of the regressions results unless specified otherwise. The independent variable of interest is household disposable labour income instrumented using the level of education and the type of occupation (white or blue collar, defining white-collar as those individuals who are employers in small or large organisations, or hold a higher managerial, higher professional, lower professional and higher technical, lower managerial and higher supervisory and intermediate position) and the self-employment dummy variable. All specifications include year and quarter fixed effects and controls for household and personal characteristics (age and number of children in a non-quadratic fashion, number of cars and rooms, type of tenure, availability of drier and central heating). All dependent and independent variables are in logs and deflated accordingly. The sample is restricted to individuals aged less than 60 years of age. All specifications are weighted using sample weights. The income-gap presented in this table are the IV estimated income-gaps. First stage F-statistics, tests on the quality of the instrument, sample sizes and the OLS specification and estimation are provided in Appendix C. The estimates provided in this table are obtained assuming the covariance between the evasion and reporting component is zero (v and u in the model).

1.5.2.3 Measurement error

As noted earlier, one of the problems with using surveys relates to the accuracy of the information reported. The nature of the survey makes it likely that expenditure is more accurately reported than income since the survey makes available instruments for making expenditure as accurate as possible. For example, in the case of food a diary is filled by each member of the family. In the case of utilities, being a recurrent expense, it is easy to recall.[32] For income, however, underreporting may not be the only reason behind the discrepancy found. Individuals are asked about their current income. Employees are usually paid on a monthly basis, so their claim of income earned should not deviate in excess from what they actually earn. That is, it should be free from recall error. For the self-employed, however, this is not the case. Recalling their last profit of their businesses may not be such an easy task. Therefore, it may be true that the discrepancy found stems from measurement error of the income variable for the self-employed due to recall. It could be potentially down to the self-employed supplying approximate figures for their income that have little to do with their true levels.

There is substantive evidence that the length of the recall period, the lapse between occurrence and report, influences the accuracy of the report. Scott and Amenuvegbe (1991) find that average daily expenditures reported fall by almost three percent for every day added to the recall period being frequently purchased items the most affected. Deaton (1997) coined the tendency to forget earlier transactions 'progressive amnesia'.[33] For income, Withey (1954) interviewed participants about their current income in one year, and then re-interviewing the participants one year later asking them about the past income, only 61% of the cases reported the same income in their current and retrospective reports. Converse evidence is also found on the literature, Marquis and Moore (1990) only found that increasing underreporting linked to an increase in the recall period for one of the eight programmes analyzed in the Survey of Income and

[32] See Bee *et al.* (2012) for a comprehensive discussion of this issue.

[33] Evidence is mixed, though. The Indian National Sample Survey which surveys households experimented with different recall periods for different groups of goods. They found that decreasing the recall period increased reported expenditure on food by 30% and total expenditure by 17% (Deaton, 2005). But this is not always the case. The NSSO Expert Group on Sampling Errors (2003), however, have noted that for certain foods a longer recall period was better than a shorter one (30 days versus 7 days) (cited in Wodon, 2007).

Programme Participation (SIPP).

In order to assess whether the income-gap found stems from simple measurement error, we have carried out two checks. In the first, we have selected those individuals who consulted any documents in order to provide their profit figures.[34] These documents could be either a Notice of Tax Assessment (form 300), their Annual Accounts (or the summary), their Tax Return or any other documents. Estimates are reported in Table 1.3 which shows that the level of underreporting is slightly larger than in the case when no document is consulted (21.8% versus 19.6%). The second check is to look at the estimate in the presence of administrative data. One of the differences between filling out a survey and reporting to the revenue service is the incentive to report the true income. In the survey, individuals may report their true income more accurately than they do to the revenue service. However, one may also argue that some of the individuals who are afraid of their data being linked to their tax records despite reassurance of the anonymity of their responses, will give the same answer to the survey as they did to the revenue service for the sake of consistency and avoiding self-incrimination.

To disentangle whether these effects are biasing the results, we repeat the estimation only for individuals who have made use of documents that have already been used with the revenue service or that originate from the revenue service: the Notice of Tax Assessment and the Tax Return. Using only individuals who have looked at these documents we find a larger value of underreporting of 32%, Table 1.3. However, we need to take these results with care as the sample size of self-employed using these documents is very low. Even so, this highlights the potential advantages of combining the use of administrative and survey data.

[34] The nudge to consult documents is a strategy from the interviewers to reduce recall errors. This procedure has proven efficient as documented by Maynes (1968) who found less errors in reports of individuals who were prompted to look at the records of the amount held in their savings accounts than in those who were not. Same results were found by Grondin and Michaud (1994) and Moore *et al.*, (1996) who found a reduction of errors in reporting of income when using tax forms (cited in Moore *et al.* 2000). Instructions where given in the LCFS for interviewers to prompt the use of documents.

1.5.3 Determinants of non-compliance

This section constructs a *profile* of the non-compliant self-employed. A summary of the results are contained in Table 1.4, full estimation and tests of the adequacy of the instruments are available in Appendix C.

1.5.3.1 Non-compliance: Age

To assess the impact of age on self-employment underreporting, we introduce the interaction of the self-employment dummy variable and different age categories. We have analysed this taking into account the age of the HRP and also the age of the self-employed individual (if different). The results in Table 1.4 show that as individuals age, they become more compliant. This is, in households were the HRP is less than 35 years old, income is underreported by on average 37.5%; if between 35 and 45, 18.2% and between 45 and 60 by 13.6%. The finding that the income-gap decreases as age increases is consistent with the findings in the literature on tax audits in the US using TCMP (Clotfelter, 1983; Feinstein, 1991).[35]

1.5.3.2 Non-compliance: Gender

Gender is another characteristic generally observed by the revenue service. We find that men underreport significantly more than women. Table 1.4 shows that households of two self-employed (*Both* in Table 1.4) underreport more than households of one self-employed. This result is in line with intuition. A self-employed individual has higher chances of underreporting their income than an employee due to the absence of third-party reporting. A household of two self-employed has even more opportunities to underreport than a household of a single self-employed individual. This results are in line with those found by Johansson (2005) for Finland but run against those found by Schuetze (2002) for Canada. However, in households of one self-employed, the self-employment dummy variable is not significant when the self-employed is the wife and thus the income-gap is higher in households where the self-employed is male. Our result

[35] Criminology studies provide evidence of this same finding for crimes in general. It seems that for all types of criminals and crimes there exists a common distribution for age which shows to be invariant to social and cultural conditions so that crime is negatively correlated with age. Though the theories for why this happen vary (Tittle and Grasmick, 1997).

Table 1.4: Estimation of the Income-gap: Profiles of Non-Compliance.

		No correction for volatility	Correction for volatility
		Point Estimate	Midpoint
Age	< 35	0.375***	0.375***
		(0.093)	(0.093)
	35-45	0.172***	0.182***
		(0.063)	(0.063)
	> 45	0.136**	0.149**
		(0.066)	(0.065)
Gender	Wife	-0.051	-0.043
		(0.132)	(0.131)
	Husband	0.188***	0.196***
		(0.049)	(0.048)
	Both	0.309***	0.331***
		(0.089)	(0.086)
Region	North	0.179*	0.188*
		(0.106)	(0.104)
	Yorkshire & The Humber	-0.224	-0.166
		(0.261)	(0.249)
	Midlands	0.122	0.137
		(0.118)	(0.116)
	East	0.315***	0.316***
		(0.08)	(0.08)
	Greater London	0.418***	0.429***
		(0.09)	(0.089)
	South	0.148*	0.156**
		(0.076)	(0.075)
	Wales & Scotland	0.067	0.069
		(0.149)	(0.149)

Notes: 2SLS regressions based on equation (1.9). Robust standard errors in parentheses. The stars indicate significance at the following levels: $*p < 0.10$,** $p < 0.05$,*** $p < 0.01$. Survey data pooled for waves 2010-2012. The dependent variable is food expenditure. The independent variable of interest is household disposable labour income instrumented using the level of education and the type of occupation (white or blue collar, defining white-collar as those individuals who are employers in small or large organisations, or hold a higher managerial, higher professional, lower professional and higher technical, lower managerial and higher supervisory and intermediate position). The self-employed dummy variable is interacted with the characteristic N in question as in equation (1.9). All specifications include year and quarter fixed effects and controls for household and personal characteristics (age and number of children in a non-quadratic fashion, number of cars and rooms, type of tenure, availability of drier and central heating). All dependent and independent variables are in logs and deflated accordingly. The sample is restricted to individuals aged less than 60 years of age. All specifications are weighted using sample weights. The income-gap presented in this table are the IV estimated income-gaps. First stage F-statistics, tests on the quality of the instrument, sample sizes and the OLS specification and estimation are provided in Appendix C. The estimates provided in this table are obtained assuming the covariance between the evasion and reporting component is zero (v and u in the model.

is in line with what has been found in studies using surveys and experiments analysing tax compliance.[36]

Vogel (1974), using a survey in Sweden, found that men think of themselves as having better illegal opportunities than women. Torgler and Valev (2010) using data from eight Western European countries from the World Values Survey and the European Values Survey that span the period from 1981 to 1999 also find that in the case of bribes, men are more frequently asked for bribes by government officials than women. They find in general women are more willing to comply.[37] Opposite results can also be found in the literature such as Baldini *et al.* (2009) who use the discrepancy between the income reported to the fiscal authorities and the income reported to the Survey of Household Income and Wealth in Italy for the region of Modena as a measure of income tax evasion. Using a probit model they find females are more likely to underreport than men. However their study refers only to a particular region of Italy and the results of the study cannot be generalized to the whole population. Schuetze (2002), however, finds that the sex of the individual yields no difference in compliance behaviour.

1.5.3.3 Non-compliance: Region

Knowing how non-compliance is distributed geographically can provide evidence on which regions are less compliant and therefore deserve more attention from the compliance activities of the tax authority.[38] Table 1.4, reports the results, with the most non-compliant regions in the UK being Greater London followed by the East of England, the North and the South.

[36] See Tittle and Grasmick (1997), Torgler and Schneider (2006); for experiments, Spicer and Becker (1980), Spicer and Hero (1985), Baldry (1987), Kleven *et al.* (2011), Friedland *et al.* (1978).

[37] Gender differences in delinquency have been largely investigated in criminology and many theories have been proposed. The fact that this difference in behaviour was due to the inequitable role in society of women and men was discarded since entrance in the labour market of women did not affect crime rates. The most accepted theory to support this difference points at self-control and opportunities to commit crimes as the drivers. Delinquent men are found to be more exposed to delinquent companions that can influence their behaviour through imitation (Mears *et al.* 1998).

[38] We have tried considering all the regions of England as well as Scotland and Wales. Restrictions on the sample size due to the Secure Access nature of the data used led to the pooling of certain regions together in order to achieve a sufficient sample size that ensured anonymity.

1.6 Concluding remarks

This paper has provided estimates for the income-gap in Great Britain using data from 2010-2012. The empirical evidence presented (with concerns of potential endogeneity issues addressed) suggests that on average the self-employed taxpayers report 80.4% of their true income (or, equivalently, the income-gap is 19.6%). Taking into account that for the period between 2010-12 self-employment income in the UK represented 5.5% of GDP, this yields an estimate of unreported taxable income during this period of 1.6% of GDP. The magnitude of the income income-gap does not vary significantly along alternative expenditure measures praising the robustness of the results.

The estimated model has relied on somewhat restrictive assumptions that, arguably, raise concerns about whether the observed discrepancy in reported income is driven by other sources different to misreporting by the self-employed. Interestingly, the robustness checks performed have shown that the income-gap is neither driven by preference heterogeneity nor by the possibility of households being financially constrained and thus using their savings to fund current consumption, nor can it be attributable to measurement error from the survey. Importantly, too, the analysis has elaborated on the characteristics of the non-compliant taxpayers analysing a range of characteristics beyond those available in tax returns. It has emerged that men are less compliant than women and households of two self-employed more non-compliant than households of one; age is inversely related to compliance; the income-gap is also found to vary across regions.

Measuring the income-gap and the range of characteristics most associated to it substantially improves our understanding of how to design compliance activities—audit strategies or public campaigns—to be more effective.

1.7 References

Ahmed, N., Brzozowski, M., Crossley, T.F. (2006). *Measurement errors in recall food consumption data*. IFS Working Paper W06/21, Institute Fiscal Studies, London.

Aguiar, M., Hurst, E. (2004). Consumption vs. expenditure. *Journal of Political Economy*, October 2005, 113(5), 919-948.

Attanasio, O., Battistin, E., Leicester, A. (2004). *Inequality, Poverty and their evolution in the US: Consumption and Income information in the Consumer Expenditure Survey*. Notes prepared for the National Poverty Center's ASPE-Initiated Workshop on Consumption among Low-Income Families. Washington DC November 5, 2004.

Baldini, M., Bosi, P., Lalla, M. (2009). Tax Evasion and Misreporting in Income Tax Returns and Household Income Surveys. *Politica Economica*, 3, 333-348.

Baldry, J. (1987). Income Tax Evasion and the Tax Schedule: Some Experimental Results. *Public Finance*, 42, 357-383.

Battistin, E. (2003). *Errors in Survey Reports of Consumption Expenditures*. IFS Working Paper W03/07, Institute Fiscal Studies, London.

Bee, A., Meyer, B.D., Sullivan, J.X. (2012). *The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?*. No. w18308. National Bureau of Economic Research.

Brewer, M., O'Dea, C. (2012). *Measuring living standards with income and consumption: evidence from the UK* (No. 2012-05). ISER Working Paper Series.

Browning, M., Crossley, T., Weber, G. (2003). Asking Consumption Questions in General Purpose Surveys. *Economic Journal*, Royal Economic Society, vol. 113(491), pages F540-F567, November.

Clotfelter, C. (1983). Tax Evasion and Tax Rates: An Analysis of Individual Returns. *Review of Economics and Statistics* 65, 363-373.

Cowell, F.A. (1990). *Cheating the government: The economics of evasion.* MIT Press Books. 1990;1.

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* Baltimore: Johns Hopkins University Press.

Deaton, A. (2005). Measuring Poverty in a Growing World (Or Measuring Growth in a Poor World). *Review of Economics and Statistics*, 87(1), pp1-19.

Dilnot, A., Morris, C. N. (1981). What do we know about the black economy?. *Fiscal Studies*, 2(1), 58-73.

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

Feinstein, J. (1991). An Econometric Analysis of Income Tax Evasion and Its Detection. *The RAND Journal of Economics*, 22(1), 14-35.

Fiscalis TGPG. (2016). *The concept of tax gaps: Report of VAT Gap estimations*, EU. Retrieved from: `https://ec.europa.eu/taxation_customs/sites/taxation/files/docs/body/tgpg_report_en.pdf`

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

Gibson, J. (2002). Why does the Engel method work? Food demand, economies of size and household survey methods. *Oxford Bulletin of Economics and Statistics*, 64(4), 341-359.

Gray, PG. (1955). The memory factor in social surveys. *Journal of the American Statistics Association.* 50:344–63

Grondin, C., Michaud, S. (1994). Data Quality of Income Data Using Computer-Assisted Interview: The Experience of the Canadian Survey of Labour and Income Dynamics. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 830-835.

HMRC (2015). *Expenses if you're self-employed*. Retrieved from: https://www.gov.uk/expenses-if-youre-self-employed/overview

Hurst, E., Li, G., Pugsley, B. (2014). Are Household Surveys Like Tax Forms: Evidence from Income Underreporting of the Self Employed. *The Review of Economics and Statistics*, 96(1), 19-33.

Johansson, E. (2005). An estimate of self-employment income underreporting in Finland. *Nordic Journal of Political Economy*, 31, 99-109.

Kleven, H., Knudsen, M., Kreiner, C., Pedersen, S., Saez, E. (2011). Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, 79(3), 651-692.

Lyssiotou, P., Pashardes, P., Stengos, T. (2004). Estimates of the Black Economy based on Consumer Demand Approaches. *The Economic Journal*, 114, 622-640.

Marquis, K. and Moore, J. (1990). Measurement Errors in SIPP Program Reports. *Proceedings of the Sixth Annual Research Conference*. Washington, DC: U.S. Bureau of the Census, March, 1990, 721-745.

Maynes, E. (1968). Minimizing Response Errors in Financial Data: The Possibilities. *Journal of the American Statistical Association*, 63, 214-227.

Mears, D., Ploeger, M., & Warr, M. (1998). Explaining the Gender Gap in Delinquency: Peer Influence and Moral Evaluations of Behavior. *Journal of Research in Crime and Delinquency*, 31, 251-266.

Meghir, C., Robin, JM. (1992). Frequency of purchase and the estimation of demand systems. *Journal of Econometrics* , 53 (1-3) pp. 53-85.

Meyer, B. and Sullivan, J. (2010), *Five Decades of Consumption and Income Poverty*. Harris School of Public Policy Working Paper 09.07, University of Chicago.

Moore, J. C., Stinson, L.L., Welniak, E.J. (2000). Income measurement error in surveys: A review. *Journal of Official Statistics*-Stockholm. 16.4 (2000): 331-362.

Neter, J., Waksberg J. (1964). A study of response errors in expenditures data from household interviews. *Journal of the American Statistical Association.* 59:18–55Palumbo, MG, 1999.

Pissarides, C. A., Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, 39(1), 17-32.

Scott, C., Amenuvegbe, B. (1991). Recall loss and recall duration: an experimental study in Ghana. *Inter-Stat*, 4(1): 31-55.

Schuetze, H. (2002). Profiles of Tax Non-compliance Among the Self-Employed in Canada: 1969 to 1992. *Canadian Public Policy*, 28(2), 219-237.

Spicer, M., Becker, L. (1980). Fiscal Inequity and Tax Evasion: An Experimental Approach. *National Tax Journal*, 171-175.

Spicer, M., Hero, R. (1985). Tax Evasion and Heuristics. A Research Note. *Journal of Public Economics*, 26, 263-267.

Sudman, S., Bradburn, N. (1973). Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association*, 68(344): 805-815.

Tittle, C. R., Grasmick, H. G. (1997). Criminal behavior and age: A test of three provocative hypotheses. *Journal of Criminology Literature*, 88, 309.

Torgler, B., Valev, N. (2010). Gender and public attitude toward corruption and tax evasion. *Contemporary Economic Policy*, 28, 554-568.

Torgler, J., Schneider, F. (2006). *What shapes attitudes toward paying taxes?*. Working Paper No. 0608, Department of Economics, Johannes Kepler University of Linz.

Vogel, J. (1974). Taxation and Public Opinion in Sweden: an Interpretation of Recent Survey Data. *National Tax Journal*, 27, 499-513.

Withey, S. (1954). Reliability of Recall of Income. *Public Opinion Quarterly*, 18, 197-204.

Wodon, Q. (2007). *Growth and Poverty Reduction. Case Studies from West Africa.* World Bank Working Paper no. 79.

Appendix A

This appendix discusses the choice of estimates: point (as in Hurst *et al.* (2014)) versus bounds (as in Pissarides and Weber (1989)). It demonstrates that obtaining a point estimate is a feature of the assumptions imposed regarding the key parameters, and that by changing the relationship between the income variables one can obtain both types of estimates of underreporting. Four cases outlining the different relationships between the variables are presented.

Hurst *et al.* (2014) have developed an abridged version of Pissarides and Weber (1989) which does not adjust for the distinct variance of income of the self-employed. This, in the present framework, is equivalent to setting $n = 1$ expressing the relationships in (1.2) and (1.3) as, respectively,

$$\ln Y_i^a = \ln k_i + \ln Y_i^r, \tag{A.1}$$

and

$$\ln Y_i^a = \ln Y_i^p + \Omega' X_{it} + v_{it}, \tag{A.2}$$

with the first term in (A.2), being permanent income, $\ln Y_i^p$, and the second transitory, $(\Omega' X_{it} + v_{it})$. As transitory income is just composed of the characteristics of the households, X_{it}, for which we already control for in equation (1.1) and an error term, v_{it}, this would just lead to a reparametrization of these two components. Ignoring all the assumptions about the differing volatility of income for the two groups one arrives at an estimation of income underreporting given by

$$\bar{k} = \exp\left(\frac{\gamma_j}{\beta_j}\right), \tag{A.3}$$

which relates to the estimate obtained by Hurst given by *et al.* (2014)

$$\bar{d} = \exp\left(-\frac{\gamma_j}{\beta_j}\right), \tag{A.4}$$

with a coefficient of underreporting given by

$$1 - \bar{d} = 1 - \frac{1}{\bar{k}}. \tag{A.5}$$

In an online appendix, Hurst *et al.* (2014) include the possibility of differing volatility of income for the two groups introducing the assumptions of the Pissarides and Weber (1989) framework (shown in equation (1.4) and the lognormality assumptions in (2.7)). They arrive at a point estimate which conforms with the lower bound of Pissarides and Weber (1989). Reaching a point estimate (in this case the lower bound) is a feature of the relationship between the parameters of permanent income and underreporting.

To see this, if one was to rewrite Hurst *et al.* (2014) in a more familiar notation where

$$\ln Y_i^a = \ln p_i + \ln Y_i^p, \tag{A.6}$$

$$\ln Y_i^r = \ln d_i + \ln Y_i^a, \tag{A.7}$$

then

$$\ln Y_i^p = \ln Y_i^r - \ln d_i - \ln Y_i^a \tag{A.8}$$

The fact that $\ln d_i$ and $\ln p_i$ (hence $\sigma_{v_S}^2$ and $\sigma_{u_S}^2$) enter the consumption equation, (1.1), with the same sign allows the substitution of equation (1.15) into (1.11) which can be rewritten as

$$\ln \overline{d} = -\frac{\gamma_j}{\beta_j} + \frac{1}{2}\left(\sigma_{v_S}^2 + \sigma_{u_S}^2 - \sigma_{u_E}^2\right), \tag{A.9}$$

obtaining a point estimate without the need of bounds

$$\ln \overline{d} = -\frac{\gamma_j}{\beta_j} + \frac{1}{2}\left(\sigma_{Y_S}^2 - \sigma_{Y_E}^2\right), \tag{A.10}$$

where the coefficient of income underreporting is just $1 - \bar{d}$. This would correspond to the first case, **Case A**.

Likewise one can obtain a point estimate, in this case equal to the upper bound if we estimate the scaling factor of underreporting k_i and introduce a random variable $P_i = 1/p_i$, that relates actual to permanent income assuming that actual income is a fraction of permanent income (**Case B**). In this case we can write

$$\ln Y_i^p = \ln P_i + \ln Y_i^a, \tag{A.11}$$

$$\ln Y_i^a = \ln k_i + \ln Y_i^r, \tag{A.12}$$

and thus

$$\ln Y_i^p = \ln k_i + \ln P_i + \ln Y_i^r, \tag{A.13}$$

This allows direct substitution using (1.15) as again $\sigma_{v_S}^2$ and $\sigma_{u_S}^2$ would have the same direction in the sign thereby eliminating the need to make assumptions about the bounds. The point estimate obtained in this case corresponds to the upper bound of the Pissarides and Weber (1989) estimate defined as

$$\ln \overline{k}_h = \frac{\gamma_j}{\beta_j} + \frac{1}{2}\left(\sigma_{Y_S}^2 - \sigma_{Y_E}^2\right). \tag{A.14}$$

On the other hand, bounded estimates are obtained in two instances. In the first, (**Case C**), if we were to estimate the scaling factor of underreporting k_i, instead of d_i which relate as $k_i = 1/d_i$, and make use of the random variable p_i to relate permanent and actual income, then

$$\ln Y_i^a = \ln p_i \ln Y_i^p, \tag{A.15}$$

$$\ln k_i + \ln Y_i^r = \ln Y_i^a, \tag{A.16}$$

and, therefore,

$$\ln Y_i^p = \ln k_i + \ln Y_i^r - \ln p_i. \tag{A.17}$$

Observe that both $\ln k_i$ and $\ln p_i$ will enter the consumption function with opposite signs, thereby not allowing the direct substitution using (1.15), since $\sigma_{v_S}^2$ and $\sigma_{u_S}^2$ will consequently have opposite signs requiring additional assumptions to reach a bounded estimate. The lower and upper bound are found by

$$\ln \overline{k}_l = \frac{\gamma_j}{\beta_j} - \frac{1}{2}\left(\sigma_{Y_S}^2 - \sigma_{Y_E}^2\right), \tag{A.18}$$

and

$$\ln \overline{k}_h = \frac{\gamma_j}{\beta_j} + \frac{1}{2}\left(\sigma^2_{Y_S} - \sigma^2_{Y_E}\right). \tag{A.19}$$

In the second instance, (**Case D**), if we are to estimate the coefficient of underreporting $1 - d_i$, and assume again that actual income is only a fraction of permanent income using P_i, then one can express

$$\ln Y_r = \ln d_i + \ln Y_i^a, \tag{A.20}$$

$$\ln Y_p = \ln P_i + \ln Y_i^a, \tag{A.21}$$

and, therefore,

$$\ln Y_p = \ln P_i + \ln Y_i^r - \ln d_i. \tag{A.22}$$

As $\ln P_i$ and $\ln d_i$ enter the consumption equation with opposite signs this means that $\sigma^2_{v_S}$ and $\sigma^2_{u_S}$ will have opposite signs and, therefore, direct substitution of (1.15) into (1.11) is not possible. This implies that a point estimate cannot be obtained and the bounded estimate are found using the expressions

$$\ln \overline{k}_l = -\frac{\gamma_j}{\beta_j} - \frac{1}{2}\left(\sigma^2_{Y_S} - \sigma^2_{Y_E}\right), \tag{A.23}$$

and

$$\ln \overline{k}_h = -\frac{\gamma_j}{\beta_j} + \frac{1}{2}\left(\sigma^2_{Y_S} - \sigma^2_{Y_E}\right). \tag{A.24}$$

Summarising the preceding discussion:

- If the random variable that refers to underreporting (k_i or d_i) and the random variable that relates permanent to actual income (p_i and P_i) enter the consumption equation with the same sign, this would allow direct substitution of the income volatility equation leading to point estimates of underreporting that would correspond to either the lower or upper bound of underreporting depending on the assumptions made (cases A and B).

- If the relationship, on the other hand, is inverse this would lead to the bounded estimates of underreporting (cases C and D).

What this all in practice means is that it is at the discretion of the researcher which assumptions about the relationship between the income variables to make.

Appendix B

Summary statistics

Table B.1: Descriptive Statitics of the covariates used in the regressions

		N	mean	sd
Employed	Age	4045	40.053	10.020
	Age Squared	4045	1704.618	824.391
	Number of children	4045	0.833	0.941
	Number of children (squared)	4045	1.579	2.318
	Local Authority Tenant	4045	0.030	0.169
	Other rented	4045	0.122	0.327
	Owner with Mortgage	4045	0.617	0.486
	Number of Cars	4045	1.322	0.711
	Central Heating	4045	0.968	0.176
	Drier	4045	0.628	0.483
	Number of Rooms	4045	5.836	1.780
Self-Employed	Age	733	42.175	9.663
	Age Squared	733	1871.998	825.402
	Number of children	733	0.956	0.984
	Number of children (squared)	733	1.882	2.554
	Local Authority Tenant	733	0.027	0.161
	Other rented	733	0.112	0.316
	Owner with Mortgage	733	0.606	0.489
	Number of Cars	733	1.294	0.758
	Central Heating	733	0.969	0.173
	Drier	733	0.669	0.471
	Number of Rooms	733	6.218	1.898

Appendix C

All results are reported in this appendix.

Table C.1: Income-gap using different measures of expenditure

	Food		Utilities		Non-Durables	
	IV	OLS	IV	OLS	IV	OLS
Dummy SE	0.090***	0.067***	0.100***	0.064***	0.122***	0.091***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Income Elasticity	0.436***	0.265***	0.417***	0.149***	0.580***	0.357***
	(0.04)	(0.01)	(0.04)	(0.02)	(0.03)	(0.02)
Variance SE	0.724		0.728		0.726	
Variance E	0.476		0.477		0.477	
Spread	0.248		0.25		0.25	
Income-gap: No correction for income volatility.						
$\bar{\kappa}$	0.187***	0.224***	0.214***	0.348***	0.190***	0.226***
	(0.043)	(0.065)	(0.045)	(0.095)	(0.031)	(0.043)
Income-gap: Correction for income volatility.						
Upper Bound						
$\bar{k}_h$	1.427***	1.495***	1.480***	1.784***	1.434***	1.501***
	(0.075)	(0.125)	(0.085)	(0.26)	(0.054)	(0.084)
$\bar{\kappa}_h$	0.299***	0.331***	0.324***	0.439***	0.303***	0.334***
	(0.037)	(0.056)	(0.039)	(0.082)	(0.026)	(0.037)
Lower Bound						
$\bar{k}_l$	1.060***	1.111***	1.094***	1.319***	1.062***	1.111***
	(0.056)	(0.093)	(0.063)	(0.192)	(0.04)	(0.062)
$\bar{\kappa}_l$	0.056	0.1	0.086	0.242**	0.058	0.100**
	(0.049)	(0.075)	(0.052)	(0.11)	(0.036)	(0.05)
Midpoint						
$\bar{k}_m$	1.243***	1.303***	1.287***	1.551***	1.248***	1.306***
	(0.065)	(0.109)	(0.074)	(0.226)	(0.047)	(0.073)
$\bar{\kappa}_m$	0.196***	0.233***	0.223***	0.355***	0.199***	0.234***
	(0.042)	(0.064)	(0.045)	(0.094)	(0.03)	(0.043)
First stage statistics						
F-stat	228.551		228.875		229.235	
Hansen J	0.284		1.736		0.456	
p-value	0.868		0.420		0.796	
Observations						
Self-Employed	730		728		731	
Employed	4026		4020		4036	
Total	4756		4748		4767	

Notes:

1. The stars indicate significance at the following levels: $*p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$

2. Robust standard errors in parenthesis.

3. The estimates provided in this table are obtained assuming the covariance between the evasion and reporting component is zero (v and u in the model).

Table C.2: Income-gap for different measures of income

	Profit		Drawings		Proxy	
	IV	OLS	IV	OLS	IV	OLS
Dummy SE	0.081**	0.043	0.169***	0.087	0.090***	0.067***
	(0.03)	(0.03)	(0.06)	(0.05)	(0.02)	(0.02)
Income Elasticity	0.405***	0.198***	0.368***	0.163***	0.436***	0.265***
	(0.04)	(0.02)	(0.04)	(0.01)	(0.04)	(0.01)
Variance SE	0.833		0.964		0.724	
Variance E	0.547		0.61		0.476	
Spread	0.286		0.354		0.248	
Income gap: No correction for income volatility						
$\bar{\kappa}$	0.180***	0.195	0.368***	0.412**	0.187***	0.224***
	(0.066)	(0.123)	(0.099)	(0.198)	(0.043)	(0.065)
Income gap: Correction for income volatility						
Upper Bound						
$\bar{k}_h$	1.486***	1.514***	2.089***	2.248***	1.427***	1.495***
	(0.12)	(0.231)	(0.327)	(0.758)	(0.075)	(0.125)
$\bar{\kappa}_h$	0.327***	0.339***	0.521***	0.555***	0.299***	0.331***
	(0.054)	(0.101)	(0.075)	(0.15)	(0.037)	(0.056)
Lower Bound						
$\bar{k}_l$	1.002***	1.021***	1.197***	1.288***	1.060***	1.111***
	(0.081)	(0.156)	(0.187)	(0.434)	(0.056)	(0.093)
$\bar{\kappa}_l$	0.002	0.02	0.165	0.224	0.056	0.1
	(0.081)	(0.15)	(0.131)	(0.262)	(0.049)	(0.075)
Midpoint						
$\bar{k}_m$	1.244***	1.267***	1.643***	1.768***	1.243***	1.303***
	(0.1)	(0.194)	(0.257)	(0.596)	(0.065)	(0.109)
$\bar{\kappa}_m$	0.196***	0.211*	0.391***	0.434**	0.196***	0.233***
	(0.065)	(0.121)	(0.095)	(0.191)	(0.042)	(0.064)
First stage statistics						
F-stat	189.417		180.080		228.551	
Hansen J	1.431		2.053		0.284	
p-value	0.489		0.358		0.868	
Observations						
Self-Employed	310		131		730	
Employed	4239		4431		4026	
Total	4549		4562		4756	

Notes:

1. The stars indicate significance at the following levels: $^{*}p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$
2. Robust standard errors in parenthesis.
3. The estimates provided in this table are obtained assuming the covariance between the evasion and reporting component is zero (v and u in the model).

Table C.3: Alternative definition of a self-employed household

	Self-Employed Head		Any Self-Employed	
	IV	OLS	IV	OLS
Dummy SE	0.105***	0.079***	0.105***	0.083***
	(0.02)	(0.02)	(0.02)	(0.02)
Income Elasticity	0.436***	0.267***	0.433***	0.266***
	(0.04)	(0.01)	(0.04)	(0.01)
Variance SE	0.692		0.687	
Variance E	0.473		0.469	
Spread	0.219		0.218	
Income gap: No correction for income volatility				
$\bar{\kappa}$	0.215***	0.257***	0.215***	0.267***
	(0.038)	(0.056)	(0.037)	(0.054)
Income gap: Correction for income volatility				
Upper Bound				
$\bar{k}_h$	1.447***	1.528***	1.444***	1.547***
	(0.069)	(0.116)	(0.067)	(0.113)
$\bar{\kappa}_h$	0.309***	0.346***	0.308***	0.353***
	(0.033)	(0.049)	(0.032)	(0.047)
Lower Bound				
$\bar{k}_l$	1.121***	1.184***	1.123***	1.202***
	(0.054)	(0.089)	(0.052)	(0.088)
$\bar{\kappa}_l$	0.108**	0.155**	0.109***	0.168***
	(0.043)	(0.064)	(0.042)	(0.061)
Midpoint				
$\bar{k}_m$	1.284***	1.356***	1.284***	1.375***
	(0.061)	(0.102)	(0.06)	(0.101)
$\bar{\kappa}_m$	0.221***	0.263***	0.221***	0.272***
	(0.037)	(0.056)	(0.036)	(0.053)
First stage statistics				
F-stat	228.072		231.037	
Hansen J	0.206		0.190	
p-value	0.902		0.909	
Observations				
Self-Employed	899		976	
Employed	3857		3780	
Total	4756		4756	

Notes:

1. The stars indicate significance at the following levels: $^*p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

2. Robust standard errors in parenthesis.

3. The estimates provided in this table are obtained assuming the covariance between the evasion and reporting component is zero (v and u in the model).

Table C.4: Use of documents

	Any documents		HMRC documents	
	IV	OLS	IV	OLS
Dummy SE	0.099**	0.085**	0.159**	0.152**
	(0.04)	(0.04)	(0.08)	(0.07)
Income Elasticity	0.431***	0.271***	0.430***	0.284***
	(0.04)	(0.02)	(0.04)	(0.02)
Variance SE	0.762		0.764	
Variance E	0.477		0.477	
Spread	0.285		0.288	
Income gap: No correction for income volatility				
$\bar{\kappa}$	0.206**	0.268**	0.310**	0.414***
	(0.081)	(0.111)	(0.127)	(0.153)
Income gap: Correction for income volatility				
Upper Bound				
$\bar{k}_h$	1.503***	1.630***	1.732***	2.040***
	(0.154)	(0.246)	(0.319)	(0.531)
$\bar{\kappa}_h$	0.335***	0.386***	0.422***	0.510***
	(0.068)	(0.093)	(0.106)	(0.128)
Lower Bound				
$\bar{k}_l$	1.055***	1.145***	1.212***	1.427***
	(0.108)	(0.173)	(0.223)	(0.372)
$\bar{\kappa}_l$	0.053	0.126	0.175	0.299
	(0.097)	(0.132)	(0.152)	(0.182)
Midpoint				
$\bar{k}_m$	1.279***	1.387***	1.472***	1.733***
	(0.131)	(0.21)	(0.271)	(0.451)
$\bar{\kappa}_m$	0.218***	0.279**	0.320**	0.423***
	(0.08)	(0.109)	(0.125)	(0.15)
First stage statistics				
F-stat	210.586		205.027	
Hansen J	0.609		1.081	
p-value	0.737		0.583	
Observations				
Self-Employed	175		57	
Employed	4036		4036	
Total	4211		4093	

Notes:

1. The stars indicate significance at the following levels: $*p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$

2. Robust standard errors in parenthesis.

3. The estimates provided in this table are obtained assuming the covariance between the evasion and reporting component is zero (v and u in the model).

Table C.5: Other alternative explanations for the income-gap

	Not Financially Constrained		No claim of Business Expenses	
	IV	OLS	IV	OLS
Dummy SE	0.098**	0.09**	0.105***	0.076***
	(0.04)	(0.04)	(0.03)	(0.02)
Income Elasticity	0.394***	0.328***	0.441***	0.158***
	(0.08)	(0.03)	(0.04)	(0.02)
Variance SE	0.744		0.71	
Variance E	0.462		0.476	
Spread	0.283		0.233	
Income gap: No correction for income volatility				
$\bar{\kappa}$	0.220***	0.239**	0.212***	0.381***
	(0.085)	(0.098)	(0.05)	(0.096)
Income gap: Correction for income volatility				
Upper Bound				
$\bar{k}_h$	1.520***	1.558***	1.457***	1.855***
	(0.166)	(0.201)	(0.092)	(0.288)
$\bar{\kappa}_h$	0.342***	0.358***	0.314***	0.461***
	(0.072)	(0.083)	(0.043)	(0.084)
Lower Bound				
$\bar{k}_l$	1.081***	1.108***	1.105***	1.406***
	(0.118)	(0.143)	(0.069)	(0.218)
$\bar{\kappa}_l$	0.075	0.098	0.095*	0.289***
	(0.101)	(0.116)	(0.057)	(0.11)
Midpoint				
$\bar{k}_m$	1.300***	1.333***	1.281***	1.630***
	(0.142)	(0.172)	(0.08)	(0.253)
$\bar{\kappa}_m$	0.231***	0.250***	0.219***	0.387***
	(0.084)	(0.097)	(0.049)	(0.095)
First stage statistics				
F-stat	56.91		222.099	
Hansen J	9.46		1.623	
p-value	0.009		0.444	
Observations				
Self-Employed	151		531	
Employed	979		4002	
Total	1130		4533	

Notes:

1. The stars indicate significance at the following levels: $*p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$
2. Robust standard errors in parenthesis.
3. The estimates provided in this table are obtained assuming the covariance between the evasion and reporting component is zero (v and u in the model).
4. For the estimation regarding business expenses, the dependent variable is expenditure on utilities.

Table C.6: Income-gap by age of the HRP

	IV			OLS		
	< 35	$35-45$	> 45	< 35	$35-45$	> 45
Dummy SE	0.205***	0.082**	0.064*	0.182***	0.058**	0.037
	(0.07)	(0.03)	(0.03)	(0.06)	(0.03)	(0.03)
Income Elasticity		0.437***			0.254***	
		(0.04)			(0.01)	
Variance SE	0.475	0.732	0.758			
Variance E		0.477				
Spread	-0.002	0.255	0.282			
Income-gap no correction for volatility						
$\bar{\kappa}$	0.375***	0.172***	0.136**	0.511***	0.205**	0.136
	(0.093)	(0.063)	(0.066)	(0.124)	(0.089)	(0.102)
Income-gap correction for income volatility						
Upper Bound						
$\bar{k}_h$	1.597***	1.409***	1.377***	2.042***	1.468***	1.377***
	(0.238)	(0.108)	(0.105)	(0.518)	(0.165)	(0.162)
$\bar{\kappa}_h$	0.374***	0.290***	0.274***	0.510***	0.319***	0.274***
	(0.093)	(0.054)	(0.055)	(0.124)	(0.077)	(0.085)
Lower Bound						
$\bar{k}_l$	1.601***	1.035***	0.972***	2.046***	1.079***	0.973***
	(0.238)	(0.079)	(0.074)	(0.519)	(0.121)	(0.115)
$\bar{\kappa}_l$	0.375***	0.034	-0.028	0.511***	0.073	-0.028
	(0.093)	(0.074)	(0.079)	(0.124)	(0.104)	(0.121)
Midpoint						
$\bar{k}_m$	1.599***	1.222***	1.174***	2.044***	1.273***	1.175***
	(0.238)	(0.093)	(0.09)	(0.519)	(0.143)	(0.138)
$\bar{\kappa}_m$	0.375***	0.182***	0.149**	0.511***	0.215**	0.149
	(0.093)	(0.063)	(0.065)	(0.124)	(0.088)	(0.1)
First stage statistics						
F-stat		228.410				
Hansen J		0.389				
p-value		0.823				
Observations						
Self-Employed	64	342	325	64	342	325
Employed			3619			
Total		4756			4756	

Notes:

1. The stars indicate significance at the following levels: $^{*}p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$
2. Robust standard errors in parenthesis.
3. The estimates provided in this table were obtained assuming the covariance between the evasion and reporting component was zero (v and u in the model).

Table C.7: Income-gap by gender

	IV			OLS		
	Wife	Husband	Both	Wife	Husband	Both
Dummy SE	-0.022	0.091***	0.162***	-0.028	0.066***	0.095**
	(0.05)	(0.03)	(0.06)	(0.06)	(0.02)	(0.04)
Income Elasticity		0.438***			0.255***	
		-0.04			-0.01	
Variance SE	0.682	0.711	0.856	0.682	0.711	0.856
Variance E			0.477			
Spread	0.205	0.235	0.379	0.205	0.235	0.379
Income-gap: No correction for income volatility						
$\bar{\kappa}$	-0.051	0.188***	0.309***	-0.116	0.227***	0.311***
	(0.132)	(0.049)	(0.089)	(0.247)	(0.07)	(0.119)
Income-gap: Correction for income volatility						
Upper Bound						
$\bar{k}_h$	1.072***	1.416***	1.864***	1.009***	1.488***	1.869***
	(0.135)	(0.085)	(0.241)	(0.223)	(0.136)	(0.322)
$\bar{\kappa}_h$	0.067	0.294***	0.464***	0.009	0.328***	0.465***
	(0.117)	(0.042)	(0.069)	(0.219)	(0.061)	(0.092)
Lower Bound						
$\bar{k}_l$	0.845***	1.072***	1.124***	0.795***	1.126***	1.127***
	(0.106)	(0.064)	(0.145)	(0.176)	(0.103)	(0.194)
$\bar{\kappa}_l$	-0.184	0.067	0.11	-0.257	0.112	0.113
	(0.149)	(0.056)	(0.115)	(0.278)	(0.081)	(0.153)
Midpoint						
$\bar{k}_m$	0.958***	1.244***	1.494***	0.902***	1.307***	1.498***
	(0.12)	(0.075)	(0.193)	(0.199)	(0.119)	(0.258)
$\bar{\kappa}_m$	-0.043	0.196***	0.331***	-0.109	0.235***	0.332***
	(0.131)	(0.048)	(0.086)	(0.245)	(0.07)	(0.115)
First stage statistics						
F-stat		228.801				
Hansen J		0.268				
p-value		0.874				
Observations						
Self-Employed	80	518	133	80	518	133
Employed			3619			
Total		4756			4756	

Notes:

1. The stars indicate significance at the following levels: $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
2. Robust standard errors in parenthesis.
3. The estimates provided in this table are obtained assuming the covariance between the evasion and reporting component is zero (v and u in the model).

Table C.8: Estimates of self-employment underreporting by region, 2010-2012. Pooled data

	IV							OLS						
	North	Yorkshire & The Humber	Midlands	East	Greater London	South	Wales & Scotland	North	Yorkshire & The Humber	Midlands	East	Greater London	South	Wales & Scotland
Dummy SE	0.084	-0.086	0.055	0.161***	0.230***	0.068*	0.029	0.024	-0.146**	0.024	0.153***	0.239***	0.051	0.015
	(0.06)	(0.09)	(0.06)	(0.05)	(0.06)	(0.04)	(0.07)	(0.05)	(0.07)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)
Income Elasticity				0.425***							0.248***			
				(0.04)							(0.01)			
Variance SE	0.725	0.924	0.773	0.591	0.792	0.706	0.614	0.725	0.924	0.773	0.591	0.792	0.706	0.614
Variance E				0.477							0.477			
Spread	0.438	0.637	0.486	0.304	0.505	0.419	0.327	0.438	0.637	0.486	0.304	0.505	0.419	0.327
Income-gap: No correction for income volatility														
$\bar{\kappa}$	0.179*	-0.224	0.122	0.315***	0.418***	0.148*	0.067	0.091	-0.802*	0.094	0.459***	0.617***	0.187	0.059
	(0.106)	(0.261)	(0.118)	(0.08)	(0.09)	(0.076)	(0.149)	(0.189)	(0.486)	(0.18)	(0.116)	(0.084)	(0.116)	(0.201)
Income-gap: Correction for income volatility														
Upper Bound														
$\bar{k}_h$	1.414***	1.118***	1.370***	1.552***	2.098***	1.344***	1.155***	1.276***	0.759***	1.327***	1.964***	3.193***	1.408***	1.146***
	(0.182)	(0.238)	(0.185)	(0.181)	(0.326)	(0.12)	(0.184)	(0.265)	(0.205)	(0.264)	(0.42)	(0.703)	(0.201)	(0.244)
$\bar{\kappa}_h$	0.293***	0.105	0.270***	0.356***	0.523***	0.256***	0.134	0.217	-0.317	0.246	0.491***	0.687***	0.290***	0.127
	(0.091)	(0.191)	(0.098)	(0.075)	(0.074)	(0.066)	(0.138)	(0.163)	(0.355)	(0.15)	(0.109)	(0.069)	(0.101)	(0.186)
Lower Bound														
$\bar{k}_l$	1.050***	0.597***	0.947***	1.373***	1.405***	1.025***	0.994***	0.948***	0.406***	0.917***	1.738***	2.139***	1.074***	0.986***
	(0.135)	(0.127)	(0.128)	(0.16)	(0.218)	(0.092)	(0.159)	(0.197)	(0.109)	(0.183)	(0.372)	(0.471)	(0.153)	(0.21)
$\bar{\kappa}_l$	0.047	-0.675*	-0.056	0.272***	0.288***	0.025	-0.006	-0.055	-1.465**	-0.09	0.425***	0.533***	0.069	-0.014
	(0.122)	(0.357)	(0.142)	(0.085)	(0.11)	(0.087)	(0.161)	(0.219)	(0.665)	(0.217)	(0.123)	(0.103)	(0.133)	(0.216)
Midpoint														
$\bar{k}_m$	1.232***	0.857***	1.159***	1.463***	1.751***	1.185***	1.074***	1.112***	0.583***	1.122***	1.851***	2.666***	1.241***	1.066***
	(0.158)	(0.183)	(0.156)	(0.171)	(0.272)	(0.106)	(0.171)	(0.231)	(0.157)	(0.223)	(0.396)	(0.587)	(0.177)	(0.227)
$\bar{\kappa}_m$	0.188*	-0.166	0.137	0.316***	0.429***	0.156**	0.069	0.101	-0.717	0.109	0.460***	0.625***	0.194*	0.062
	(0.104)	(0.249)	(0.116)	(0.08)	(0.089)	(0.075)	(0.149)	(0.187)	(0.463)	(0.177)	(0.116)	(0.083)	(0.115)	(0.2)
First stage statistics														
F-stat				224.033										
Hansen J				0.174										
p-value				0.917										
Observations														
Self-Employed	95	57	103	90	90	207	89	95	57	103	90	90	207	89
Employed				4036							4036			
Total	4131	4093	4139	4126	4126	4243	4125	4131	4093	4139	4126	4126	4243	4125

Notes:
1. The stars indicate significance at the following levels: $^{*}p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$
2. Robust standard errors in parenthesis.
3. The estimates provided in this table are obtained assuming the covariance between the evasion and reporting component is zero (v and u in the model).

Chapter 2

Moonlighters: An Estimation of their Income-Gap.

2.1 Introduction

As tax evasion is intimately linked to *visibility*, wage income is typically found to have the highest compliance rates with a reported 99% in the US Tax Gap estimation for 2008-2010 (IRS, 2016). However, this does not entail that employees cannot underdeclare their *true* labour income. In spite of the fact that withholding prevents employees from having the *opportunity* and therefore *choice* of how much to declare from their main job, wilful tax evaders can perform a second job and conceal the income derived from it, failing to register this second source of labour income with the tax administration. These individuals are going to be referred to as moonlighters.[1]

The lack of *visibility* and paper trail makes it difficult for tax administrations to uncover moonlighting practices. Using a survey of undeclared work conducted by Pedersen (2003), HMRC estimates the moonlighters tax gap in the UK to amount to 2 billion in 2014-2015 (HMRC, 2016).[2]

[1] The term 'moonlighter' is generally used with the connotation of concealment of the second job, this is informal second job holders. However, the term has been used in the literature about multiple job holding to refer to formal second jobs as well. In the paper the term is used to refer to informal moonlighters. Note that HMRC in their Tax Gap report use the term 'moonlighters' with the informal meaning. When referring to multiple-job holding, we will label it formal moonlighting for ease of explanation.

[2] The methodology that HMRC uses to estimate the moonlighters' tax gap appears as illustrative

The lack of continuity of undeclared work surveys difficults tracking the phenomenon. The present paper aims at making headway into this question and proposes an innovative method to measure income from moonlighting practices that relies on publicly available survey data that is released on a timely manner. This will enable the update of the estimate with a higher frequency and at a lower cost than running a specific survey. Using a survey of general scope allows us to shelter from a common shortcoming of specific surveys: the reliability of self-reported non-compliance.

The proposed method exploits the fact that the non-observed economy is concentrated in a few sectors (OECD, 2002; NAO, 2008). Those industries in which non-compliance is typically concentrated are going to be considered high-risk industries—as in high-risk of underreporting—and the rest of the industries low-risk. All employed individuals—self-employment is not considered[3]—belong to one of the two groups. This implies that while expenditure is accurately reported by both groups, the two groups differ in their possibility of underreporting their incomes: the low-risk group's formal labour income is subject to withholding taxes and as they do not have opportunities to perform second jobs their income is reported fully. Practically this implies that given a certain level of expenditure, any discrepancies between the income level reported by the high-risk and the low-risk is indicative of a second source of labour income.

Two approaches are proposed for the definition of high-risk industries: A *factual approach* based on the information available from the intelligence of revenue agencies, surveys and the literature about which industries typically concentrate informal suppliers; and a *data-driven approach* which aims at not predetermining industries as high-risk but extracting from the data which industries evidence a discrepancy against the baseline (the rest). On *the data-driven approach*, a subset selection method is used repetitively in order to reach the occupations in which a discrepancy is observed without arbitrarily selecting a baseline occupation as low-risk. The occupations selected in this way by the

in their 'Measuring the Tax Gap report'. The IMF (2013) encouraged further research to provide new measurements of undeclared work to feed into the Tax Gap. As the methodology used by HMRC is based on a survey of undeclared work, the estimates of participation that stem from it would correspond to a lower bound estimate of informal moonlighting as it relies on self-reported non-compliance. Typically, supply is measured with errors due to informal moonlighters failing to answer truthfully about their participation to avoid self-incrimination.

[3] Note that we are interested in measuring moonlighting and thus we ony need to consider those employed by a third-party.

model, this is occupations that observe levels of expenditure that correspond to higher levels of incomes than those reported, are going to be classified as high-risk industries. The industries consistently selected as exhibiting non-compliance are: Construction, Distribution (Wholesale, retail and repair of motor vehicles) and Professionals. Within *the factual approach* two alternative strategies are pursued: a single-industry baseline (manufacturing) and a multi-industry baseline (low-risk—EH for reference). Under the factual approach, the industries highlighted as high-risk and the income-gap estimated are comparable to those obtained using the data-driven method.

Using data from the Living Costs and Food Survey (Secure Access)[4] for the period 2010-13, we find that employees working in Construction exhibit an income-gap of 19.6%, those working in the Real Estate and Professionals industry of 22.9% and those working in the Distribution sector of 19.8%. This translates into an estimate of the income-gap of 2.6% of GDP for the years 2010-2013. The research finds that these estimates over time remain statistically indistinguishable. The fact that informal moonlighting is relatively stable over time maps the behaviour of the formal moonlighting market in the UK.

The remainder of the paper is organised as follows. Section 2.2 provides a brief literature review, section 2.3 outlines the model to estimate the income-gap by moonlighters. Section 2.4 describes the data and the empirical strategy. The results are discussed in Section 2.5 and Section 2.6 concludes.

2.2 Brief Literature Review

The informal economy's hard to measure nature has led researchers to find innovative ways of understanding what lies beneath the informal economy and how it is to be measured.[5] These can be classified in two broad categories: macro methods (indirect), that use indicators to proxy its size using macroeconomics variables; and micro methods

[4] Office for National Statistics, Department for Environment, Food and Rural Affairs. (2014). Living Costs and Food Survey, 2006-2012: Secure Access. [data collection]. 5th Edition. UK Data Service. SN: 7047

[5] We use the informal economy as a broad term that includes all activities that are productive and legal but that are concealed to avoid obligations with any tax authorities or other institutions. It would also include home production and activities of unincorporated enterprises. Generally, illegal activities are taken off the framework and studied separately. The reason for not using the term Non-Observed Economy as defined by the OECD (2002) is to not account for illegal acts.

(direct) that use individual's data. As the method to be used in the present study falls into the second category, no further comment is pursued on macro methods.[6]

Micro methods use mainly individual survey data either from *specific* (black activities surveys) or g*eneral purpose surveys*; and *administrative data* typically from audits or tax returns.

Surveys on black activities have a longer tradition in Scandinavian countries, and particularly Denmark, Norway and the Netherlands. Pedersen (2003) conducted a survey of this kind in Scandinavia (Denmark, Norway, Sweden), Germany and Great Britain finding that the extent of black activities in this countries was around 1.8%, 1.1%, 1%, 1.3% and 0.6% of GDP respectively. Despite the existence of a substantial number of surveys of this kind, the discrepancy in their methodologies and their circumscription to a certain moment in time and country hampers comparability.[7]

The Eurobarometer survey of undeclared work expanded the format of Pedersen (2003) to all EU countries in two waves—2007 and 2013. This survey offers a common framework to study undeclared work across a wide group of countries addressing the issue of comparability. Undeclared work practices were addressed from the supply and the demand side. Survey participants were asked about their purchases, participation and whether they knew someone who participated.[8]

Smith and Adams (1987) on a report for the Internal Revenue Service (IRS) conducted a survey of income flows in informal markets. The survey consisted on a two stage process where households were questioned about their purchases of informal goods and months after on a second phase, they were asked about their informal sales. Despite addressing supply and demand, this survey concentrated on measuring the size of the hidden economy from a demand side as the supply side is measured with more error. A

[6] For a review of methods see Gemmell and Hasseldine (2012) and OECD (2002).

[7] For evidence of other survey studies on undeclared practices see Van Eck and Kazemier (1988) for Holland; Feld and Larsen (2005), Merz and Wolff (1993) for West Germany; Isachsen and Strøm (1985) for Norway; and Pedersen (2003) for Germany, Great Britain and Scandinavia. We focus on Pedersen (2003) as it presents a good outset of the methodology used for the survey and careful interpretation of the results.

[8] These direct surveys contain a high volume of information linked to participation. See Cabral *et al.* (2015) that uses the Eurobarometer Survey of 2013 to explore the links between classical deterrents and intrinsic motivations to comply in the decision of participation.

particular caveat to this survey is that as recognised by the authors, the survey relies on the ability of interviewees to distinguish informal from formal vendors. Information that the interviewees hold about whether vendor is indeed registered or not can be vague or limited making the judgement (and consequently the data to the survey) highly subjective. They are able to provide a size of the informal economy in 1985-1986 of $83.7 billion which translates into a 2.1% of GNP. Note also that this figure represents the value of gross receipts. According to the Tax Compliance Measurement Programme (TCMP) 1981, reported net income represented around 51% of gross receipts. Applying this figure means that net unrecorded income is estimated to be around $42.7 billions.

Undeclared work surveys provide granular information regarding the profile of the individual, the industry they are involved at on their (in)formal employment and provide direct information on the reasons for participation. However, there are measurement concerns that hover over specific surveys on sensitive topics, typically the reliability of self-reports. The sensitivity of the topic can lead to non-responses, partial responses or deviations from the truth that can affect measurement.[9]

Alm and Erard (2016) exploit the use of *general purpose national surveys* and *administrative data*. The authors estimate the informal supplier gap comparing data on earnings from the Current Population Survey, the Consumer Expenditure Survey and the Revenue Service statistics for certain predetermined industries where informal activities are believed to be concentrated.

The IRS defines informal suppliers as "individuals who provide products or services through informal arrangements which frequently involve cash-related transactions or

[9] In Pedersen (2003), the highest refusal rates were found in Great Britain (24.7%), Norway (18.9%) and Germany (17.2%). Item non-response ranged from 0.4-2.8%, this is refusing to answer a particular question on black activities. These responses rates are thought to be usual for the type of survey. As is the case of the Eurobarometer survey, these surveys are usually constructed to ensure the maximum reliability of self-reports. There is a typical format in undeclared work surveys. As the questions dealt with are of a sensitive nature, a technique called careful priming is used to build the questionnaire. This entails that the most compromising questions lie at the end (Bradburn *et al.*, 2004). There is typically a statement ensuring of the anonymity of the answers and specific examples of occupations in the informal economy to minimise misconceptions. A particular advantage of the Eurobarometer Survey in this setting is that despite being the questions specific to undeclared work, the survey has a bigger scope of topics. As an example Eurobarometer 79.2 addressed questions on topics as diverse as the internal market, cultural activities, non-urban road use and science and technology. This reduces the salience of the questions on undeclared work although not completely eliminating the bias.

'off the books' accounting practice."(IRS, 1996, p.43). Informal suppliers are often individuals who file a tax return and comply only partially with their tax obligations. Non-filing is also common among informal suppliers. Non-filers are what in the jargon are referred to as *ghosts*. Ghosts together with moonlighters are the two main figures within the informal economy for which evidence is scarce. Erard and Chin-Chin Ho (2001) using a unique data source on audits on non-filers together with audited data on filers, study the decision of filing, analyse the characteristics of non-filers and estimate the amount of tax liabilities due for the ghosts.

Another strand of the literature, in which this study lies, has relied on the analysis of the expenditure-income pattern of households to elicit black economy activities. Dilnot and Morris (1981) using household expenditure data flagged households for which the discrepancy between the two magnitudes was large and could not be explained by household circumstances. Pissarides and Weber (1989) used a more refined approach to extract the amount of underreporting of self-employed households. The method hinges on the assumption that the employees, as their income is subject to withholding taxes, report their income correctly. Self-employment income not being subject to third-party reporting is likely to be misreported. Expenditure is correctly reported for both groups—self-employed and employed. The relationship between income-expenditure of the employed is thought to be correct and therefore can be used as a benchmark to assess the level of misreporting of income of the self-employed. This approach has been used to address self-employment underreporting in different countries.[10]

Unlike previous papers, we focus on moonlighters rather than self-employed. In order to do so, we relax the assumption of a correct reporting of labour income by the employed to embrace the figure of *moonlighters*. This is, formal labour income is subject to withholding taxes and thus is accurately reported, but income from an undeclared second source of labour income may exist.[11] As Alm and Erard (2016), the model is

[10] See Schuetze (2002), Johansson (2005), Hurst *et al.* (2014), Cabral *et al.* (2016).

[11] Our focus is not the self-employed who act as informal suppliers (filers or nonfilers) but the moonlighters, individuals employed on a formal job who hold a secondary job that is not declared. For clarification purposes, if an entrepreneur files a return, any misreporting of income will belong to the category of *evasion* in the HMRC Tax Gap framework whereas non-filers either entrepreneurs or employed are considered as *ghosts*. In this research the problem of non-filers is not encountered because moonlighters are conceptually defined as employed whose first formal job is through PAYE but that

built to recognise the higher propensity of certain occupations to concentrate black activities.

Cowell (1990) addressed the moonlighters problem from a theoretical perspective. The decision to moonlight was laid out as a mere labour supply model with a fixed number of formal hours and a decision between leisure and an informal job source. Thus, we concentrate on the empirical estimation of their income-gap.

2.3 The model

The identifying assumption this paper uses is that the relationship between expenditure and income can be expressed by a log-linear Engel curve of the form,

$$\ln C_{ij} = \beta_j \ln Y_i^p + X_i\alpha_j + \varepsilon_{ij}, \tag{2.1}$$

where C_{ij} represents expenditure on good j which is assumed to be correctly reported by all households i, Y_i^p is permanent income and thus β_j is the elasticity of income for good j, X_i is a vector of households controls, and ε_{ij} is a white noise error.

Actual income and permanent income can differ due to the transitory component of income and so,

$$\ln Y_i^a = \ln p_i + \ln Y_i^p, \tag{2.2}$$

where p_i is a random variable that will capture the effect of transitory shocks.

Income reported in the survey Y_i^r can also differ from actual income if misreporting is present. All households in the model derive their main source of income from employment by a third party, hence, self-employment is not considered.[12] Misreporting of income from their main job is not possible as income is subject to withholding taxes, and therefore displays high-visibility. Income reported from their formal job, $Y_i^{r\mathrm{f}}$, then

do not record their second job. Informally, one can talk of ghosts as those individuals "off-the-books" and moonlighters as those working "on-the-side" with the requisite of being employed in their formal.

[12] A household is defined as employed if it derives less than 25% of their income from self-employment sources. Robustness checks have been conducted assuming different definitions of an employed household thereby not allowing any self-employment income source and the results are robust. The criteria of choosing a cut-off in the importance of self-employment sources of income is to avoid misclassification of households due to the self-report of their employment status.

equals their true income from their formal job, Y_i^{af}. Thus, for all households in their formal job,

$$\ln Y_i^{rf} = \ln Y_i^{af} \tag{2.3}$$

The employed however might be willing to take on a second job, at the expense of leisure, and avoid declaring it to the tax authorities. There exists two broad categories of industries in the economy: those in which moonlighting is typically observed, which are going to be referred to as high-risk of moonlighting industries; and low-risk which are industries where the converse occurs. Each household belongs to one of these categories according to the industry the main earner works at in their main formal job.[13] The possibility of moonlighting is only available therefore to those in the high-risk group. Thus the total true income available to a household in the high-risk sector where moonlighting opportunities are concentrated, can be composed of their formal labour income (Y_i^{rf}) and informal moonlighting income (Y_i^m) which is by definition not declared. Therefore, for moonlighters—individuals that belong to the high-risk group—their formal reported labour income needs to be scaled to capture income from moonlighting as,

$$\ln Y_i^a = \ln k_i + \ln Y_i^{rf} \tag{2.4}$$

where k_i is a random variable that will capture by how much does formal reported income by the employed in the high-risk needs to be scaled up to reach their total actual income (capturing then moonlighting income). Note that for households in low-risk occupations, for whom moonlighting opportunities are not available, $\ln k_i = 0$ and hence, $\ln Y_i^a = \ln Y_i^{af} = \ln Y_i^{rf}$.

As all households are in employment, it is assumed that transitory shocks affect all households on the same manner independent of the industry group they work on.[14]

[13] The classification of households into industries is going to be discussed in more details in Section 2.4.1. Note that participation in an undeclared second job is not observable in the survey. The classification into high or low risk industries is made on the basis that skills are transferable from the formal into the informal sector and thus those participating in industries that typically concentrate black activities are more prone to be moonlighters. Evidence on this is discussed on Section 2.4.1.

[14] Different asymmetry of the impact of a shock for the high-risk group could be allowed assuming a higher variance of the transitory shock, however, in principle there is no reason why this should be the case so we simplify. Taking up this assumption would yield a bounded estimate of the income-gap

Thus,

$$\bar{p}_{HR} = \bar{p}_{LR} \quad ; \quad \sigma^2_{uHR} = \sigma^2_{uLR} \tag{2.5}$$

Both k_i and p_i are assumed to be log-normally distributed and therefore can be written as deviations from their means as,

$$\ln p_i = \mu_p + u_i \quad ; \quad \ln k_i = \mu_k + \nu_i, \tag{2.6}$$

where u_i and ν_i are random variables with zero mean and (constant) variances σ^2_u and σ^2_ν, respectively.

When combined to equation (2.1), measurement equations (2.2) and (2.4) assuming the log-normality in equation (2.6), yield equation (2.7),

$$\ln C_{ij} = \beta_j \ln Y_i^{rf} - \beta_j \left(\mu_p - \mu_{k_i}\right) - \beta_j \left(u_i - v_i\right) + X_i\alpha_j + \epsilon_{ij}. \tag{2.7}$$

As the means and variance of p_i is the same for both groups, consequently the mean of its log, μ_p, would be the same for both high and low-risk occupations. The difference in expenditure conditional on the level of income between the two groups would be captured by the term $\beta_j\mu_k$. Equation (2.7) can be estimated using a dummy variable to capture this difference in the intercept, this is,

$$\ln C_{ij} = \beta_j \ln Y_i^{rf} + \gamma_j HR_i + X_i\alpha_j + \eta_i, \tag{2.8}$$

where HR_i is a dummy variable that takes the value 1 if the household belongs to a high-risk occupation where there are moonlighting opportunities and 0 otherwise.[15] The error term η_i is defined as $\eta_i = \beta_j(u_i - v_i) + \epsilon_{ij}$.

Consequently following (2.7) and (2.8),

$$\gamma_j = \beta_j\mu_k \tag{2.9}$$

where the estimation obtained in (2.15) would correspond to an upper bound estimate and a lower bound could be found where the correction term of the volatility $1/2(\sigma^2_{Y\,HR} - \sigma^2_{Y\,LR})$ will appear with a negative sign.

[15] Note that we have simplified the notation in the model at this stage using the high risk as a group. Effectively, there could be different dummy variables for each high-risk industry identified. This just entails that we can identify the income-gap separately for each industry.

As income is measured with error, an instrumental variable approach is used. First stage regressions of income on a set of instruments Z_i, household controls X_i are of the form,

$$\ln Y_i^r = X_i \delta_1 + Z_i \delta_2 + \phi_i. \tag{2.10}$$

First-stage regressions allow us to obtain independent estimates of the income variance which is a composite of three errors: unexplained variation in permanent income and deviations from actual to permanent income, u_i, which are assumed to be the same for both groups; and ν_i which is allowed to differ. The variance of income for moonlighters (high-risk) is bound to be larger than that of those in low-risk due to the underreporting component. Defining $var\varsigma_{HR} \equiv \sigma^2_{Y_{HR}}$ and $var\varsigma_{LR} \equiv \sigma^2_{Y_{LR}}$, it is the case that

$$\begin{aligned} var\varsigma_{HR} - var\varsigma_{LR} &\equiv var(u - \nu)_{HR} - var(u)_{LR}, & (2.11) \\ &= \sigma^2_{\nu_{HR}} - 2cov(u\nu)_{HR}. & (2.12) \end{aligned}$$

From the log-normality of k_i, the mean of k_i and the mean of its log, μ_k can be related as follows,

$$\ln \bar{k} = \mu_k + 1/2\sigma^2_\nu. \tag{2.13}$$

Substituting μ_k from (2.9) into (2.13), one can arrive at[16]

$$\ln \bar{k} = \frac{\gamma_j}{\beta_j} + \frac{1}{2}\sigma^2_\nu. \tag{2.14}$$

Using (2.12) into (2.14) one can arrive at an estimate of the average scaling factor $\bar{k}_i$ given by,

$$\ln \bar{k} = \frac{\gamma_j}{\beta_j} + \frac{1}{2}(\sigma^2_{Y_{HR}} - \sigma^2_{Y_{LR}}). \tag{2.15}$$

The income-gap from moonlighting, $\bar{\kappa}$, derived by this methodology relates to the average scaling factor, $\bar{k}$, as

$$\bar{\kappa} \equiv 1 - \frac{1}{\bar{k}}. \tag{2.16}$$

[16] It is assumed for simplicity that underreporting and income volatility are uncorrelated and thus the covariance is set to zero. The lack of observability of the distributions of the two random variables makes it not possible to calculate this covariance.

Section 2.5 report the income gap as derived by (2.16).

2.4 Empirical strategy

2.4.1 Data

Microdata on household expenditure and income from the Secure Access version of the Living Costs and Food Survey (LCFS) is used. The period covered in this study is 2010-2013. Data is deflated using the consumer price index base 2010. In order to make consistent comparisons across households, the sample is constrained to households of two-adults, either cohabitees/married or civil partners who live in Great Britain, whose household reference person (HRP) is in employment.[17] Due to concerns about a different expenditure behaviour found after retirement, we cap the age of the HRP to be less than 60, Aguiar and Hurst (2005).

To analyse the issue of moonlighters, the data is constrained to households where the main income source comes from employment by a third-party. Employed households are defined as those who draw less than 25% of their income from self-employment sources. Using this rule allows a consistent definition that elopes from the possibility of misclassification of households as 'employed' by using self-reported employment status.[18]

Once the sample is restricted to employed households, they are classified according to the industry they belong to. It will be given by the industry the HRP works at in their main formal job. Industry SICCODES are available and we keep SICCODES to the first division in order to maintain a reasonable sample size per industry group. A list of the industry codes relevant to each industry category is available in Appendix A in Table A.1 along with sample sizes displayed in Table A.2.

[17] The HRP is the householder who: either owns the household accommodation or 1) is legally responsible for the rent of the accommodation, or 2) has the household accommodation as an emolument or perquisite, or 3) has the household accommodation by virtue of some relationship to the owner who is not a member of the household. In the event of joint householders, HRP is the one with the highest income. Be it the same, the eldest individual is taken.

[18] Alternative definitions of an employed household are used but the definition does not alter significantly the results. As an example, one of the alternative specification restricts the income of the household to come solely from employment sources for the household to be classified into the sample of study.

As the industry the individual works at during their moonlighting practices is unobservable, the formal industry is used to classify households into high or low-risk. This is a reasonable assumption if we consider that the informal activity tends to occur within the same sector as the formal job. The transferability of the skills from their formal to informal job enables moonlighters to use their knowledge and skills to perform their informal job without any extra investment of effort. Their formal occupation also allows them to form a network of individuals that exposes them to more opportunities to be informally employed within the same sector. The first Eurobarometer Survey of undeclared work in 2007 enables us to explore the link between employment in the formal and the informal job. We can see from Table A.3 on Appendix A that the informal activity does in fact tend to remain within the same sector as the formal. In Construction, almost half of the individuals who perform undeclared work operate within the same sector. This seems to be the case also for household and personal services, agriculture and hospitality. In cases such as Industry, the informal job concentrates more in Construction. This is due to the high capital intensive nature of industry and the impossibility of participation in informal activities. Regulatory frameworks might operate a high restriction in this sense as well.

Three measures of expenditure are going to be considered: food expenditure, a basket of non-durable goods and total expenditure (including durables). Food expenditure is used for several reasons. Firstly, food is one of the items of expenditure with the best coverage (80%) on the LCFS when compared to the National Accounts (Brewer and O'Dea, 2012).[19] Secondly, food is not an item of expenditure for which consumption can be postponed to future periods and issues of infrequency of purchase are not present. Lastly, food expenditure is recorded in detail using a diary for each individual of the household where expenditure needs to be recorded for a fortnight.

A basket of non-durable goods was included for comparison purposes. The basket consists of expenditure on food, alcohol and tobacco, clothing, utilities, non-durable expenditure on recreation, transportation, communication, health, education and other

[19] This finding is in line with the same analysis that was conducted by Meyer and Sullivan (2010) in the US.

miscellaneous nondurable expenditures.[20] Total expenditure including durables expenditure is also included for the same purpose although it is well known that the coverage of durable goods is lower in the survey and suffer from issues of infrequency of purchase. The coverage ratio of durables in the survey is very volatile ranging between 55-80%. Summary statistics of the income and expenditure variables can be found in Table A.4 in Appendix A.[21]

Income is measured with error as a measure of reported income is used instead of a measure of permanent income which is the income that influences consumption. Therefore, income is going to be instrumented using educational attainment variables (whether the household reference person and spouse had higher education) and whether the household reference person of the household is in a white-collar or blue-collar occupation.

The control variables to be used in the estimation include: age of the HRP in a linear and quadratic specification, number of children, type of tenure of housing, a range of wealth variables (availability of cars, central heating and driers, number of rooms in the house) as well as controlling for quarter and year in which the interview took place. Summary statistics of the covariates are also available in Table A.5 and A.6 in Appendix A.

2.4.2 Methodology

One of the most important and challenging tasks in this method is to embed the structure of the black economy activities within the framework. On this, one can quote the OECD (2002): "A point that is not always understood is that the non-observed economy is heavily concentrated in a small number of sectors. In power generation, heavy industry, rail and air transport, government services, banking and telecommunications, for example, there is little scope for a 'shadow' economy. Shadow activities are confined to a relatively small number of 'susceptible' sectors, such as home repairs, retail trade, taxis, trucking, cafés and restaurants."[22]

[20] Items such as clothing, tobacco and alcohol are known to have a lower coverage rate in the LCFS around 40%.

[21] Further comment on the Summary Statistics is provided in Section 2.5.4.

[22] We recognise that technology might have a role on creating new avenues in which hidden activities can be developing and that careful trace of this new activities is vital for tax administrations.

Following this observation, we are going to recognise moonlighting activities to be concentrated within certain industries that allow for these opportunities hence introducing the structure of the informal market in the estimation. To map the terminology used in the theoretical model, these industries prone to concentrating moonlighting opportunities are going to be referred to as high-risk (of moonlighting) industries. We would refer to low-risk (of moonlighting) industries otherwise.

Assuming that the relationship between income and expenditure for those in the low-risk industry group is accurate since their expenditure is correctly reported and so is there income due to the absence of moonlighting opportunities and the withholding of labour income on their formal job; one can map the amount of *true income* that those in high-risk industries have in order to sustain their reported level of expenditure. The difference between the *true* and the *reported income* would yield an estimation of the income-gap for those in high-risk industries. In short, given a certain level of expenditure, the low-risk group will give an indication of the amount of income needed to access that level of expenditure and thus serves as a benchmark against which the income-gap of the high-risk can be assessed. In principle, it does not seem apparent that there might be a link between food preferences and industry choice.

One can very easily anticipate that the crucial part of this research is the definition of which industries are considered high and low-risk. Two alternative approaches are proposed: a factual approach, based on evidence from surveys, revenue agencies intelligence, and the literature; and a data driven method to outline which industries are to be considered within the high-risk group.

2.4.2.1 Factual Approach

The factual approach collects information available from *revenue agencies* along with *academic research* and *surveys* in order to gather evidence about which sectors are susceptible to concentrate a higher amount of hidden economy.

- **Literature**

Erard and Chin-Chin Ho (2003) analyse the tax compliance behaviour of different occupational groups in the US using a micro-simulation with administrative sources. They combine the use of audit data from 1988 on filers (TCMP Filer data) and on nonfilers (detected through the IRS TCMP Phase IX Nonfiler Survey). The resulting tax liability uncovered by the auditors is adequately adjusted to account for partial detection and income from tips and informal suppliers are imputed. Simple summary statistics of the average level of non-compliance (in dollars) leads them to find that ranked amongst the most compliant occupations are military; administrative support; retired or disabled and production/manufacturing and a mixed category called 'other'. The occupations that seem to concentrate more non-compliance are vehicle sales, investors, informal suppliers, construction and extraction, real estate and financial insurance, other sales occupations, agriculture and farming, lawyers and judges and doctors and dentists.

Alm and Erard (2016) following previous literature on the type of goods supplied by informal suppliers identify 12 broad industry categories where they generally concentrate: 1) Food catering and roadside stands; 2) Direct sales; 3) Building Maintenance/Landscaping; 4) Forestry, Fishing, Hunting and Trapping; 5) Arts and Entertainment; 6) Construction; 7) Teaching/Lessons; 8) Care of Children and Elderly (including Home Health Services); 9) Personal Services; 10) Auto Repair and Maintenance; 11) Other repair and maintenance; 12) Transportation and Moving.

- **Surveys on Undeclared Work**

Smith and Adams (1987) address the informal economy from a demand side first asking purchasers about the product and information about the vendors. They investigate the categories of goods and services in which informal suppliers typically operate finding that purchases of informal goods are concentrated in goods such as food, home repairs (which ranged from minor repairs to carpenters moonlighting and the construction of buildings), vehicle repairs, personal care (unlicensed, treated as not being employed through a company), flea markets and sidewalk vendors and services in general. This gives us an insight into the supply of informal goods. Some assumptions were made to classify these transactions into the informal economy, such as offering services directly and not through a company in the case of cleaning services. Measurement of the size of the informal economy from a demand side amounted to $81.8 billions.[23]

[23]From the demand side of this survey,valuable information about the occupation of the sellers is

From a supply side, 21% of the households report a member with second employment (not formal) on the past year which amounts to $30.3 billions income received from these activities. The sort of occupation in the informal economy includes mainly home repairs, personal care, gardening actitivies, vehicle repairs and arts and crafts. The gap between the estimated income figure from purchases and sales is not surprising. Apart from sampling issues, underdeclaration seems like the more evident factor behind this discrepancy.

Pedersen (2003) provides a cross-country survey. This required tailoring the questionnaire to the legislation of the countries in order to account for which activities were taxable in each country.[24] A break-down by industry of the proportion of black to white hours show that for all countries, black activities seems to be highly concentrated in the Construction sector. The estimates vary: The proportion of black to formal hours in Denmark is 25% being around 17% for Norway, 13% in Germany, 9% in Sweden and 5% in the UK. The agricultural sector seems to be more predominant for Germany (15%) ranking first in the concentration of black economy activities whereas it is second in the rest of the countries. These two sectors are followed by Hotel and Restaurants and Wholesale and Retail Trade however the extent of penetration of these types of activities vary. The sector that less black economy activities concentrates turns out to be the manufacturing sector.

Another interesting survey on undeclared work in the European Union is the Eurobarometer Survey which counts with two waves, in 2007 and 2013. The questionnaire of the survey suffered modifications from the first to the second wave making the results not directly comparable. The first Eurobarometer survey in 2007 identifies the sector within which the informal activity lies. Informal work appeared to be concentrated within construction, household services, personal services, repair and transportation.

obtained. About 54-72% of the sellers had regular jobs, between 12-31% were unemployed or not in the labour market and 7-23% were said to perform only work "on-the-side". This is interesting in recognising that the largest proportion of informal suppliers were individuals who are on formal employment. Note that the range of the percentages shown stems from the fact that the survey separates activities by type (e.g. home repair, lesson) and then the employment status of the informal supplier is asked. The ranges show the upper and lower bound of the percentages given for each type of activity.

[24] It turns out that a lower number of activities were taxable in Great Britain compared to the rest of countries, so special questions were performed to try to capture better black activities. 31% of activities were not possible to be coded to an industry in Great Britain.

In the second wave of the survey in 2013, particular activities were recorded. These included: repairs and renovations, gardening, cleaning, babysitting, waitering, administrative and IT assistance, moving houses, tutoring, car repairs, assistance for dependant, food selling, selling goods and services. Undeclared activities seemed to be concentrated within repair and renovations (19%) and selling services (15%).

- **Revenue Agencies Intelligence**

According to NAO (2008), revenue agencies target broadly similar risk sectors with their compliance activities: construction, real state, services (hairdressing, child minding, taxis, legal and consultancy services), retail stores and motor vehicle repairing and retailing.

In our case as the country of interest is the UK, activities pursued by the competent revenue authority, HM Revenues and Customs (HMRC), are observed. HMRC's enforcement and compliance activities follows a 'promote, prevent, respond' strategy promoting a closer (customer) relationship with the taxpayer making instruments available to help them comply.

HMRC makes use of public campaigns in order to incentivise compliance, e.g. the "eyes" campaign reminds individuals that HMRC is vigilant and encourage them to comply with their obligations. Campaigns are also launched to invite individuals to bring their tax affairs up to date, declaring voluntarily what is owed. These types of voluntary disclosure campaigns gives us good insights to which occupational groups HMRC consider to concentrate non-compliance.

An instrument that can provide direct insight on the occupations targeted are HMRC task forces. A Taskforce is a specific HMRC Enforcement and Compliance programme that uses information to target high-risk evaders within a certain geographical area or industry sectors. Table A.7 on Appendix A provides a non-exhaustive list of occupations that have been targeted either through encouraging voluntary compliance or through enforcement providing first hand information of which occupations typically appear to concentrate informal suppliers.

Broadly, the occupations targeted by HMRC with its compliance activities indeed match NAO (2008) report: doctors and therapy, plumbers and electricians, property, building work, catering and hospitality, personal services, motor trade and scrap metal dwellers, retail and the legal profession.

- **Factual approach: Baseline Selection—Low-risk industry**

A natural way of approaching the selection of an industry towards high or low risk of moonlighting would be to select as low-risk those industries in which moonlighting opportunities do not typically appear. In practice, as we can see from the survey evidence, no industry is free of moonlighting but we need to choose a baseline with a low occurrence of moonlighting and measure moonlighting in relative terms.[25]

Based on the literature, the activities of revenue agencies and the surveys, we propose two strategies to select low-risk industries to use as a baseline against which to measure moonlighting. The first one uses a *single-industry approach* as a baseline. It seems apparent that certain occupations are less popular to concentrate informal workers. There seems to be a convergence in the findings that manufacturing is an occupation where informal opportunities are very limited due to the characteristics of the job (Erard and Ho, 2003; Pedersen, 2003). Even if individuals are skilled to work in the industry, its capital intensive nature and the infrastructure needed impedes informal activities. This industry is going to be selected as the baseline low-risk industry against which to assess non-compliance in the first strategy.

The second strategy uses a *multi-industry approach*, this means that the baseline would be composed of a group of industries allowing for a broader control group for identification purposes. Still we need to pursue having a baseline where moonlighting opportunities are low in order to approximate the amount of non-compliance in the rest of the industries. For this purpose, the second strategy utilises the industries recognised by Erard and Ho (2003) to be the ones that concentrate the least non-compliance using administrative sources. This baseline is composed of the following occupations:

[25] Note that the estimate we obtain for the rest of the industries would represent a lower bound as we are assuming in the baseline some amount of moonlighting. The aim is to try to choose a baseline where the amount of moonlighting is minimal so that the estimates obtained for the rest of the industries are more accurately estimated.

military; manufacturing; protective services, accountants, administrative and support services. For reference this baseline is going to be referred to as *low-risk (EH).*

Once the baseline is selected, either a single industry or multiple, we proceed to estimate the Engel curve given in equation (2.8) to obtain the parameters of interest to compute the income-gap as in equation (2.16).

2.4.2.2 Data-driven approach

The present approach introduces more flexibility by letting the data select the indsutries exhibiting a higher level of expenditure for the level of income reported. Under the factual approach, we classified industries as low-risk using prior evidence from the literature and the intelligence of tax administrations. In this data-driven approach we aim at freeing the model as much as possible from prior information. This is particularly interesting to fight possible misconceptions on informal work practices and to identify new sectors informal activities are developing at.

Subset selection methods, to be more precise stepwise regressions, are performed to elicit which industries are consistently shown to exhibit non-compliance. The advantage of subset selection methods is that it will allow sifting through the industry variables leaving in the model only the ones that report a significant discrepancy in their income-expenditure pattern against the baseline.

Practically, the aim is to estimate equation (2.8) through stepwise regression, this is an Engel curve that includes dummy variables for each of the industries identified in Section 2.7. The estimation method will then select the industries that signal underreporting with respect to the baseline. First, introducing all industry dummy variables at the same time in the regression inevitably leads to multicollinearity. One of the industry must be chosen as a baseline against which to assess the rest. However, this would mean imposing some structure to the model already as we did in the factual approach. The choice of the baseline can also influence the estimations of the income-gaps obtained.[26] In order to circumvent this choice of an initial baseline, we perform the stepwise regression repetitively using a distinct industry baseline each time.

[26] If the baseline category happens to be a high-risk occupation, imagine construction, then the measurement of the rest of the variables is not correctly interpreted as non-compliance. We could in

Second, underreporting is captured in the estimation by a positive industry dummy variable. This indicates a higher level of expenditure than it corresponds to the amount income reported. The observation of the results obtained from the repetitive stepwise regressions—from alternating the baseline industry—leads us to identify which industries consistently display a positive coefficient and thus are consistently highlighted as high-risk. Note that in this data-driven method the focus is on determining the industries where non-compliance is concentrated. Once these industries are chosen, then the rest would be considered part of the low-risk industry group. The industries relegated to the low-risk group are going to be referred to as *low-risk (ST).* We can then proceed to the estimation of equation (2.8) as under the factual approach to obtain the parameters of interest necessary to compute the income-gap as in equation (2.15).

Effectively, both the factual and the data-driven approach pursue the same objective: determining which industries exhibit low opportunities of moonlighting to use them as a baseline category in order to assess the moonlighting income-gap of the rest—the high risk. The factual approach as we have presented it focuses on using prior evidence to identify the low-risk occupations while the data-driven approach focuses on allowing the data to identify the high-risk. Since the aim of both methods is ultimately common, similar underreporting results are expected to be found.

2.5 Results

Regardless of the approach used to choose the low-risk baseline industries—factual or data-driven—the results exposed in this section come from the estimation of equation (2.8). To address the measurement error of the income variable, income is instrumented using educational variables (whether the household reference person and spouse had higher education) and whether the household reference person of the household is in a white-collar or blue-collar occupation (defining white-collar as those individuals who are employers in small or large organisations, or hold a higher managerial, higher profes-

this case encounter actually negative and significant industry coefficients which only highlight that the baseline is not adequate. Graphically this will mean that the Engel curve of the considered low-risk group lies above the rest of the Engel curves for the rest of occupations. This implies a pattern of overreporting which is only led by the misspecification of the baseline as low risk.

sional, lower professional and higher technical, lower managerial and higher supervisory and intermediate position). The first-stage estimation is given by equation (2.10). The coefficients of interest for the estimation of the income-gaps as evidenced by equation (2.15), are the elasticity of income with respect to consumption, β_j, the coefficients of each high risk industry γ_j and the income variances for both groups from the first-stage, $\sigma^2_{Y_{HR}}$ and $\sigma^2_{Y_{LR}}$. The results section reports the results separately for the factual and data-driven baseline. A summary of results is available within the section. Full results are available in Appendix B.

2.5.1 Factual Approach

Firstly, using the *single-industry approach*—treating manufacturing as the baseline low-risk industry against which the income-gap of the high-risk can be assessed—Table 2.1 shows that using the three different measures of expenditure, construction, distribution and professionals consistently appear as industries where moonlighting practices are concentrated.[27]

Using food expenditures, Table 2.1 reports an estimated income-gap of 23.1% for Construction and Distribution and 26.5% for Professionals. These figures are lower in the case of the other two measures of expenditure as the elasticity of consumption is larger being the income gaps around 20.5% for those in construction, 23.1% for those in distribution and 18% for professionals.[28] The other two measures of expenditure also capture some degree of non-compliance within the Financial insurance and Real Estate and Administration and Support Services sectors. The public sector, defence and education, the accommodation and food sector; and the arts and entertainment sector are industries where some degree of non-compliance is found by the non-durables basket expenditure measure.

The *multi-industry approach*— the second baseline proposed by the factual approach—which uses industries where evidence non-compliance is found to be low using administrative sources (Low-Risk—EH for reference), obtains results similar to those under

[27] Table B.1 in Appendix B contains the full estimation of the Engel curves.

[28] Full estimation of the Engel curves has been relegated to Appendix B, Table B.1 contains the estimation for the case of the single-industry baseline.

the single-industry baseline.[29] The construction and distribution income-gap amounts to 21.5% and the professionals income-gap is of 25.9%.[30] Using total expenditure, the same industries as using food are highlighted as concentrating misreporting, except that some non-compliance appears for the Financial and Real Estate sector. In the case of non-durables Financial and Real Estate and Public Administration and Education; and Arts and Entertainment also appear as concentrating non-compliance in a similar magnitude as under the previous single-industry baseline.

Using different baselines and expenditure variables, we can see that Construction, Distribution and Professionals are consistently flagged as uncovering non-compliance. Non-compliance in other industries is not robust to the expenditure variable used and some variability is observed. This could partially be attributed to the possible measurement error present on these expenditure variables. Food expenditure for the reasons outlined in Section 2.4.1 represents a more reliable measure.

[29] The group of low-risk industries named 'Low-Risk (EH)' are composed of: military; manufacturing; protective services, accountants, administrative and support services.

[30] Table B.2 in Appendix B contains the full estimation of the Engel curves using the multiple-industry baseline.

Table 2.1: Estimation of the Income-gap: Factual Approach.

Baseline	Manufacturing			Low-risk (EH)		
	Food	ND	TX	Food	ND	TX
CONS	0.231 (0.058)	0.205 (0.04)	0.151 (0.039)	0.215 (0.055)	0.174 (0.037)	0.13 (0.037)
DISTR	0.231 (0.051)	0.231 (0.033)	0.191 (0.033)	0.215 (0.047)	0.202 (0.031)	0.171 (0.031)
PROF	0.265 (0.059)	0.18 (0.045)	0.166 (0.042)	0.259 (0.056)	0.146 (0.043)	0.143 (0.067)
FI-RE	n.s.	0.2 (0.046)	0.168 (0.043)	n.s.	0.167 (0.044)	0.051 (0.067)
ADMIN-SUP	n.s.	0.174 (0.05)	0.184 (0.048)	n.s.	n.s.	n.s.
PUB-DEF-EDU	n.s.	0.071 (0.035)	n.s.	n.s.	0.061 (0.033)	n.s.
ARTS	n.s.	0.182 (0.071)	n.s.	n.s.	0.151 (0.072)	n.s.
ACCO	n.s.	0.182 (0.061)	n.s.	n.s.	n.s.	n.s.

Notes:

'ND' stands for non-durables; 'TX' stands for Total Expenditure. Two baselines are used for the factual approach: single and multi-industry. The single-industry baseline is Manufacturing. The multi-industry baseline—Low-Risk EH industries—are those which rank lower in terms of non-compliance following administrative sources (Erard and Ho, 2003): military, manufacturing, protective services, accountants, administrative and support services. This corresponds to the second baseline within the factual approach. Thus, PROF will not include accountants and likewise PUB-DEF-ED does not include military and protective services in the definition of this baseline. Standard errors in parenthesis are calculated using the delta method. Income-gaps are calculated for the industries that display misreporting, 'n.s.' imply no significant misreporting was found on that industry.

2.5.2 Data-driven Approach

The data-driven approach differs from the factual approach in that the focus is on determining the industries where there is a discrepancy and treating the rest as low-risk and therefore the base. The result is a broader low-risk base which allows for more identification.

On a first stage, all industries were included in the estimation of the Engel curve laid in equation (2.8) and each industry was chosen as the baseline industry once. Stepwise regression were run on the Engel curves using a different baseline industry at a time.[31] Each time occupations for which a positive and significant discrepancy was found, were marked as high-risk. The repetition taking into account different baselines to look for consistency in the industries that display a discrepancy allows us to not predefine an industry as low-risk (as was done in the case of the factual approach), but to extract this information from the data. The subset selection method allows to choose the best model out of all possible combinations of the industry variables available.

Table B.3 in Appendix B summarizes under the heading Low-Risk (ST) the industries that consistently appear as positive and significant from the stepwise regressions in every baseline category used. One can straightforward relate that the industries obtained through this method are comparable to those obtained using the factual approach. In this case however as opposed to the strategy followed within the factual approach, no preliminary assumptions were imposed about which industry to choose as a baseline. Note that once the high-risk industries have been identified, the low-risk industries are the rest.[32]

For the case of expenditure on food, construction, distribution, professionals and financial and real estate were found to consistently exhibit non-compliance (positive coefficient) after applying the stepwise procedure repetitively for the different baseline categories. These industries are thus labelled as high-risk and enter the estimation of equation (2.8) using the IV procedure to account for measurement error on income.

[31] The results from running the stepwise regressions are omitted but available upon request.

[32] Practically, this means of all the industries that appear in table A.1 of Appendix A and are not present in table B.4 of Appendix B.

The rest of the industries that form the low-risk (ST) baseline in this case are primary sector, manufacturing, energy, transportation, hospitality, information and communication, administration and support, public administration, education and defence, health and social workers, arts and entertainment and others.[33]

Table 2.2 shows that the estimates of the income-gap following the data-driven approach are comparable to those found in the factual approach. Using food expenditure, an income gap for construction of 19.6%, distribution of 19.8% and Professionals of 22.9% is revealed.

Table 2.2: Estimation of the Income-gap by Industry: Data-driven Approach.

	Low-risk (ST)		
	Food	ND	TX
CONS	0.196 (0.049)	0.146 (0.034)	0.117 (0.035)
DISTR	0.198 (0.043)	0.178 (0.028)	0.158 (0.028)
PROF	0.229 (0.048)	0.115 (0.039)	0.133 (0.037)
FI-RE	-	0.134 (0.04)	0.136 (0.038)
ADMIN-SUP	-	0.117 (0.046)	0.151 (0.047)

Notes:
'ND' stands for non-durables; TX stands for Total Expenditure. Standard errors in parenthesis are calculated using the delta method. The low-risk (ST) baseline is composed of the following industries. For the case of food: primary sector, manufacturing, energy, transportation, hospitality, information and communication, financial and real estate, administration and support, public administration, education and defence, health and social workers, arts and entertainment and others. For the case of non-durables and total expenditure the same baseline applies except for financial and real estate and administration and support for which some degree of misreporting is uncovered.

[33] Note that it may occur that after instrumenting for income significance of the industry is weakened as is the case of Financial Insurance and Real Estate for food expenditure and the regression result is refined to treat the industry as a low-risk in a second stage. For the final estimation result only industries that are signalled by the stepwise procedure and stay significant after correcting for the measurement error on income stay as high-risk.

2.5.3 Estimation of the income-gap

The estimations obtained following this method can serve to provide an estimation of the moonlighting income-gap for Great Britain in monetary terms.

Although different measures of expenditure are considered, for the reasons outlined in Section 2.4.1, food expenditures stands as the preferred item of expenditure and therefore these estimations are going to be used. All the expenditure measures coincide on three sectors as the ones that concentrate non-compliance: construction, distribution and real estate and professionals. Thus these three sectors as there is consistency across measures of expenditure and baselines are the ones for which an income-gap is going to be translated to monetary terms.

All three baselines: manufacturing, low-risk (EH) and the low-risk baseline that stems from the use of the subset selection method (ST) yield similar results in terms of the income-gap of the employees in the three sectors. Regarding food expenditures and low-risk ST industries as the baseline, the construction income-gap is estimated to be 19.6%, 19.8% for distribution and 22.9% for professionals.[34]

Using the figures for compensation of employees (COE) in the National Accounts for the period 2010-12, one can estimate the income-gap in monetary terms by applying the income-gap coefficients to the amount of COE in the pertinent sectors. In the case where the sector to be considered was grouped with other sectors to produce the estimation of COE, the share of COE belonging to each sector was apportioned using the Gross Value Added (GVA). From Table B.5, we can observe that the income-gap is estimated to be around 2.6% of GDP for the years 2010-2013. The estimation vary only slightly when considering different baselines from 2.6% to 3% of GDP in the case manufacturing is used as the baseline low-risk industry.[35]

[34] Table B.5 on Appendix B contains the estimations of the income-gap for food expenditure under the estimation of the income-gap obtained using both the factual and data-driven approaches.

[35] We have tested whether the estimates vary over time but found no statistically significant changes. This maps the behaviour of the formal moonlighting market which shows an acyclical behaviour in the UK.

2.5.4 Summary Statistics and Graphical Evidence

The central identification in this model is the fact that there exists a discrepancy in the relationship between the expenditure and income for certain industry types. Suggestive evidence of underreporting can be found by observing the unconditional means of the expenditure and income variables.[36] Three baselines are considered in the analysis: following the factual approach, *manufacturing* and a composite of industries called *low-risk (EH)*; and following the data-driven approach a broader composite of industries referred to as *low-risk (ST)*.[37]

Appendix C presents the distribution of expenditure and income for the selected high-risk—Construction, Distribution and Professionals—against each of the low-risk baselines. One can observe that generally the distribution of expenditure is shifted to the right of the low-risk baseline indicating higher levels of expenditure than the baseline with the density of income lying either on top of the distribution of income for the low-risk baseline—indicating similar levels of income while sustaining higher levels of expenditure—or slightly to the left of it indicating lower income streams. This is particularly patent in the case of Construction. In the case of Distribution, the distribution of expenditure is not too dissimilar from that of the low-risk baseline while exhibiting much lower incomes.[38] This observation, even though we are referring to unconditional summary statistics are suggestive of non-compliance.

Summary statistics of the relevant variables for the baselines and industries marked as concentrating underreporting are also available in Appendix A, Table 1.1.

2.6 Conclusion

The high visibility that employment income enjoys due to the withholding of taxes, has led the literature to disregard other ways in which the employees—as opposed to the self-employed—can avoid their tax liability. Moonlighting—holding a second source of

[36] Note that it may seem unorthodox to present summary statistics after the estimation but only after estimation do we know which industries are high-risk, especially for the data-driven method.

[37] See 2.4.2.1

[38] It can be noted that in the case of professionals both distributions lie to the right displaying higher levels of expenditure and income before controlling for other characteristics.

undeclared labour income while on formal employment—can offer wilful evaders the possibility to reduce their tax liability.

We contribute to the literature on tax compliance by specifically addressing the estimation of the extent of misreporting due to moonlighting practices. The method uses micro-level data on expenditure and income from a general purpose survey.

The proposed method builds on three assumptions. First, we embed the observation that informal activities are typically concentrated within certain sectors which are going to be labelled as *high-risk*. This creates two groups for the model: those working in high-risk and those in low-risk occupations. Second, expenditure for all groups—high and low risk—is correctly reported. Third, income for those in low-risk industries is correctly reported due to the low-occasion of moonlighting opportunities and the withholding of taxes on their formal income. Income for those in high-risk occupations may be misreported due to the existence of a second source of undeclared labour income. The relationship between expenditure and income for the low-risk industry group is going to allow the identification of anomalous levels of expenditure for a given level of reported income for those in the high-risk industry group.

Two approaches are proposed to select industries into low or high risk. The first, the factual approach uses prior information from revenue agencies practices, surveys and the literature to propose two baselines: a single industry and a multi-industry baseline. The second method, is a data driven method that allows the data to choose the high-risk industries minimising the amount of prior information introduced. Both methods are consistent in flagging the same industries as concentrating moonlighting practices: Construction, Distribution (Wholesale, Retail and Repair of Motor Vehicles) and Professionals and Real Estate.

Using food expenditures and taking the data-driven approach to construct the baseline, it is found that individuals working in Construction display an income-gap of 19.6%, Professionals of 22.9% and those working on the Distribution sector display an income-gap of 19.8%. Using national accounts data on compensation of employees by sectors, this translates into approximately 2.6% of GDP.

2.7 References

Alm, J., Erard, B. (2016). Using Public Information to Estimate SelfEmployment Earnings of Informal Suppliers. *Public Budgeting & Finance*, 36(1), 22-46.

Aguiar, M., Hurst, E. (2004). Consumption vs. expenditure. *Journal of Political Economy*, October 2005, 113(5), 919-948.

Bradburn, N., Sudman, S., Wansink, B. (2004). *Asking Questions The Definitive Guide to Questionnaire DesignFor Market Research, Political Polls, and Social and Health Questionnaires*, Revised Edition. Jossey-Bass. USA.

Brewer, M., O'Dea, C. (2012). *Measuring living standards with income and consumption: evidence from the UK* (No. 2012-05). ISER Working Paper Series.

Cabral, A.C.G.; Kotsogiannis, C.; Myles, G.D. (2016). *Self-Employment Underreporting: Who and How Much?*. TARC DP 010-14. Available at: `http://tarc.exeter.ac.uk/publications/discussionpapers/`.

Cabral, A.C.G.; Kotsogiannis, C.; Myles, G.D. (2015). *The Hidden Economy in Europe: Extrinsic and Intrinsic Motivations*. TARC DP 014-15. Available at: `http://tarc.exeter.ac.uk/publications/discussionpapers/`.

Cowell, F.A. (1990). *Cheating the government: The economics of evasion.* MIT Press Books. 1990;1.

Dilnot, A., Morris, C. N. (1981). What do we know about the black economy?. *Fiscal Studies*, 2(1), 58-73.

Erard, B., 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.

Erard, B., Ho, C. C. (2003). Explaining the US income tax compliance continuum. *eJournal of Tax Research*, 1(2), 93-109.

Feld, L., Larsen, C. (2005). *Black Activities in Germany in 2001 and 2004: A comparison based on survey data.* The Rockwool foundation Research Unit, Study No. 12, Copenhagen.

Gemmell, N., Hasseldine, J. (2012). The tax gap: a methodological review. *Advances in Taxation* (pp. 203-231). Emerald Group Publishing Limited.

HMRC (2016). *Measuring tax gaps 2016 edition. Tax gap estimates for 2014-15.* Retrieved from: `Measuringtaxgaps2016edition.Taxgapestimatesfor2014-15.`

Hurst, E., Li, G., Pugsley, B. (2014). Are Household Surveys Like Tax Forms: Evidence from Income Underreporting of the Self Employed. *The Review of Economics and Statistics*, 96(1), 19-33.

Internal Revenue Service (1996). *Federal Tax Compliance Research: Individual Income Tax Gap Estimates for 1985, 1988, and 1992.* Internal Revenue Service Publication 1415. Washington, D.C.

Internal Revenue Service (2016). *Tax Gap Estimates for Tax Years 20082010.* Retrieved from: `https://www.irs.gov/PUP/newsroom/tax%20gap%20estimates%20for%202008%20through%202010.pdf`

IMF (2013). *United Kingdom: Technical Assistance Report— Assessment of HMRC's Tax Gap Analysis.* Retrieved from: `https://www.imf.org/external/pubs/ft/scr/2013/cr13314.pdf`.

Isachsen, A. J., Strøm, S. (1985), The size and growth of the hidden economy in Norway, *Review of Income and Wealth*, 31(1): 21–38.

Johansson, E. (2005). An estimate of self-employment income underreporting in Finland. *Nordic Journal of Political Economy*,31, 99-109.

Merz, J., Wolff, K. G. (1993). The shadow economy: illicit work and household production – a microanalysis of West Germany, *Review of Income and Wealth*, 39(2): 177–94.

Meyer, B. and Sullivan, J. (2010), *Five Decades of Consumption and Income Poverty.* Harris School of Public Policy Working Paper 09.07, University of Chicago.

NAO (2008). *Comparing how some tax authorities tackle the hidden economy.* Retrieved from: `https://www.nao.org.uk/wp-content/uploads/2008/04/0708341_comparing.pdf`

OECD (2002). *Measuring the non-observed economy*. Retrieved from: `https://www.oecd.org/std/2389461.pdf`.

Pedersen, S. (2003). *The Shadow Economy in Germany, Great Britain and Scandinavia: A Measurement Based on Questionnaire Service*. Study No. 10, The Rockwool Foundation.

Pissarides, C. A., Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, 39(1), 17-32.

Schuetze, H. (2002). Profiles of Tax Non-compliance Among the Self-Employed in Canada: 1969 to 1992. *Canadian Public Policy*, 28(2), 219-237.

Smith, J. D., Adams, T. K. (1987). *The Measurement of Selected Income Flows in Informal Markets, 1981 and 1985-1986*. Report prepared for the Internal Revenue Service, Contract No. TIR 84-0084. Ann Arbor, MI: Economic Behavior Program, Survey Research Center, Institute for Social Research, The University of Michigan.

Van Eck, R. Kazemier, B. (1988). Features of the hidden economy in the Netherlands, *Review of Income and Wealth*, 34(3): 251–73.

Appendix A

Table A.1: Equivalence in SICCODES.

	SICCODES
Primary (PRI)	01100-09900
Manufacturing (MANUF)	10110-33200
Energy (ENER)	35100-39000
Construction (CONS)	41000-43999
Distribution (DISTR)	45000-47999
Transportation (TRANS)	49100-53202
Accommodation and Food Services (ACCO)	55000-56302
Information and Communication (INFO-COM)	58000-63990
Financial Insurance and Real Estate (FIN-RE)	64000-68320
Professionals (PROF)	69000-75000
Professionals, no accountants (PROF NO ACC)	69000-69109; 70100-75000
Administration and Support (ADMIN)	77000-82990
Public Administration, Education and Defence (PUB-DEF-ED)	84000-85600
Public Administration, Education (PUBADMIN-EDU)	84000-84210; 84300
Protective Services (PROT SS)	84230-84250
Military (MIL)	84220
Human Health (HH)	86000-88990
Arts and Entertainment (ARTS)	90000-93290
Others (OTH)	94000-99000

Table A.2: Distribution of household by industry.

	2010	2011	2012	2013	Total
Primary	21	14	26	20	81
Manufacturing	214	198	194	154	760
Energy	16	28	34	28	106
Construction	101	116	100	95	412
Distribution	140	157	146	136	579
Transportation	74	77	90	73	314
Hospitality	28	44	37	34	143
Information and Communication	66	79	69	76	290
Financial Insurance	78	79	86	84	327
Professionals	84	98	105	95	382
Administration and Support	44	56	48	46	194
Public Admin., Defence	230	249	280	209	968
* Public Admin. and Education	159	181	202	156	698
* Defence	26	28	33	24	111
* Protective Services	45	40	45	29	159
Health and Social Workers	127	149	121	149	546
Arts and Entertainment	22	19	26	21	88
Others	20	18	22	19	79
Total	1265	1381	1384	1239	5269

Table A.3: Matrix of formal and informal employment using Eurobarometer Survey (2007).

	Formal Employment Sector																					
	Construction		Industry		Household services		Transport		Personal services		Retail		Repair services		Hotel, restaurant, cafes		Agriculture		Other		Total	
Informal Sector	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Construction	90	49.7	31	30.7	7	18.9	6	13.6	10	7.5	7	13.7	11	21.6	1	3.6	5	11.9	22	15.5	190	23.4
Industry	4	2.2	10	9.9	4	10.8	2	4.5	4	3	4	7.8	3	5.9	1	3.6	2	4.8	5	3.5	39	4.8
Household services	11	6.1	5	5	13	35.1	2	4.5	22	16.4	3	5.9	4	7.8	2	7.1	8	19	17	12	87	10.7
Transport	22	12.2	16	15.8	1	2.7	13	29.5	2	1.5	0	0	1	2	0	0	2	4.8	6	4.2	63	7.8
Personal services	8	4.4	5	5	5	13.5	7	15.9	44	32.8	7	13.7	4	7.8	3	10.7	0	0	32	22.5	115	14.2
Retail	3	1.7	5	5	0	0	0	0	8	6	6	11.8	1	2	2	7.1	0	0	3	2.1	28	3.5
Repair services	13	7.2	12	11.9	1	2.7	3	6.8	2	1.5	4	7.8	14	27.5	0	0	1	2.4	11	7.7	61	7.5
Hotel, restaurant, cafs	0	0	5	5	1	2.7	0	0	6	4.5	3	5.9	0	0	13	46.4	1	2.4	4	2.8	33	4.1
Agriculture	9	5	1	1	1	2.7	1	2.3	7	5.2	2	3.9	1	2	0	0	12	28.6	6	4.2	40	4.9
Other	26	14.4	14	13.9	6	16.2	10	22.7	29	21.6	11	21.6	10	19.6	2	7.1	5	11.9	30	21.1	143	17.6
Refusal	17	9.4	15	14.9	5	13.5	8	18.2	9	6.7	13	25.5	7	13.7	8	28.6	10	23.8	13	9.2	105	12.9
Don't Know	13	7.2	3	3	2	5.4	4	9.1	11	8.2	3	5.9	3	5.9	1	3.6	3	7.1	7	4.9	50	6.2

Table A.4: Summary Statistics of Expenditure and Income by Industry.

		Disposable Labour Income	Food Expenditure (ln)	Non-Durables basket	Total Expenditure
Construction	N	412	410	412	412
	Mean	6.403	4.442	5.799	6.128
	S.D.	0.522	0.498	0.484	0.472
Distribution	N	575	578	579	579
	Mean	6.154	4.29	5.638	5.996
	S.D.	0.669	0.549	0.51	0.504
Professionals	N	382	378	382	382
	Mean	6.642	4.513	5.889	6.261
	S.D.	0.75	0.525	0.591	0.55
Financial Insurance-Real Estate	N	326	327	327	327
	Mean	6.738	4.479	5.973	6.325
	S.D.	0.707	0.527	0.592	0.55
Admin and Support	N	194	194	194	194
	Mean	6.187	4.268	5.614	6.021
	S.D.	0.673	0.453	0.525	0.55
Manufacturing	N	759	757	760	760
	Mean	6.404	4.316	5.676	6.044
	S.D.	0.511	0.543	0.517	0.496
Low-Risk (EH)	N	1245	1243	1246	1246
	Mean	6.407	4.326	5.7	6.069
	S.D.	0.519	0.54	0.513	0.489
Low-Risk (ST)	N	3368	3369	3375	3375
	Mean	6.432	4.328	5.718	6.081
	S.D.	0.554	0.555	0.511	0.497

Notes:

1. Manufacturing is the first baseline to be used in the factual approach. 2. Low-Risk EH industries are those which rank lower in terms of non-compliance following administrative sources (Erard and Ho, 2003): military, manufacturing, protective services, accountants, administrative and support services. This corresponds to the second baseline within the factual approach. 3. Low-Risk ST are those which are not selected by the data-driven method as displaying a significant discrepancy between their income and expenditure pattern. These are: primary sector, manufacturing, transportation, accommodation, information and communication, energy; public administration, education and defence; human health and social work and arts and entertainment. Signalled as high risk using the different expenditure variables are: construction, distribution, professionals, financial insurance and real estate and administration and support.

Table A.5: Summary Statistics of continuous variables.

	N	Mean	S.D.
Age	5269	41.516	9.923
Number of Children	5269	0.896	0.965
Number of Rooms	5269	5.982	1.776
Number of Cars	5269	1.346	0.707
Central Heating	5269	0.968	0.175
Drier	5269	0.648	0.478

Table A.6: Summary Statistics of categorical included variables.

Ownership	Frequency	Percentage
Local Authority Tenant	154	2.9
Rented Accommodation	600	11.4
Owner with Mortgage	3,258.00	61.8
Other	1,257.00	23.9
Total	5,269.00	100

Table A.7: Activities targeted by HMRC compliance activities (Volutary Disclosures and Task Forces)

	Type of compliance activity
Primary Sector	
Fishing industry	TF
Hospitality	
Restaurants	TF
Fast food	TF
Flea Markets	TF
Holiday Industry	TF
Construction	
Building industry	TF
Plumbers Tax Safe Plan	VD
Electricians Tax Safe Plan	VD
Real Estate	
Property rentals	TF/ VD
Property transactions	TF/VD
Motor trade	TF
Transport	
Taxi	TF
Haulage industry	TF
Personal Services	
Hairdressers and beauticians	TF
Jewellery trade	TF
Security guards/ bouncers	TF
Tax Catch Up Plan for tutors and coaches	VD
Health	
Tax Health Plan (Doctors and Dentists)	VD
Health and Wellbeing Campaign	VD
Distribution	
E Marketplaces	VD
Direct Selling	VD
Professionals	
Solicitors Tax Campaign	VD
General	
Fraudulent repayments	TF
Tax evasion	TF
Hidden wealth	TF
Offshore Disclosure Facility	VD
Second Incomes	VD
VAT repayments	TF/VD
Tax Catch Up Plan (Non-filers)	VD
Tax Returns Initiative	

Notes:

1. The list provided here is non-exhaustive. It captures "HMRC 2010 to 2015 government policy: tax evasion and avoidance", media outbreaks on task forces activities, HMRC's taskforces site `http://hmrcdigitalpilots.com/taskforces/`. 2. TF=Task Force; VD= Voluntary Disclosure campaign. 3. Note: The Health and Wellbeing campaign includes (physiotherapists, chiropractors, chiropodists, osteopaths, occupational therapists, those working in homeopathy, acupuncture, nutritional therapy, reflexology, nutrition, as well as psychology, speech therapy, arts therapy. 4. Second Incomes Campaign: Examples of second income include tutoring, consulting, taxi driving, fitness training, hairdressing, crafts, buying and selling goods in stalls or car boots. 5. Tax Returns Initiative targets high-income taxpayers who did not filed SA return when was advised to in order to declare extra income.

Appendix B

Table B.1: Factual Approach: Estimation of the Engel Curve using Manufacturing as a baseline industry.

	IV			OLS		
	Food	Non-durables	Total Expenditure	Food	Non-durables	Total Expenditure
Disposable Labour Income	0.454***	0.593***	0.618***	0.274***	0.391***	0.401***
	(0.04)	(0.04)	(0.04)	(0.01)	(0.01)	(0.01)
Primary	0.053	0.046	-0.004	0.061	0.053	0.014
	(0.06)	(0.05)	(0.04)	(0.06)	(0.05)	(0.05)
Energy	0.063	0.027	0.067	0.073	0.051	0.096**
	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
Construction	0.110***	0.123***	0.088***	0.102***	0.112***	0.085***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
Distribution	0.095***	0.124***	0.099***	0.061**	0.087***	0.064***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Transportation and Storage	-0.035	-0.004	0.008	-0.039	-0.017	-0.001
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Accommodation and Food	0.024	0.081*	0.056	-0.038	0.003	-0.006
	(0.06)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)
Information and Communication	-0.008	0.03	0.012	0.028	0.077***	0.068**
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Financial Insurance and Real Estate	0.015	0.088***	0.067**	0.058*	0.139***	0.123***
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Professional, Scientific and Technical	0.116***	0.087***	0.079***	0.141***	0.114***	0.118***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Administration and Support	0.049	0.090**	0.101***	0.034	0.069**	0.080**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
Public Administration and Defence	0.019	0.047**	0.006	0.036	0.050**	0.02
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Human Health and Social Work	-0.035	0.024	0.02	-0.028	0.026	0.031
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
Arts and Entertainment	0.07	0.106**	0.066	0.05	0.095**	0.06
	(0.07)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)
Others	0.063	0.059	0.039	0.021	0.023	-0.016
	(0.07)	(0.05)	(0.07)	(0.06)	(0.05)	(0.05)
Controls for household characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5244	5257	5257	5244	5257	5257
Baseline	Manufacturing					

Notes:
1.The stars indicate significance at the following levels: $*p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$
2. Robust standard errors in parenthesis

Table B.2: Factual Approach: Estimation of the Engel Curve using Low-Risk Industries as in EH (2003).

	IV			OLS		
	Food	Non-durables	Total Expenditure	Food	Non-durables	Total Expenditure
Disposable Labour Income	0.454***	0.598***	0.619***	0.274***	0.391***	0.400***
	(0.04)	(0.04)	(0.04)	(0.01)	(0.01)	(0.01)
Primary	0.044	0.023	-0.019	0.05	0.032	-0.003
	(0.06)	(0.04)	(0.04)	(0.06)	(0.05)	(0.05)
Energy	0.054	0.006	0.052	0.062	0.03	0.079*
	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
Construction	0.101***	0.102***	0.073***	0.091***	0.092***	0.068***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Distribution	0.086***	0.104***	0.084***	0.050**	0.066***	0.047**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Transportation and Storage	-0.043	-0.025	-0.006	-0.05	-0.038	-0.017
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Accommodation and Food	0.015	0.061	0.04	-0.049	-0.019	-0.024
	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Information and Communication	-0.016	0.007	-0.003	0.017	0.056**	0.052*
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Financial Insurance and Real Estate	0.006	0.065**	0.052*	0.047	0.118***	0.107***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Professionals (not accountants)	0.112***	0.063**	0.063**	0.135***	0.094***	0.102***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Public Admin and Education	0.018	0.036*	0.001	0.03	0.034*	0.007
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Human Health and Social Work	-0.044	0.003	0.005	-0.038	0.006	0.015
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Arts and Entertainment	0.061	0.086*	0.051	0.039	0.074	0.043
	(0.07)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Others	0.054	0.039	0.024	0.01	0.002	-0.032
	(0.07)	(0.05)	(0.07)	(0.06)	(0.05)	(0.05)
Controls for household characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5244	5257	5257	5244	5257	5257
Baseline	LR (EH)					

Notes:
1.The stars indicate significance at the following levels: $*p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$
2. Robust standard errors in parenthesis.

Table B.3: Summary of Industries that signal non-compliance by industry, type of expenditure and estimation method.

Food			Nondurables			Total Expenditure		
Manufacturing	EH	ST	Manufacturing	EH	ST	Manufacturing	EH	ST
CONS	CONS	CONS	CONS	CONS	CONS	CONS	CONS	CONS
DISTR	DISTR	DISTR	DISTR	DISTR	DISTR	DISTR	DISTR	DISTR
PROF	PROF NO ACC	PROF	PROF	PROF NO ACC	PROF NO ACC	PROF	PROF NO ACC	PROF
		FI-REAL	FI-REAL	FI-REAL	FI-REAL	FI-REAL	FI-REAL	FI-REAL
			PUB-DEF-EDU	PUBADMIN-EDU		PUBADMIN-EDU		
			ARTS	ARTS	ARTS			
			ADMIN-SUPPORT			ADMIN-SUPPORT		ADMIN-SUPPORT
			ACCO					(ENERGY)
								(INFO-COM)

Notes:

1. Industries between brackets turn insignificant when instrumenting income to control for measurement error.

2. ST stands for 'Stepwise'. The industries contained in the table are the ones that appeared significant and positive highlighting some degree of non-compliance in the stepwise regressions with alternative baselines.

Table B.4: Data-driven Approach: Estimation of the Engel Curve.

	IV			OLS		
	Food	Non-durables	Total Expenditure	Food	Non-durables	Total Expenditure
Disposable Labour Income	0.462***	0.614***	0.616***	0.281***	0.397***	0.405***
	(0.04)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)
Construction	0.103***	0.094***	0.074***	0.088***	0.083***	0.064***
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Distribution	0.089***	0.099***	0.084***	0.049**	0.059***	0.045**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Professionals	0.107***	0.053**	0.065***	0.125***	0.083***	0.096***
	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Financial-Real Estate	-	0.052*	0.086**	-	0.108***	0.060**
		(0.03)	(0.03)		(0.02)	(0.03)
Admin and Support	-	0.062*	0.054**	-	0.041	0.101***
		(0.03)	(0.03)		(0.03)	(0.02)
Controls for household characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5244	5257	5257	5244	5257	5257
Baseline	LR (ST)					

Notes:
1.The stars indicate significance at the following levels: $*p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$
2. Robust standard errors in parenthesis.

Table B.5: Estimation of the income-gap for GB.

	Construction	Distribution	Professionals	Rest of the industries
Compensation of Employees				
2010	45126	106256.4	52859.74	614934.8
2011	44528.67	102078.7	54781.31	593915.3
2012	44304.63	98801.25	54362.19	593321
2013	43586.36	100635.2	53232.72	595076.3
Average	44386.41	101942.9	53808.99	599311.8
Income-gap: Baseline- Manufacturing				
Estimated Income-gap (%)	23.10%	23.10%	26.50%	
Income-gap (£)	10253.26	23548.81	14259.38	
Total Income-gap (£)	48061.45			
Income-gap (% GDP)	3.07%			
Income-gap: Baseline- EH				
Estimated Income-gap (%)	21.50%	21.50%	25.90%	
Income-gap (£)	9543.079	21917.72	13936.53	
Total Income-gap (£)	45397.33			
Income-gap (% GDP)	2.90%			
Income-gap: Baseline-ST				
Estimated Income-gap (%)	19.60%	19.80%	22.90%	
Income-gap (£)	8699.737	20184.69	12322.26	
Total Income-gap (£)	41206.69			
Income-gap (% GDP)	2.63%			

Notes:
1. Compensation of Employees are as in ONS Blue Book 2016. All figures are in 2010 million pounds.
2. The variable of expenditure for which calculations of the income-gap are presented are food expenditure.
3. An average of GDP for 2010-2013 was used in the last row.

Appendix C: Graphical Annex

Figure C.1: Distribution of Expenditure and Income by industry. Baseline: Manufacturing.

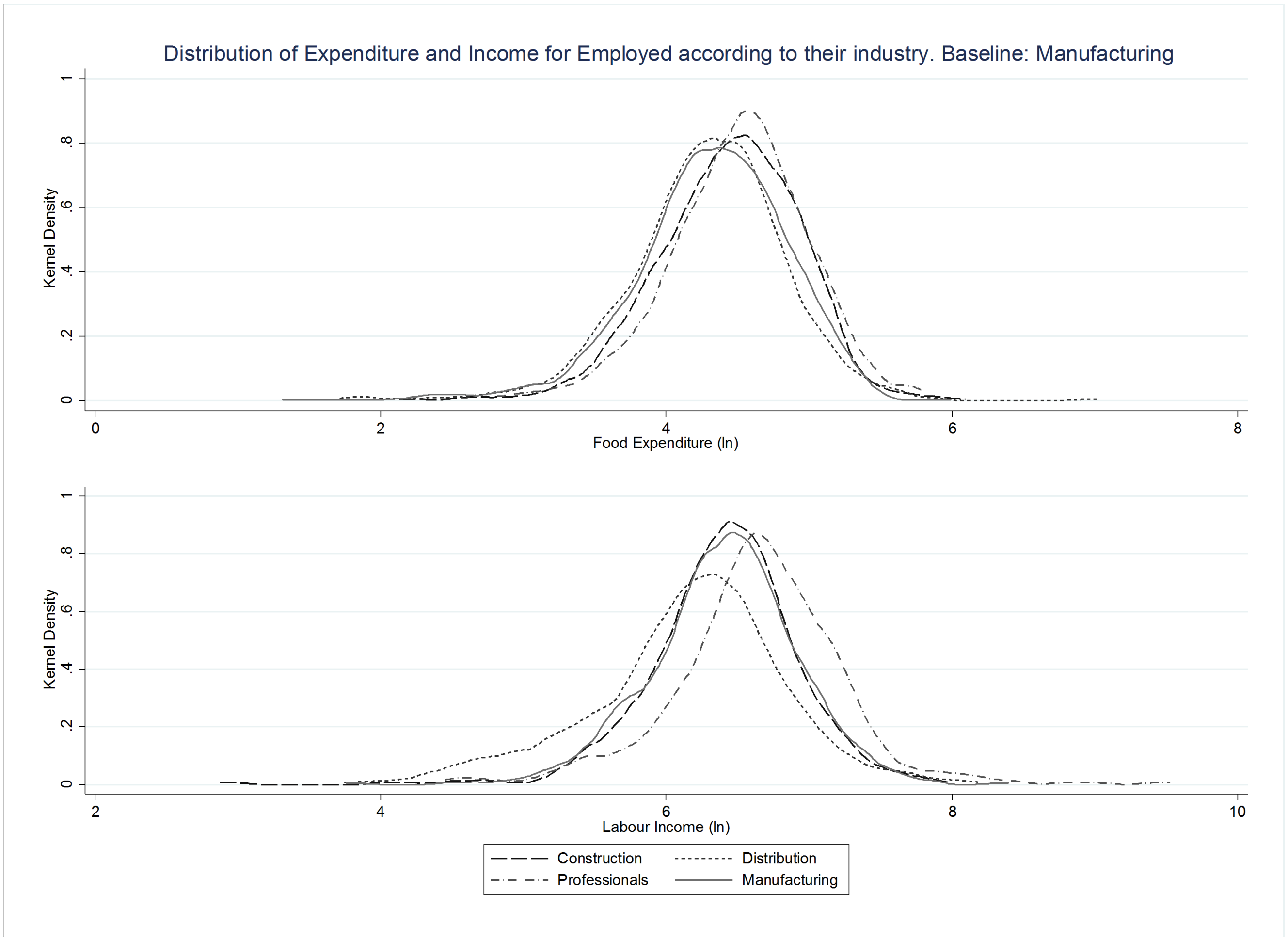


Figure C.2: Distribution of Expenditure and Income by industry. Baseline: Low-risk (EH).

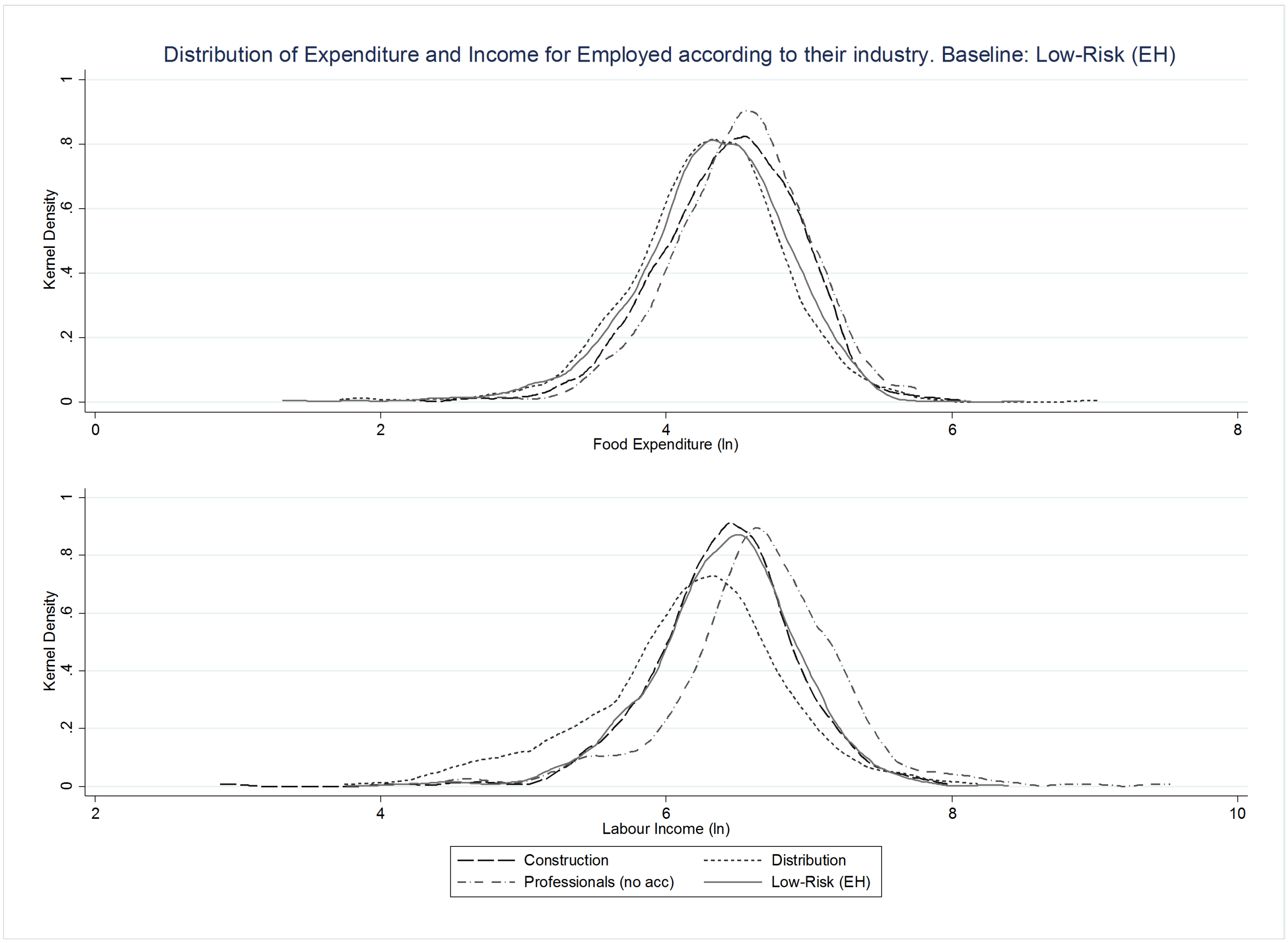
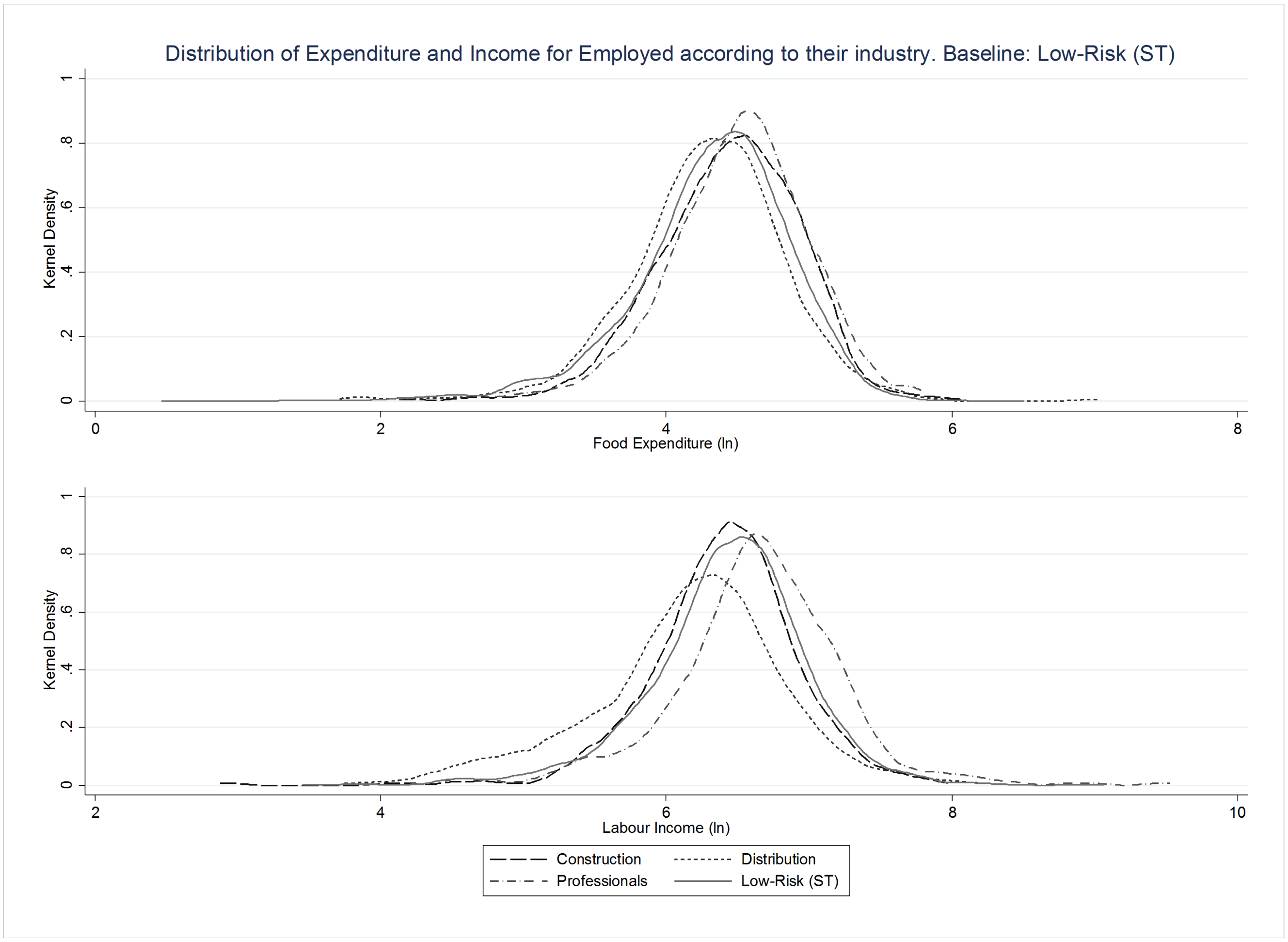


Figure C.3: Distribution of Expenditure and Income by industry. Baseline: Low-risk (ST).



Chapter 3

The Hidden Economy in Europe: Extrinsic and Intrinsic Motivations.

3.1 Introduction

The hidden economy is a pervasive phenomenon that erodes tax and social security bases causing public finance deficits and misleading employment figures. European countries recognise this and mark the reduction of the size of the hidden economy as one of the main strategies to meet the Europe 2020 employment goals (EU, 2015).[1]

Different methods have been presented to quantify the hidden economy, but if the aim is how to reduce it, then as Cowell (1990) states in his book 'sheer size is not the point'. Understanding whether compliance is extrinsically or intrinsically motivated becomes central to the question. Traditional tax compliance theory that frames the taxpayer as a utility maximizer, such as Alligham and Sandmo (1972), fails to be successful in explaining the high level of compliance observed in modern tax systems given the low probabilities of audit and moderate sanctions.[2] This observation outlines the non-trivial role of non-pecuniary factors in explaining compliance. Those non-pecuniary factors

[1]Fighting undeclared work stands as one of the strategies in Guideline 7 to enhance the functioning of the labour markets.

[2]Alm *et al.* (1992), calibrate the classical model of Allingham and Sandmo (1972) using probabilities of detection, fines and risk aversion that mimic those observed in modern tax systems and find that the level of risk aversion needs to be exceptionally high in order to obtain levels of compliance closer to what is observed. Note also that Alligham and Sandmo themselves were the first to recognise the ignorance of non-pecuniary factors in the model.

(high civic virtues, social norms, reciprocity or shame) that summarise the intrinsic motivation to comply are going to be labelled under the term 'tax morale'.[3]

We consider this compliance puzzle within the context of the hidden economy.[4] Using micro-level data from a publicly available survey on engagement in the Hidden Economy in Europe, we are able to observe individuals' participation, their perceptions about the classical deterrents of the tax administration, this is the degree of extrinsic motivation, and we can elicit their intrinsic motivation. We model the decision for participation across the 28 member states of the European Union using a multi-level specification that accounts for regional and country differences. This survey provides a very unique dataset that analyses undeclared work within a comparable framework across all European countries. The itemised nature of the data means that we can relate both motivations with the behaviour directly, as opposed to recent literature that uses macro approaches to proxy non-compliance (see Luttmer and Singhal, 2014).

Our main findings are as follows. First, we find that individuals respond to extrinsic motivations rationally. This is, participation is lower when the perceived extrinsic incentives—probability of detection and sanctions—are higher. However, this alone does not explain the low participation observed. Intrinsic motivations play a substantial role on constraining hidden economy participation. Second, we are able to analyse any cross effects of intrinsic and extrinsic incentives, this speaks to the literature on crowding effects. Our findings are consistent with Dwenger *et al.*, (2016) and Boyer *et al.* (2016) that, within the tax context, the crowding-out effect typically observed in

[3]A further discussion of the term is presented in Section 3.2.

[4] The OECD Handbook of the Non-Observed identifies four types of areas where non-observed activities tend to concentrate: underground, illegal and informal production and production of households for own final use. The activities we refer to in this paper fall under underground production—'activities that are productive and legal but that have been concealed from the authorities to avoid payment of taxes or complying with the regulations (OECD, 2002). The Eurostat and Istat framework divides underground production into two different categories according to the reason for being underground: because the enterprise conducting the activities is not registered or because the enterprises conducting them underreport. Note that in the literature several terms are used to refer to non-observed activities, e.g. shadow, grey or hidden economy. Van Eck (1987), as cited in OECD(2002), identifies at least 30 different terms. The differences in the meaning between these terms is very subtle and the definition of what they entail (underground, illegal, informal production) is necessary for a correct interpretation. In our case, we will use the term 'hidden economy' to refer to underground production in its broader sense. Thus the term hidden economy, underground production and undeclared work are used interchangeably. This term is also equivalent to the term 'shadow economy' as defined by Schneider and Williams (2013).

voluntary contributions is absent. This finding entails that there is no evident 'slippery slope' between the intrinsic and extrinsic motivations as proposed by Kirchler *et al.* (2008). Thirdly, we are able to rank compliance in different scenarios of intrinsic and extrinsic motivations and as expected compliance is at its highest when both intrinsic and extrinsic motivations are high, and the converse occurs when both are at their lowest. Lastly, a profile of the typical individual that performs undeclared work is presented showing an association of engagement in undeclared work and financial strain.

Our paper contributes to the literature on Hidden Economy reviewed in Schneider and Williams (2013) and to a wider literature on tax compliance reviewed in Andreoni, (1998). Despite the substantial amount of work on compliance, evidence on the impact of extrinsic and intrinsic motivations on participation in the hidden economy is scarce and inconclusive (Schneider and Williams, 2013). Previous studies aimed at analysing the impact of intrinsic motivations on participation used country level estimates of evasion. This approach however fails to reflect whether self-reported intrinsic motivation translates into compliance (Luttmer and Singhal, 2014). The present paper seeks to make headway into this question by using a unique set of individual data that allows to pin down attitudes, perceptions and behaviour.

The remainder of the paper is organised as follows. Section II provides a review of the relevant literature, Section III discusses the survey, the empirical strategy and presents a discussion of the variables included. Section IV presents the results and Section V discusses the robustness of our findings. Finally, Section VI concludes.

3.2 Literature Survey

In this section, we review the progress in the literature regarding the effect of motivations in engagement in undeclared work.[5] Undeclared work has not yet received enough recognition as a subject of analysis on its own. Until very recently surveys specifically targeted to analyse undeclared work are anecdotal, country specific and referring to a

[5]Most of the literature has been occupied with the measurement of the hidden economy. Reviewing this strand of literature goes beyond the scope of the present paper. For a critical evaluation of direct and indirect methods proposed to assess the size of the hidden economy refer to Gemmell and Hasseldine (2012).

certain moment in time. In fairness, the strongest tradition in using surveys to unravel the undeclared economy can be found in Denmark, Norway and the Netherlands.[6] The lack of homogeneity and thus comparability of the questionnaires, the territorial limitation and their lack of contemporaneity really challenges the development of a strong literature on the decision of participation in the hidden economy.

3.2.1 Extrinsic Motivation

Although the predictions from economic theory seem very straightforward, the evidence on the influence of the classical deterrents in compliance is mixed (Andreoni *et al.*, 1998).

Regarding the hidden economy literature, Van Eck and Kazemier (1988) find a negative effect of the probability of auditing on participation in the hidden economy in the Netherlands in 1982/1983. Pedersen (2003) corroborates the finding for men in Denmark, Norway, Sweden and Germany; and for women in Germany and marginally in Sweden. However, no evidence is found for Great Britain.[7]

The role of sanctions is understudied. Feld and Larsen (2009) in their study of the hidden economy in Germany fail to document an effect of fines in constraining participation in the hidden economy in any wave of their survey (2004-07), nor in the pooled sample. In the experimental tax evasion literature, some studies have documented the significance of both audits and sanctions (Alm *et al.*, 1992; Blackwell, 2007).

3.2.2 Intrinsic Motivation

The term 'tax morale' has been coined as an umbrella for all non-pecuniary factors affecting compliance. It is defined generally as the 'intrinsic motivation to pay tax' (Alm and Torgler, 2006; Feld and Frey, 2002; Braithwaite, 2003; Braithwaite and Ahmed, 2005). Taxpayers are endowed with a considerable amount of civic virtue and tax morale that explains why most individuals are found to comply even when incentives are low (Frey, 2003). Dwenger *et al.* (2016) show that in a very particular setting, a

[6]See Van Eck and Kazemier (1988) for Holland; Feld and Larsen (2005, 2009), Merz and Wolff (1993) for West Germany; Isachsen and Strøm (1985) for Norway; and Pedersen (2003) for Germany, Great Britain and Scandinavia.

[7]This may be attributed to the sample size for this particular country.

church tax where there was no enforcement, still compliance was about 20%. These individuals were clearly intrinsically motivated to comply as no outside incentive was present.

Evidence on the effect of tax morale in compliance is scarce due to the difficulty in measuring and studying it empirically (Frey and Jegen, 2001). In order to test the impact of morale, the hidden economy literature has used a country level approach where morale, proxied from survey variables is introduced as a regressor on estimations of the size of the hidden economy (Feld *et al.*, 2011; Torgler and Schneider, 2009; Torgler *et al.*, 2007). Leaving estimation issues about the dependent variable aside (reviewed in Slemrod and Weber (2011)), this approach eliminates the individual dimension of the compliance decision. Most variance occurs at the individual level: *individuals'* perceptions and intrinsic motivation is what shapes the behaviour and the link between self-reported attitudes and behaviour is then lost through aggregation (see Luttmer and Singhal, 2014).

Individual survey data on participation is more scarce but findings are consistent. Feld and Larsen (2005, 2009) and Van Eck and Kazemier (1988) also find a negative effect of tax morale on the probability of participation. Torgler *et al.* (2007), using lab experiments are also successful in finding a role of the intrinsic motivation in inhibiting non-compliance.

3.2.3 Crowding in or out?

Motivation crowding theory studies the link between the intrinsic and extrinsic motivations. Both the economic and psychological literature have evolved to recognise that external incentives can influence the intrinsic motivation (Frey and Jegen, 2001; Feld and Frey, 2007). Crowding-out effects have been found related to voluntary contributions such as blood donation (Mellström and Johannesson, 2008); services (Gneezy and Rustichini, 2000) and public goods allocation (Frey and Oberholzer-Gee, 1997).[8]

Within the tax compliance context, crowding effects are far from conclusive. A strand of literature has emerged from psychology that aims at shedding light over the reasons

[8] The literature on empirical findings of a crowding effects is surveyed by Gneezy *et al.* (2011).

for compliance and the interplay of the motivations. The slippery slope framework presented by Kirchler, Hoelzl and Wahl (2008) presents a descriptive approach to understanding tax compliance.[9] The framework outlines that taxpayers differ in their motivations to comply. It describes the importance of both classical deterrents, referred to in this framework as *power of the authorities*, and non-pecuniary factors, in their case *trust in the authorities*, in understanding compliance.[10] Whether taxpayers comply as a response to the power of the authorities or to the trust in the authority, implies a different type of compliance, enforced in the first case and voluntary in the second case. The framework describes that both strategies, building a trusting relationship with the taxpayer and exerting power, are necessary to guarantee high compliance but that the interaction of both dimensions is crucial. They recognise that both dimensions could act as complements or substitutes. To illustrate this, increases in the power of the authorities might be interpreted by taxpayers as a sign of mistrust and therefore lower compliance, evidencing a 'slippery slope' between the two dimensions; while it could also be interpreted as a sign that the tax authority is making taxpayers pay their 'fair share' and boost compliance. KHW simply outline that both scenarios are possible in the interaction of the two dimensions, a crowding in and out effect. Evidence of the 'slippery slope' main predictions has been found in lab experiments such as Kogler *et al.* (2012) and survey studies as in Muehlbacher *et al.* (2011).[11] Empirical evidence however about the incidence of power on trust is scarce and far from conclusive.[12]

[9] Note the description of the framework was formalized in Prinz, Muehlbacher and Kirchler (2014).

[10] Trust has consistently been found as a determinant of tax morale. See Daude *et al.* (2012). Note it is beyond the scope of this paper to study the determinants of tax morale. For this literature see for example Daude *et al.* (2012) and Alm and Torgler (2006).

[11] Muehlbacher *et al.* (2011) find the primary predictor of enforced compliance is power and trust of voluntary compliance. They find that increases in trust reduce enforced compliance but trust and power do not seem to be related in terms of voluntary compliance.

[12] Fischer and Schneider (2009) address the question and find suggestive evidence of a crowding-in effect. As their design does not allow to observe participation, Fischer and Schneider (2009) use tax morale as their dependent variable to proxy voluntary compliance. The shortcoming of this approach is that tax morale indicates attitudes not behaviour. Also, the empirical application casts some doubts about the validity of the results. On a separate study, Lisi (2012) use the size of the hidden economy as a proxy of compliance obtained from Schneider *et al.* (2010) and regresses it over power, trust and tax morale and GDP. It seems worrying the overlap between the covariates used to estimate the size of the shadow economies and the model proposed by Lisi (2012) which may be the source of the collinearities that are found on the analysis. It also seems from the paper that when addressing the interaction effect between trust and power the level effect is not included and the interpretation of the interaction then becomes unclear. Unlike both these studies, we are able to observe the effect of the intrinsic motivation (morale) and extrinsic motivation in actual participation in undeclared practices and therefore we circumvent the issue of proxying compliance. Note that we use the term

Our objective in this study can link and be interpreted in the light of this framework as empirically testing some of its hypothesis. First, we empirically test the relevance of both extrinsic and intrinsic motivations in understanding compliance, in our case participation in undeclared activities. Second, we also seek to understand how both types of motivations interact shedding light over whether there is in fact a case for a 'slippery slope' between the intrinsic and extrinsic motivation. We however leave open another possibility: the absence of an interaction between the two. While we share commonalities we also differ in two aspects that does not enable us to present this paper as a full test of the slippery slope framework. First, our measure of non-pecuniary factors that affect non-compliance differs from that chosen in KHW. We build a multi-dimensional proxy of the intrinsic motivation using the individual's attitudes towards non-compliant tax and non-tax practices while trust in the authorities is the dimension outlined by KHW to measure the individual's intrinsic motivation. Trust in the authorities is a dimension not available in the survey we use but as we will argue in Section 3.3.2.3 and further in Section 3.5.1, we believe our measure of the intrinsic motivation to be an accurate reflection of a complex construct. Second, we are not able to link the type of compliance observed (enforced vs. voluntary) to the type of motivation (extrinsic and intrinsic) as is one of the predictions of KHW. We can ascertain whether both motivations matter and interact but we are not able to establish the link to enforced and voluntary compliance as was done in a experimental setting by Muelhbacher *et al.* (2011).

Natural field experiments have provided an exciting new avenue to analyse the interplay of extrinsic and intrinsic incentives. Dwenger *et al.* (2016) implement a field experiment on a local church levy in Germany where variations of the enforcement parameters and rewards are studied to analyse their effect on compliance. The almost zero deterrence baseline previous to the intervention allows an identification of those who are intrinsically motivated, this is that comply with the levy in the case of no

shadow to preserve the original nomenclature in the paper. Shadow economy has been broadly used in the literature with the definition of Schneider and Williams (2013) which is the one for underground production given by the OECD (2002). In this sense for our case hidden economy and shadow economy amount to the same concept.

enforcement. Focusing on the deterrence treatment, where the level of the enforcement parameters are manipulated, strong effects on compliance are found for the baseline evaders—those not intrinsically motivated—but no effect is found for the intrinsically motivated. Therefore, they document an absence of crowding effects between the two types of motivations.[13]

Boyer *et al.* (2016) in a field experiment also on the context of church levy in Germany investigate whether extrinsic incentives can backfire crowding out the intrinsic motivation. They identify the level of intrinsic motivation by analysing the contributions on the baseline on a context of low deterrence. Letters phrasing the church tax as either a voluntary contribution or a compulsory tax were sent as part of the design. Individuals who are strongly intrinsically motivated—defined as those who are regularly contributing in the baseline—do not respond to the information that the church levy is a tax. This is, an increase in the extrinsic incentives does not crowd out intrinsic motivation in the presence of a binding tax norm for those who are intrinsically motivated. This finding is ratified for those who are occasional contributors. For baseline non-contributors, those with low intrinsic motivation, the tax context increases compliance, as expected as they are only driven by extrinsic incentives. The authors also evaluate the scenario where the levy is treated as a voluntary contribution. In this case, the strongly motivated do not react to this information but a crowding-out effect is found in the case of those occasional and not intrinsically motivated. In essence, the moment the tax context is set, any crowding effects disappear.

3.3 Data, Model and Empirical Method

3.3.1 Data

The data for this study comes from the Eurobarometer Survey (EBS) on Undeclared work conducted upon request of the European Commission in 27 member states of the European Union and Croatia between April and March 2013. The survey constitutes a unique dataset as it provides a common framework to analyse undeclared work

[13] We defer the comment on the results about how rewards affect compliance as it is not relevant in this setting.

in a comparable way for the different European economies.[14] Face-to-face interviews were performed on a sample of 27,563 interviewees which are residents in one of the aforementioned countries and are aged 15 or over. All individuals are presented with a common definition of undeclared work which excludes criminal activities. Further details of the survey are provided in Appendix A.

3.3.2 The variables

We present in this section a description of the variables used. A description of the variables and summary statistics are available on Appendix B.

3.3.2.1 Dependent variable

The dependent variable captures participation in the hidden economy by the interviewee. The question posed in the survey reads:

"Apart from a regular employment, have you yourself carried out any undeclared paid activities in the last 12 months?"

This variable is a dichotomous variable that takes the value 1 if the individual has engaged in undeclared work practices and 0 otherwise. There is a very small number of individuals who refused to answer the question (around 2% of the sample).[15]

The reliability of using self-reported measures of deviant behaviour in general has been highly debated in the social science literature. Self-reports may be affected by three factors: self-presentation, awareness of the behaviour and fear of being identified. Self-presentation refers to individuals trying to project an image that is related to their 'ideal' self (closer to commonly held social norms) rather than their 'real' self. The second, awareness, refers to the individual not considering their acts as deviant or simply forgetting about having engaged in undeclared work. Several studies have accounted for the fact that the longer the time between the occurrence of the deviant act and the questioning, the lower the awareness (Hessing *et al.*, 1992). However, awareness is

[14] Until now, studies have been atomistic, referred to a certain country at a certain period of time. This has not enabled a one-to-one comparison of estimates across studies.

[15] See Section 3.3.3 and Appendix D for the treatment of missing observations.

affected by salience. If undeclared work is performed frequently or a substantial income is raised from it, then the individual is more likely to remember. Lastly, responses can be affected by fear that the survey will not preserve anonymity and it could potentially lead to identification by the tax authority of those engaging in undeclared work.

The survey is carefully constructed to minimise the impact of these issues and increase the quality of self-report. First to note is a great peculiarity of the EBS survey: individuals are interviewed about a varied amount of topics. In this case, for the EBS 79.2 the topics included are the internal market, cultural activities, non-urban road use and quality and tolls, science and technology, and lastly undeclared work and tax fraud. This acts as an advantage for this dataset against surveys that directly target undeclared work as it tames the salience of the topic. Within the undeclared work section, the questionnaire is built using careful priming techniques in order to ensure that the sensitivity of the questions increase as the questionnaire evolves.[16] This method is advised to maximise responses to sensitive questions (Bradburn *et al.*, 2004). Second, the interviewee is reminded constantly about the anonymity of the answers.[17] Third, the interviewees are provided a common definition of undeclared work where it is presented as a socially spread phenomenon which seeks to encourage individuals to respond more honestly.[18] Last, the questions refer to the previous 12 months reducing the incidence of recall errors.

As can be ascertained, many mechanisms are in place in order to ensure the maximum reliability of the self-reports. The opportunities that this singular survey brings providing a common framework to analyse undeclared work participation, its ability to directly link engagement to the intrinsic and extrinsic motivations and the rich bank of demographic information that could only be obtained from a direct method, clearly

[16]Within the undeclared work section of the questionnaire, the interviewer asks first about knowing someone who participates in the hidden economy, follows by asking about purchases and finally enquires about own participation. The number of individuals who recognise knowing someone who is engaged in undeclared practices is significantly larger than the number of positives for own participation. This enables us to use another dependent variable to proxy own participation. We return to this observation in Section 3.5.

[17]Along the survey, the following statement is mentioned throughout: "The following questions are of a sensitive nature and I would like to confirm you that all the information collected is handled in strict confidentiality and anonymity. Your answers to the following questions therefore will remain absolutely ANONYMOUS."

[18]The definition is presented in Appendix A.

outweigh the concerns for any residual effect of the aforementioned issues.

In Section 3.5 we perform a robustness exercise where an alternative dependent variable is used.

3.3.2.2 Extrinsic Motivation

Extrinsic motivations are captured using the individual's perception about the probability of being audited and the perceived severity of the sanctions. Given that deterrence strategies usually remain confidential, there is evidence that taxpayers do not have an accurate idea of the actual probability of being audited (Scholz and Pinney, 1993; Del Carpio, 2014), but that this perception may be influenced by previous experiences with the tax administration (Hessing *et al.*, 1992; Sheffrin and Triest, 1992).

Williams and Hawkins (1986) make the case for using perceptions of the extrinsic incentives rather than the actual values: 'Deterrence theory implies a psychological process whereby individuals are deterred from committing criminal acts only if they *perceive* legal sanctions as certain, swift and/or severe'. Considering objective measures of the deterrents does not recognise the fact that each taxpayer is unique and that we are observing a behavioural response to their subjective beliefs of the probabilities of auditing and severity of the sanctions.[19]

The perception about the extrinsic motivations are captured in the survey using two separate questions. First, interviewees are inquired about their perceived risk of detection resulting from performing undeclared work and their answers can range from very small to very high. We allow 'don't know' as a valid response. Second, individuals are asked about the severity of the sanctions to be received if caught. These responses range from returning the amount due of tax and social security contributions, paying a fine, going to prison or other sanctions. 'Don't know' is also allowed as a valid response to reflect uncertainty. The uncertainty about the sanction faced is larger than

[19]Feld *et al.* (2011) study the effect of deterrence in the shadow economy, trying to disentangle the causality between the two across time. They use MIMIC currency demand estimates and penalties per investigation and firms per audit as proxies for deterrence. These measures are objective measures of deterrence and therefore ignore the fact that individuals' perceptions might be completely misguided. Using objective measures to proxy perceptual measures roots on the belief that there is a positive correlation between the objective reality of the tax administration deterrence policy and the subjective perception of the taxpayers.

the uncertainty about the risk of being detected (10% of valid responses in the case of sanctions and 0.8% for risk of detection).[20]

3.3.2.3 Intrinsic Motivation

Behaviour is both personal and situational (Yinger, 1963), therefore it may seem rather unnatural to aggregate motivations at a country level. Contrasting with previous literature that aggregate morale at a country level, we recognise the link between the self-reported intrinsic motivation and the behaviour (Luttmer and Singhal, 2014).

We construct a composite measure of morale that includes perceptions about a wide variety of dishonest acts: free-riding on public transport or the welfare system, concealment of income from undeclared work by firms or individuals and tax evasion.[21] Interviewees are asked to rank these acts from 1 "absolutely unacceptable" to 10 "absolutely acceptable". In this paper we invert the order of the variable so as to ease interpretation. The scores given by each individual for each of the items named above are averaged to obtain a proxy for the individual's morale. For ease of reference, this variable that includes a proxy for each interviewee's intrinsic motivation is going to be referred to as 'Broad Morale' (BM) and is going to be the leading measure used through the paper. Alternative variables to proxy the intrinsic motivation are presented as a robustness check in Section 3.5.

As opposed to most of the literature, we use a multi-item measure of the intrinsic motivation in order to better capture the complexity of the construct by allowing more variability in the measure. Single-item measures tend to be sensitive to random errors in measurement which tend to be averaged out in multi-item measures.

The unidimensionality of the construct was tested through Exploratory Factor Analysis (EFA). A one-factor solution is appropriate which makes it legitimate to average the scores as has been done to compute our variables for the intrinsic motivation.[22] The

[20] Individuals can also refuse to answer these question. We address the issue of missing data on Section 3.3.3 but the number of refusals are minimal around 1.6% in the case of risk of detection and about 3% for sanctions.

[21] Full description of this variable is available in Appendix B.

[22] We have also used factor loadings to consider each item's contribution to the factor and the results are invariable.

scale internal reliability has been tested using Cronbach's alpha. The alpha coefficient is 0.8728, suggesting a relatively high internal consistency of the items.[23] Further information about the EFA can be found in Appendix C.

3.3.2.4 Other relevant variables

A range of individual and household characteristics are controlled for. The individual characteristics we use are: age, gender, education, occupation, and marital status. We include gender to capture if there exists a different participation rate for men and women as there is in the formal sector. Age will determine the different behaviours along the life-cycle. Occupation captures whether different groups tend to participate more intensively than others which will have an impact on the type of policy required to reduce engagement. Marital status will capture possible differences in behaviour that may be linked to the type of household. Regarding household composition, we control for the number of children and the size of the household in general.

Among factors that can affect engagement in undeclared practices, the literature has identified that those experiencing financial strain and those that perceive their situation as comparatively worse are more willing to engage in evasion (Wärneryd and Walerud, 1982; Webley *et al.*, 1991). Due to the lack of a variable measuring income directly in the survey, we use two distinct groups of variables: to capture financial strain, we use the frequency with which the interviewees struggle to pay their bills; and to capture social comparison, we use a self-classification of the individual into social classes and a score of what *level* they think they hold in society. For further details on these variables see Appendix B.

3.3.3 Empirical strategy

As we observe participation, we can investigate the extensive margin responses to the intrinsic and extrinsic motivation. The geographical variation in the data allows us to locate each individual i in a particular region j of country k. In order to account for this multilevel structure, we are going to fit a three-level logistic regression which allows for

[23]Conventionally, an alpha higher than 0.7 is considered as acceptable.

the fact that individuals pertaining to a certain region and country are more similar to one another than individuals in different regions and countries. There are many reasons why individuals living in the same regions may be more similar to each other. First, there are substantial cultural differences between regions. Tax policy could also introduce between regions variation if some regions are more intensively targeted by the tax administration with their compliance activities—be it audits or public campaigns—increasing the overall perception of risk in the region. The institutional framework differs from country to country as well as the tax environment making individuals more dissimilar across countries and social norms in general will differ.[24] All these effects are captured by two group random errors at the regional and country level. A group random error at the country level will capture differences in participation that endure across time. Certain countries or regions can tolerate or condemn undeclared practices to a varying degree both from a tax enforcement perspective or from a cultural perspective (norms). This will cause unobserved variance that will be wrongly attributed at the individual level. Given that countries are not homogeneous, regional variation is necessary. We find both between country variation and within-country between-region variations to be relevant in our framework. Using LR tests, the multilevel model is superior to the lineal model and the model with three-levels superior to the model with two levels.

We can define the probability of supplying undeclared work for each individual i that lives in region j of country k, as $\pi_{ij} \equiv Pr(y_{ijk}|Intrinsic_{ijk}, Extrinsic_{ijk})$, where y_{ijk} is our response variable, and x_{ijk} is a vector of variables containing individual specific covariates. Using the latent response form, there is an underlying continuous latent variable y^*_{ijk} for which we can define the three-level random intercept model as,

$$y^*_{ijk} = \beta_0 + \beta_1 Intrisic_{ijk} + Extrinsic_{ijk}\delta + v_{0k} + u_{0jk} + \epsilon_{ijk}, \tag{3.1}$$

where $Intrinsic$ is a variable representing the intrinsic motivation; $Extrinsic$ is a vector containing the individual's perception of both sanctions and audits; ϵ_{ijk} is a level-1 error

[24]Failure to take this nested structure into consideration may lead to biased results as the assumption of conditional independence is broken due to the existence of group random errors.

term with a standard logistic distribution with mean 0 and variance $\pi^2/3$; [25] and v_{0k} and u_{0jk} are the cluster and super cluster random effects which are normally distributed as,

$$v_{0k} \sim N(0, \Sigma_v) \quad ; \quad u_{0jk} \sim N(0, \Sigma_u), \tag{3.2}$$

$$\Sigma_v = \sigma_v^2 I \quad ; \quad \Sigma_u = \sigma_u^2 I. \tag{3.3}$$

Given that both pecuniary and non-pecuniary factors are important in influencing non-compliance, it is important to identify whether these two interact, that is, whether there is a crowding in or out effect of the intrinsic motivation. In order to do so, we include in the regression the interaction between the intrinsic motivation—morale—and the extrinsic—sanctions and detection—as,

$$y^*_{ijk} = \beta_0 + \beta_1 Intrisic_{ijk} + Extrinsic_{ijk}\delta + Intrinsic_{ijk} * Extrinsic_{ijk}\rho + v_{0k} + u_{0jk} + \epsilon_{ijk}, \tag{3.4}$$

where all the variables are defined as in (3.1) and $Intrinsic_{ijk} * Extrinsic_{ijk}$ represents the interaction of both types of motivation. The error terms are defined as in (3.2) and (3.3). A significant coefficient of the interaction would indicate a crowding effect between the intrinsic and extrinsic motivation.

Our estimation strategy does not allow for the unambiguous identification of the direction of the causal effect. It is fair to recognise that causality might partially run reversely. That is, lower tax morale might trigger participation but also through engaging in undeclared work practices, the interaction with other agents might lower morale as well.[26] In this sense, our results should be interpreted as an association rather than

[25] The level one variance for logit models is fixed at 3.29, if there is any variation in the inclusion of explanatory variables this would be reflected in the level 2 and 3 variance.

[26] Previously in the literature, Torgler and Schneider (2007) have attempted to address the problem of endogeneity of the intrinsic motivation on the size of the hidden economy in their case. They use instrumental variables techniques where morale is instrumented using the index of cloudiness arguing that cloudiness has been found to impact negatively attitudes. The authors have run the corresponding instrumental variable tests to assess the validity of the instruments. They still find a negative association between higher levels of morale and the size of the shadow economy. However, the weather conditions might also affect the size of the underground production. We considered this approach but refrained from addressing causality due to the difficulty in finding sound instruments using this route and the inability to find a natural experiment to aid identification.

looking for the causality. In our discussion, we focus on the direction and the robustness of the results rather than on the size of the coefficients.

Missing data in the survey[27] was diagnosed and the appropriate techniques were applied to address the missingness mechanism. Multiple imputation and sensitivity analysis were conducted to assess the robustness of the results. 91% of the records of interviewees were complete for all variables included in the estimation and the missing observations corresponding to the more sensitive questions regarding engagement in undeclared work represented 3% in the case of own participation, our leading dependent variable.[28] The results of multiply imputing the missing observations did not significantly differ from the results using complete case analysis due to the low number of missing observations and therefore for computational purposes we decided to use only complete records.[29]

3.4 Results and Discussion

In this section, we discuss the results of our main estimation strategy. Long regression tables and robustness checks for brevity have been relegated to Appendix E, only summary tables of the main results, these are those relevant to the motivations and their interactions, are included within the text to aid the discussion.[30]

Table E.1 in Appendix E contains the full results of the estimation of equation (3.1) that analyses the impact of four sets of covariates in the decision to non-comply: sociodemographics, financial constraints, extrinsic incentives and intrinsic motivation. For that purpose, the table has been conveniently divided into four panels to be discussed within

[27]Refer to Appendix D for further information on missing data.

[28]Note that as quoted in Section 3.3.2.1, 2.2% of these missing observations were refusals to respond and the other 0.8% corresponds to individuals who answer ‘don’t know’ to whether they engage or not in undeclared work. Bradburn *et al.* (2004) indicate that ‘don’t know’ in the case of sensitive questions about behaviour can in fact be a polite refusal. Thus, in this case, ‘do not know’ is treated as a missing observation which is supported by the low occasion of this answer and that does not allow us to classify it as a positive or negative response. A multinomial multilevel model was considered at first but discarded due to the extra computational power required and the lack of additional interpretation given due to the low occasion of ‘do not know’ as an answer.

[29]For more information on missing data and the procedure undertaken on this part, please refer to Appendix D.

[30]Note also that as discussed in Section 3.3.3 we do not interpret the magnitude of the effects but rather the direction of the results and thus we defer from referring to the specific figures on the tables discussed but rather on the sign of the relationship.

this section. Panel A and B contain the results that refer to the demographics and the financial situation of the participants are to be discussed in Section 3.4.1. For convenience, Panel C and D as they refer to the core results of the estimation are shown within the text in Table 3.1. We report weighted and unweighted results to test for the robustness of the weighting procedure as suggested by Carle (2009).[31]

3.4.1 Sociodemographics and Financial Constraints

Panel A in Table E.1 discusses the main demographics of the participants in the hidden economy. It suggests that the employed and the retired appear to participate less in the hidden economy than individuals who are self-employed (see *Occupation*).[32] The variable that indicates the unemployed is not significant which suggests that the level of participation of the self-employed and the unemployed is not statistically different. There are several reasons why the self-employed might be more active in undeclared practices. A high burden of taxes, a high regulatory environment and difficult financial access have been identified as determinants of informality (Dabla-Norris *et al.*, 2008). The variability of the income-flow of the self-employed can also stand as one of the apparent reasons for recurring to undeclared work as a way of smoothing their income flow. According to those self-employed that admitted supplying undeclared work in the survey, the high burden of taxes and social security contributions, the lack of regular jobs in the market, the low salaries in the formal market and the lack of control from the authorities stand as the four primary reasons that prompted their participation.[33] This result is in line with Lobo (1990), that finds that a third of the informal work in Spain

[31] As discussed in Anderson *et al.* (2013) there are different tendencies in terms of the inclusion or not of weights. A model approach defends that the introduction of weights can lead to more inefficiency if the weighting procedure is independent of the probability modelled, but the design approach defends the introduction of weights if the sampling weights carry relevant information, that is they vary a lot across observations. Also, in multilevel models ideally weights should be introduced at each level of the multilevel structure. In our case, levels are only available at level 1 which is the same as assuming that all observations at level 2 and 3 have a weight of 1. Rabe-Hesketh and Skrondal (2006) discuss that multilevel models for dichotomous responses can be affected by the scaling of level 1 weights which are the ones considered in the present paper. In our case we have produced unweighted and weighted results as suggested by Carle (2009) in order to test for the importance of the introduction of weights. Note that we only do so for the main estimation result discussed in Table 3.1 but in the subsequent specifications only weighted results are displayed. We make this explicit in the footnotes of the tables discussed.

[32] Variable names are presented in italics.

[33] The reasons for supplying undeclared work are explicitly asked in the survey.

is carried out by self-employed individuals, Pahl (1988) for the UK and Williams and Horodnic (2016) for an earlier wave of the Eurobarometer Survey. However, evidence on the link of occupations and engagement in undeclared work is far from conclusive (Pedersen *et al.*, 2003; Haigner, 2011).

Participation seems to be more widespread among men than women (*Male* in Panel A). This feature appears as the occupations of the individuals who report participation are male dominated in the hidden economy. We have tested this result by analysing the occupations that individuals take in the hidden economy separately. In the first group we considered individuals who perform babysitting, cleaning, and ironing tasks as part of their hidden economy activities. The second group was composed of individuals who did gardening, repairs and renovations, car repairs, or help moving house. The third group was composed of individuals who gave administrative or IT assistance and tutoring and the fourth included the selling of goods and services. We find that the categories are male dominated with the exception of the first one.

We find no conclusive evidence of an age pattern (*Age* in Panel A) and no significance in the education variable, which may be because the effect is likely to be correlated with other variables in the model such as occupational choice or some of our wealth indicators.

Individuals who face higher financial strain are found to engage more widely in undeclared work as exhibited in Panel B. A higher struggle to pay their bills (*Bills*) makes individuals more prone to participate. This is also marginally captured by the variables social class (*Social Class*) which shows that individuals who are in the working class participate more than individuals in higher social classes.

3.4.2 Extrinsic and Intrinsic Motivation

The discussion around the impact of motivations upon compliance can be structured into two parts. First, the discussion of the impact of intrinsic and extrinsic incentives in the decision for participation which can be interpreted from the estimation of equation (3.1). Second, the analysis of whether there exists an interaction between both types of motivations making use of (3.4). Table 3.1 summarizes the main estimation results to be discussed in this section.[34]

Column (1) and (2) of Table 3.1 summarize the main results, unweighted and weighted respectively, from estimating equation (3.1) to analyse the impact of the different motivations on participation. Participation is responsive to the extrinsic incentives—the *perceived* severity of the sanctions and the *perceived* risk of detection—on the expected direction. The negative coefficient of the risk of detection, shows that as the perceived probability of detection increases, the odds of participation decreases (baseline category is very low risk of detection).[35] The odds of participation almost halve when the perceived risk of detection is raised. The perceived severity of the sanctions is also inversely related to the probability of performing hidden activities judging also by the negative coefficient of the sanctions variable. Participation is lower when the supply of undeclared work is believed to be punished with a fine than when no fine is believed to be due (omitted category).

As is common in the tax evasion literature, the deterrence effect of sanctions is less conclusive than the effect of the probability of detection. This could well be related to the way sanctions are measured. The relevance of the severity of sanctions in their role of deterring undeclared work is directly related to two matters. The first one is individuals' perceptions about their chances of being caught: even if they think the sanction is quite severe, if the chances of being caught are minimal then the deterrent power of sanctions is lessened.[36] The second is enforcement: if there are severe sanctions established but not enforced, then this will cause the sanctions to also lose their relevance, this is regarded as the tipping point or the threshold effect. The significance of prison as a believed sanction for evading taxes is lost most likely due to the low number of occasions with which this sanction is mentioned.[37]

[34]Note however that full estimation results of equation (3.1) are available in the Appendix E. The results presented in the text in Table 3.1 are equivalent to Panel C and D in Table E.1 in Appendix E. Within text results are only a summary of the main estimation coefficients to be discussed. Full results for equation (3.4) are not included in the paper but available upon request.

[35]Reference to the baseline categories and further comment on the construction and specification of the variables is available in Appendix B.

[36]Williams and Hawkins (1986) outline the importance of the link between perceived certainty and perceived severity of sanctions. Grasmick and Bryjak (1980) found a significant negative effect of

Table 3.1: Summary of Estimation Results for the Supply of Undeclared Work.

	(1) Unweighted	(2) Weighted	(3) Unweighted	(4) Weighted
Extrinsic Motivation: Tax Administration				
Sanction				
Due plus fine	-0.174**	-0.141*	0.507	1.115***
	(0.0795)	(0.0757)	(0.313)	(0.416)
Prison	-0.198	0.0346	1.174*	2.598***
	(0.174)	(0.316)	(0.654)	(0.516)
Other	0.390***	0.346***	-0.296	0.348
	(0.138)	(0.106)	(0.475)	(0.992)
DK	-0.220	-0.0491	-1.232***	-1.101*
	(0.137)	(0.166)	(0.474)	(0.621)
Risk of Detection				
Fairly Small	-0.381***	-0.426***	-0.0825	-0.567
	(0.0848)	(0.0945)	(0.333)	(0.519)
Fairly High	-0.885***	-0.903***	-0.879**	-1.429***
	(0.101)	(0.152)	(0.381)	(0.382)
Very High	-0.777***	-0.736***	-0.548	-2.410***
	(0.168)	(0.257)	(0.653)	(0.925)
DK	-0.879***	-1.171***	-0.793	-1.342
	(0.170)	(0.349)	(0.628)	(0.968)
Intrinsic Motivation				
Morale	-0.367***	-0.413***	-0.327***	-0.365***
	(0.0185)	(0.0417)	(0.0427)	(0.0821)
Interactions				
Morale* Sanction				
Fine*Morale			-0.0889**	-0.163***
			(0.0405)	(0.0569)
Prison*Morale			-0.182**	-0.335***
			(0.0853)	(0.0671)
Other*Morale			0.0982	0.00493
			(0.0632)	(0.131)
DK*Morale			0.141**	0.136
			(0.0612)	(0.0829)
Morale*Risk				
FS*Morale			-0.0404	0.0169
			(0.0427)	(0.0707)
FH* Morale			-0.00203	0.0679
			(0.0492)	(0.0468)
VH* Morale			-0.0306	0.205**
			(0.0819)	(0.101)
DK*Morale			-0.0126	0.0216
			(0.0806)	(0.0982)
Sociodemographics and household			Yes	Yes
Financial Constraints			Yes	Yes
Number of Observations			24,623	24,623

Notes:
Column (1) and (2) present the results of estimating equation (3.1) and column (3) and (4) the result of estimating equation (3.1). The dependent variable is Own-participation in undeclared practices. Coefficients are expressed in log odds metric. To attain the odds ratio interpretation exponentiation of the coefficients is required. Full estimation results are provided in Table E.1 in Appendix E. Weighted estimation includes design and post-stratification weights. The stars indicate significance at the following levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$

The intrinsic motivation acts as a constraint for participation as shown by its significant coefficient at 1% in columns (1) and (2) Table 3.1 (see *Morale*). Individuals with higher morale have lower odds of joining the hidden economy.

For the second part of this discussion, we turn to whether there exists evidence of a crowding effect between the intrinsic and extrinsic motivation by estimating equation (3.4). The main results of the level effects and the interaction terms are reported on column (3) and (4) of Table 3.1 for the unweighted and weighted specification respectively. The interaction term of risk and morale is jointly not significant indicating the absence of an interrelationship between the two types of motivation (in both the weighted and unweighted specification). The only term that exhibits significance in this interaction is for those with very high perception of being caught in the weighted specification. However, this result as we will reinforce in our robustness checks of the crowding effects, is driven for this particular group by the low count of individuals reporting for this category which drives noisy estimates when analysing more complex relationships.[38]

As previous research has noted and as we document on the previous set of results when no crowding effects are included, the significance of the risk of detection is more salient than that of sanctions. In the case of analysing the crowding effects of sanctions we also run into the caveat of small cells that can muddle the results making this especially salient in the case that we are considering a continuous covariate as morale is at this stage. Interpreting the result with caution in the case of sanctions, we find a slight crowding-in effect between the two judging from the negative direction of the interaction between the two that reinforces the negative effect of the intrinsic motivation in participation as the level of severance of sanctions increases. We note that the interaction term of sanction and morale is jointly significant at 5%. Within this interaction term no significance is achieved for the categories of 'Other' and 'DK'. These are

---

sanctions which was greater when the level of perceived certainty was higher.

[37] As a general comment, results that refer to the categories 'Prison' in the perceived severity of the sanctions variable and 'Very High' for the perception of being caught should be interpreted with caution as they represent categories with a low occasion of observations. In Section 3.5 we introduce a robustness check where we reduce the number of dimensions within the extrinsic motivation variables to test whether the reduction in the number of interaction terms and the increase in the cell count of each type of motivation had an impact on the conclusions obtained.

[38] We treat with caution any results that relate to those selecting 'Very High' perceived risk of detection and 'Prison' as the perceived sanction if caught as the count of observations relating to both is rather low as can be ascertained in Table B.5 in Appendix B.

however black-box categories and no interpretation can be ascertained from them.

In essence, our results suggest an absence of robust evidence of a crowding effect between the intrinsic and extrinsic motivations which we test for robustness thoroughly in Section 3.5. The absence of an interaction between both types of incentives sheds light over whether the manipulation of the extrinsic incentives, that is the classical deterrence parameters—sanctions and audits—affect the taxpayers' intrinsic motivation to comply. Our results suggest that changes in the enforcement parameters—that is, increases in the perceived severity of the sanctions or risk of detection—do not influence the individual's intrinsic motivation. This means that if we could separate the taxpayers population between those intrinsically and those extrinsically motivated to comply, from a policy perspective, increasing the extrinsic incentives will raise compliance among the extrinsically motivated while not affecting the compliance of the intrinsically motivated. If anything, our results suggest a mild crowd-in of the intrinsic motivation when extrinsic incentives are raised.

Our result lines up well with the evidence from field experiments such as Dwenger *et al.* (2016) and Boyer *et al.* (2016) that find intrinsically motivated individuals' compliance decisions do not respond significantly to changes in the extrinsic incentives under a tax setting.[39] Similarly, Wenzel (2004) finds that personal norms moderates the effects of sanction severity, yielding only a deterrence effect when tax morale is lax and no effect of deterrence on tax evasion when tax morale is strong. Smith (1990) also found that deterrence has a stronger effect on individuals who regarded tax evasion as acceptable than between those who regarded it as unacceptable.

In general, this result has a clear policy implication. A policy of increasing the presence of extrinsic incentives (or the perception of it) would drive those non-intrinsically motivated individuals towards compliance while not affecting the compliance of those honest and intrinsically motivated individuals. In general, the presence of legal sanctions are essential in a society not only because of their deterrent effect, but as a way of condemning socially wrong behaviours and to create a social cost (Williams and

[39]Note that as referred in the Literature Survey, Section 3.2, in the treatment in which the church levy was phrased as a tax rather than a voluntary contributions, no evidence of crowding effects were found.

Hawkins, 1986). The tax administration's objective should be to boost and create a stronger tax culture increasing the intrinsic motivation to comply as it stands a less resource intensive way of promoting compliance, while at the same time maintaining a portfolio of activities that would increase the perception of the classical deterrents to ensure the compliance of the least motivated.

Our findings can also be interpreted within the Kirchler *et al.* (2008) 'slippery slope' framework. We find no evidence of a 'slippery slope' between the intrinsic and extrinsic motivations. However, we coincide in that the extrinsic incentives are specially relevant for those with a low intrinsic motivation, and thus their compliance is *enforced*. However, for those intrinsically motivated, although extrinsic incentives still matter, their compliance could be labelled more *voluntary*. However, our setting does not allow for a distinction between voluntary and enforced as in Muehlbacher *et al.* (2011).

3.5 Robustness

We next discuss the robustness of our empirical estimation. We report two main robustness checks. First, in subsection 3.5.1 we test the robustness of our results to alternative specifications of the intrinsic motivation variable and to a different specification of the extrinsic motivations variables. Second in subsection 3.5.2 we address the possibility of using an alternative dependent variable to proxy own participation in the hidden economy. We test for the robustness of the main predictions obtained in the paper: first, the relevance of the intrinsic and extrinsic incentives in shaping participation in the hidden economy, and second, the absence of a crowding effect between the two.

3.5.1 Alternative Measurement of the Intrinsic Motivation

The intrinsic motivation so far is captured using the perception of the interviewees about a variety of dishonest acts. This is recorded in the survey using a ten-point scale where 1 stands for completely acceptable and 10 for completely unacceptable. In the main estimations, we followed the strategy of constructing a broad morale variable that aggregated the score for all dishonest acts and using a continuous scale in order to allow more heterogeneity of the scores in predicting compliance. We labelled this variable 'Broad Morale' (BM).

We recognise that asking particularly about the acceptability of tax evasion and undeclared work might bias the response for two main reasons exposed by Andreoni *et al.* (1998). On one hand, individuals could use this question as a way to justify their non-compliant behaviour or on the other hand, individuals may over-report their morality for self-presentation concerns. Although this bias is natural in a survey of these characteristics, we address it by constructing a measure of the intrinsic motivation that does not use the individual's perceptions about questions directly related to undeclared work or evasion (MNO-for reference in Appendix B). This new measure of morale will be an aggregate measure that takes into account only two items: the perception about free-riding on public transport and on the benefit system. This variable uses a continuous scale that ranges from 1-"completely acceptable" to 10-"completely unacceptable". Table E.2 column (2) on Appendix E reports the results. We can see that the direction of the relationship between the intrinsic motivation and the probability of participation is maintained. In column (1) we report the results of the impact of the intrinsic motivation using the broad morale (BM) variable which is the baseline specification used throughout the paper for the purpose of comparison.

Other empirical studies such as Alm and Torgler (2006) or Doerrenberg and Peichl (2013) use single-item measures of morale typically associated with the perception of cheating on taxes. The use of multi-scale items versus single-item scales is highly debated in the literature on different fields, see Bergkvist and Rossiter (2009) in marketing or Loo (2002) and Hoepnner *et al.* (2011) for psychology. Single-item scales may be easier and cheaper to operationalize but are more vulnerable to random errors in measurements and biases of interpretation or meaning. These random errors are typically averaged out in multi-item scales. There seems to be agreement that multiple-item scales perform better when dealing with more complex constructs.[40] Multiple-scale items also allow to test for the internal consistency of the scale whereas this is not

[40]More technically, the literature argues that single-item scales perform as well as multiple-scales when measuring 'doubly concrete' constructs, this is constructs for which the attribute of measurement and the object of measurement are undoubtedly clear to those rating it (Rossiter, 2002).

allowed for in single-item scales. We do however construct a single-item scale extracting one of the items from the multiple-item scale which is similar to those used in the previous literature: the perception of cheating on taxes. Column (3) in Table E.2 of Appendix E reports the results to be consistent with those obtained when using multi-scale measures.

Alternatively, the literature has rescaled the variable that relates to the intrinsic motivation to be a dichotomous variable. The ten-point scale is recoded to be 1 if the individual scores the item as completely unacceptable indicating high morale, and 0 otherwise (this is for the rest of the scale). This seems to be restricting the amount of information we receive from the questionnaire as the variability in the variable is completely eliminated. Choosing the cut-off of what is considered high morale is quite arbitrary. It may seem very restrictive to just consider high morale those scoring only 1 on a ten-point scale. As a test of robustness of our results, we recode the different types of morale named above: broad morale (BM), restricted morale (MNO) and evasion (EVA), to a bivariate variable (HBM, HMNO, HEVA respectively) using the criteria outlined in the literature just for comparison purposes. As reported in columns (3), (4) and (6) for each of the variables respectively of Table E.2 on Appendix E, this alternative way of identifying the intrinsic motivation using discrete as opposed to continuous variables captures a more constraining effect of the intrinsic motivation while rendering the direction of the results invariant.

Thus, the use of all the different specifications of the intrinsic motivation, be it continuous or discrete, ratifies our previous findings: the intrinsic motivation is a key factor in explaining the low occasion of participation.

In order to confirm our second finding about the absence of consistent crowding effects between the two types of motivation, we estimate equation (3.4) to analyse the interaction of risk and sanction but this time using the alternative discrete measures of the intrinsic motivation (HBM, HMNO, HEVA) as a opposed to a continuous variable. Table E.3 reports the main level effect of sanctions and risk of detection and their interaction with the different specifications of the variable morale. One of the first things to ascertain is that there is a consistent absence of significance in the crowding effects for all specifications. The regression result in column (1) that uses the variable HBM, reports no significance in the interaction between sanctions and morale expect for those reporting prison as their perceived sentence if caught that exhibit a negative interaction. Although this might point to a crowding-in effect we defer from interpreting it further due to the scarce representativeness of this choice of sanction. We interpret the interaction term of sanction as overall not significant. In terms of the effect of risk and morale, we can ascertain the absence of any significance crowding between the two. The result is consistent for Columns (2) using HMNO and column (3) using HEVA.

This result therefore corroborates the second of our findings: the absence of a robust crowding effect between both types of motivation.

In order to further contrast our findings, we introduce a new specification of the extrinsic incentives variables by reducing the number of categories within each extrinsic motivation. That is, we aggregate the categories of risk and sanctions in order to increase the count per cell to test whether the reduction of the number of interaction terms yields more significance. The caveat of this approach is that we lose dimensionality and that is why it was not pursued as a desired specification. We create only three categories for risk: High risk (that results from aggregating the categories of very high perceived risk and fairly high perceived risk); Low risk (Fairly low and very low perceived risks) and a category for those responding 'Don't Know'. We establish three categories for the perceived sanction: Amount Due; Due plus a fine; Other (that aggregates the categories Prison, Other and Dont Know).

Results using the recoded version of the extrinsic motivation variables are reported in Table E.4. Column (1) reports the results using the continuous broad morale variable used as the preferred measure across the paper in the absence of crowding effects as a control regression. The new specification of risk and sanctions corroborates previous findings. Those with higher intrinsic motivation have lower odds of participating and individuals respond rationally to the increase of extrinsic incentives by reducing their participation. In column (2) we report the results of introducing the interaction between the recoded sanction and risk with the continuous variable of morale and as previously found, there is a mild crowding-in in the case of sanctions judging by the negative effect of the increase severity of sanctions with increases in the intrinsic motivation. In the case of detection, the interaction term is insignificant. We see a small significant positive effect in the interaction term. However, due to its small magnitude it is not economically meaningful and particularly not large enough to offset (even minimally) the effect of an increased perceived level of detection and the constraining effect of the intrinsic motivation in participation. The probability of participation falls faster for those with lower detection perceptions as their intrinsic motivation is increased.

In column (3), (4) and (5) we report the result for the effect of risk of detection and sanctions using the discrete variables HBM, HMNO and HEVA. As expected, there is no evidence of crowding effects as the interaction terms of sanction and risk with each of the morale specifications are jointly not significant expect for the case of sanction in column (3) where a slight crowding-in is found.

In essence, we can document that both our main findings: the significance of the role of intrinsic and extrinsic incentives in constraining participation in the hidden economy and the absence of a crowding effect between the two types of motivations are robust to alternative specifications of both the intrinsic and extrinsic incentives.

3.5.2 Alternative Dependent variable

It is not difficult to recognise that interviewees might be wary of answering truthfully about their own participation due concerns about self-presentation or self-incrimination.[41] However these concerns relax when the questions refer to a third party and this is apparent in the survey as the positive responses for knowing someone who participates are substantially higher than that of own participation. The survey then allows a natural way to proxy own participation. Interviewees are asked:

"Do you personally know any people who work without declaring their income or part of their income to tax or social security institutions?"

Table E.5 in Appendix E, reports the results of estimating equation (3.1) in column (1) and (3.4) to investigate any crowding effects in column (2). Using an alternative

[41] We discussed previously the reliability and validity of self-reports in Section 3.3.2.1.

dependent variable to proxy own participation, the results in column (1) ratify those obtained previously using own-participation. Both intrinsic motivation and extrinsic incentives shape the decision to participate. This alternative dependent variable also ratifies that there does not seem to be robust evidence of a crowding effect between intrinsic and extrinsic incentives as reported in column (2).

Overall, this section has corroborated the robustness of our findings to alternative specification of the intrinsic and extrinsic incentives and to alternative specifications of the dependent variable.

3.6 Conclusion

The present paper contributes firstly to the hidden economy literature in understanding participation in undeclared practices for which evidence is scarce (Schneider and Williams, 2013). Secondly, our paper contributes to the emerging literature on the intrinsic motivation to comply (Luttmer and Singhal, 2014), and on broad terms to the literature on tax compliance.

Using a unique European micro-level survey of engagement in undeclared practices, we investigate the role of intrinsic and extrinsic motivations in shaping the decision to engage in undeclared practices. The survey allows us to link directly attitudes to behaviour and to directly measure participation rather than using macroeconomic approaches to estimate the dependent.

We find that participation in undeclared activities is lower among those intrinsically motivated. We also document that imposing extrinsic incentives—sanctions and audits—have a significant effect in discouraging participation. Following the literature on crowding effects, we investigate whether harder deterrence can in fact crowd-out the intrinsic motivation to comply. We fail to find substantive evidence of a crowding effect, and if anything the results point at a slight crowding-in of the intrinsic motivation. This finding is robust to different specifications of the intrinsic and extrinsic incentives and to an alternative specification of the dependent variable. Our results are consistent with findings from field experiments such as Dwenger *et al.* (2016) and Boyer *et al.* (2016) when applying a tax setting.

This finding suggests that given two types of taxpayers, those intrinsically motivated and those not intrinsically motivated, higher deterrence will pool the non-intrinsically motivated towards compliance while not affecting the compliance level of the intrinsically motivated. Our results highlight the importance of building a tax administration that understands the profile of the taxpayers.

The tax administration should promote the creation of a tax culture and should educate taxpayers in order to increase the number of intrinsically motivated individuals while at the same time maintaining a high profile in terms of deterrence. Higher deterrence does not need to solely translate into resource intensive activities such as audits. Third-party information reporting and tax withholding have proven to be effective in reducing the amount of evasion (Kleven *et al.*, 2011). Although the aim of this paper has not been to analyse the determinants of the intrinsic motivation, research should further steer into this direction in order to shed light over what influences the taxpayers' attitudes and how policy interventions, actions and communications from the tax office can affect the baseline motivation to comply.

We also find a strong link between engagement in the hidden economy and financial strain. We acknowledge this may not be the sole motivation for participation but it is the one that our data clearly outlines. This makes individuals who struggle to face their financial burden and individuals affected by an economic shock, such as losing their job more prone to participation. This brings about the importance of having well-functioning processes that enable easy access for the individuals back to the formal economy and that ensures good quality employment so as to reduce having to recur to underground practices.

The importance of analysing undeclared activities is paramount. They are a source of loss of revenues, causes inequalities being the tax burden shifted from those who are working undercover to those who do not, and affects competition between the informal and formal producers. It also has social consequences, individuals working in the hidden economy do not contribute towards social security reducing their pensions rights in the future; the access to healthcare may also be restricted as contributions are not made; their rights may be violated and the health and safety conditions in their workplace may not be met. Undeclared work has even more negative consequences than tax evasion and therefore further investigation is required.

Although the subject of undeclared work is elusive in nature, the creation of quality data to help monitor undeclared practices is essential. The discontinuity of the surveys is detrimental as the behaviour cannot be tracked along time. Having a continuous survey or what is best a panel of individuals, would help track participation and the size of the hidden economy along time and what is more important to be able to analyse how the motivations and perceptions of the individuals are evolving and are contributing to shaping the hidden economy. Also, field experiments such as Dwenger *et al.* (2016) and Boyer *et al.* (2016) present new avenues to analyse the effect of policy interventions. What is clear is that despite claims that the hidden economy needs to be reduced and despite a great focus on measurement, research on the motivations to participate provides the key to creating effective policies.

3.7 References

Abayomi, K., Gelman, A., Levy, M. (2008). Diagnostics for multivariate imputations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57(3), 273-291.

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

Alm, J., Jackson, B. R., McKee, M. (1992). Estimating the determinants of taxpayer compliance with experimental data. *National Tax Journal*, 45(1), 107114.

Alm, J., Torgler, B. (2006). Culture differences and tax morale in the United States and in Europe. *Journal of economic psychology*, 27(2), 224-246.

Anderson, C. J., Kim, J. S., Keller, B. (2013). Multilevel modeling of categorical response variables. *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*, 481-519.

Andreoni, J., Erard, B., Feinstein, J. (1998). Tax Compliance, *Journal of Economic Literature* 36, 818-860.

Blackwell, C. (2007). *A Meta-Analysis of Tax Compliance Experiments.* International Studies Program Working Paper 07-24 December 2007.

Bergkvist, L., Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research*, 44 (2), 175-184.

Boyer, P.C., Dwenger N., Rincke, J. (2016). Do norms on contribution behavior affect intrinsic motivation? Field-experimental evidence from Germany, *Journal of Public Economics*, Volume 144, Pages 140-153.

Bradburn, N., Sudman, S., Wansink, B. (2004). *Asking Questions The Definitive Guide to Questionnaire DesignFor Market Research, Political Polls, and Social and Health Questionnaires.* Revised Edition. Jossey-Bass. USA.

Braithwaite, V. (2003). Dancing with tax authorities: Motivational postures and non-compliant actions. *Taxing democracy*, 15-39.

Braithwaite, V., Ahmed, E. (2005). A threat to tax morale: The case of Australian higher education policy. *Journal of Economic Psychology*, 26 (4), pp. 523540

Carle, A. (2009). Fitting multilevel models in complex survey data with design weights: Recommendations. *BMC Medical Research Methodology.*

Council Decision No. 2015/1848(EU), 2015. OJ L 268, p.28.

Cowell, F.A. (1990). *Cheating the government: The economics of evasion.* MIT Press Books. 1990;1.

Dabla-Norris, E., Gradstein, M., Inchauste G. (2008). What causes firms to hide output? The determinants of informality. *Journal of development Economics*, 85(1), 1-27.

Daude, C., Gutiérrez, H., Melguizo, Á. (2012). *What drives tax morale?.* OECD Working Paper No. 315.

Del Carpio, Lucia. 2014. *Are the Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru.* Unpublished.

Doerrenberg, P., Peichl, A. (2013). Progressive taxation and tax morale. *Public Choice*, 155(3-4), 293-316.

Dwenger, N., Kleven, H.J., Rasul, I., Rincke, J. (2016). Extrinsic and intrinsic motivations for tax compliance: evidence from a field experiment in Germany. *American Economic Journal: Economic Policy*, 8(3), 203-232.

Eddings, W., Marchenko, Y. (2012). Diagnostics for multiple imputation in Stata. *Stata Journal*, 12(3), 353.

Feld, L.P., Frey, B.S. (2002). Trust breeds trust: how taxpayers are treated. *Economics of Governance*, 3(2), 87-99.

Feld, L. P., Frey, B. S. (2007). Tax compliance as the result of a psychological tax contract: The role of incentives and responsive regulation. *Law & Policy*, 29(1), 102-120.

Feld, L., Larsen, C. (2005). *Black Activities in Germany in 2001 and 2004: A comparison based on survey data.* The Rockwool foundation Research Unit, Study No. 12, Copenhagen.

Feld, L. P., Larsen, C. (2009). *Undeclared Work in Germany 2001–2007 –Impact of Deterrence, Tax Policy, and Social Norms: An Analysis Based on Survey Data*, Berlin: Springer

Feld, L., Schmidt, A. J., Schneider, F. (2011). *Deterrence Policy and the Size of the Shadow Economy in Germany: An Institutional and Empirical Analysis*, ch. 12 in , Handbook on the Shadow Economy, Edward Elgar Publishing. Retrieved from: $http : //EconPapers.repec.org/RePEc : elg : eechap : 13432_12$.

Fischer, J., Schneider, F. (2009). *The puzzle of tax compliance revisited: testing the 'slippery slope' hypothesis for trust and power against field data.* Conference paper.

Frey, B. S., Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *The American Economic Review*, 87(4), 746-755.

Frey, B. S., Jegen, R. (2001). Motivation crowding theory. *Journal of economic surveys*, 15(5), 589-611.

Frey, B. S. (2003). Deterrence and tax morale in the European Union. *European Review*, 11(03), 385-406.

Gemmell, N., Hasseldine, J. (2012). *The Tax Gap: A Methodological Review.* Victoria University of Wellington School of Business Working Paper No. 09/2012. Retrieved from SSRN: `http://ssrn.com/abstract=2199200orhttp://dx.doi.org/10.2139/ssrn.2199200`

Gneezy, U., Rustichini, A. (2000). Fine is a price. *The Journal of Legal Studies*, 29(1), 1-17. 1.

Gneezy, U., Meier, S., Rey-Biel, P. (2011). When and Why Incentives (Dont) Work to Modify Behavior. *Journal of Economic Perspectives*, 25, 191209.

Grasmick, H. G., Bryjak, G. J. (1980). The deterrent effect of perceived severity of punishment. *Social forces*, 471-491.

Haigner, S., S. Jenewein, F. Schneider and F. Wakolbinger (2011), *Dissatisfaction, fear and annoyance: driving forces of informal labor supply and demand*. Discussion paper, Department of Economics, University of Linz, Linz, Austria. Paper presented at the European Public Choice Meeting, Rennes, 28 April 1 May.

Hessing, D. J., Elffers, H., Robben, H. S. J., Webley, P. (1992). Does deterrence deter? Measuring the effect of deterrence on tax compliance in field studies and experimental studies. In J. Slemrod (Ed.), *Why people pay taxes: Tax compliance and enforcement* (pp. 291–305). Ann Arbor: University of Michigan Press.

Hoeppner B.B., Kelly J.F., Urbanoski K.A., Slaymaker V. (2011). Comparative Utility of a Single-Item vs. Multiple-Item Measure of Self-Efficacy in Predicting Relapse among Young Adults. *Journal of Substance Abuse Treatment*, 41(3), 305-312.

Isachsen, A. J., Strøm, S. (1985). The size and growth of the hidden economy in Norway. *Review of Income and Wealth*, 31(1), 21–38.

Kenward, M. G., Carpenter, J.R. (2007). Multiple imputation: Current perspectives. *Statistical Methods in Medical Research*, 16, 199218.

Kirchler, E., Hoelzl, E., Wahl, I. (2008). Enforced versus voluntary tax compliance: The slippery slope framework. *Journal of Economic Psychology*, 29, 210-225.

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

Kogler, C., Batrancea, L., Nichita, A., Pntya, J., Belianin, A., Kirchler, E. (2012). Trust and Power as Determinants of Tax Compliance: Testing the Assumptions of the Slippery Slope Framework in Austria, Hungary, Romania and Russia. *Journal of Economic Psychology*, 34, 169-180.

Lisi, G. (2012). Testing the slippery slope framework. *Economics Bulletin*, 32 (2) , pp. 13691377.

Lobo, F.M. (1990). *Irregular Work in Spain, in Underground Economy and Irregular Forms of Employment*, Final Synthesis Report, Brussels: Office for Official Publications of the European Communities.

Loo, R. (2002). A caveat on using singleitem versus multipleitem scales. *Journal of Managerial Psychology*, Vol. 17 Iss: 1, pp.68 - 75.

Luttmer, E. F., Singhal, M. (2014). Tax morale. T*he Journal of Economic Perspectives*, 28(4), 149-168.

Mellström, C., Johannesson, M. (2008). Crowding out in blood donation: was Titmuss right?. *Journal of the European Economic Association*, 6(4), 845-863.

Merz, J., Wolff, K. G. (1993). The shadow economy: illicit work and household production – a microanalysis of West Germany. *Review of Income and Wealth*, 39(2), 177-94.

Muehlbacher, S., Kirchler, E., Schwarzenberger, H. (2011). Voluntary versus enforced tax compliance: Empirical evidence for the "slippery slope" framework. *European Journal of Law and Economics*, 32(1), 89-97.

Pahl, R.E. (1988). *The Black Economy in the UK, in Research on the Black Economy in Europe*. Final Report. Programme of Research on the Black Economy in Europe. Retrieved from: `$http://aei.pitt.edu/75651/1/28.pdf$`

Pedersen, S. (2003). *The Shadow Economy in Germany, Great Britain and Scandinavia: A Measurement Based on Questionnaire Service*. Study no. 10, Copenhagen: Rockwool Foundation Research Unit.

Pigott, T. D. (2001). A review of methods for missing data. *Educational research and evaluation* .7.4, 353-383.

RabeHesketh, S., Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(4), 805-827.

Rossiter, J.R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19(4), 305-335.

Royston, P. (2004). Multiple imputation of missing values. *Stata journal*, 4(3), 227-41.

Royston, P. (2005). Multiple imputation of missing values: update. *Stata Journal*, 5(2), 188. Chicago.

Royston, P., White, I. R. (2011). Multiple imputation by chained equations (MICE): implementation in Stata. *Journal of Statistical Software*, 45(4), 1-20.

Rubin, D.B. (1978) Multiple imputations in sample surveys a phenomenological Bayesian approach to nonresponse. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 2034.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Schneider, F., Buehn, A., Montenegro, C. (2010). *Shadow Economies All over the World: New Estimates for 162 Countries from 1999 to 2007.* Policy Research Working Paper Series, 5356.

Schneider, F., Williams, C. (2013). *The Shadow Economy.* Institute of Economic Affairs.

Scholz, J. T., Pinney, N. (1995). Duty, Fear, and Tax Compliance: The Heuristic Basis of Citizenship Behavior. *American Journal of Political Science*, 39(2): 490512.

Sheffrin, S., Triest, R.K. (1992). Can Brute Deterrence Backfire? Perceptions and Attitudes in Taxpayer Compliance, in Joel Slemrod, ed. 1992. *Why People Pay*

Taxes: Tax Compliance and Enforcement. Ann Arbor: U. of Michigan Press, 1992.

Slemrod, J., 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(1), 25-53.

Smith, K. W. (1990). Integrating three perspectives on noncompliance: A sequential decision model. *Criminal Justice and Behavior*, 17, 350-369.

Torgler, B., Schaffner, M., Macintyre, A. (2007). *Tax compliance, tax morale, and governance quality*. International Studies Program Working Paper, Andrew Young School of Policy Studies, Georgia State University.

Torgler, B., Schneider, F. (2007). *Shadow economy, tax morale, governance and institutional quality: a panel analysis. Governance and Institutional Quality: A Panel Analysis*. School of Economics and Finance Discussion Papers and Working Papers Series 210, School of Economics and Finance, Queensland Uni- versity of Technology.

Torgler, B., Schneider, F. (2009). The impact of tax morale and institutional quality on the shadow economy. *Journal of Economic Psychology*, 30(2), 228-245.

TNS Infratest, Rockwool Foundation and Regioplan (2006). *Feasibility Study on a Direct Survey about Undeclared Work*. VC/2005/0276, Directorate General of Employment and Social Affairs, European Commission, Brussels.

Van Eck, R., Kazemier, B. (1988). Features of the hidden economy in the Netherlands'. *Review of Income and Wealth*, 34(3), 251-273.

Wärneryd, K.-E., Walerud, B. (1982). Taxes and economic behavior: Some interview data on tax cheating in Sweden. *Journal of Economic Psychology*, 2, 187-211.

Webley, P., Robben, H. S. J., Elffers, E., Hessing, D. J. (1991). Tax evasion: An experimental approach. Cambridge, England: Cambridge University Press.

Wenzel, M. (2004). The social side of sanctions: personal and social norms as moderators of deterrence. *Law and Human Behaviour*, 28(5), 547.

White, I. R., Royston, P., Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30, 377399.

Williams, K. R., Hawkins, R. (1986). Perceptual Research on General Deterrence: A Critical Review. *Law and Society Review*, 545-572.

Williams, C., Horodnic, I.A. (2016). Tackling the undeclared economy in the European Union: an evaluation of the tax morale approach. *Industrial Relations Journal*, 47(4) 322-340.

Yinger, J. (1963). Research Implications of a Field View of Personality. *American Journal of Sociology*, 68(5), 580-592. Retrieved from `http://www.jstor.org/stable/2774610`

Appendix A: The Survey

The Eurobarometer Survey is used by the European Commission or other EU Institutions to investigate the public opinion in the different member estates on certain topics of interest. It was carried out by the TNS Opinion & Social network on request of the European Comission in 27 Member States of the European Union and Croatia between April and March 2013. This special Eurobarometer 402 survey corresponds to the 79.2 wave. Face-to-face interviews were performed on a sample of 27,563 interviewees which are resident in each of the aforementioned states and are 15 years or over of age. This means the sample size is an average of 1,000 individuals per country.[42,43]

A multi-stage random probability method was used to obtain the sample of the study. In each country a number of sampling points with a probability proportional to the population size and density was drawn from each of the administrative regional units after stratifying by individual unit and type of area. The sample is therefore representative of the territory of the countries surveyed and of the distribution of the population. In each of the sampling points, the first address was drawn at random and then every Nth address was selected. In each household the respondent was chosen randomly according to the "closest birthday rule".

The Eurobarometer Survey on undeclared work provides a common framework to analyse undeclared work in a comparable way for the different countries in the EU. This survey is a follow-up survey for the one in 2007 although comparison is not possible as the questionnaire underwent modifications. The questionnaire was originally created following the proposal of the Danish Rockwool Foundation Research Unit which had previously applied it in a number of studies in Denmark and later a modified version was used in studies in Sweden, Germany, the United Kingdom and Norway. The EBS

[42] This sample size is quite small per country to provide an independent analysis on a country basis. As it was advised in the Feasibility study performed by TNS Infratest, Rockwool Foundation and Regioplan (2006; p.90) estimations based on a sample size of less than 2,000 individuals may not provide a reliable analysis of the structure of undeclared work.

[43] Some may argue that a face-to-face interview may not be the best method to use in order to address such sensitive topics. However, face-to-face interviews eliminate the problem of self-selection which plague online and mail surveys. Face-to-face interviews have even proved superior to telephone interviews in cases such as Germany (TNS Infratest *et al.*, 2006).

is a very complete survey that records information on individuals involvement in undeclared work practices in different roles, as users and as suppliers, and it surveys their perceptions and reasons for doing so.

At the start of the survey, the individuals are all faced with a common definition of undeclared work that reads:

"It is widely known that part of the population is engaged in undeclared work, in the sense of activities which avoid partly or entirely declaration to tax authorities or social security institutions, but which are otherwise legal. This could be people working in certain sectors of activity like construction, transport or agriculture for example but also in hotels, restaurants and cafs. Often it concerns only part of their income from work like remuneration of overtime or other extras. Undeclared work is also common in a whole range of household services - such as gardening, babysitting and elderly care -, personal services - like hairdressing,cosmetic or medical treatment - and repair services for cars, clothes, or computers."

This approach to undeclared work does no include any criminal activities such as drug dealing or smuggling. It only considers activities which being legal, they are just not reported to the tax office. This will include individuals such as ghosts, this is, individuals who perform an activity and are not in the tax administration records at all, but also moonlighters, individuals who are registered for their first employment but fail to declare their income from their secondary employment. However, we may also consider that some illegal production might be picked up in the sense that it could be a perfectly legal activity but that turns unlawful not because of the activity itself but because of the individual performing the activity not being authorised to do so.

Appendix B: The variables

Table B.1: Definition of the variables used in the regression.

Dependent variables	
Own Participation in the Hidden Economy	Dummy variable equal 1 if individual participates, 0 otherwise.
Knowing someone who participates in the hidden economy	Used as a proxy of own participation. Provides a robustness check for the results presented on this paper. It takes the value 1 if individuals know someone who participates, 0 otherwise.
Sociodemographics	
Male	Dummy variable equal 1 if individual is a male, 0 if a female.
Age	Continuous variable
Age Squared	Continuous variable
Education	Categories: Less than 15 years of full-time education (o); 16-19 years; More than 20; Still studying; No full-time education
Occupation	Categories: Self-Employed (o); Employed; Retired; Unemployed; Other Non-Active.
Marital Status	Categories: Single (o); Single with partner; Married; Other civil status.
Region	Categories: Rural (o); Small-Medium size town; Large town.
Area	Uses geographical classification from the UN. Northern Europe: Dummy variable equal 1 if the individual belongs to: Denmark, Ireland, Great Britain, Northern Ireland, Finland, Sweden, Estonia, Latvia and Lithuania.

Continued on next page

Table B.1 – continued from previous page

	Western Europe: Dummy variable equal 1 if belongs to: France, Belgium, The Netherlands, Germany, Luxembourg and Austria. Southern Europe: Italy, Greece, Spain, Portugal, Cyprus, Malta, Slovenia and Croatia Eastern Europe: Czech Republic, Hungary, Poland, Slovakia, Bulgaria and Romania.
Size of the Household	Continuous variable
Number of Children	Continuous variable
Financial Constraints	
Struggle with Bills	Categories: Never struggle (o); Occasionally struggle; Almost never struggle.
Social Class*	Categories: Working class (o); Middle class; High Class; Other; None; Don't know.
Level in Society*	On a 11-point scale where 1 corresponds to the 'lowest level in society' and 10 to the 'highest level in society'.
Wealth Variables	Dummy variables for: Ownership of a car, TV, DVD, CD, PC, Internet Connection. House: Paid is a dummy variable equal to 1 if the house is already paid and 0 otherwise. House: Paying is a dummy variable equal to 1 if the house is still being paid and 0 otherwise.
Extrinsic Motivation	
Sanctions (Perceptions of severity)	Categories: Amount due (o); Amount due plus fine; Prison; Other; Do not Know.
Risk of Detection (Perceptions of)	Categories: Very Small (o); Fairly Small; Fairly High; Very High; Do not Know

Continued on next page

	Intrinsic Motivation
	Broad Measure (BM): Multi-item composite scale resulting from the average score of the interviewee's perception of various dishonest acts. Ten-point scale where 1=Low morale(Completely Acceptable) and 10=High morale (Completely Unacceptable). The items considered respond to the question: 'How acceptable is it to...?': • Receive welfare payments without entitlement • Use public transport with no ticket • An individual not declaring income despite being hired • A firm not declaring an income despite being hired by a household • A firm is hired by another firm and it does not declare its activities • A firm hires an individual and all or a part of their wage is not declared • Evade taxes

Continued on next page

Table B.1 – continued from previous page

	Morale not related to undeclared work (MNO): Multi-item composite scale resulting from the average score of the interviewee's perception of dishonest acts not related to taxes and undeclared work. Ten-point scale where 1=Low morale and 10=High morale. Refers to the items: • Receive welfare payments without entitlement • Use public transport with no ticket
	Evasion (EVA): Single-item measure resulting from each interviewee's score of the perception of evading taxes. Ten-point scale where 1=Low morale and 10=High morale. Refers to the item: • Evade taxes
	High Morale - Broad Morale (HBM): Discrete measure of morale resulting from recoding BM. HBM equals 1 if BM=10, this is if morale is high and 0 otherwise.
	High Morale- Restricted Morale (MNO): Discrete measure of morale resulting from recoding MNO. HMNO equals 1 if MNO=10, this is if morale is high and 0 otherwise.
	High Morale- Evasion (HEVA): Discrete measure of morale resulting from recoding EVA. HEVA equals 1 if EVA=10, this is if morale is high and 0 otherwise.

Notes:

1. * Individuals are asked to classify themselves.

2. (o) Indicates that this category is omitted in the regression.

Summary Statistics

Table B.2: Summary Statistics of the dependent variables.

Own-Participation	Frequency	Percent
No Participation	23,493	95.41
Participation	1,130	4.59
Total	24,623	100
Participation of a third-party	Frequency	Percent
No Participation	15,370	63.58
Participation	8,805	36.42
Total	24,175	100

Table B.3: Summary Statistics of continuous and dichotomous variables.

	N	Mean	S.D.	Min	Max
Broad Morale (BM)	24623	8.599244	1.539392	1	10
Age	24623	48.99773	18.03676	15	98
Size of the Household	24623	2.482151	1.094991	1	4
Number of Children	24623	0.430248	0.844885	0	12
Level in society	24623	5.296105	1.645029	1	10
Car	24623	0.717987	0.449989	0	1
TV	24623	0.980587	0.137973	0	1
DVD	24623	0.668237	0.470856	0	1
CD	24623	0.589814	0.491877	0	1
PC	24623	0.729562	0.444195	0	1
Internet	24623	0.70223	0.457287	0	1
Car	24623	0.717987	0.449989	0	1
House: Paid	24623	0.50262	0.500003	0	1
House: Still paying	24623	0.255696	0.436261	0	1

Table B.4: Summary statistics of categorical included variables.

Employment status	Frequency	Percentage
Self-Employed	1,665	6.76
Employed	10,007	40.64
Unemployed	2,116	8.59
Retired	7,353	29.86
House person and students	3,482	14.14
Total	24,623	100
Education	**Frequency**	**Percentage**
Up to 15	4,167	16.92
16-19	10,695	43.43
20+	7,759	31.51
Still Studying	1,812	7.36
full-time education	190	0.77
Total	24,623	100
Marital Status	**Frequency**	**Percentage**
Single	4,455	18.09
Single with Partner	2,557	10.38
Married	12,975	52.69
Other civil Class	4,636	18.83
Total	24,623	100
Bills	**Frequency**	**Percentage**
Bills: Struggle Most of the Time	3,529	14.33
Bills: Struggle occasionally	7,169	29.12

Continued on next page

Table B.4 – continued from previous page

Bills: Struggle almost never	13,925	56.55
Total	24,623	100
Social Class	Frequency	Percentage
Social Class: Working Class	10,931	44.39
Social Class: Middle Class	12,408	50.39
Social Class: High Class	595	2.42
Social Class: Other	161	0.65
Social Class: None	265	1.08
Social Class: DK	263	1.07
Total	24,623	100
Type of Community	Frequency	Percentage
Rural area or village	8,644	35.11
Small/middle town	9,227	37.47
Large town	6,752	27.42
Total	24,623	100
Region	Frequency	Percentage
Northern Europe	7,539	30.62
Southern Europe	6,300	25.59
Western Europe	5,530	22.46
Eastern Europe	5,254	21.34
Total	24,623	100

Table B.5: Perception of the Extrinsic Motivation by Participation Status

Severity of the Sanction by Participation in the HE			
	Do not participate	Participate	Total
Amount Due	5,759	335	6,094
Fine	13,280	557	13,837
Prison	1,193	46	1,239
Other	944	96	1,040
DK	2,317	96	2,413
Total	23,493	1,130	24,623

Risk of Detection by Participation in the HE			
	Do not participate	Participate	Total
Very High	3,759	292	4,051
Fairly High	9,419	512	9,931
Fairly Low	6,787	224	7,011
Very Low	1,610	50	1,660
DK	1,918	52	1,970
Total	23,493	1,130	24,623

Table B.6: Summary statistics of variables indicating Intrinsic Motivation.

	N	Mean	S.D.	Min	Max
Broad Morale (BM)	24623	8.599	1.539	1	10
Restricted Morale (MNO)	24572	8.814	1.624	1	10
Evasion	24266	8.631	1.999	1	10
High Morale: BM	24623	0.366	0.482	0	1
High Morale: MNO	24572	0.540	0.498	0	1
High Morale: Evasion	24266	0.546	0.498	0	1

Appendix C: Exploratory Factor Analysis

The intrinsic motivation to comply is unmeasurable and therefore we need to rely on indicators (measured and observed) to account for it. The EBS questions the interviewees about their perceptions of a variety of dishonest acts ranging from free-riding on public transport or the benefits system to the performance of undeclared work by firms or individuals.[44]

Before we can construct a measure of the intrinsic motivation using these items, we need to know whether all the indicators (variables in the survey) that we have from the survey regarding their perceptions about dishonest acts measure only one factor, in our case the intrinsic motivation.[45]

The unidimensionality of the construct is tested using Exploratory Factor Analysis (EFA). Factor analysis is a technique that enables us to construct an index for an unknown/unmeasurable factor based on the pattern of correlation of different items that we have measured on the survey. Factor analysis considers that each of our observed items can be represented as a linear combination of factors which are underlying constructs not directly measured. Each observed variable is the result of the sum of each factor (underlying construct) multiplied by a coefficient (loadings) and some unique factor (measurement error—the variability that is unique to the variable and cannot be explained by factors). We run a factor analysis on the items relating to the perceptions of the different dishonest acts.[46] After running factor analysis, the number of relevant factors that are ascertained can be determined using the Guttman-Kaiser

[44] See the definition of the variable Broad Morale in Table B.4 in Appendix B for a complete list of the dishonest acts.

[45] Note that if different factors are elicited then aggregating all these variables into a measure of morale would be wrong, as we would be mismeasuring the construct.

[46] As noted on Appendix B, the interviewee is asked to rank on a ten-point scale from acceptable ("1") to unacceptable ("10") its perception about how acceptable the following dishonest acts are: 1) Receive welfare payments without entitlement; 2) Use public transport with no ticket; 3) An individual not declaring income despite being hired; 4) A firm not declaring an income despite being hired by a household; 5) A firm is hired by another firm and it does not declare its activities; 6) A firm hires an individual and all or a part of their wage is not declared ; 7) Evade taxes. These are the inputs of our factor analysis. We have used the Kaiser-Meyer-Olkin test of sample adequacy. An overall value of 0.8767 is obtained which supports the appropriateness of factor analysis. The measure of sample adequacy is greater than 0.83 for all variables indicating that the correlation of the variables is enough to warrant factor analysis.

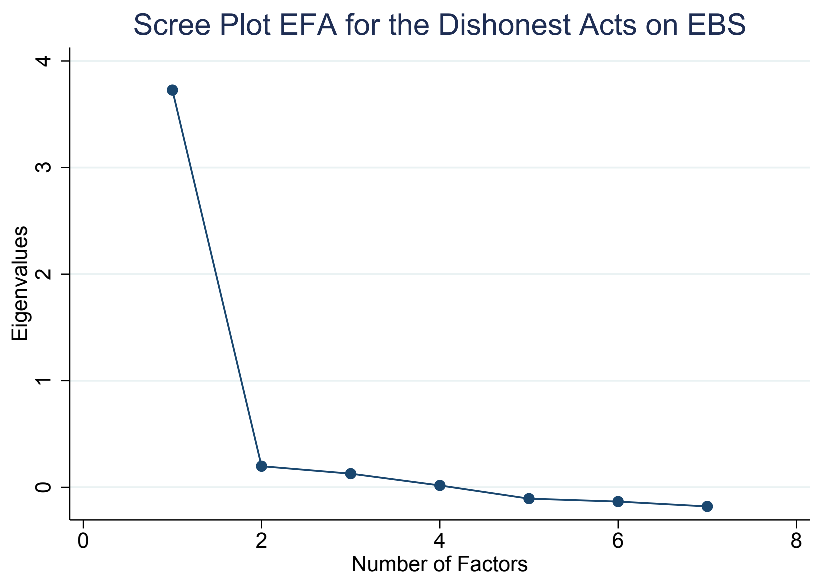


Figure C.1: Contrasts of margins of the Perceived Risk of Detection according to the Intrinsic Motivation.

rule. Relevant factors are those that have eigenvalues greater than one. In our case, we can see clearly that only one factor has eigenvalues greater than one and the rest are significantly lower. This is confirmed by using a scree plot that involves a graphic representation of the relationship between factors and eigenvalues. As shown in Figure C.1, it is clear from our case that there is only one factor that is significant with eigenvalues higher than 1. A sharp drop is observed after it, indicating that a one-factor solution is appropriate.[47]

Apart from the finding of using the Guttman-Kaiser rule that points out that only one factor is significant, we can confirm this by observing that factor loadings (correlations between the variables and the factor) are larger than 0.57 for all variables and they load mainly on the first factor. This one factor solution is going to represent our construct of tax morale and our variable 'Broad Morale' is going to be composed of the average score from the 7 different items scored.

In order to retain certainty that the measure we construct of tax morale relates to the unobserved intrinsic motivation, we can test the internal reliability of the scaled using

[47]Alternatively, we have used Principal Component Analysis (PCA) which is another data reduction method and the results are identical. One factor is to be retained only with an eigenvalue larger than 1 indicating the unidimensionality of the construct.

Cronbach's alpha. The scale internal reliability has been tested using Cronbach's alpha. The alpha coefficient is 0.8728, suggesting a relatively high internal consistency of the items.[48]

Table C.1: Exploratory Factor Analysis (EFA) Loadings (Rotated solution) for the 7 items dishonest acts.

Variables	Rotated EFA Eigenvalues >= 1
Q1: Receiving welfare without entitlement	0.5776
Q2: Use of public transport without ticket	0.5765
Q3: Individual to household UW	0.5997
Q4: Firm to household UW	0.8379
Q5: Firm to firm UW	0.8325
Q6: Individual to firm UW	0.8177
Q7: Tax evasion	0.7986

[48] Conventionally, an alpha higher than 0.7 is considered as acceptable.

Appendix D: Missing data

The literature identifies three types of mechanisms underlying missing data: Missing At Random (MAR), Missing Completely At Random (MCAR) and Missing Not At Random (MNAR). Identifying the type of missingness is essential since the appropriate technique to handle missing data needs to be employed. In our case, the data is believed to suffer from two types of missingness. On one hand, missing observations on some of the covariates such as marital status, education or social class are believed to be Missing At Random (MAR) and therefore Multiple Imputation stands as the appropriate way of imputing these missing observations.[49]

On the other hand, the dependent variables, own-participation and participation of a third-party, contain missing values due to individuals either refusing to respond or responding 'do not know'.[50] The mechanism of missingness of these variables, we consider it to be missing not at random (MNAR). The reason being that the missing value depends on the value of the missing observation itself. This will happen if individuals who refuse to respond are more likely to participate than individuals who do respond.

Multiple imputation has stood as the most preferred choice among researchers when facing MAR data.[51] When the structure is MAR, complete case analysis or listwise deletion leads to a bias in the estimation. Using listwise deletion in general may lead to a large number of observations being deleted which can lead to inefficient results or can even leave the researcher with a sample no longer representative of the population

[49] To distinguish whether the missing data mechanism is Missing at Random (MAR) or Missing Completely At Random (MCAR) we have performed a logistic regression of the missing-data indicator for each imputed variable on other explanatory variables to test for associations. We find strong associations between the two which should not be the case if data are MCAR. The missing mechanism is not ignorable and therefore listwise deletion is in principle not an appropriate method.

[50] The reason for treating 'Don't Know' as a missing answer here is due mainly to the low occasion with which this option is chosen (0.8% of respondents) which could not justify the additional computational power required to run a multinominal logit. This estimation was performed as a check and the results from estimating a multilevel multinomial logit are comparable to those obtained under the current multilevel logit model as expected. Thus, we continued with the multilevel logit specification and treated those observations as missing.

[51] Since Rubin (1978, 1987) first introduced the method of multiple imputation as a way to go beyond complete case analysis, a large strand of literature specially stemming from medical research has been developed supporting multiple imputation to other methods. See Kenward and Carpenter (2007), Royston (2004, 2005).

leading to a bias in the estimation.[52]

In the case of MNAR data, the definition of a model for the missingness mechanism is hardly feasible. Hence, MAR techniques together with sensitivity analysis to test for the robustness of our conclusions are the advised methods. Once some information about the missigness mechanism based on the observed data is introduced, little room is left for information that can only be contained in the unobserved data.

Therefore we proceed to use multiple imputation to address the missingness in the dataset together with some sensitivity analysis for the suspected MNAR data. In short the idea of multiple imputation is based on obtaining a distribution of possible values of the missing observation and create a number M of different complete datasets in which to run the estimation.[53]

The next step after the imputation and the estimation is the pooling of the estimates of the M datasets using Rubin's rules to provide a single estimate and standard errors that incorporate missing data uncertainty. The choice on the number of imputations necessary is still under debate. Rubin (1987) asserts that to obtain a relative efficiency of 90% with 50% of the data missing only two imputations are enough but that five imputations will provide a relative efficiency of 95%. Others suggest that a minimum of twenty imputations. However, it seems to be dependent on the number of missing observations in the dataset. White et al. (2011) suggest that the rule of thumb is that the number of imputations should be equal to the proportion of missing cases in the dataset. This is, if 90% of the cases are complete, then the number of imputations M should be equal to 10. In our case 9% of the cases have missing observations. According to the rule of thumb this should be our number of imputations. We are going to choose 10 imputations as it is computationally possible to do so.[54]

[52]See Kenward and Carpenter (2007) and Piggot (2001) for a detailed analysis of the missingness mechanisms and a review of the methods to use.

[53]The imputation methods typically available in Stata obtain the imputations by simulating from a Bayesian posterior distribution of missing data.

[54]Different number of imputations have been added in order to test for the robustness of the estimates and they are robust across estimations when the number of imputations differ. Multiple imputation builds on the assumption of MAR which cannot be tested. However following Abayomi, Gelman, and Levy (2008) and Eddings and Marchenko(2012) we can check that the imputation model fits the data well which we have done graphically and also that the summary statistics of the imputation are reasonable. All these tests are available from the author.

We have used all predictors for the imputation that will appear as explanatory and outcome variables in the regression and the data has been appropriately weighted for the sample to be representative of the population. We have used Multiple Imputation by Chained Equations (MICE) as it allows different variable types to be imputed in the same command. See Royston and White (2011) for an explanation of how MICE operates.

The importance of assuming the different missingness mechanism of the data relies on how sensitive our conclusions are to the assumption behind the models we are using. In this case, and due to the small number of observations missing, the estimations and conclusions are robust. These practices are widely spread in the medical research literature (specially in Randomised Controlled Trials literature), however the questionability of the missingness mechanism needs to be addressed also in the empirical economics literature where we do find a tendency for assuming data is missing completely at random and therefore listwise deleting the data.

In cases such as ours, the gain from multiply imputing the data due to the low count of missing observations did not outweigh how computationally costly it is to run. The results as we have mentioned using multiple imputation are very comparable to those using listwise deletion but again this is due to the low number of missing observations as reported in Table D.1.

Table D.1: Summary statistics of missing observations.

Variable	% of missing observations on the total.
Education	1.38%
Marital Status	0.16%
Bills	1.70%
Social Class	0.90%
Level in Society	2.11%
Sanction	3.08%
Risk of Detection	1.70%
Broad Morale: Items	
Item 1: Welfare Payments	2.06%
Item 2: Public Transport	1.78%
Item 3: Individual Undeclared Work (UW)	3.51%
Item 4: Firm-to-Household UW	3.09%
Item 5: Firm-to-firm UW	3.11%
Item 6: Individual-to-firm UW	2.84%
Item 7: Evasion	2.80%
Third-party participation	5.18%
Own Participation	3.08%

Appendix E: Regression tables

Table E.1: Estimation results of the three-level logistic regression of Own Participation.

	(1)	(2)
	Unweighted	Weighted
Panel A: Sociodemographics		
Male	0.671***	0.393**
	(0.0705)	(0.194)
Age	0.00413	-0.00287
	(0.0147)	(0.0174)
Age squared	-0.000363**	-0.000404**
	(0.000167)	(0.000174)
Education		
Education: 16-19	-0.0796	-0.166
	(0.122)	(0.119)
Education: More 20	-0.101	-0.311*
	(0.137)	(0.171)
Education: Still studying	-0.244	-0.336
	(0.223)	(0.253)
Education: No Education	-0.484	-0.850
	(0.618)	(0.708)
Occupation		
Employed	-0.584***	-0.711***
	(0.117)	(0.208)
Unemployed	-0.0640	-0.251
	(0.141)	(0.312)
Retired	-0.922***	-0.880***
	(0.173)	(0.190)
Other non-active	-0.282	-0.363*
	(0.183)	(0.197)
Marital status		
Single with partner	0.208*	0.120
	(0.110)	(0.110)

continued on next page

continued from previous page

	(1)	(2)
	Unweighted	Weighted
Married	0.113	-0.121
	(0.109)	(0.148)
Other Civil Status	0.285**	0.266
	(0.116)	(0.226)
Household		
Size	-0.124***	-0.145**
	(0.0440)	(0.0678)
Number of children	-0.00781	-0.00216
	(0.0473)	(0.0765)
Region		
Small-Medium size town	-0.107	0.0518
	(0.0804)	(0.161)
Large town	-0.265***	-0.0124
	(0.0994)	(0.202)
Geographical Area		
Northern Europe	0.496	0.707***
	(0.310)	(0.189)
Southern Europe	-0.208	-0.231*
	(0.320)	(0.133)
Western Europe	0.479	0.290
	(0.331)	(0.214)
Panel B: Financial Constraints		
Bills		
Struggle occasionally	-0.429***	-0.501***
	(0.0963)	(0.164)
Never struggle	-0.836***	-0.909***
	(0.105)	(0.170)
Social Class		
Middle Class	-0.173**	-0.143
	(0.0813)	(0.125)
High Class	-0.300	-0.300
	(0.247)	(0.460)

continued on next page

continued from previous page

	(1)	(2)
	Unweighted	Weighted
Other	0.504	0.873
	(0.319)	(0.670)
None	-0.790**	0.353
	(0.367)	(0.910)
DK	-0.319	-0.0461
	(0.335)	(0.571)
Level in society	0.00366	-0.0354
	(0.0245)	(0.0467)
Wealth Variables		
Still paying house	-0.114	-0.0347
	(0.0947)	(0.140)
House already paid	-0.100	0.0213
	(0.0909)	(0.167)
Internet	0.00711	-0.250
	(0.150)	(0.202)
Car	-0.0860	0.108
	(0.0889)	(0.132)
TV	-0.394**	0.0472
	(0.187)	(0.410)
DVD	-0.0911	-0.0563
	(0.0919)	(0.236)
CD	0.193**	0.313***
	(0.0865)	(0.102)
PC	0.0658	0.137
	(0.159)	(0.147)
Panel C: Extrinsic Motivation		
Sanction		
Due plus fine	-0.174**	-0.141*
	(0.0795)	(0.0757)
Prison	-0.198	0.0346
	(0.174)	(0.316)
Other	0.390***	0.346***
	(0.138)	(0.106)

continued on next page

continued from previous page

	(1)	(2)
	Unweighted	Weighted
DK	-0.220	-0.0491
	(0.137)	(0.166)
Risk of Detection		
Fairly Small	-0.381***	-0.426***
	(0.0848)	(0.0945)
Fairly High	-0.885***	-0.903***
	(0.101)	(0.152)
Very High	-0.777***	-0.736***
	(0.168)	(0.257)
DK	-0.879***	-1.171***
	(0.170)	(0.349)
Panel D: Intrinsic Motivation		
Broad Morale	-0.367***	-0.413***
	(0.0185)	(0.0417)
Constant	2.227***	2.760***
	(0.511)	(0.693)
Random effects	0.543	0.418
Region	(0.062)	(0.097)
	0.497	0.259
Country	(0.093)	(0.075)
Number of observations	24,623	24,623

Notes:

This is the result of estimation equation (3.1). The dependent variable is own-participation. The weighted estimation includes design and post-stratification weights. The stars indicate significance at the following levels: * p<0.10 ** p<0.05 *** p<0.01

Robustness checks

Table E.2: Regression results for the decision of own participation on the Hidden Economy at different specifications of the intrinsic motivation.

	(1)	(2)	(3)	(4)	(5)	(6)
	Continuous measures			Discrete measures		
Intrinsic Motivation	Broad Morale (BM)	Restricted Morale (MNO)	Evasion (EVA)	High: Broad (HBM)	High: Restricted (HMNO)	High: Evasion (HEVA)
	-0.413***	-0.198***	-0.275***	-1.483***	-0.597***	-1.187***
	(0.042)	(0.0294)	(0.0196)	(0.218)	(0.125)	(0.112)
Sociodemographics	Yes	Yes	Yes	Yes	Yes	Yes
Financial Constraints	Yes	Yes	Yes	Yes	Yes	Yes
Tax Administration	Yes	Yes	Yes	Yes	Yes	Yes
N	24623	24572	24266	24,623	24,572	24,266

Notes:
The regression results shown here are the result of estimating (3.1) using alternative measures of the intrinsic motivation in each column. The dependent variable for all regressions is own participation. Column (1), (2) and (3) uses alternative continuous specifications of the morale variables. The regression results of column (1) estimates the intrinsic motivation using the specification Broad Morale (BM) which is the baseline specification throughout the paper. This column is included for comparison purpose. The second column uses a measure of the intrinsic motivation that leaves out the questions that relate to undeclared work and evasion (MNO- for reference in Appendix B). This is, only the perception about free riding on public transport and the benefit system is taken into account. Column (3) includes uses as a proxy for the intrinsic motivation the acceptability of the evasion behaviour (EVA- for reference in Appendix B). Columns (4), (5) and (6) use discrete specifications of the intrinsic motivation. The variables in columns (1), (2) and (3) are recoded for these regressions to discrete variables equal to 1 if morale is high and 0 otherwise. Morale is defined as high when the act is considered completely unacceptable which means 1 on a ten-point scale. The regression results of column (4) refers to recoding the broad variable of morale (HBM in Appendix B), column (5) refers to recoding morale not related to undeclared work or evasion (HMNO) and column (6) refers to recoding the perception of evasion. A full description of the variables is available in Appendix B. Results are weighted using design and post-stratification weights. The unweighted results are comparable but omitted for brevity. They are available upon request. The stars indicate significance at the following levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table E.3: Estimation results of the crowding effects using the discrete variables of morale.

	(1)	(2)	(3)
Extrinsic Motivation			
Sanction			
Due plus fine	-0.205***	-0.0882	-0.102
	(0.0754)	(0.110)	(0.108)
Prison	0.272	0.299	0.749*
	(0.351)	(0.292)	(0.384)
Other	0.479***	0.612***	0.538**
	(0.139)	(0.166)	(0.215)
DK	-0.143	-0.318	-0.218
	(0.146)	(0.206)	(0.208)
Risk of Detection			
Fairly Small	-0.410***	-0.427**	-0.640***
	(0.117)	(0.174)	(0.143)
Fairly High	-0.868***	-0.840***	-1.085***
	(0.164)	(0.163)	(0.220)
Very High	-0.861***	-0.886**	-1.430***
	(0.303)	(0.384)	(0.354)
DK	-1.007***	-1.082***	-1.222***
	(0.314)	(0.349)	(0.309)
Intrinsic Motivation			
Morale	High: Broad Morale (HBM)	High: Restricted (HMNO)	High: Evasion (HEVA)
	-1.139***	-0.205	-1.251***
	(0.263)	(0.171)	(0.182)
Interactions			
Morale* Sanction			
Fine*Morale	-0.0563	-0.528*	-0.405
	(0.365)	(0.277)	(0.251)
Prison*Morale	-2.900***	-1.229***	-2.541***
	(0.990)	(0.407)	(0.806)
Other*Morale	0.145	-0.231	-0.0783
	(0.953)	(0.279)	(0.434)
DK*Morale	0.615	0.561	0.490
	(0.570)	(0.419)	(0.440)
Morale*Risk			
FS*Morale	-0.660	-0.0756	0.418
	(0.465)	(0.313)	(0.280)
FH* Morale	-0.124	-0.205	0.498
	(0.383)	(0.302)	(0.320)
VH* Morale	0.369	0.0672	1.403***
	(0.534)	(0.393)	(0.544)
DK*Morale	-2.079**	-0.243	-0.0249
	(0.981)	(0.379)	(0.389)
Sociodemographics and household	Yes	Yes	Yes
Financial Constraints	Yes	Yes	Yes
Number of Observations	24,623	24,572	24,266

Notes:

The regression results shown here are the result of estimating (3.4) using alternative discrete measures of the intrinsic motivation in each column. The dependent variable in all the regressions is own participation. The variables representing the intrinsic motivation are recoded in this regression to discrete variables equal to 1 if morale is high and 0 otherwise. Morale is defined as high when the act is considered completely unacceptable which means 1 on a ten-point scale. The regression results of column (1) refers to recoding the broad variable of morale (HBM in Appendix B), column (2) refers to recoding morale not related to undeclared work or evasion (HMNO) and column (3) refers to recoding the perception of evasion. A full description of the variables is available in Appendix B. Results are weighted using design and post-stratification weights. The unweighted results are comparable but omitted for brevity. They are available upon request. The stars indicate significance at the following levels: * p<0.10 ** p<0.05 *** p<0.01

Table E.4: Robustness check: Alternative Specification of the Extrinsic Incentives (Risk and Sanction) at different specifications of the intrinsic motivation. Level and crowding effects.

	(1)	(2)	(3)	(4)	(5)
	Log Odds	Log Odds	Log Odds	Log Odds	Log Odds
Extrinsic Motivation					
Sanction					
Fine	-0.150*	1.099***	-0.222***	-0.0941	-0.123
	(0.0787)	(0.401)	(0.0782)	(0.110)	(0.115)
Prison, Other, DK	0.0830	0.528	0.148	0.168	0.292
	(0.152)	(0.444)	(0.146)	(0.128)	(0.180)
Risk of Detection					
High	-0.577***	-1.107***	-0.570***	-0.530***	-0.648***
	(0.101)	(0.386)	(0.131)	(0.134)	(0.173)
DK	-0.938***	-1.385	-0.845***	-0.949***	-0.952***
	(0.333)	(1.046)	(0.321)	(0.340)	(0.352)
Intrinsic Motivation					
Morale	Broad Morale	Broad Morale	High Morale: Broad	High Morale: Restricted	High Morale: Evasion
	-0.416***	-0.351***	-1.395***	-0.217	-0.934***
	(0.0405)	(0.0577)	(0.350)	(0.182)	(0.197)
Interactions					
Morale* Sanction					
Fine*Morale		-0.162***	-0.0862	-0.539**	-0.405
		(0.0559)	(0.388)	(0.270)	(0.249)
Prison, Other, DK*Morale		-0.0583	-0.374	-0.238	-0.497
		(0.0600)	(0.529)	(0.219)	(0.359)
Morale*Risk					
High*Morale		0.0687*	0.251	-0.164	0.302
		(0.0415)	(0.296)	(0.211)	(0.252)
DK* Morale		0.0615	-1.399	0.0796	0.0516
		(0.105)	(0.972)	(0.338)	(0.354)
Sociodemographics and household		Yes	Yes	Yes	Yes
Financial Constraints		Yes	Yes	Yes	Yes
Number of Observations		24,623	24,623	24,572	24,266

Notes:
The regression results shown in column (1) are the result of estimating (3.1) with Broad Morale (BM) as the intrinsic motivation which is the baseline specification and is included for comparison purposes. Column (2) estimates (3.4), this is it investigates crowding effects using Broad Morale. Column (3), (4) and (5) estimates (3.4) using the discrete morale variables: High Broad Morale (HBM), High Morale Restricted (HMNO) and High Morale Evasion (HEVA) respectively. The dependent variable is own participation in the Hidden Economy. Results are weighted using design and post-stratification weights. The stars indicate significance at the following levels: * $p<0.10$ ** $p<0.05$ *** $p<0.01$

Table E.5: Robustness check: Alternative dependent variable—participation of a third-party.

	(1)	(2)
Extrinsic Motivation: Tax Administration		
Sanction		
Due plus fine	-0.0685	0.870***
	(0.0709)	(0.164)
Prison	-0.578***	0.481
	(0.109)	(0.705)
Other	0.589***	0.961
	(0.218)	(0.633)
DK	-0.265**	0.427
	(0.117)	(0.389)
Risk of Detection		
Fairly Small	-0.366***	-0.838
	(0.0713)	(0.626)
Fairly High	-1.025***	-2.116*
	(0.0614)	(1.126)
Very High	-0.724***	-1.782*
	(0.110)	(1.034)
DK	-1.175***	-2.136*
	(0.116)	(1.240)
Intrinsic Motivation		
Morale	-0.218***	-0.220***
	(0.0410)	(0.0714)
Interactions		
Morale* Sanction		
Fine*Morale		-0.111***
		(0.0233)
Prison*Morale		-0.125
		(0.0814)
Other*Morale		-0.0448
		(0.0750)
DK*Morale		-0.0839**
		(0.0409)
Morale*Risk		
FS*Morale		0.0533
		(0.0730)
FH* Morale		0.127
		(0.125)
VH* Morale		0.120
		(0.120)
DK*Morale		0.111
		(0.136)
Sociodemographics and household	Yes	
Financial Constraints	Yes	
Number of Observations	24,175	

Notes:
The regression results shown here are the result of estimating (3.1). The dependent variable is knowledge of the participation of a third-party. The morale variable included is Broad Morale (BM) as defined in Appendix B. Results are weighted using design and post-stratification weights. The stars indicate significance at the following levels: * p<0.10 ** p<0.05 *** p<0.01

Graphical Annex

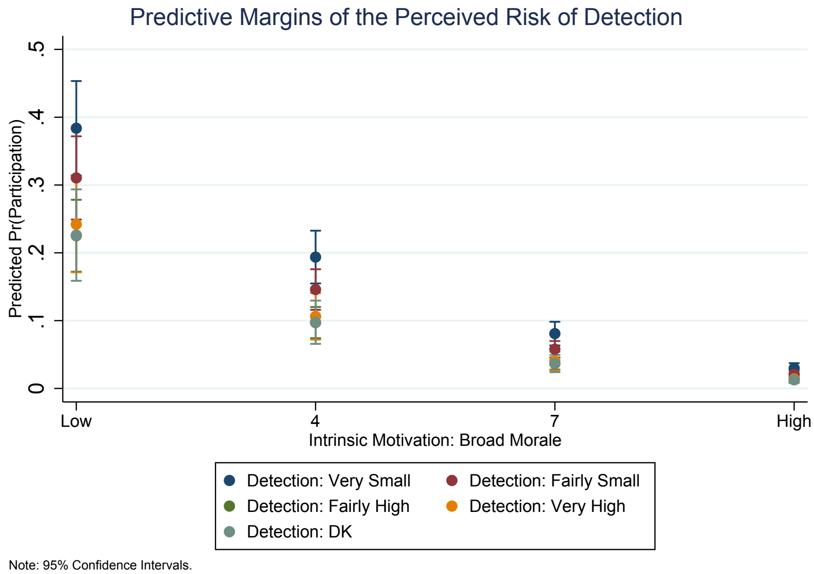


Figure E.1: Margins of the Perceived Risk of Detection according to the Intrinsic Motivation.

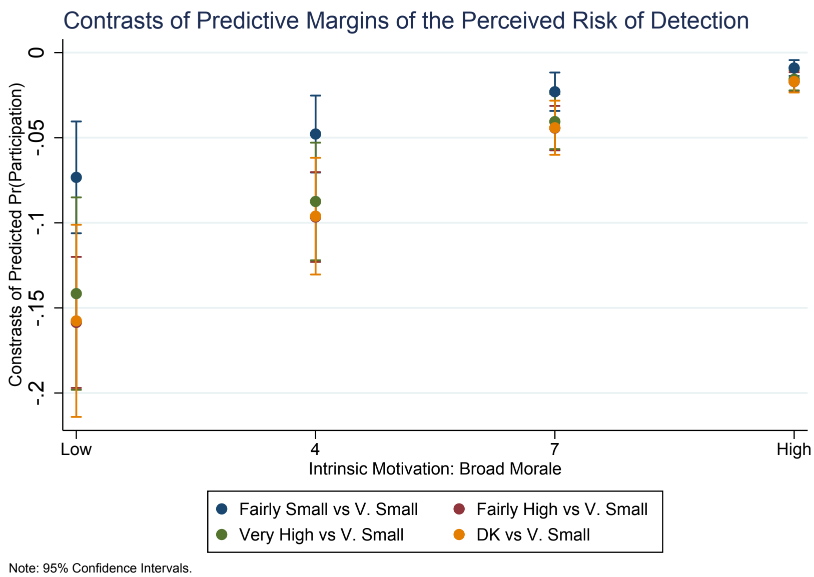


Figure E.2: Contrasts of margins of the Perceived Risk of Detection according to the Intrinsic Motivation.

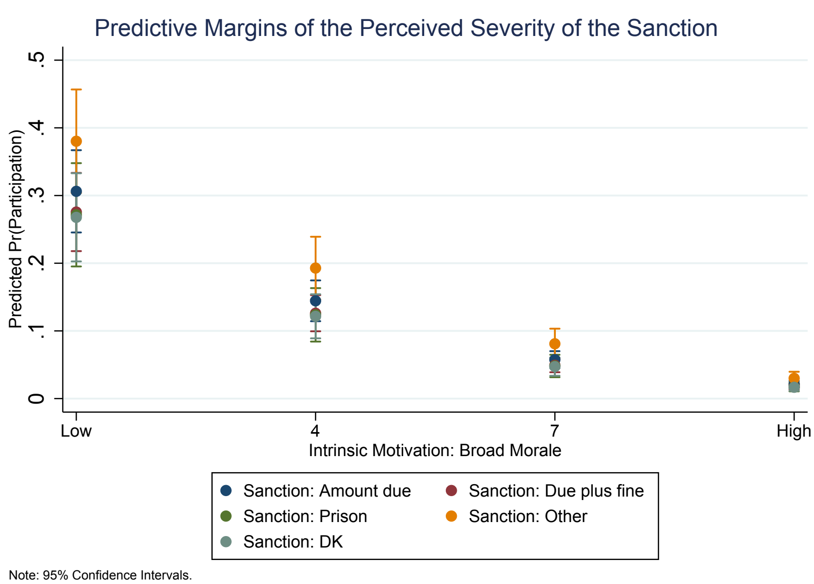


Figure E.3: Margins of the Severity of the Sanctions according to the Intrinsic Motivation.

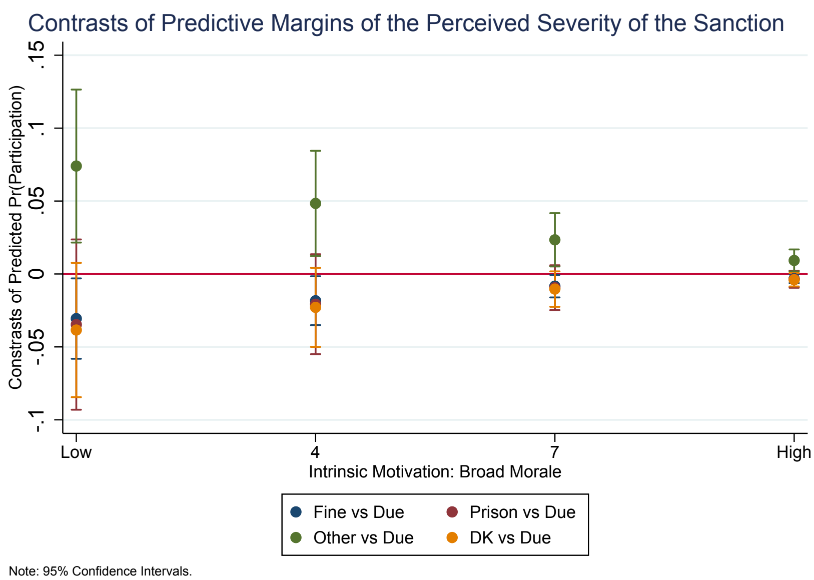


Figure E.4: Contrasts of margins of the Severity of the Sanctions according to the Intrinsic Motivation.

Final Remarks

The present thesis contributes broadly to the literature on Tax Compliance. More specifically, the first and second chapters contribute to the literature on the estimation of concealed income proposing an expenditure-based approach to the estimation of income-gaps for the self-employed and the moonlighters—individuals who are formally employed and their income is subject to withholding taxes but that have a second source of undeclared labour income. The last chapter explores the role of the intrinsic and extrinsic motivation in explaining engagement in the hidden economy, contributing firstly to the Hidden Economy literature and to an emerging literature on the crowding effects between the intrinsic and extrinsic motivations on compliance (Luttmer and Singhal, 2014).

The first chapter provides an estimation of the income-gap for the self-employed, critically assesses the link of the estimation to non-compliance and concludes with the study of the heterogeneity of the evasion behaviour. The second chapter, on the other hand concentrates on the figure of moonlighters—employees whose formal labour income is subject to withholding taxes but have a second source of undeclared labour income.

Both estimation methods are based on a expenditure-based approach where income and expenditure of a benchmark group—that can be argued to concentrate less non-compliance—is used to proxy the amount of unrecorded income of the non-compliant group. On the first estimation, the non-compliant group is the self-employed as they have a more salient opportunity to evade due to the absence of third-party reporting. The employees are chosen as the benchmark group as their income is subject to withholding of taxes and hence their opportunity for misreporting their income is lower. The second chapter, that pursues the estimation of the income-gap for moonlighters, only takes into account the employees and introduces the structure of the non-observed economy into the model. Non-observed activities are typically concentrated within certain sectors. The non-compliant group is composed of those working on these industries where concentration of non-observed activities is high; and the benchmark group is the rest of the employees.

These two methods use an expenditure survey-based data set very rich in observable characteristics to estimate the income-gap for the self-employed and the employed respectively. These methods provide estimations of the income-gaps in a relative inexpensive way as the survey is constructed for purposes other than measuring non-compliance, making this estimation a by-product. Being a general purpose survey, one does not need to rely on self-reported non-compliance dissipating the concerns that typically affect specific surveys. The use of a method constructed upon an expenditure survey provides flexibility, in that it allows the observation of variability of the income-gap for the different characteristics of the individual and different industries; frequency, in that the survey is released on a timely manner (typically yearly) as opposed to lower frequency hidden economy surveys; opportunity to obtain an estimation for the income-gap for those tax administrations that do not have advanced administrative data to do so or even if they do, the method provides an estimate to triangulate those obtained relying solely administrative data.

Measurement is essential to provide a sense of the magnitude but tackling non-compliance requires enhancing the understanding of what drives individuals to comply. The third piece of work on this thesis explores the role of the intrinsic and extrinsic incentives in explaining participation in hidden economy activities. We find that individuals react rationally to extrinsic incentives—audits and sanctions—by lowering their participation when these are higher. Their degree of intrinsic motivation to pay taxes matters, this is individuals with a stronger intrinsic motivation exhibit much lower odds of participation. These incentives are also shown not to interact with each other, that is there is no crowding effect between the intrinsic and extrinsic motivation, which has important policy implications. There are individuals that comply because they are intrinsically motivated to do so, but for those who are not, extrinsic incentives are paramount to ensure compliance. The absence of crowding effects means that a strong presence of the tax administration in terms of enforcement drives those extrinsically motivated towards compliance while not affecting the compliance behaviour of the intrinsically motivated that will continue to comply.

Although there are caveats associated to the use of survey data e.g. non-response, low response rates; there are substantial advantages with respect to other data sources. First of all, questionnaires are carefully built to minimise measurement error. Secondly, surveys provides a richness of characteristics that enables analysis not available under other sources. When it comes to specific surveys, such as the one used in the third chapter, they provide very valuable information to shed light over the reasons for participation which may evidence structural problems in the economy or the system, e.g. weak formal labour market with low quality employment or low pay.

Tax non-compliance, due to its nature, is a complicated subject of study. However, this should not be seen as a problem but as an opportunity to encourage creative approaches to measuring it. For the future, as more administrative data is becoming available, the combination of survey and administrative data sources could provide a good avenue to exploit the best of both worlds. Field and natural experiments are also good tools to analyse the impact of policy decisions and to deepen our understanding of how taxpayers react to incentives and to evaluate the effectiveness of our policy interventions. As Albert Einstein once wrote on his book *Ideas and Opinions*: "All knowledge of reality starts from experience and ends in it"(Einstein, 1960, p. 271)[55].

[55] Einstein, A. (1960). *Ideas and Opinions*. Crown Publishing Group.