LIS
Working Paper Series

No. 723

Taxation and self-employment

Zsófia L. Bárány

November 2017

Luxembourg Income Study (LIS), asbl

Taxation and self-employment*

Zsófia L. Bárány†

November 2017

Abstract

In this paper I theoretically show that if the self-employed evade income taxes, then the choice of being self-employed is more sensitive to the tax rates on wages than to tax rates on income from self-employment. Using variation in the statutory tax rates across countries, industries, and occupations, I find evidence that supports the predictions of the model. This suggests that those who choose self-employment, partly do so to take advantage of the technology it offers in evading taxes. This extensive margin of adjustment – between employment and self-employment – should be taken into account when considering the effects of tax rates on labor income, on taxable income and on welfare.

*I thank Ghazala Azmat, Moshe Buchinsky, Tom Cunningham, Quoc-Anh Do, Roberto Galbiati, Yuriy Gorodnichenko, Sergei Guriev, Emeric Henry, Henrik Kleven, Guy Laroque, as well as seminar participants at the Institute for International Economic Studies and Sciences Po for comments and suggestions. I gratefully acknowledge funding from the Scientific Advisory Board of Sciences Po.

†Sciences Po. Email: zsofia.barany@sciencespo.fr

1 Introduction

One of the central questions in tax policy is the response of tax evasion to tax rates. On the one hand, the intensive response of evaded income to the (marginal) tax rate is theoretically ambiguous, and while there is no consensus in the empirical literature, the most recent evidence suggest a small positive response.[1] On the other hand, there is overwhelming evidence that mostly the self-employed income – which is not subject to third party reporting – is evaded. This implies that there is a potential extensive response of tax evasion to tax rates, coming from a change in the fraction of self-employed workers.

In this study I analyze whether individuals respond to the tax rates they face on the extensive margin, by adjusting their status between dependent employment and self-employment. Through a simple model of the decision between dependent employment and self-employment, I show that in the absence of tax evasion opportunities, the tax rate faced as an employee should have a positive impact, while the tax rate faced as a self-employed worker should have a negative impact on the attractiveness of self-employment of roughly the same magnitude. I also show that if the self-employed underreport their income, then the negative impact of the tax rate faced as a self-employed on the attractiveness of self-employment significantly decreases.

Using cross-country household level data from the Luxembourg Income Studies (LIS), I document some novel patterns of self-employment across countries. I show that there are large differences across countries in the extent of self-employment.[2] I document that within each country there are substantial differences between occupations and between industries in the propensity of self-employment, and that the ranking of occupations/industries based on their self-employment rate is very similar across countries. Despite these systematic differences between occupations and industries, I show that cross-country differences in industry and occupation employment

[1] In general there is both an income and a substitution effect of tax rates, which of the two dominates depends on the assumed penalties and detection probabilities, see for example Allingham and Sandmo (1972) and Yitzhaki (1974). Empirically while Clotfelter (1983) finds a positive, Feinstein (1991) finds a negative effect of marginal tax rates on evasion, Kleven, Knudsen, Kreiner, Pedersen, and Saez (2011) find a positive effect. For surveys see Andreoni, Erard, and Feinstein (1998) and Sandmo (2005).

[2] This has been documented in the literature, see for example Acs, Audretsch, and Evans (1994), Blanchflower (2000), Parker and Robson (2004).

structure are not large enough to account for the differences in self-employment rates between countries.[3] These observations suggest that (i) a key determinant of the self-employment decision is the individual's occupation and industry, but that nonetheless (ii) there are large cross-country differences in self-employment rates within industries and occupations. I argue that these differences, beyond country-level shifters, are partly explained by the differences in tax schemes across countries.

I test the predictions of the model at the occupation-industry level using the LIS database combined with the complete national income tax and social security schedule for nine countries in two years. Exploiting variation in the tax rate both within and across countries, industries and occupations, I find that while the tax rate faced as an employee has a significant positive relation to the self-employment rate, the tax rate faced as a self-employed has a near zero effect. This provides supportive evidence of the predictions of the model with tax evasion. The large difference between the magnitude of the coefficient on the employee and the self-employed tax rate suggests that when taxes on dependent employment income are high, more individuals choose self-employment in order to take advantage of the technology it offers in evading taxes.

In my model I build on the literature which has documented that it is mostly the self-employed individuals who evade income taxes, and that they underreport their income by a substantial amount. This is because the income of the self-employed is not reported to the tax authority by a third party, which drastically reduces the probability of detection. The papers that document this fall into one of two categories. Papers in the first category rely on data from tax audits, which provides direct measures of uncovered tax evasion, or adjustments in income reporting in response to a threat of audit.[4] Papers in the second category build on the assumption that if employed and self-employed individuals are similar in their level of consumption or bank loans, but

[3] Acs et al. (1994), Blanchflower (2000), Parker and Robson (2004) have noted the different self-employment rates for broad sectors of the economy (agriculture vs non-agriculture, manufacturing vs services, public vs private), and Torrini (2005) for finer industries, but this is a novel finding in terms of occupations. Most papers suggest that the structure of the economy across these sectors is a key determinant of economy-wide self-employment (Blau (1987), Acs et al. (1994), Blanchflower (2000), Parker and Robson (2004)), while Torrini (2005) reaches the same conclusion as I do for industries.

[4] Andreoni et al. (1998), Bloomquist (2003), and Slemrod (2007) use the Taxpayer Compliance Measurement Program of the US Internal Revenue Service, which is a thorough tax audit on a stratified random sample of income tax returns. Kleven et al. (2011) use data from a tax enforcement field experiment in Denmark.

the self-employed have lower declared earnings, then this is due to income underreporting.[5] While these papers use different methodologies, and data from different countries and different time periods, they all find that the underreporting of income from self-employment is between 20 and 50%, with most studies finding numbers closer to 50%.

These observations suggest that the tax rates, which determine both the net income difference and the gains from income underreporting, should have an impact on the choice of becoming self-employed.[6] My estimation strategy, similarly to Bruce (2000) and Parker (2003), relies on the gains from working as a self-employed rather than as an employee, taking into account the possibility of evasion by the self-employed. My paper contributes to the literature in a number of ways. First of all, I show the importance of controlling for the individual's occupation and industry. Given the systematic difference across industries and occupations in terms of average self-employment rates that I document, these characteristics seem to play a key role in the self-employment decision. Second, following the model of the self-employment decision I explicitly acknowledge that the reported income of the self-employed is not their true income, and I allow for the employee and the self-employed tax rate to have effects of different magnitudes. In line with the model, I find that the tax rate faced as an employee has a large positive impact, while the tax rate faced as a self-employed has a small negative impact on the self-employment decision. Both my estimation strategy and my findings are different from those in Bruce (2000) and Parker (2003). Bruce (2000) controls for the difference in the employed and self-employed tax rate, constraining the coefficients to be of the same magnitude, but of different sign. He therefore implicitly assumes that the self-employed fully comply in tax payments. He finds that higher marginal tax rates in employment reduce the probability of going into self-employment, whereas

[5] Pissarides and Weber (1989), Baker (1993), Apel (1994), Schuetze (2002), Johansson (2005), Kim, Gibson, and Chung (2009), Hurst, Li, and Pugsley (2014) for example use data on consumption and income, while Artavanis, Morse, and Tsoutsoura (2016) uses data on loans and income from a large Greek bank.

[6]Several papers analyze the effects of the aggregate tax climate on self-employment rates. These papers, instead of relying on the gains of working as a self-employed rather than as an employee, look at the correlation between the tax rate faced by the employed (typically at a single income level) and the self-employment rate, and tend to find a positive correlation. See for example Long (1982), Blau (1987),Parker (1996), Robson and Wren (1999), Schuetze (2000), Parker and Robson (2004).

higher average tax rates mildly increase the probability. This latter result has similarities with what I find, but as he constrains the tax rates to have an effect of the same magnitude he cannot find the asymmetric result, which is a key finding in my paper. Parker (2003) controls for the gains in terms of net income, and assumes that the self-employed report their earnings truthfully, but do not pay a fraction of taxes due. This is at odds with the evidence, which suggests that the self-employed evade income taxes by reporting a lower income both in surveys and to the tax authority (Hurst et al. (2014)). This assumption together with not controlling for occupations and industries might be the reason that – as opposed to my results – he does not find any evidence of a tax evasion motive in the self-employment choice.

My findings imply that the extensive margin of adjustment should be taken into account when considering the effects of tax rates on labor income, on taxable income and on welfare. Since Feldstein (1999) showed that taxable income is a sufficient statistic to evaluate the welfare gains from reducing taxes, several papers have revisited this question. In particular Chetty (2009) and Gorodnichenko, Martinez-Vazquez, and Peter (2009) point out that in the presence of tax evasion or sheltering, when these activities entail not only resource, but transfer costs as well, taxable income might not be a sufficient statistic for assessing the impact of tax changes on welfare. They show that the welfare costs in such a case depend on the elasticity of both labor income and taxable income. My results suggest that these models should be enriched with a choice on the extensive margin between employment and self-employment in order to calculate the welfare costs of taxation.

The rest of the paper is organized as follows: section 2 lays out the simple model of the self-employment decision, section 3 describes the data and the descriptive analysis, section 4 the results on tax rates and self-employment, section 5 concludes.

2 A simple model of the self-employment decision

This section provides a overview of the effect of tax rates on the decision between employment and self-employment. The tax rates influence the employment – self-employment margin through their effect on disposable income. First, I sketch a model without tax evasion, where everyone is assumed to report their earnings truthfully. Second, I augment this model by taking into account the possibility of tax evasion by the self-employed.

Assume that everyone reports their income truthfully, and that utility is a logarithmic function of disposable income. Denote the expected income of individual j in employment by y_{jE}, and in self-employment by y_{jS}. Further let $t^E(y)$ denote the average tax rate that an employee faces at income level y, and $t^S(y)$ denote the average tax rate that a self-employed faces at income level y. Under these assumptions the utility from net income in the two statuses can be expressed as:

$$U_E = U(y_{jE}, t^E) = \log y^d_{jE} = \log(y_{jE}(1 - t^E(y_{jE}))),$$
$$U_S = U(y_{jS}, t^S) = \log y^d_{jS} = \log(y_{jS}(1 - t^S(y_{jS}))).$$

The marginal effects of the average tax rates on the utility difference from consumption are:

$$\frac{\partial U_S - U_E}{\partial t^E(y_{jE})} = \frac{1}{1 - t^E(y_{jE})},$$
$$\frac{\partial U_S - U_E}{\partial t^S(y_{jS})} = -\frac{1}{1 - t^S(y_{jS})}.$$

Thus, in case of no evasion, one would expect the tax rates in employment and self-employment to have opposite effects of similar magnitude.

Now consider the case when tax evasion is possible. I assume that while employees report their entire income truthfully, self-employed only report $\kappa \leq 1$ fraction of their income, i.e. $\widetilde{y}_{jS} = \kappa y_{jS}$, where $\widetilde{y}_{jS}$ is the reported and y_{jS} is the true income of the self-employed. As discussed earlier, this assumption is widely supported by the empirical evidence.[7] For simplicity assume that the fraction κ is given, and that people never get caught.[8] Under these assumptions the log utility from consumption for the self-

[7]See for example Andreoni et al. (1998), Bloomquist (2003), Slemrod (2007), and Kleven et al. (2011). The papers estimating the extent of income underreporting by the self-employed (denoted by κ here) also start from this assumption.

[8]See the discussion in the appendix for a case where these assumptions are relaxed.

employed can be written as:

$$U_S = \log(y_{jS} - \kappa y_{jS} t^S(\kappa y_{jS})) = \log(y_{jS}(1 - \kappa t^S(\widetilde{y}_{jS})).$$

When the self-employed shelter $(1 - \kappa)$ fraction of their income the marginal effects of the average tax rates on the utility difference from consumption are:

$$\frac{\partial U_S - U_E}{\partial t^E(y_{jE})} = \frac{1}{1 - t^E(y_{jE})},$$
$$\frac{\partial U_S - U_E}{\partial t^S(\widetilde{y}_{jS})} = -\frac{\kappa}{1 - \kappa t^S(\widetilde{y}_{jS})}.$$

The magnitude of the effect of the tax rate on reported self-employed income on the utlility difference between self-employment and employment is decreasing in the amount of income evaded. This can be seen in the expression above, as the numerator is a decreasing function, while the denominator is an increasing function of κ. A smaller κ means a smaller fraction of income reported to the tax authority. Thus, in case of evasion, one would expect the tax rates in employment and self-employment to have opposite effects, and a smaller absolute magnitude for the tax rate in self-employment.

To summarize, through this simple model of the decision between employment and self-employment I showed that in the absence of tax evasion, the tax rate on wage earnings and on income from self-employment should have opposite effects of roughly the same magnitude, whereas if the self-employed can evade income taxes, then the magnitude of the effect of tax rates on self-employed income should be smaller.

3 Data and descriptive analysis

3.1 Employment data

The employment data comes from the Luxembourg Income Study (LIS) Database, which is a harmonized collection of microdatasets for upper- and middle-income countries. Depending on the country, the Luxembourg Income Study Database contains data from either a tax register or a household survey. The LIS dataset has been collected in waves, the initial, Wave 1 contains data from around 1980, while the most recent Wave 9 contains data from around 2013. This paper uses data from Wave 5 and Wave 6, from the years 1999-2002 and 2003-2005.

I restrict the sample to 18-65 year old working individuals, who are either household heads or the partner of the household head. The most important variables are the *status in employment* and the *personal labor income* variables. All persons whose *status in employment* is self-employed, employer, own-account worker, or contributing family worker are counted as self-employed in the empirical specification, while all others are considered employees. The *personal labor income* is the gross yearly earnings of the individual, before personal income tax and social security contribution deductions. When using the *personal labor income* variable, it is important to keep in mind the limited reliability of the income data of the self-employed. For some countries the LIS microdata comes directly from the tax registers. For these countries, if the self-employed indeed tend to underreport their income, then the income data for the self-employed is not the true one. Moreover, household surveys are similar to tax registers in this sense, i. e. the self-employed also tend to underreport their income in household surveys, as demonstrated for example in Hurst et al. (2014). The fact that in household surveys a large fraction of the self-employed do not report their personal labor income at all just exacerbates the problem.

Two other important variables are the industry and the occupation of workers. The industry of a worker is from the following 9-category industry classification: (1) agriculture, forestry and fishing, (2) mining and quarrying, manufacturing, utilities, (3) construction, (4) wholesale and retail trade, repair, hotels and restaurants, (5) transport, storage and communication, (6) financial intermediation, (7) real estate, renting and business activities, (8) public administration, education, health and social work, (9) community, social and personal services; activities of households. The occupation of the worker is one of the following 10-category International Standard Classification of Occupations (ISCO) recode: (1) managers, (2) professionals, (3) technicians and associate professionals, (4) clerical support workers, (5) service and sales workers, (6) skilled agricultural, forestry and fishery workers, (7) craft and related trades workers, (8) plant and machine operators, and assemblers, (9) elementary occupations, (10) armed forces.

3.2 Building the tax functions

The main hypothesis this paper aims to test is that people respond to higher tax rates by adjusting their employment status from employment to self-employment in order to be able to underreport their income and hence evade some of their tax payments. In order to test this, I rely on both cross-country and within-country variation in statutory tax rates. However, accurate information on the tax rates on wages and on income from self-employment at different income levels for several countries is not readily available.

The Organization for Economic Cooperation and Development (OECD) publishes Taxing Wages annually (OECD (2000, 2001, 2002, 2004, 2005)).[9] This publication provides information for all OECD countries on the precise rules governing the income tax schedule, as well as the social security contributions paid by the employees, by the employers, and by the self-employed, as well as all family benefits paid as cash transfers or as tax credits.[10]

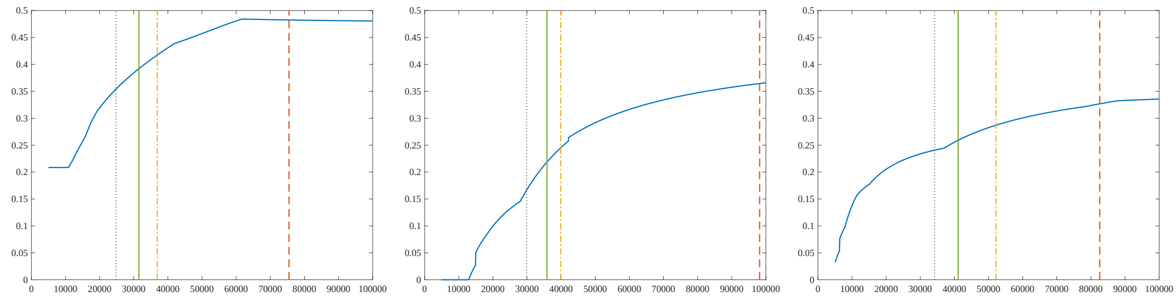
Using the information contained in the Taxing Wages publications I create two tax functions for each year, and for each country of interest: one for the employees and one for the self-employed. In these tax functions beyond carefully coding the standard tax schedules, I also take into account all the tax credits and tax reliefs, the different rates and limits of all social security contributions made by the employee and the self-employed, as well as any cash transfers.[11] These tax functions return for any reported

[9]Before 1996 it was called The Tax/Benefit Position of Production Workers, between 1996-1998 it was called The Tax/Benefit Position of Employees. From 1998 it is published as Taxing Wages.

[10]Additionally Taxing Wages reports the average and marginal tax rates, the total labor costs and benefits for eight household types, which differ by income level and household composition. However, this information cannot be directly applied in the setting of this paper, as the tax rates are needed for several income levels.

[11]As an example of the employee social security contributions: in Ireland in 2004 each employee was exempt from paying health insurance contributions if their weekly earnings did not exceed 356 EUR, but they needed to pay the 2% contribution rate on all of their earnings if their weekly earnings exceeded 356 EUR. For social insurance, employees were exempt if they earned less than 287 EUR, above this limit, but below 42160 EUR a year they had to pay 4% with a weekly exemption of 127 EUR of earnings, and above 42160 EUR per year they had to pay 4% of 42160 EUR.

gross income the average tax rate including social security contributions made by the employee or the self-employed. To my knowledge, this is the first paper that creates these functions and uses tax rates corresponding to the actual income level of individuals in a cross-country setting.



(a) Germany (b) Ireland (c) USA

Figure 1: Average tax rates as a function of annual gross labor income

This figure plots average tax rates inclusive of social security contributions of employees as a function of annual income in Germany, Ireland, and the USA in 2004, authors own calculations based on OECD Taxing Wages (2004 and 2005 editions). The vertical lines show the average earnings of workers in the same year in the following occupation-industry pairs: managers in financial intermediation (dashed red), managers in wholesale and retail trade, repair, hotels and restaurants (dashed-dotted yellow), craft and related trades workers in mining and quarrying, manufacturing, utilities (solid green), craft and related trades workers in wholesale and retail trade, repair, hotels and restaurants (dotted grey), authors own calculation from Wave 6 of the LIS.

Figure 1 shows the average tax rates as a function of annual gross labor income for three selected countries in the year corresponding to Wave 6 of the LIS. The vertical lines show the average earnings of employed workers in the given year and country calculated from Wave 6 of the LIS for four occupation-industry pairs. These are the following: managers in financial intermediation (dashed red), managers in wholesale and retail trade, repair, hotels and restaurants (dashed-dotted yellow), craft and related trades workers in mining and quarrying, manufacturing, utilities (solid green), craft and related trades workers in wholesale and retail trade, repair, hotels and restaurants (dotted grey). There are a few things to note from this Figure. First of all, there is quite a big variation in average tax rates faced by the average earner across occupation-industry cells, within a country. Second, comparing the tax rates within a country and an occupation (red dashed and yellow dashed-dotted for managers, green solid and grey dotted for craft and related trades workers), there is also quite a large difference. The substantial variation in average tax rates within countries suggests that when assessing the effect of average tax rates on the choice of status in employment, controlling for only the average realized value of the average tax rate in the economy as a whole might not be sufficient.

3.3 Descriptive Analysis

Table 1 shows the self-employment rate in the whole working population and among the non-agricultural workers for several countries as calculated from the LIS data. This Table shows that self-employment rates vary substantially across countries, and that it is relatively stable within countries.[12] The rates vary from 6.6% in Denmark to 33.9% in Greece for the entire working population, and between 5.4% in Denmark and 26.3% in Greece among the non-agricultural workers.

The differences across countries can be potentially due to differences in the economic structure. As pointed out by Torrini (2005) there is a large and systematic variation in the self-employment rates across industries. This is true in the LIS data as well, as shown in the top panel of Figure 2. A new finding of this paper, demonstrated in the bottom panel of Figure 2 is that the self-employment rates across occupation groups differ significantly as well.[13] The cross country average of industry or occupation self-employment rates are shown with the black squares, the other symbols show the self-employment rate for the industry or occupation in the given country. This Figure shows on the one hand that there is a systematic difference in the propensity of self-employment both across industries and across occupations, which is common to countries. Perhaps not surprisingly the highest self-employment rate tends to be in the *agriculture, forestry and fishing* industry and in the *skilled agricultural, forestry and fishery* occupations. The lowest self-employment rates tend to be in the industry of *public administration, education, health and social work*, and in the occupation of *clerical support workers*. This suggests that perhaps some of the cross-country differences in overall self-employment rates are due to differences in the structure of employment either across industries or across occupations. On the other hand this Figure also shows

[12]Acs et al. (1994), Blanchflower (2000), Parker and Robson (2004) (among others) establish the large dispersion and document the patterns of self-employment over time within a countries.

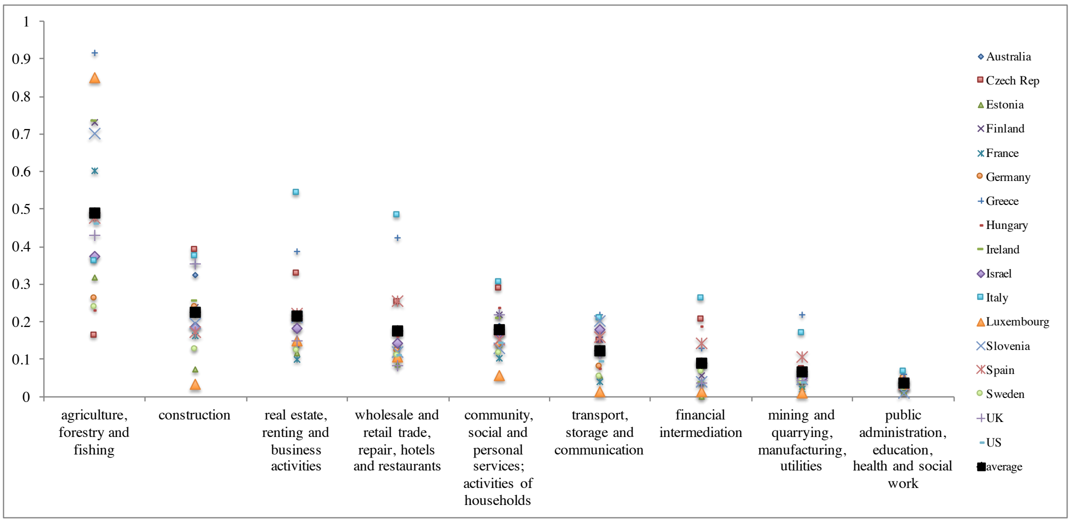
[13]See Table 4 and Table 5 in the Appendix for the cross-country average self-employment rates and employment shares, as well as their standard deviation.

Table 1: Self-employment rates across countries

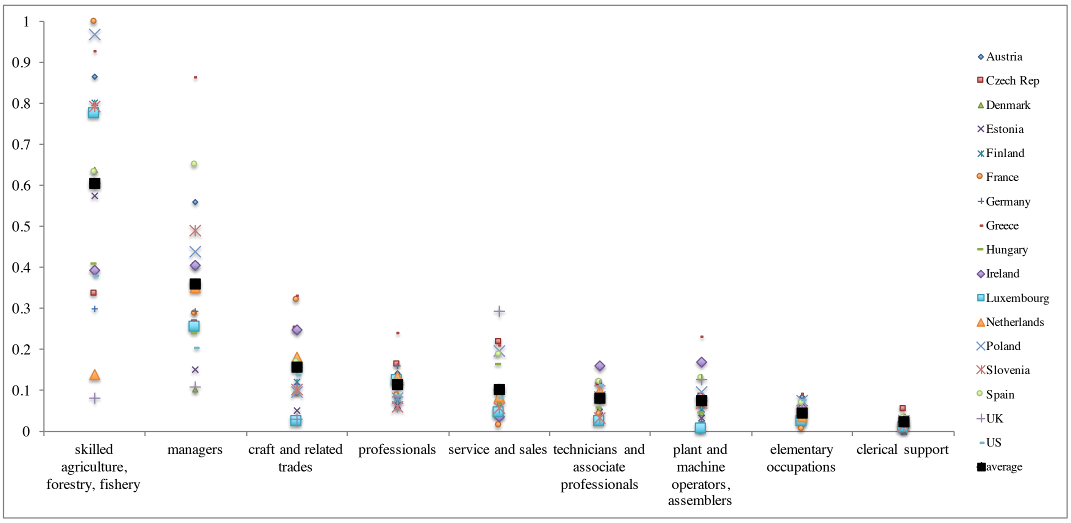
	Share of self-employed	Share of self-employed in non-agriculture
Australia	0.129	0.115
Austria	0.119	0.091
Canada	0.158	-
Czech Republic	0.160	0.159
Denmark	0.066	0.054
Estonia	0.072	0.056
Finland	0.144	0.111
France	0.095	0.070
Germany	0.097	0.094
Greece	0.339	0.263
Hungary	0.112	0.103
Ireland	0.157	0.139
Israel	0.111	0.107
Italy	0.267	0.261
Luxembourg	0.071	0.052
Netherlands	0.116	0.129
Norway	0.070	-
Poland	0.268	0.117
Slovenia	0.122	0.090
Spain	0.160	0.149
Sweden	0.098	-
Switzerland	0.108	-
UK	0.107	0.104
US	0.108	0.103

Author's own calculations from Wave 6 of the Luxembourg Income Study Database (LIS). Share of self-employed, employer, own-account worker and contributing family worker, in the working age (18-65) total employed population in the first column, and among the non-agricultural workers in the second column.

that the actual industry/occupation self-employment rates in specific countries vary quite a bit around the cross-country averages. For example Figure 2a shows that Italy, which has one of the highest overall self-employment rates in Wave 6 of the LIS data at 26.7%, has one of the highest self-employment rates in every industry except *agriculture, forestry and fishing*, well above each industry's cross-country average. In the other panel 2b we see that Luxembourg, which has a very low overall self-employment rate at 7.1%, tends to have low self-employment rate in all occupations, except in the occupation of *skilled agricultural, forestry and fishery workers*.



(a) Industry



(b) Occupation

Figure 2: Self-employment rates

Author's own calculations from Wave 6 of LIS. This figure shows the self-employment rates in each industry in the left panel, and in each occupation in the right panel for various countries. The industries and occupations are ordered based on their cross-country average self-employment rate (shown with black squares).

In what follows, I compare two counterfactual self-employment rates to gauge the importance of the structure of employment and of the industry/occupation self-employment rates for the variation in overall self-employment rates across countries.

In each country the overall self-employment rate can be written as:

$$\eta = \sum_{j=1}^{J} \lambda_j \eta_j,$$

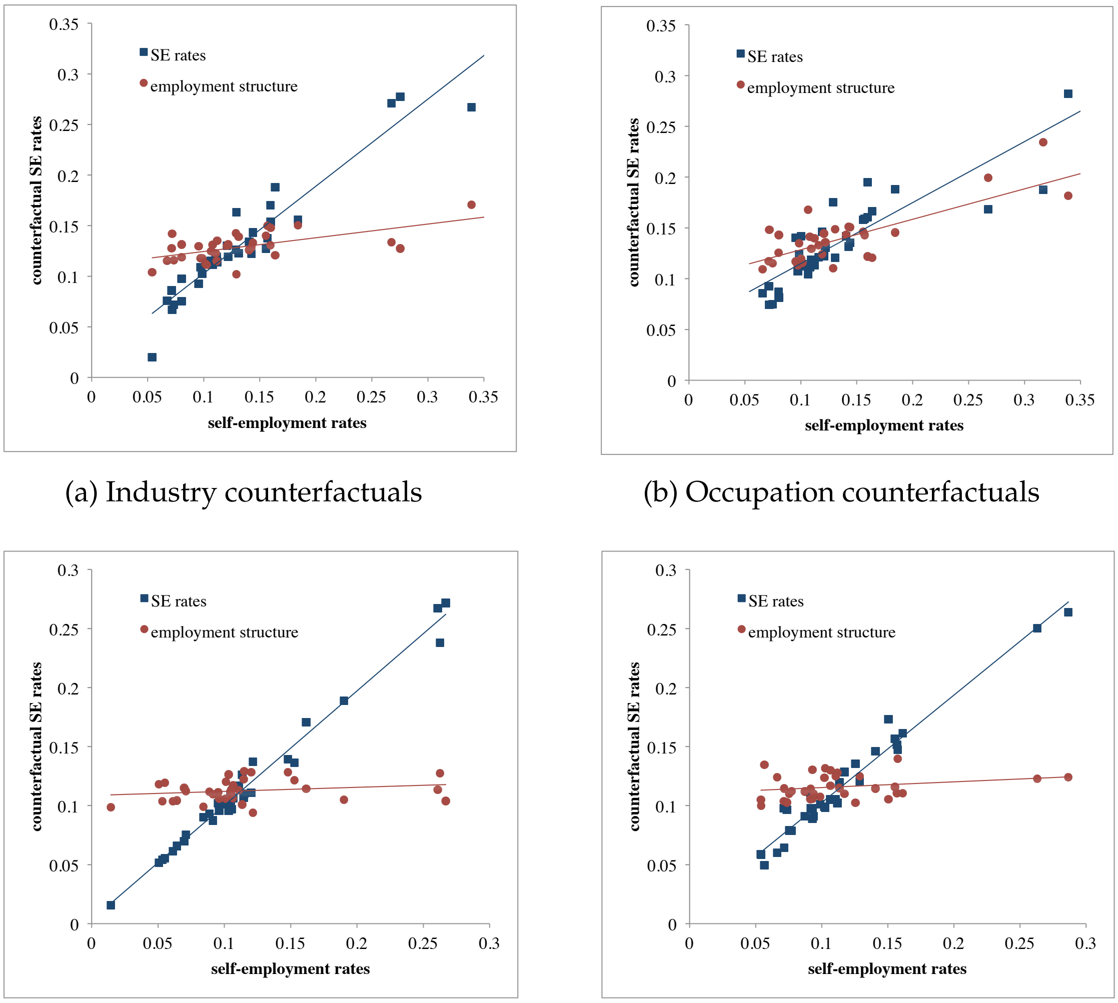
where the aggregation j can be over industries or occupations, η is the overall self-employment rate, λ_j is the employment share in industry (or occupation) j, and η_j is the self-employment rate in industry (or occupation) j. Given this formulation, one can compute two counterfactual overall self-employment rates for each type of aggregation category:

$$\widehat{\eta}^{SE} = \sum_{j=1}^{J} \overline{\lambda}_j \eta_j, \tag{1}$$

$$\widehat{\eta}^{emp} = \sum_{j=1}^{J} \lambda_j \overline{\eta}_j. \tag{2}$$

In the above formulation $\widehat{\eta}^{SE}$ captures the importance of cross-country differences in industry or occupation self-employment rates: it uses the country's self-employment rate in each category, while using the cross-country average of each category's employment share. Conversely $\widehat{\eta}^{emp}$ captures the importance of cross-country differences in the structure of employment: it uses the country's employment share structure across categories, but uses the cross-country average of the self-employment rate in each category.

Figure 3 shows the counterfactual self-employment rates based on industries in the left panels, and based on occupations in the right panels. The top row is based on all working individuals, whereas the bottom row is based on those outside of agricultural industries (in the left), and occupations (in the right panel).The dots correspond to countries, showing on the horizontal axis the actual self-employment rates (η), and on the vertical axis the counterfactual one. The blue squares show the counterfactual self-employment rates based on the country's actual self-employment rates in each industry or occupation ($\widehat{\eta}^{SE}$), whereas the red circles show the counterfactual self-employment rates based on the country's actual employment structure ($\widehat{\eta}^{emp}$). In these graphs, the closer is the fitted line to the 45 degree line, the better that predictor is



(a) Industry counterfactuals

(b) Occupation counterfactuals

(c) Industry cfs non-agriculture

(d) Occupation cfs non-agriculture

Figure 3: Counterfactual self-employment rates

Author's own calculations from Wave 5 and 6 of LIS. In this figure each dot corresponds to a country, showing the counterfactual self-employment rates on the vertical axis, against its actual value on the horizontal axis, as well as linear fits. In all panels the red circles correspond to $\widehat{\eta}^{emp}$ from (2), and the blue squares correspond to $\widehat{\eta}^{SE}$ from (1). The top row is based on all working individuals, the bottom row is based on those working in non-agricultural sectors or occupations. The left panels show these based on the industry aggregation, while the right based on occupations. The average employment share ($\overline{\lambda}_j$) and self-employment rate ($\overline{\eta}_j$) in each category is calculated by wave.

for the actual self-employment rate. In general these graphs show that while both the actual structure of employment and the actual self-employment rates in different categories are decent predictors of a country's actual self-employment rate, the latter provides much better predictions. Moreover, looking at the self-employment rate among non-agricultural workers, the fit using the actual employment structure ($\widehat{\eta}^{emp}$) deteriorates, while the fit using the actual self-employment rates across categories ($\widehat{\eta}^{SE}$)

improves.[14] This implies that the predictive power of the structure of employment for the cross-country differences in self-employment rates mainly comes from the size of the country's agricultural sector and agricultural occupations. These observations imply that to understand cross-country differences in self-employment rates, one has to understand what drives cross-country differences within occupations and industries in self-employment rates.

In the next sections I investigate whether a country's tax schedule plays a role in the share of self-employed in specific industry-occupation cells.

4 Tax rates and self-employment

In what follows, I test the predictions of the simple model outlined in Section 2. Recall that the model predicts that in the case of no evasion the marginal impact of the tax rate on wages and of the tax rate on income from self-employment should be similar in magnitude but of opposite sign. It also predicts that if only the self-employed can evade their income taxes, then the effect of the tax rate on income from self-employment should be smaller in magnitude.

Besides the expected differences in the utility from consumption other factors can impact the decision to work as self-employed. This decision likely depends on non-pecuniary factors such as age, education, marital status, number of children, industry, occupation and country of work (all captured in X_j), and an idiosyncratic utility from working as a self-employed ($\varepsilon_j \sim F(\cdot)$ iid with mean zero). Using the formulation allowing for tax evasion by the self-employed, the selection equation can be written as:[15]

$$I_j^* = 1 \quad \text{if} \quad \gamma_1(\log y_{jS} - \kappa t^S(\widetilde{y}_{jS}) - \log y_{jE} + t^E(y_{jE})) + \gamma_2 X_j + \varepsilon_j \geq 0, \tag{3}$$

where $I_j^* = 1$ if the individual works as a self-employed, and is zero otherwise. There are two difficulties in estimating the above equation. The first problem, especially for

[14] Torrini (2005) calculates what is here called $\widehat{\eta}^{SE}$ for the industry aggregation for non-agricultural workers for several European countries, and reaches a similar conclusion as I do in stating that the employment structure across industries does not seem to be the major determinant of cross-country differences in the self-employment rate.

[15] This formulation relies on the approximation $\log(1 - t^E(y_{jE})) \approx -t^E(y_{jE})$ and $\log(1 - \kappa t^S(\widetilde{y}_{jS})) \approx -\kappa t^S(\widetilde{y}_{jS})$.

the self-employed, is that I observe the reported income ($\widetilde{y}_{jS}$), and not the true income (y_{jS}). I do not correct for income underreporting, instead in the empirical specification I allow the coefficients on income and on tax rates to be different between the employed and the self-employed.[16] The second issue is that I observe the income for each individual in the employment status that he is currently in, but not in the alternative employment status. In this paper I use the average realized reported earnings of the employed (y^E) and of the self-employed ($\widetilde{y}^S$) as an approximation for the expected (reported) earnings.[17] Given the expected earnings in each employment status, I can calculate the average tax rate at the relevant income level for the employed $t^E(y^E)$ and the self-employed $t^S(\widetilde{y}^S)$ using the tax functions specific to the country and employment status described in section 3.2.

In order to implement this, I create country-industry-occupation-year cells, and I calculate the average reported labor income of the employed and the self-employed and the respective tax rates for each cell. I create these cells based on the evidence in section 3.3, which shows that there are significant differences in the propensity of self-employment across occupations and across industries, as well as significant cross-country differences. The fraction of self-employed in a given cell can then be approximated as:

$$\eta_{SE} = 1 - F\left(-\gamma_1 \log \widetilde{y}^S + \gamma_1 \log \kappa + \gamma_1 \kappa t^S(\widetilde{y}^S) + \gamma_1 \log y^E - \gamma_1 t^E(y^E) - \gamma_2 X\right), \quad (4)$$

where X contains the average of the personal characteristics of individuals in the cell, and cell characteristics, and $F(\cdot)$ is the cumulative distribution function of the idiosyncratic utility of working as a self-employed. Denoting its probability density function by $f(\cdot)$, it can be seen that the fraction of self-employed in a given cell is increasing in $t^E(y^E)$ at rate $\beta_1 f(\cdot)$, while it is decreasing in $t^S(\widetilde{y}^S)$ at rate $\kappa\beta_1 f(\cdot)$.

[16]The papers estimating the degree of income underreporting by the self-employed all use information on a different measure (consumption for example in Pissarides and Weber (1989) or bank loans as in Artavanis et al. (2016)), which is assumed to 1) be correctly reported for the self-employed as well, and 2) have a similar relation to actual income for the employed and the self-employed. This type of information is not available in the LIS data.

[17]This procedure does not address the issue of selection bias in earnings. In this context, since selection depends on net income, which is a non-linear function of gross income, a selection bias correction as in Heckman (1979) is not possible.

To test for this relation I run several versions of the following regression:

$$\eta_{ciot} = \beta_0 + \beta_1 t^E_{ciot} + \beta_2 t^S_{ciot} + \beta_3 \log y^E_{ciot} + \beta_4 \log \widetilde{y}^S_{ciot} + \beta_5 X_{ciot} + \beta_6 FE + \varepsilon_{ciot}.$$

Table 2 summarizes the baseline results. The dependent variable in each column is the fraction of self-employed in the country-industry-occupation-year cell, and the main explanatory variables are the log of the average reported labor income for the employees ($\log y^E$) and for the self-employed ($\log \widetilde{y}^S$), and the average tax rates that these income levels are subject to (t^E and t^S respectively). The second and the fourth columns also include the square of the log of the average earnings. All columns control for demographic characteristics of the cell in X_{ciot} (average age, education, fraction of women, fraction of married, number of children), as well as country-year fixed effects. The columns differ in whether and how they control for occupations and industries. The first two columns include both occupation and industry fixed effects, the third and fourth columns include occupation and industry fixed effects and their interaction, while the fifth column does not include neither occupation, nor industry fixed effects, the sixth column only includes industry fixed effects, and the last column only includes occupation fixed effects.

The main result of this Table is that out of the main regressors, only the coefficient on the tax rate faced as an employee is significantly different from zero, and has a value of around 0.9, which is robust across different specifications which control for both occupation and industry. None of the other explanatory variables are statistically significant, but they all have the expected sign. This confirms the hypothesis of the model which allows for tax evasion by the self-employed. The tax rate faced as an employee has a significant positive correlation with the probability of self-employment, while the tax rate faced as a self-employed has a near zero negative point estimate. This suggests two things: first that a larger tax rate faced as an employee pushes individuals towards self-employment, and second – since the point estimate on t^S is near zero – this is partly due to the benefits of income tax evasion, because individuals expect to underreport their earnings when they are self-employed.

The last three columns show the importance of including industry and to a lesser

Table 2: Cell level regression results

DV: share SE	1	2	3	4	5	6	7
$t^E(y^E)$	0.887 ***	0.870 ***	0.901 ***	0.885 ***	0.164	0.822***	0.064
	(0.226)	(0.222)	(0.219)	(0.214)	(0.425)	(0.158)	(0.230)
$t^S(\widetilde{y}^S)$	-0.035	-0.031	-0.036	-0.032	0.016	-0.059	-0.013
	(0.081)	(0.074)	(0.052)	(0.048)	(0.073)	(0.105)	(0.045)
$\log y^E$	-0.233	-0.073	0.030	-0.012	-0.013	0.064*	-0.119**
	(0.059)	(0.144)	(0.067)	(0.143)	(0.073)	(0.031)	(0.053)
$(\log y^E)^2$		0.002		0.002			
		(0.006)		(0.005)			
$\log \widetilde{y}^S$	0.007	-0.032	0.004	-0.024	0.013	0.014	0.009
	(0.009)	(0.046)	(0.006)	(0.036)	(0.016)	(0.012)	(0.008)
$(\log \widetilde{y}^S)^2$		0.002		0.001			
		(0.002)		(0.002)			
Controls	X	X	X	X	X	X	X
cntry x time FE	X	X	X	X	X	X	X
occupation FE	X	X					X
industry FE	X	X				X	
occ x ind FE			X	X			
R-squared	0.665	0.666	0.803	0.803	0.179	0.606	0.446
Observations	976	976	976	976	976	976	976

OLS regressions, with robust standard errors in brackets. Dependent variable: share of self-employed in the given country-occupation-industry-wave cell among working age, employed population, authors own calculations from LIS Wave 5 and Wave 6. Independent variables: average tax rate at the average earnings of employed/self-employed and log average income (and its square) of employed/self-employed within a cell. Controls: cell average of age, education, marital status, gender, number of children. Countries: Czech Republic, Germany, Finland, Hungary, Ireland, Luxembourg, Spain, UK, US. Occupations: 9 ISCO categories, industries: 9 categories see section 3.1. Average tax rates for each income level: authors own calculations based on country tax codes from OECD.

extent occupation fixed effects. Excluding both industry and occupation fixed effects all of the main explanatory variables lose significance, and the explanatory power of the model decreases significantly. Including occupation fixed effects improves the R-squared quite a bit, and all explanatory variables take the expected sign, but only employee income becomes significant. It seems that it is enough to control for industry to get a precise estimate of the coefficient on the employee tax rates. However, employee income is also significant, but with the opposite sign then expected, and the R-squared of the model is lower than when controlling for both occupation and industry, and especially relative to when controlling for their interaction.

In Table 3 I show a different set of regressions. The dependent variable is a self-

Table 3: Individual level regression results

DV: dummy SE	1	2	3	4	5	6	7
$t^E(y^E)$	1.004 ***	1.006 ***	1.039 ***	1.039 ***	0.215	0.727**	0.691*
	(0.314)	(0.316)	(0.216)	(0.216)	(0.418)	(0.304)	(0.383)
$t^S(\widetilde{y}^S)$	-0.037	-0.033	-0.054	-0.048	0.006	-0.103	0.059
	(0.079)	(0.060)	(0.040)	(0.031)	(0.083)	(0.094)	(0.073)
$\log y^E$	-0.031	-0.066	-0.009	-0.069	0.010	0.022	-0.068
	(0.054)	(0.163)	(0.048)	(0.128)	(0.534)	(0.034)	(0.061)
$(\log y^E)^2$		0.002		0.003			
		(0.008)		(0.006)			
$\log \widetilde{y}^S$	0.005	-0.069	0.006	-0.041	0.008	0.005	0.009
	(0.009)	(0.049)	(0.006)	(0.035)	(0.015)	(0.011)	(0.011)
$(\log \widetilde{y}^S)^2$		0.004		0.002			
		(0.002)		(0.002)			
Controls	X	X	X	X	X	X	X
cntry x time FE	X	X	X	X	X	X	X
occupation FE	X	X					X
industry FE	X	X				X	
occ x ind FE			X	X			
R-squared	0.122	0.123	0.153	0.153	0.032	0.111	0.066
Observations	277024	277024	277024	277024	277024	277024	277024

OLS regressions, with robust standard errors in brackets, clustered at the cell level. Dependent variable: self-employment dummy. Independent variables: average tax rate at the average earnings of employed/self-employed and log average income (and its square) of employed/self-employed within the country-industry-occupation-wave cell. Controls: age, education, marital status, gender, number of children. Data, countries, occupation and industry categories as in Table 2.

employment indicator, which takes the value 1 if the individual is self-employed, and 0 otherwise. The main independent variables are the same as in Table 2: the average tax rate at the average earning of the employed and self-employed, and the log (and its square) of the average income of the employed and self-employed in the cell. I control for the individual's age, education, gender and number of children, rather than their cell average. I also use the same set of fixed effects as in Table 2.

The results from this Table confirm the cell level regression results: the coefficient on the tax rate faced by the average employee is around 1, it is highly statistically significant, and its value is robust across specifications, while none of the other coefficients are. The coefficient on the self-employed tax rate is close to zero. This again supports the predictions of the model where only the self-employed are able to evade some of their income tax. The last three columns again demonstrate the importance of including both occupation and industry fixed effects. When neither is included, none of the explanatory variables are significant, while when including either industry or occupation fixed effect the coefficient on the employee tax rate becomes significant, albeit smaller than when including both and/or their interaction.

The main result from these regressions is that the magnitude of the coefficients on the two tax rates is very different: the employee tax rate has a large positive coefficient, while the self-employed tax rate has a near zero negative coefficient. These two results – given the simple model of the self-employment decision – jointly suggest that individuals are more likely to become self-employed if taxes on employees are higher, and this is partly driven by the possibility of tax evasion when self-employed. It is worth to note that the model does not suggest a near zero coefficient on the tax rate on self-employed income, it merely suggests a negative coefficient of smaller magnitude. The near zero coefficient could be rationalized by a behavior where the self-employed always report an income level that leads to the payment of a fixed amount of taxes. This behavior could be optimal in avoiding tax audits if those were triggered by different amount of tax payments in consecutive years. In such a case the tax rate would not affect their disposable income, and thus have no impact on the decision whether to become self-employed or not.

These regressions show that there is a strong positive correlation between the tax rate faced as an average employee and the self-employment rate (or the self-employment indicator) in a given occupation–industry–country cell. Since the tax scheme is not exogenous, this cannot be treated as a causal relation. In particular there are two issues: reversed causality and omitted variables. It seems unlikely that there is reversed causality. Reversed causality would imply that a high incidence of self-employment in a given occupation-industry cell (relative to the cross-country average in the given occupation-industry, and relative to the country-wave average) leads the government to increase taxes for the (relatively few) employed workers in this cell, but not for the self-employed. Such a policy does not sound very plausible, and it is also hard to implement, as the tax rates depend on the income, and not on the industry and occupation of the individual. Another possibility is that there is an omitted variable that leads to high self-employment and high taxes on the employees (but not on the self-employed). While it is not possible to exclude this possibility the fact that I control for country-wave and occupation-industry fixed effects greatly limits its scope.

5 Conclusion

Using a simple model I showed that if evasion is possible only for the self-employed, then the self-employment decision is less responsive to the tax rate on self-employed income than to the tax rate on income from employment. I test this prediction using cross country household level data from the Luxembourg Income Studies. In the empirical specification – driven by patterns I document in the data – I use cross-country variation in self-employment and tax rates across occupations and industries. I argue that controlling for occupations and industries is important because there are systematic differences in self-employment rates not only across industries, but across occupations, which are common across countries. The regressions robustly confirm the predictions of the model. When taxes on income from dependent employment are high, more people choose self-employment, but people do not respond symmetrically to high taxes on self-employed income. This is evidence that there is adjustment on the extensive margin of tax evasion to tax rates, by people choosing their status in employment.

This implies that when analyzing the welfare costs of taxation, the extensive margin of adjustment should be taken into account. The literature thus far has neglected the analysis of this margin, and instead has focused on the role of the type of costs tax evasion entails. It has been shown that if tax evasion has a transfer cost as well as a resource cost, then the welfare costs are a weighted average of the elasticity of both labor income and taxable income, and empirically these two elasticities are quite different. My results suggest that these models should be further enriched with a choice on the extensive margin between employment and self-employment in order to calculate the welfare costs of taxation.

References

Zoltan J Acs, David B Audretsch, and David S Evans. Why does the self-employment rate vary across countries and over time? CEPR Discussion Papers 871, C.E.P.R. Discussion Papers, January 1994.

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

James Andreoni, Brian Erard, and Jonathan Feinstein. Tax compliance. *Journal of Economic Literature*, 36(2):818–860, 1998.

Mikael Apel. An expenditure-based estimate of tax evasion in Sweden. Working papers, Uppsala - Working Paper Series, 1994.

Nikolaos Artavanis, Adair Morse, and Margarita Tsoutsoura. Measuring income tax evasion using bank credit: Evidence from Greece *. *The Quarterly Journal of Economics*, 131(2):739–798, 2016.

Paul Baker. Taxpayer compliance of the self-employed: estimates from household spending data. Oct 1993.

David G. Blanchflower. Self-employment in OECD countries. *Labour Economics*, 7(5): 471 – 505, 2000.

David M. Blau. A time-series analysis of self-employment in the United States. *Journal of Political Economy*, 95(3):pp. 445–467, 1987.

Kim M Bloomquist. Tax evasion, income inequality and opportunity costs of compliance. In *Proceedings of the National Tax Association, 96th Annual Conference, Chicago*. Citeseer, 2003.

Donald Bruce. Effects of the United States tax system on transitions into self-employment. *Labour Economics*, 7(5):545–574, September 2000.

Raj Chetty. Is the taxable income elasticity sufficient to calculate deadweight loss? the implications of evasion and avoidance. *American Economic Journal: Economic Policy*, 1(2):pp. 31–52, 2009.

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

Luxembourg Income Study (LIS) Database. (Australia, Austria, Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Luxembourg, Netherlands, Norway, Poland, Russia, Slovenia, Spain, Sweden, Switzerland, UK, US; December 2011 to November 2017). URL `http://www.lisdatacenter.org`.

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

Martin Feldstein. Tax avoidance and the deadweight loss of the income tax. *The Review of Economics and Statistics*, 81(4):pp. 674–680, 1999.

Yuriy Gorodnichenko, Jorge Martinez-Vazquez, and Klara Sabirianova Peter. Myth and reality of flat tax reform: Micro estimates of tax evasion response and welfare effects in Russia. *Journal of Political Economy*, 117(3):pp. 504–554, 2009.

James J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1): 153–161, 1979.

Erik Hurst, Geng Li, and Benjamin Pugsley. Are household surveys like tax forms? evidence from income underreporting of the self-employed. *The Review of Economics and Statistics*, 96(1):19–33, 2014.

Edvard Johansson. An estimate of self employment income underreporting in Finland. *Nordic Journal of Political Economy*, 31(1):99–109, 2005.

Bonggeun Kim, John Gibson, and Chul Chung. Using panel data to exactly estimate income under-reporting by the self employed. Labor Economics Working Papers 22996, East Asian Bureau of Economic Research, January 2009.

Henrik Jacobsen Kleven, Martin B. Knudsen, Claus Thustrup Kreiner, Sren Pedersen, and Emmanuel Saez. Unwilling or unable to cheat? evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3):651–692, 2011.

JamesE. Long. Income taxation and the allocation of market labor. *Journal of Labor Research*, 3(3):259–276, 1982.

OECD. Taxing wages 2000. *OECD Publishing*, 2000.

OECD. Taxing wages 2001. *OECD Publishing*, 2001.

OECD. Taxing wages 2002. *OECD Publishing*, 2002.

OECD. Taxing wages 2004. *OECD Publishing*, 2004.

OECD. Taxing wages 2005. *OECD Publishing*, 2005.

Simon C. Parker. A time series model of self-employment under uncertainty. *Economica*, 63(251):pp. 459–475, 1996.

Simon C. Parker. Does tax evasion affect occupational choice? *Oxford Bulletin of Economics and Statistics*, 65(3):379–394, 2003.

Simon C. Parker and Martin T. Robson. Explaining international variations in self-employment: Evidence from a panel of OECD countries. *Southern Economic Journal*, 71(2):pp. 287–301, 2004.

Christopher A. Pissarides and Guglielmo Weber. An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, 39(1):17 – 32, 1989.

Martin T. Robson and Colin Wren. Marginal and average tax rates and the incentive for self-employment. *Southern Economic Journal*, 65(4):pp. 757–773, 1999.

Agnar Sandmo. The theory of tax evasion: A retrospective view. *National Tax Journal*, 58(4):643–63, 2005.

Herb J Schuetze. Taxes, economic conditions and recent trends in male self-employment: a Canada - US comparison. *Labour Economics*, 7(5):507 – 544, 2000.

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

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

Roberto Torrini. Cross-country differences in self-employment rates: the role of institutions. *Labour Economics*, 12(5):661 – 683, 2005.

Shlomo Yitzhaki. Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 3(2):201–202, 1974.

A Appendix

A.1 Descriptive analysis

Table 4 and 5 complement Figure 2 from the main text. These two tables show the average cross-country patterns of self-employment rates and employment shares of different industries and of different occupations. Even though the general cross-industry (cross-occupation) patterns are similar across countries, there are significant differences across countries as can be seen from the standard deviations.

Table 4: Self-employment rates and employment shares across industries

	SE rate	Emp share
Agriculture, forestry and fishing	0.49(0.23)	0.04(0.03)
Mining and quarrying; manufacturing; utilities	0.07(0.05)	0.20(0.07)
Construction	0.23(0.11)	0.08(0.02)
Wholesale and retail trade, repair; hotels and restaurants	0.18(0.12)	0.18(0.03)
Transport, storage and communication	0.12(0.06)	0.06(0.01)
Financial intermediation	0.09(0.07)	0.04(0.02)
Real estate, renting and business activities	0.22(0.11)	0.09(0.03)
Public admin; education; health and social work	0.04(0.01)	0.25(0.05)
Community, social/personal serv; activities of households	0.18(0.07)	0.07(0.02)

Average self-employment rates and employment shares (and their standard deviation in brackets) in different industries calculated for Australia, Czech Republic, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Luxembourg, Slovenia, Spain, Sweden, UK, US. Author's own calculations from Wave 6 of LIS. Definition of self-employed as in the main text.

Table 5: Self-employment rates and employment shares across occupations

	SE rate	Employment share
Managers	0.37(0.20)	0.09(0.04)
Professionals	0.12(0.05)	0.14(0.03)
Technicians and associate professionals	0.08(0.04)	0.15(0.05)
Clerical support workers	0.02(0.02)	0.11(0.04)
Service and sales workers	0.12(0.08)	0.13 (0.02)
Skilled agricultural, forestry and fishery workers	0.62(0.28)	0.05(0.05)
Craft and related trades workers	0.14(0.10)	0.14(0.04)
Plant and machine operators, and assemblers	0.08(0.06)	0.09(0.04)
Elementary occupations	0.05(0.03)	0.09(0.03)

Average self-employment rates and employment shares (and their standard deviation in brackets) in different occupations calculated for Austria, Czech Republic, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Luxembourg, Poland, Slovenia, Spain, UK, US. Author's own calculations from Wave 6 of LIS. Definition of self-employed as in the main text.

A.2 Self-employment decision

In the main body of the paper I assume that the fraction of income reported, κ, is exogenously given. In general, however, κ is chosen to maximize the expected utility from net income, and therefore can depend on the entire tax schedule that the self-employed face, $\tau^S(\cdot)$:

$$U_S = \max_{\kappa}(1-p(\kappa))\log(y_{jS} - \tau^S(\kappa y_{jS})) + p(\kappa)\log(y_{jS} - \tau^S(y_{jS}) - z(y_{jS}, \kappa, t^S)) - g(\kappa).$$

In the above equation $p(\kappa)$ is the probability of getting caught for underreporting κ fraction of income, $g(\kappa)$ is the utility cost of sheltering this fraction, and $z(y_{jS}, \kappa, t^S)$ is the penalty one has to pay when caught evading, which can depend on the level of income, the fraction of income evaded, and the average tax rate schedule t^S. Note that this penalty is on top of paying the actual taxes due, $\tau^S(y_{jS})$. However, if κ is indeed chosen optimally, then $\partial U_S/\partial\kappa = 0$. This implies that for the expected utility difference from self-employment and employment, only the direct effect of taxes matter, the one working through κ does not. Therefore, instead of using $\tau^S(\kappa y_{jS})$ for the total post-evasion tax payment, one can use $\kappa y_{jS} t^S(\kappa y_{jS})$, where $t^S(\cdot)$ denotes the average tax rate as a function of declared income. For the utility comparison one can work with the following, simplified version, where κ^* denotes optimal income reporting:

$$U_S = (1-p(\kappa^*))\log(y_{jS}(1-\kappa^* t^S(\kappa^* y_{jS}))) + p(\kappa^*)\log(y_{jS}(1-t^S(y_{jS})) - z(y_{jS}, \kappa^*, t_S)) - g(\kappa^*).$$

The effect of a change in average tax rate t^S (at every income level) on the maximized utility is:

$$\frac{\partial U_S}{\partial t^S} = -\frac{(1-p(\kappa^*))\kappa^*}{1-\kappa^* t^S(\kappa^* y_{jS})} - \frac{p(\kappa^*)\left(y_{jS} + \frac{\partial z}{\partial t^S}\right)}{(y_{jS}(1-t^S(y_{jS})) - z(y_{jS}, \kappa^*, t_S))}.$$

The first term in the above sum is very similar to the one derived in the main text, it just has a multiplier $1-p(\kappa^*) < 1$, and is thus smaller. The second term is also negative, therefore increasing the magnitude of the effect of the tax rate on maximized utility. It is important to note that this term is very close to zero, as $p(\kappa^*)$, the probability of getting caught at the optimal level of underreporting κ^* is likely to be very small. This implies that even if κ^* is endogenous, and people can get caught and can be fined if they evade taxes, the impact of taxes on the maximized utility of the self-employed is likely to be smaller than for the employees.