

Taxation and Self-Employment

Zsofia Barany

▶ **To cite this version:**

Zsofia Barany. Taxation and Self-Employment. 2018. hal-03393144

HAL Id: hal-03393144
https://sciencespo.hal.science/hal-03393144

Preprint submitted on 21 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

SciencesPo
DEPARTMENT OF ECONOMICS

Discussion Paper

TAXATION AND SELF-EMPLOYMENT

Zsófia Bárány

SCIENCES PO ECONOMICS DISCUSSION PAPER

No. 2018-02

Taxation and self-employment*

Zsófia L. Bárány†

March 2018

Abstract

In this paper I study the relation between self-employment and the tax rates on wages and on self-employment income. Using variation in the statutory tax rates across countries, industries, and occupations, I find that while the share of self-employed is strongly positively correlated with the tax rate on wage income, it is weakly negatively correlated with the tax rate on self-employed income. The asymmetry between the effects of the tax rates suggests that those who choose self-employment partly do so in order to evade 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.

Keywords: tax rates, self-employment, tax evasion

*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 and CEPR. 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.[2] 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 paper I study the relation between the share of self-employed and the tax rates on salaried and on self-employment income. My paper contributes to the literature in a number of ways. First, I document a novel pattern in self-employment rates across occupations. Second, I create tax schedules for the employed and the self-employed for each country in my sample, which allows me to analyze the effects of taxes at a more micro level than previous studies. Finally, I explore the relation between self-employment and tax rates calculated at the industry–occupation–country level. I find that while the share of self-employed in an industry–occupation is strongly positively correlated with the salaried average tax rate, it is either weakly negatively or not correlated with the self-employed average tax rate. The asymmetry between the correlation with the employed and the self-employed tax rate suggests that those who choose self-employment, partly do so to take advantage of the technology it offers in evading taxes.

Using cross-country household level data from the Luxembourg Income Studies (LIS) I document an important novel pattern of self-employment across countries. I show that there are systematic differences across occupations in the propensity of self-employment: for example among *managers* the average self-employment rate in

[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] See for example Kleven et al. (2011) and Artavanis, Morse, and Tsoutsoura (2016), this literature is discussed in further detail below.

my sample is 37 percent, among *clerical support workers* this fraction is only 2 percent. While it is known that the self-employment rate varies across different industries of the economy,[3] to my knowledge the systematic differences across occupations have received little attention. However, even after controlling for these systematic differences between occupations and industries there is a large variation in self-employment rates across countries. 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 schedules across countries.

I build detailed income tax schedules, including employee and self-employed social security schedules for each country in my sample. I show that the average tax rate faced by employees and by self-employed vary substantially within countries both across occupations and across industries. For example in Finland the average of the average tax rate on salaried income is 31.8%, across occupations this varies between 26.3 and 41.8%, and across industries between 29.4 and 35.3%. This suggests that even in a cross-country setting using a single tax rate to capture the effect of taxes is probably insufficient.

In a simple model I show the intuitive result that if only the self-employed can evade income tax, then the employee and the self-employed tax rate have an asymmetric effect on the propensity of self-employment. I test the predictions of the model at the industry–occupation level using the LIS database combined with the income tax and social security functions I built 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 much smaller, less precisely estimated negative effect. These results provide supportive evidence of the predictions of the model with tax evasion. The large difference between the magni-

[3]For example Acs, Audretsch, and Evans (1994), Blanchflower (2000), Parker and Robson (2004) document the differences in self-employment rates between agriculture vs non-agriculture, manufacturing vs services, public vs private sectors, and Torrini (2005) documents that there are systematic differences across finer industry categories as well.

tude 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.

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 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. Most of the previous literature on the effects of taxes on self-employment estimate reduced form equations, where they control for a single tax rate faced by the employed – among other variables – and tend to find a positive correlation.[6]

My estimation strategy is closer to that of Bruce (2000) and Parker (2003), in that it relies on the gains from working as a self-employed rather than as an employee. I

[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.

[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 et al. (2016) uses data on loans and income from a large Greek bank.

[6] See for example Long (1982), Blau (1987), Parker (1996), Robson and Wren (1999), Schuetze (2000), Parker and Robson (2004). The additional variables typically used are country level variables such as the minimum wage, social security coverage, unemployment rate and benefit, interest rate, total factor productivity, share of manufacturing, private services and the public sector, or demographic characteristics such as age, gender, education.

explicitly take into account the possibility of evasion by the self-employed, and in line with the model's predictions, 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 the type of data, my estimation strategy and my findings are different from those in Bruce (2000) and Parker (2003). While both of these papers use individual earnings and tax rates in the employed and self-employed status in a single country, I use employed and self-employed earnings and tax rates at the industry–occupation level, and I use not only within, but between country variation in the analysis. In terms of methodology, 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 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 presents the data and the descriptive analysis, section 3 describes the construction of the tax functions, section 4 shows the results on tax rates and self-employment, and section 5 concludes.

2 Self-employment, occupations and industries

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. The benefit of using this data is that it contains detailed information on income for both employed and self-employed individuals. 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. I restrict the sample to 18-65 year old working individuals, and 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. This Table shows that self-employment rates vary substantially across countries;[7] 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.

A new finding of this paper, demonstrated in the top panel of Figure 1 is that there is a large and systematic variation in self-employment rates across occupation

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

Table 1: Self-employment rates across countries

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

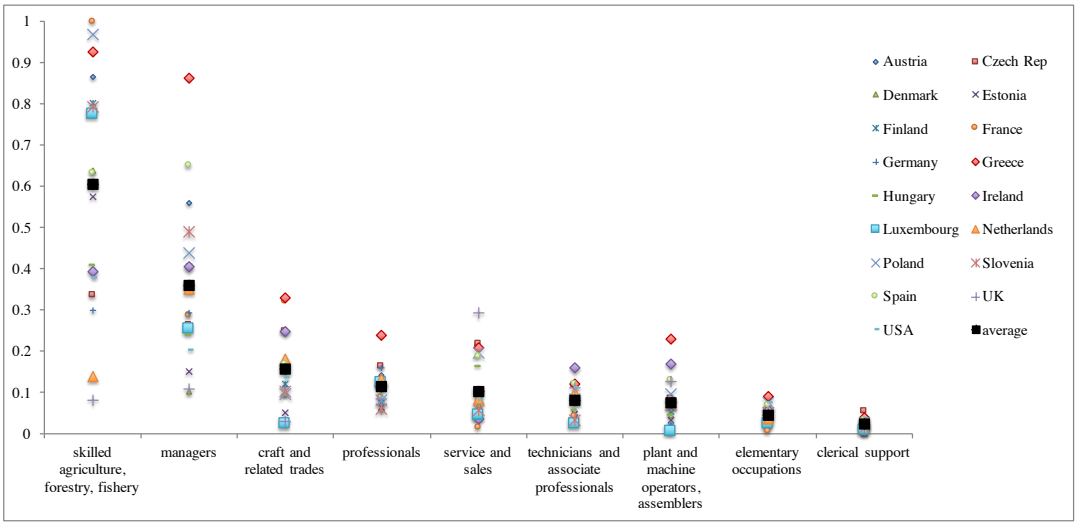
Author's own calculations from Wave 6 of the Luxembourg Income Study Database (LIS). Share of self-employed, employer, own-account worker and contributing family worker, in the working age (18-65) total employed population in the first column, and among the non-agricultural workers in the second column.

groups.[8] The bottom panel of Figure 1 confirms the findings of Torrini (2005) that the self-employment rates across industries differ significantly as well.[9] The cross country average of occupation or industry self-employment rates are shown with the black squares, the other symbols show the self-employment rate for each occupation and each industry in various countries.[10] This figure demonstrates that there is a systematic difference in the propensity of self-employment both across occupations and across industries which is common to countries. Perhaps not surprisingly the highest self-employment rate tends to be in the *skilled agricultural, forestry and fishery* occupations and in the *agriculture, forestry and fishing* industry. The lowest self-employment

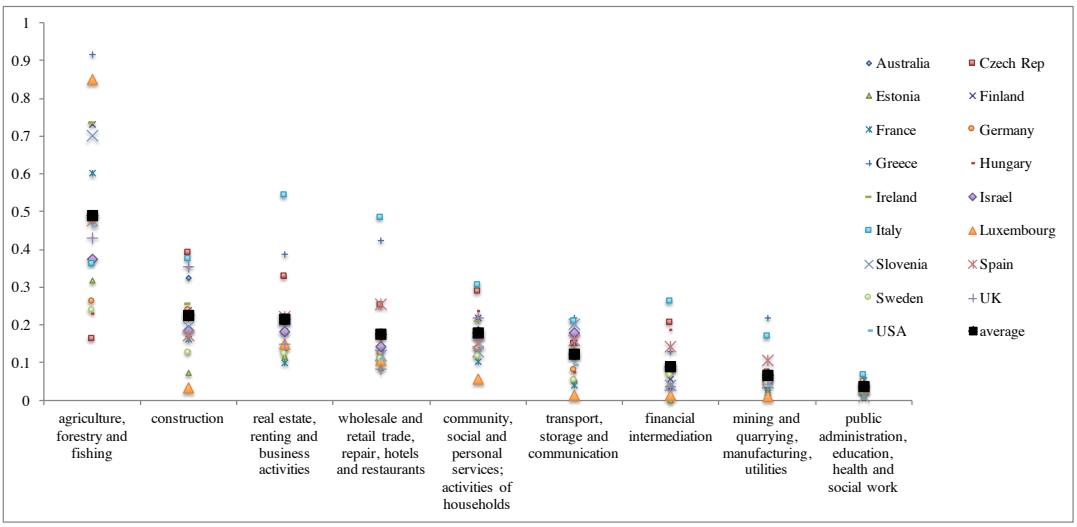
[8]The occupation of the worker is one of the 10-category International Standard Classification of Occupations (ISCO) recode: (1) managers, (2) professionals, (3) technicians and associate professionals, (4) clerical support workers, (5) service and sales workers, (6) skilled agricultural, forestry and fishery workers, (7) craft and related trades workers, (8) plant and machine operators, and assemblers, (9) elementary occupations, (10) armed forces.

[9]The industry of a worker is one 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.

[10]See Table 6 and Table 5 in the appendix for the cross-country average self-employment rates and employment shares, as well as their standard deviation by industry and by occupation.



(a) Occupation



(b) Industry

Figure 1: Self-employment rates

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

rate tends to be in the occupation of *clerical support workers* and in the industry of *public administration, education, health and social work*. 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 occupations or across industries. However, as shown in the appendix this is not the case. This can be understood by observing in

Figure 1 that the actual occupation/industry self-employment rates in specific countries vary quite a bit around the cross-country averages, and this variation also seems to be systematic. For example in Figure 1a we see that Luxembourg, which has a very low overall self-employment rate at 7.1%, tends to have a low self-employment rate in all occupations, except in the occupation of *skilled agricultural, forestry and fishery workers*. The other panel 1b 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. These findings suggest on the one hand that a crucial determinant in the decision to become self-employed is the occupation and industry of the individual. On the other hand there are large cross-country differences in self-employment rates within industries and occupations.

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. For the empirical analysis I retain those countries where occupation and industry codes are available in both Wave 5 and 6 in the LIS data. These countries are the Czech Republic, Finland, Germany, Hungary, Ireland, Luxembourg, Spain, the United Kingdom and the United States of America. Table 2 shows summary statistics for both waves and all countries considered. This table shows that there is a difference in the distribution of employed and self-employed individuals across occupations and across industries. Among the self-employed there tend to be more managers and agricultural workers and less clerical support workers than among the employees. In terms of industries the self-employed are more likely to work in agriculture, construction, real estate, renting and business activities, and are less likely to work in mining and quarrying and public administration than employees. This table also confirms previous findings that the self-employed are a little bit older, a little bit less educated, less likely to be females, are slightly more likely to be married, and tend to have slightly more children.[11]

[11] See for example Blau (1987), Schuetze (2000), Wen and Gordon (2014) among others.

Table 2: Summary statistics for employed and self-employed

	Employed		Self-employed	
	Mean	SD	Mean	SD
Age	41.54	10.46	45.16	9.83
Education	2.23	0.79	2.17	0.83
Married	0.71	0.46	0.79	0.41
Female	0.50	0.50	0.34	0.47
Number of children	1.17	1.16	1.29	1.27
Occupations				
Managers	0.13	0.33	0.24	0.43
Professionals	0.18	0.38	0.13	0.33
Technicians & associate professionals	0.15	0.35	0.12	0.32
Clerical support workers	0.13	0.34	0.03	0.18
Service & sales workers	0.12	0.32	0.11	0.31
Skilled agricultural, forestry & fishery workers	0.01	0.12	0.13	0.34
Craft & related trades workers	0.12	0.32	0.15	0.35
Plant & machine operators & assemblers	0.09	0.28	0.05	0.22
Elementary occupations	0.08	0.27	0.05	0.21
Industries				
Agricultural, forestry & fishing	0.01	0.11	0.15	0.36
Mining & quarrying, manufacturing, utilities	0.20	0.40	0.08	0.27
Construction	0.06	0.24	0.15	0.36
Wholesale & retails trade, repair, hotels & restaurants	0.16	0.36	0.19	0.39
Transport, storage & communications	0.06	0.24	0.05	0.23
Financial intermediation	0.05	0.22	0.03	0.16
Real estate, renting & business activities	0.10	0.30	0.17	0.38
Public admin., education, health & social work	0.30	0.46	0.09	0.29
Community, social & pers. serv., activities of the hhs	0.05	0.23	0.08	0.27
Number of observations	246,609		39,475	

Author's own calculations from Wave 5 and 6 of LIS for the countries where both occupation and industry variables are available in both waves. The values for occupations and industries show the fraction of employed and self-employed working in each category.

3 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.[12] 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.[13]

Using the information contained in the Taxing Wages publications I carefully code the standard tax schedules, and 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.[14] This allows me to calculate a separate income tax and social security contribution function for the employees and the self-employed.[15] Based on these functions the average income tax rate, excluding or including the social security contributions at any gross income level and in both employment statuses can be calculated.[16] To my knowledge, this is the first paper that creates these functions and uses tax rates corresponding to the actual income at the industry–occupation level in a cross-country setting.

Figure 2 shows the average tax rates excluding (solid line) and including (dashed line) social security contributions as a function of annual gross labor income for employees in the top row and for self-employed in the bottom row for three selected coun-

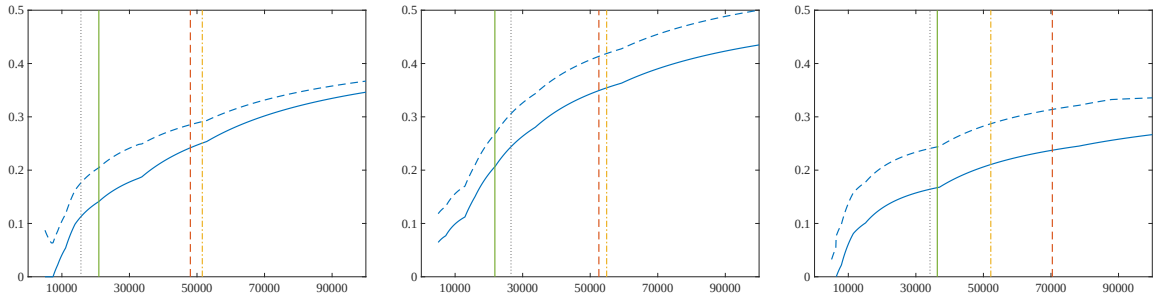
[12]Before 1996 it was called The Tax/Benefit Position of Production Workers, between 1996-1998 it was called The Tax/Benefit Position of Employees. From 1998 it is published as Taxing Wages. In creating the tax functions I use the following releases: 2000, 2001, 2002, 2004, 2005.

[13]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.

[14]I calculate the income tax and social security functions for a single individual without children. I also ignore any within country differences stemming from local taxes.

[15]The income tax schedules (i. e. the rates and bands) are the same in all of these countries for the employed and the self-employed. The only reason for any difference in the income tax functions between employees and self-employed is that allowances might differ due to the difference in social security contributions.

[16]While the marginal income tax rates can also be calculated, it is the average tax rate that matters for net income differences and the self-employment decision, therefore I do not use them in the regression analysis.



(a) Employee, Spain (b) Employee, Finland (c) Employee, USA

(d) Self-employed, Spain (e) Self-employed, Finland (f) Self-employed, USA

Figure 2: Average tax rates as a function of annual gross labor income

This figure plots average tax rates excluding (solid) and including (dashed) social security contributions for employees in the top row and for self-employed in the bottom row as a function of annual income in Spain, Finland, 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 and employment status in four industry–occupation pairs, see text for details.

tries in 2004. It is worth to note that for Finland and the USA the average income tax rate function excluding social security contributions (solid blue lines) is the same for employees and self-employed. For Spain there is a slight difference, as the income tax allowances are based on the social security payments and these differ across employment statuses. The tax rate functions including the social security contributions are substantially different even at the same income level between the employed and the self-employed. The vertical lines in Figure 2 show the average earnings of employed and respectively self-employed workers in the country calculated from Wave 6 of the LIS for four industry–occupation pairs. The average earnings of the employed and of the self-employed in a given industry–occupation cell are calculated from the *personal labor income* variable from the LIS, which records the gross yearly earnings of the individual, before personal income tax and social security contribution deductions.[17] The

[17]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

industry–occupation pairs marked in Figure 2 are managers in real estate, renting and business activities (dashed red), managers in wholesale and retail trade, repair, hotels and restaurants (dashed-dotted yellow), craft and related trades workers in real estate, renting and business activities (solid green), craft and related trades workers in wholesale and retail trade, repair, hotels and restaurants (dotted grey). Therefore tax rates faced by an employee and a self-employed in a given industry–occupation–country cell are different because the incomes and potentially the tax functions are different in the two statuses.

This figure shows the types of variation I use in the empirical analysis. First, within countries there is a large difference in the earnings and the average tax rate across industry–occupation cells, which are not the same across countries. Second, even within country and within occupations (comparing dashed red to dotted-dashed yellow and solid green to dotted grey) or within industries (dashed red to solid green and dashed-dotted yellow to dotted grey) there is quite a bit of variation in income and the corresponding tax rates. Finally, across countries there is variation in the tax rate differentials (between employees and self-employed) within industry–occupation cells.[18]

The substantial variation in average tax rates within countries[19] 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.

4 Tax rates and self-employment

In what follows I first sketch a model to demonstrate that the decision to become self-employed depends in an asymmetric way on the income tax rate faced by the

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). In Figure 2 the average self-employed income is significantly lower in Spain and Finland than the average employed income for each industry–occupation pair, this could suggest underreporting.

[18] In section 4 I use only the average tax rates including social security contributions. The results are similar for average tax rates excluding social security contributions and are available on request.

[19] See Table 7 in the appendix for summary statistics by country.

employed and by the self-employed if the self-employed evade some of their income taxes. Then using the LIS data combined with the tax schedules I test and find supportive evidence for the predictions of the model at the industry–occupation–country–year level.

Denote the exogenous expected income of individual j in employment by y_j^E, and in self-employment by y_j^S. 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. In line with the existing literature and the empirical evidence I assume that while employees report their entire income truthfully, self-employed underreport their income.[20] I further assume that the self-employed report $\kappa \leq 1$ fraction of their income, such that $\widetilde{y}_j^S$ the reported income can be expressed as $\widetilde{y}_j^S = \kappa y_j^S$. For simplicity I assume that the fraction κ is exogenously given, and that people never get caught.[21] Under these assumptions the log utility from income for the employed and the self-employed can be written as:

$$U_E = \log(y_j^E(1 - t^E(y_j^E))) = \log y_j^E + \log(1 - t^E(y_j^E))$$
$$U_S = \log(y_j^S - \kappa y_j^S t^S(\kappa y_j^S)) = \log(y_j^S(1 - \kappa t^S(\widetilde{y}_j^S))) = \log y_j^S + \log(1 - \kappa t^S(\widetilde{y}_j^S)).$$

It is important to note that the utility difference between the two statuses depend on the average tax rate, rather than on the marginal tax rate that the individual faces.[22] The marginal effects of the average tax rates on the utility difference from income are:

$$\frac{\partial(U_S - U_E)}{\partial t^E(y_j^E)} = \frac{1}{1 - t^E(y_j^E)},$$
$$\frac{\partial(U_S - U_E)}{\partial t^S(\widetilde{y}_j^S)} = -\frac{\kappa}{1 - \kappa t^S(\widetilde{y}_j^S)}.$$

[20] 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 also build on the assumption that the employees truthfully report, while the self-employed underreport their income.

[21] In the appendix I relax these assumptions and show that the predictions of the model are very similar.

[22] The marginal tax rates are probably more important for the amount of income under-reported. I discuss the case of endogenous κ in the appendix, and as I show there, due to the optimality of the amount of under-reporting, only the direct effect of taxes matter for the utility difference, and hence it is enough to consider the effect of the average tax rate even in that case.

There are two things to note in the above expressions. First, the effect of the self-employed tax rate on the utility difference is increasing in magnitude in κ. This can be seen in the expression above, as in absolute terms the numerator is an increasing function, while the denominator is a decreasing function of κ. Put differently, the more the self-employed under-report their income, the smaller κ is, and the less important is the self-employed tax rate in the self-employment decision. Second, a lower κ also implies a larger asymmetry between the employed and the self-employed tax rate, if only the self-employed are able to evade taxes. Given that the empirical evidence suggests that $\kappa << 1$, 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.

In what follows, I test the predictions of this simple model at the level of industry–occupation–country–year cells, based on the evidence in section 2, 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. Taking as given an individual's occupation, industry and country, I ask how the tax rate on the expected (reported) income from employment and from self-employment affect the decision to become self-employed. Besides the expected differences in the utility from disposable income other factors can impact the decision to work as self-employed. This decision likely depends on non-pecuniary factors such as age, education, gender, marital status, number of children, and based on the evidence in the section 2 also on industry, occupation and country of work (all captured in Z_j), and an idiosyncratic utility from working as a self-employed ($\nu_j \sim F(\cdot)$ iid with mean zero). Using the formulation allowing for tax evasion by the self-employed, the selection equation can be written as:

$$I_j^* = 1 \quad \text{if} \quad \gamma_1(\log y_j^S + \log(1 - \kappa t^S(\widetilde{y}_j^S)) - \log y_j^E - \log(1 - t^E(y_j^E))) + \gamma_2 Z_j + \nu_j \geq 0, \quad (1)$$

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 the self-employed, is that I observe the reported income ($\widetilde{y}_j^S$), and not the true income (y_j^S). 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.[23] It is important to note from (1) that using the tax rate differential between the two employment statuses, $t^E(y_j^E) - t^S(\widetilde{y}_j^S)$, as an independent variable in case $\kappa < 1$ is not correct and would lead to biased estimates.[24] 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 ($\widetilde{y}^S$) within the industry–occupation–country–year cell as an approximation for the expected (reported) earnings.[25] Given the calculated average reported 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-year and employment status described in section 3.

The fraction of self-employed in a given cell can be approximated as:

$$\eta_{SE} = 1 - F\left(-\gamma_1[\log \widetilde{y}^S + \log \kappa - \log(1 - \kappa t^S(\widetilde{y}^S)) + \log y^E + \log(1 - t^E(y^E))] - \gamma_2 Z\right), \tag{2}$$

where Z 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 $\gamma_1 f(\cdot)/(1 - t^E(y^E))$, while it is decreasing in $t^S(\widetilde{y}^S)$ at rate $\gamma_1 \kappa f(\cdot)/(1 - \kappa t^S(\widetilde{y}^S))$.

[23] 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.

[24] Even if one used the approximation that $\log(1 - t) \approx -t$ implying that $\log(1 - \kappa t^S(\widetilde{y}_j^S)) - \log(1 - t^E(y_j^E)) \approx t^E(y^E) - \kappa t^S(\widetilde{y}_j^S)$, thus $t^E(y_j^E) - t^S(\widetilde{y}_j^S)$ is not a good approximation for $\kappa << 1$.

[25] 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. Parker (2003) linearizes the average tax rates on a set of income bands, and assumes that the error correction does not modify the tax rate. Wen and Gordon (2014) run a reduced form probit ignoring taxes to extract the correction terms.

To test for this relation I run several versions of the following regression:[26]

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

Table 3 summarizes the baseline results. The dependent variable in each column is the natural logarithm of 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 ($\log y^E_{ciot}$) and for the self-employed ($\log \widetilde{y}^S_{ciot}$), and the country-year and employment status specific average tax rates that these income levels are subject to ($t^E_{ct}(y^E_{ciot})$ and $t^S_{ct}(\widetilde{y}^S_{ciot})$ respectively). The fourth and the sixth 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 fixed effects. The columns differ in whether and how they control for occupations and industries. The first column does not control for neither industries nor occupations, the second column only includes industry fixed effects. The third and the fourth columns include both occupation and industry fixed effects, while the fifth and the sixth columns include occupation and industry fixed effects and their interaction.

Comparing the first three columns shows the importance of including industry and to a lesser extent occupation fixed effects. Excluding both industry and occupation fixed effects the explanatory power of the model is very low, and neither tax rates have an effect significantly different from zero. The sign of the income effect is as expected, the magnitude is similar across specifications, and the magnitude is similar between the log of the employee and the reported self-employed income (as predicted by the simple model). Including industry fixed effects increases the R^2 of the model significantly, and the effect of the average employee tax rate takes the expected sign, and becomes highly significant, while the effect of the self-employed tax rate is not significant, but takes the expected sign.

[26] This formulation relies on the approximation $\log(1 - t^E(y^E_j)) \approx -t^E(y^E_j)$ and $\log(1 - \kappa t^S(\widetilde{y}^S_j)) \approx -\kappa t^S(\widetilde{y}^S_j)$. An alternative is to use directly $\log(1 - t^E(y^E))$ in the regression and while $\log(1 - \kappa t^S(\widetilde{y}^S))$ can not be used as κ is unknown, $\log(1 - t^S(\widetilde{y}^S))$ can be used as an approximation. I show the results of this alternative regression in Table 8 in the appendix.

Table 3: Cell level regression results

DV: log share SE	1	2	3	4	5	6
$t^E(y^E)$	-0.60	6.89 ***	4.46 ***	4.59 **	4.94 ***	4.65 ***
	(3.62)	(1.77)	(2.06)	(1.68)	(1.37)	(1.45)
$t^S(\widetilde{y}^S)$	-0.56	-1.06	-0.76	-0.68	-0.42	-0.42
	(0.98)	(0.81)	(0.69)	(0.60)	(0.74)	(0.70)
$\log y^E$	-0.23	-0.26 *	-0.26 **	-0.06	-0.12	0.04
	(0.15)	(0.13)	(0.11)	(0.90)	(0.11)	(0.79)
$(\log y^E)^2$				-0.01		0.00
				(0.04)		(0.03)
$\log \widetilde{y}^S$	0.31 **	0.30 **	0.24 **	0.56	0.09	0.42
	(0.15)	(0.12)	(0.10)	(0.50)	(0.10)	(0.45)
$(\log \widetilde{y}^S)^2$				-0.02		-0.02
				(0.02)		(0.02)
Controls	X	X	X	X	X	X
country FE	X	X	X	X	X	X
industry FE		X	X	X	X	X
occupation FE			X	X	X	X
occ x ind FE					X	X
R-squared	0.20	0.54	0.62	0.63	0.75	0.75
Observations	976	976	976	976	976	976

OLS regressions, with robust standard errors in brackets. Dependent variable: natural logarithm of the share of self-employed in the given industry–occupation–country–year 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 2. Average tax rates for each income level and employment status: authors own calculations based on country tax codes from OECD. Significance: *** at 1%, ** at 5% and * at 10%.

Columns (3) to (6) show the main results of this table: out of the main regressors, only the coefficient on the tax rate faced as an employee is significantly different from zero across all specifications, has a value of around 4.5-5, which is robust across different specifications which control for both occupation and industry. This implies that a one percentage point increase in the employee tax rate leads to a 4.5 percent increase in the self-employment rate. The point estimate of the effect of the average self-employed tax rate is smaller in magnitude and negative, although not statistically significant.[27]

[27]Using the alternative specification of $\log(1 - t^E(y^E))$ and $\log(1 - t^S(\widetilde{y}^S))$ rather than the tax rates directly the point estimate of the effect of the self-employed log net of tax is significant, with about a 4-fold difference in the magnitude with the employee log net of tax. These results are shown in Table 8

In column (3) both the log income of employees and self-employed are of the expected sign and similar magnitude, as predicted by the simple model. This is true even when including industry–occupation fixed effects, though both income variables lose significance. When including the quadratic log income terms, all income variables lose significance.

These results confirm the main 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 much smaller 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 smaller – that this is partly due to the benefits of income tax evasion, because individuals expect to underreport their earnings when they are self-employed.

Table 4: Individual level regression results

DV: dummy SE	1	2	3	4
$t^E(y_E)$	0.28	0.80 ***	0.68 ***	0.75 ***
	(0.18)	(0.13)	(0.17)	(0.12)
$t^S(\widetilde{y}_S)$	0.01	-0.14	-0.11 *	-0.10 **
	(0.10)	(0.09)	(0.07)	(0.05)
$\log y_E$	0.00	-0.01	-0.01	-0.01
	(0.02)	(0.01)	(0.01)	(0.01)
$\log \widetilde{y}_S$	0.01	0.01	0.01	0.01
	(0.02)	(0.01)	(0.01)	(0.01)
Controls	X	X	X	X
country FE	X	X	X	X
industry FE		X	X	X
occupation FE			X	X
occ x ind FE				X
Observations	277024	277024	277024	277024

Marginal effects from probit regressions, with robust standard errors in brackets, clustered at the industry–occupation–country–year cell level. Independent variables: average tax rate at the average earnings of employed/self-employed and log average income of employed/self-employed within the industry–occupation–country–year cell. Controls: age (and its square), education (and its square), marital status, gender, number of children. Data, countries, occupation and industry categories as in Table 3. Significance: *** at 1%, ** at 5% and * at 10%.

in the appendix.

In Table 4 I show a different set of regressions. The dependent variable is a self-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 3: the average tax rate at the average earning of the employed and self-employed, and the log of the average income of the employed and self-employed in the cell. I control for the individual's age (and its square), education (and its square), marital status, gender and number of children, rather than their cell average. I also use the same set of fixed effects as in Table 3.

The results from this table confirm the cell level regression results. The first columns again demonstrate the importance of including industry fixed effects and to a lesser extent occupation fixed effects. When neither is included, none of the explanatory variables are significant, while when including at least industry fixed effect the coefficient on the employee tax rate becomes significant. Once at least industry fixed effects are included, the marginal effect of the tax rate faced by the average employee is around 0.7, it is highly statistically significant, and its value is robust across specifications. This implies that a 1 percentage point higher employee tax rate implies a 0.7 percentage point increase in the probability of becoming self-employed conditional on the distribution of the independent variables in the sample. The marginal effect of the self-employed tax rate is negative and has a smaller absolute value (it is around -0.1) and is statistically significant when controlling for both industry and occupation. None of the income variables have a significant marginal effect, and the point estimates are quite small. This suggests that the other controls, namely age, education, gender, the occupation and the industry of the worker are better predictors of their respective employee and self-employed income than the cell averages. It seems nonetheless, that the average tax rate captures the tax incentives well.

The main result from these regressions is that the magnitude of the marginal effect of the two tax rates is very different: the employee tax rate has a large positive, while the self-employed tax rate has a much smaller negative effect, which is not statistically significant. These 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.

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 industry–occupation–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 industry–occupation cell (relative to the cross-country average in the given industry–occupation, and relative to the country average) leads the government to increase taxes for the (relatively few) employed workers in this cell, but not for the self-employed. While in some low income countries the policy of a high tax rate on the narrow tax base of salaried individuals might be used, it is difficult to interpret the findings of this paper in such a way. First, without controlling for industry and occupation fixed effects I do not find evidence of higher taxes being associated with higher self-employment rates. Second, such a policy is also hard to implement, as the tax rates depend on the income, and not on the industry and occupation of the individual.[28] 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 and industry–occupation fixed effects greatly limits its scope.

5 Conclusion

In this paper I analyze the relation between the share of self-employed and the tax rates on salaried and on self-employment income. Using a simple model I show 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

[28] While in some countries there might be industry or occupation specific tax breaks, I did not code these, and therefore the tax rates used in the analysis only depend on the income and employment status of the individual.

document in the data – I use variation in tax rates both within and across countries, 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 suggestive 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? Technical Report 871, CEPR 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. Technical report, 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. Technical Report W93/14, IFS Working Paper, 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):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):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 March 2018). 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):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):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.

James E. 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):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):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):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.

Jean-Francois Wen and Daniel V. Gordon. An empirical model of tax convexity and self-employment. *The Review of Economics and Statistics*, 96(3):471–482, 2014.

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

A Appendix

A.1 Descriptive analysis

Table 5 and 6 complement Figure 1 from the main text. These two tables show the average cross-country patterns of self-employment rates and employment shares of different occupations and of different industries. Even though the general cross-occupation (cross-industry) patterns are similar across countries, there are significant differences across countries as can be seen from the standard deviations.

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.

Table 6: 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.

Below 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:

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

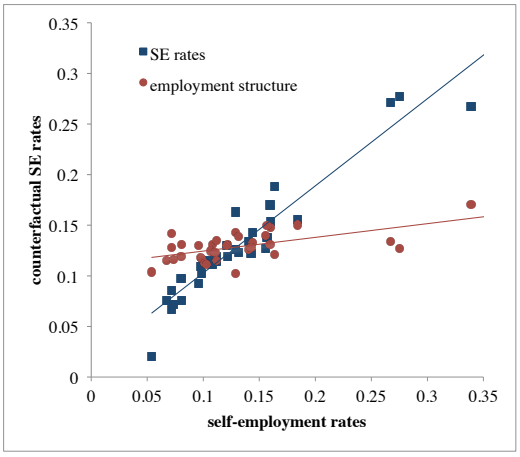
where the aggregation j can be over industries or occupations, η is the overall self-employment rate, λ_j is the employment share in industry (or occupation) j, and η_j is the self-employment rate in industry (or occupation) j. Given this formulation, one can compute two counterfactual overall self-employment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

$$\widehat{\eta}^{SE} = \sum_{j=1}^{J} \overline{\lambda}_j \eta_j, \tag{3}$$

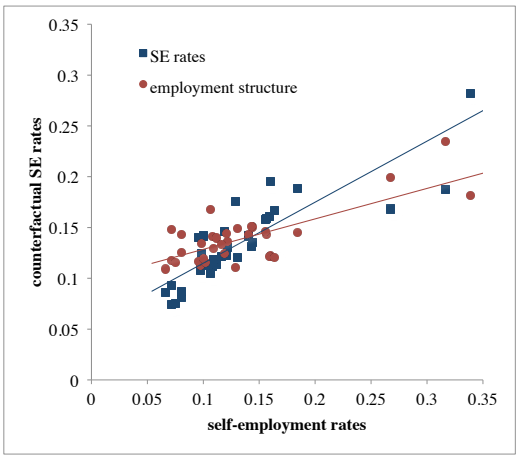
$$\widehat{\eta}^{emp} = \sum_{j=1}^{J} \lambda_j \overline{\eta}_j. \tag{4}$$

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.

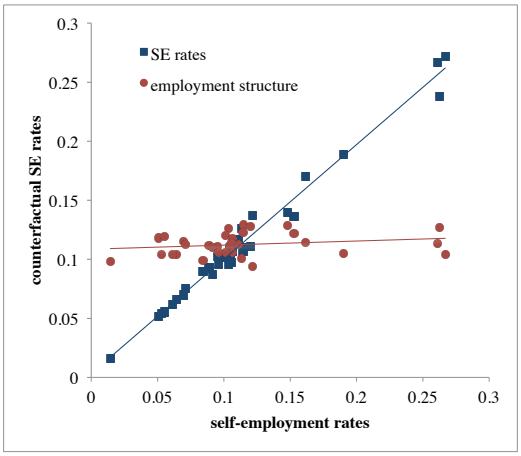
Figure 3 shows the counterfactual self-employment rates based on industries in the left panels, and based on occupations in the right panels. The top row is based on all working individuals, whereas the bottom row is based on those outside of agricultural industries (in the left), and occupations (in 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



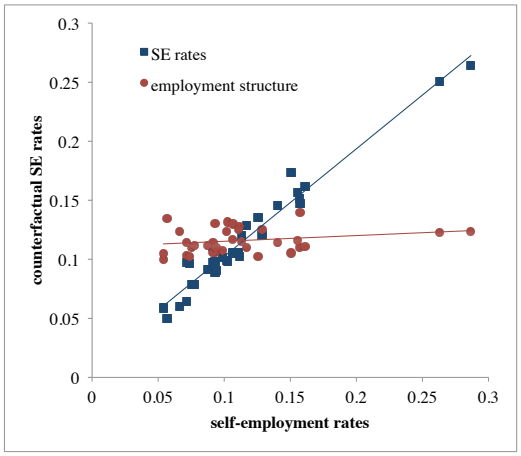
(a) Industry counterfactuals



(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 (4), and the blue squares correspond to $\widehat{\eta}^{SE}$ from (3). 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.

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 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 pro-

vides 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.[29] 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.

A.2 Summary statistics of average tax rates

Table 7 reports the average, the minimum and the maximum of the average tax rate faced by an employee in a given country across the two years used in the analysis (corresponding to Wave 5 and 6 of the LIS) and across all industry–occupation cells.

Table 7: Average and marginal tax rate variations within countries

	average tax rate		
	average	minimum	maximum
Czech Republic	0.23	0.13	0.32
Finland	0.32	0.15	0.46
Germany	0.36	0.21	0.52
Hungary	0.31	0.14	0.46
Ireland	0.17	0.00	0.37
Luxembourg	0.26	0.13	042
Spain	0.21	0.06	0.39
UK	0.22	0.00	0.32
US	0.26	0.17	0.35

[29] 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.

A.3 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, $\tau^S(\cdot)$:

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

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 $z(y_j^S, \kappa, t^S)$ 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, $\tau^S(y_j^S)$. 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 $\tau^S(\kappa y_j^S)$ for the total post-evasion tax payment, one can use $\kappa y_j^S t^S(\kappa y_j^S)$, 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:

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

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

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

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.

A.4 Alternative tax measure: log of net rate

Table 8 is the same as Table 3 in the main text, except instead of $t^E_{ct}(y^E_{ciot})$ and $t^S_{ct}(\widetilde{y}^S_{ciot})$, it uses the natural logarithm of the net of tax rate, $\log(1-t^E_{ct}(y^E_{ciot}))$ and $\log(1-t^S_{ct}(\widetilde{y}^S_{ciot}))$. As mentioned in the main text, the correct independent variables to be used would be $\log(1-t^E_{ct}(y^E_{ciot}))$ and $\log(1-\kappa t^S_{ct}(\widetilde{y}^S_{ciot}))$, but as κ in unknown, this is not possible and an approximation needs to be used. In the main text I rely on the approximation that $\log(1-t^E_{ct}(y^E_{ciot})) \approx -t^E_{ct}(y^E_{ciot})$ and $\log(1-\kappa t^S_{ct}(\widetilde{y}^S_{ciot})) \approx -\kappa t^S_{ct}(\widetilde{y}^S_{ciot})$. Here I show the results for the logarithm of the net of tax rate. Based on the simple model, one would expect the logarithm of the employee net of tax rate to have a negative, while the self-employed net of tax rate to have a positive effect of smaller magnitude. This prediction along with all other results from Table 3 in the main text is verified in Table 8, where also the coefficient on the logarithm of the self-employed net of tax is significant.

Table 8: Cell level regression results: log net of tax

DV: log share SE	1	2	3	4	5	6
$\log(1-t^E(y^E))$	-0.32	-4.88 ***	-3.36 ***	-3.49 **	-3.74 ***	-3.58 ***
	(2.29)	(1.18)	(1.09)	(1.41)	(0.89)	(1.00)
$\log(1-t^S(\widetilde{y}^S))$	0.89 **	1.01 **	0.98 **	0.92 **	0.61	0.55
	(0.47)	(0.42)	(0.38)	(0.44)	(0.43)	(0.55)
$\log y^E$	-0.29 **	-0.28 **	-0.31 **	0.13	-0.16	0.18
	(0.14)	(0.13)	(0.12)	(0.96)	(0.11)	(0.82)
$(\log y^E)^2$				-0.02		-0.01
				(0.04)		(0.03)
$\log \widetilde{y}^S$	0.36 ***	0.32 ***	0.28 **	0.30	0.12	0.26
	(0.13)	(0.12)	(0.11)	(0.63)	(0.10)	(0.58)
$(\log \widetilde{y}^S)^2$				0.00		-0.01
				0.03		0.03
Controls	X	X	X	X	X	X
country FE	X	X	X	X	X	X
industry FE		X	X	X	X	X
occupation FE			X	X	X	X
occ x ind FE					X	X
R-squared	0.20	0.54	0.63	0.63	0.75	0.75
Observations	973	973	973	973	973	973