LUND
UNIVERSITY

Cross-country differences in self-employment rates: A panel data study on the relation between entrepreneurship, taxes and institutions within the OECD.

Master Thesis

Author: Emilie Toresson Grip
Supervisor: Margareta Dackehag

August 26, 2014

Abstract

This essay provides an analysis of the relation between entrepreneurship, taxes and institutions. Based on endogenous growth theory, the entrepreneur, as a provider of innovations, is expected to have an essential impact on economic growth. For the modern welfare state this raises the question on how to optimally tax entrepreneurial activity in order to minimize distortion effects, while stimulating growth. The most widely used proxy for entrepreneurship in the tax literature is self-employment. However, a large part of the empirical literature is based on single country analysis and fewer studies have been concerned with longitudinal cross-country comparisons. Using a relatively novel data set on harmonized self-employment rates, this thesis aims at contributing to the literature by further investigating the channels by which taxes affects self-employment in order to explain the large cross-country variations in self-employment over time in OECD countries. Moreover, this study analyses, both theoretically and empirically, the joint effect of tax policies and institutional quality on self-employment, two aspects that the economic literature previously has investigated, though mostly in isolation.

The effect of taxes is estimated using a panel data set on 17 OECD countries between 1982 and 2008, using data on personal income taxes and corporate income taxes. The main contribution of the study is the use of a relatively new harmonized measure of self-employment, the use of both personal and corporate income taxes in an aggregated cross-country setting as well as the inclusion of several measures of institutional quality. The results indicate that corporate income tax is the only tax that seems to have a significantly robust negative effect on self-employment, even though the effect is very small. The results also suggest that opposing effects of taxes may be explained by the degree of corruption in a country. In general, regulation of business, labour and credit seems to have a larger impact on self-employment relative to taxes.

Keywords: tax, entrepreneurship, panel data, institutions, OECD

Contents

Chapter one

Introduction

For a long time exogenous growth models formed the basis for empirical growth research. Within this framework technology is exogenously given implying that technological progress can not be explained; hence, long-run growth does not depend on economic conditions nor is it affected by government policy. The development of endogenous growth models made it possible to take into account long-term productivity growth and its endogenous determinants (Aghion & Howitt 2009, p. 13).

In the second wave of endogenous growth models[1], technological progress is explained as innovation driven (ibid, p.14-5). Even though the importance of the entrepreneur as a provider of innovation was pointed out already by Joseph Schumpeter in the early 20th century, innovation and entrepreneurship has been reintroduced to growth models quite recently. Within this framework, entrepreneurial activity generates innovation, boosts technological progress, and is therefore predicted to have an essential impact on economic growth. As globalization trends have contributed to larger firms relocating their production to low-cost countries, the importance of entrepreneurs and small businesses, as a provider of both innovation and employment, in high-cost countries have been stressed even further (Hansson 2008).

The most widely used proxy for entrepreneurship is self-employment (ibid).[2] During the last ten years, in light of endogenous growth theory, the interest in self-employment as a proxy for innovation has grown and so has the empirical research trying to identify its determinants. For the modern welfare state this raises the question on how to optimally tax entrepreneurial activity in order to minimize distortion effects, while stimulating growth. However, until only a few years ago, relatively few studies had considered taxation as a factor influencing self-employment (Hansson 2012). Other institutional factors have only been considered even more recently (Nyström 2008). Both tax policy and other institutional factors, such as economic freedom, enforcement of property rights, rule of law and corruption have been investigated as they are assumed to be determinants of entrepreneurial incentives. The rational is that institutions, including taxes, are assumed to affect the expected profit of the entrepreneur (Hansson 2008). As a consequence, they will affect the individual labour choice between paid employment and self-employment.

A large part of the empirical literature on self-employment is based on studies using national self-employment rates and fewer studies have been concerned with cross-country comparisons. Still there are considerably large differences in self-employment rates even within the OECD (see figure 1 below), both as cross-section "snapshots" and over time, which becomes difficult to explain without cross-country comparisons (Robson and Parker 2004). Furthermore, previous research has failed to offer consistent conclusions on how taxes affect self-employment (Stenkula 2009). Personal income taxes have turned out both positive and negative and not always significant, whereas corporate income tax has not until recently been considered. In addition, an unharmonized measure of self-employment has been used in several of the cross-country studies.

This essay aims at contributing to this cross-country literature that investigates and tries to explain cross-sectional differences in self-employment rates within the OECD. It

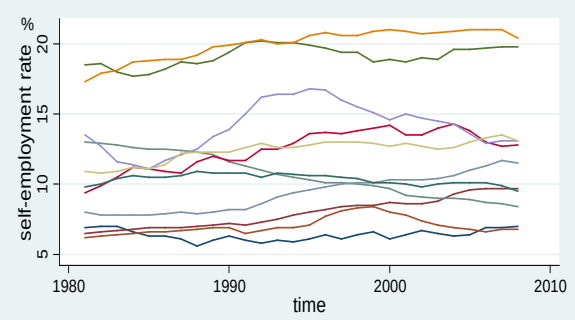
[1] The first wave is usually referring to the AK-theory, see more in Aghion and Howitt (2009, p. 13).

[2] In the following essay 'self-employed' will be used interchangeably with 'business owners' and 'entrepreneurs'. However, a brief discussion on these possibly different concepts will follow in section 1.2.

does so by providing an analysis of the relation between self-employment, taxes and institutions. The main contribution to current research is the use of improved data with less measurement errors and a comprehensive analysis of different taxes as well as a joint analysis of taxes and the quality of institutions. As far as the author is aware of and has been able to found, no other aggregated cross-country study has previously investigated the effects of both personal and corporate income taxes. Furthermore, the use of an improved and harmonized measure of self-employment, that so far does not seem to have been used in tax research, gives motivation to revisit a few different channels of impact of taxes that the empirical literature have investigated with other measures.

The first proposed hypothesis is that the data may reduce measurement errors present in previous research, and thereby be able to contribute to the literature on aggregate determinants. The second hypothesis to be investigated is that by acknowledging the impact of institutions, the same tax policy might generate different effects if the institutional conditions differ between countries. The first hypothesis will be considered in the baseline regression and the following extensions of the model, whereas the second hypothesis will be considered in one of the particular specifications which include an interaction term between taxes and an indicator on institutional quality.

Figure 1. Self-employment rates in different OECD countries 1981-2008[3].



Source: COMPENDIA (2013)

1.1 Aim, general method and disposition

Given the above, this thesis is both descriptive and explanatory. More specifically, *this thesis will evaluate how taxes affect the level of entrepreneurial activity, while taking in to account institutional quality, in 17 OECD countries between the years 1982-2008.*

Entrepreneurial activity, the dependent variable, will be measured by the rate of non-agricultural self-employed, as a percentage of the total labour force. In the next section 1.2, the concepts of entrepreneurship and self-employment will be briefly introduced. The dependent variable will be regressed on the independent variables measuring personal income tax and corporate income tax together with indicators on institutions and other theoretically plausible determinants of entrepreneurial activity. Other taxes may also influence self-employment, however, as a delimitation of the thesis, only personal income taxes and corporate income taxes will be considered.

Empirically, this thesis builds on previous studies of the determinants of self-employment in order to construct a solid empirical model. One of the main contributions of this thesis is the use of a relatively new dataset on self-employment. In most previous cross-

[3] Self-employment rates refer to the non-agricultural self-employed as a percentage of total labour force in each country. Further, the top rates refer to Italy, Greece and Portugal, whereas the lower rates are found in Denmark and Switzerland. Note that not all countries in the employed sample are represented in this figure (some countries with rates in the middle are omitted for clarity of presentation).

country studies the measure of self-employment used, has been drawn from the OECD Labour Statistic database (used in Robson and Wren (1999), Fölster (2002) and Parker and Robson (2004) for example). However, this data has been gathered from countries with different definitions of self-employment, hence this indicator is not suitable for cross-country comparisons, as it may introduce severe measurement errors (Fölster 2002). To address this, this thesis will use self-employment data from the COMPENDIA database (EIM Business and Policy Research 2013) which has been constructed with the specific purpose of constructing comparable statistics within the OECD. Nevertheless, this new and harmonized data has yet not been used in order to confirm the findings of earlier aggregate studies of taxes on self-employment, as far as the author has been able to find.

Econometrically, panel data allows the researcher to take into account of both individual heterogeneity and dynamic effects which is not possible in cross-country studies. A least square dummy variable (LSDV) model will be employed as a baseline model as to control for this and further investigation in the relation between self-employed, taxes and institutions will be pursued by allowing for dynamic, non-linear effects and interaction terms.

The paper is laid out as follows: the first chapter ends with a short introduction to the key concepts self-employment and entrepreneurship. Chapter two contains an outline of the theoretical and empirical determinants of self-employment. Based on this theoretical and empirical framework, the empirical part of the thesis, chapter three, begins with a description of the chosen data which is followed by a discussion of the empirical strategy as well as other methodological considerations. The results are presented and elaborated on in chapter four whereas a discussion of the results and concluding remarks are found in chapter five. Descriptive statics and regression outputs are found in the appendix.

1.2 Defining the concepts- what is an entrepreneur?

For a long time there were almost no entrepreneurs in the models used by neoclassical economists. Recent years of theoretical and empirical research has shown that this now is changing and as a parallel development there are many policymakers who have come to see entrepreneurs as the solution to both unemployment and weak economic performance (Henrekson 2007). For a formal presentation, assume that the domestic output can be described by the Cobb-Douglas production function GDP = Y = $AK^{\alpha}L^{(1-\alpha)}$, where output per capita implies:

$$y = Ak^{\alpha} \qquad \text{where } k = K/L \qquad (1)$$

Given that economic growth is the growth rate of output per person and assuming equal growth rate in labour and population the growth rate (g) can be expressed as[4]:

$$g = \frac{\dot{y}}{y} = \dot{a} + \alpha \frac{\dot{k}}{k}, \qquad \text{where } \dot{a} = \dot{A}/A \qquad (2)$$

The importance of entrepreneurs will formally be captured within the term $\dot{a}$, as entrepreneurial activity is expected to generate innovations that increase productivity within the economy[5]. Hence, taxation related to entrepreneurial activity would also be captured by

[4] Given by taking the natural logarithm of (1) and then differentiating with respect to time, where $\frac{\dot{y}}{y}$ equal $\frac{\frac{dy}{dt}}{y}$.

[5] In the Schumpeterian endogenous growth model the innovation process is, in short, described as the outcome of each period where there is an entrepreneur that has the opportunity to attempt an innovation. Given a successful innovation, a new version of an intermediate product (used to produce final goods) is created that is more

this term (Aghion and Howitt 2009, p. 87-88, 106, Lee and Gordon 2005). However, research on entrepreneurship and innovation is still fairly new in academia and this may be the reason for why entrepreneurship, despite its possible importance, still represents an elusive concept (Henrekson 2007). Yet, I will try to make a very short and general summary of the concepts in order to guide the discussion in the following chapters.

Usually, different definitions of entrepreneurship typically refer to (the creation of) new, usually small, businesses. These businesses are owned and managed by individuals that usually are assumed be innovative in their business approach and accept some degree of risk in their business venture (Bruce and Schuetze 2004). In chapter three the data will be described, including the dependent variable. However, related to the definition of self-employed, is that the measure employed in this study includes both unincorporated and incorporated business owner/managers.

Henrekson (2007) argues that entrepreneurship "almost always entails an ambition to grow". He defines entrepreneurship as individuals and organizations that actively contribute to renewal and change in the economy which implies both creation of opportunity and response to the existing environment while embracing risk of an uncertain outcome. Entrepreneurs can hence be described both as rent-seekers and risk-takers. Given the broad definition, individuals that choose to enter entrepreneurship, may cause activities to be "productive, unproductive or even destructive from a societal perspective" (ibid, p. 3), that is, not necessary positive innovative ventures.

Lastly, related to the discussion of elusive definitions is the discussion of measuring such innovative activity generated by an entrepreneur. In most empirical studies the measure that has been used is the number of individuals who report working for themselves, which quite closely matches the "self-employed" classification. Nonetheless, self-employed includes individuals engaged in widely varying activities, including those who operate as franchisers, as an example. Franchiser may encounter some risk but as they are following a stylized approach and strategy within the chain they are most likely not to be as innovative as other "self-made" entrepreneurs (Bruce and Schuetze 2004). Other non-entrepreneurial motives to be self-employed may be to pursue a certain lifestyle, to circumvent discrimination on the labour market or to be able to avoid taxes or heavy regulation. This is why the measurement may overstate the amount of entrepreneurial activity. On the other hand, this measure excludes individuals engaged in innovative behavior within established firms, hence the measure may also be too narrow. Moreover, illegal activities as motives to be self-employed, may be ambiguous, as evasive activities might seem necessary for individuals or firms that strives to be productive, in circumstances where the market is excessively regulated or taxed (Henrekson 2007).

However, in keeping with the notation of most of the previous literature, I use the terms entrepreneurship, self-employed and business owners interchangeably throughout the remainder of the paper.

---

productive than the previous version. The innovation process is random and hence will growth be random. As the growth rate can be shown to be proportional to the productivity affected by the innovations of entrepreneurs, in the long run the economy's average growth rate is equal to the frequency times the size of the innovations. As new innovations that cause economic growth also makes the previous technologies obsolete, this refers to the Schumpeterian "creative destruction" (Aghion and Howitt 2009, p. 85-87).

Chapter two

Theoretical and empirical framework

This chapter is divided in two main parts. The first part, section 2.1 to 2.4, will review the theoretical literature on the determinants of self-employment in order to shape the theoretical framework of the thesis. The second section, section 2.5, is devoted to previous empirical research which aims at complementing the discussion on theoretical determinants.

Even though the interest in entrepreneurship as a unique factor in the modern economy has grown, the concept of entrepreneurship is of a multifaceted nature (Nyström 2008) and this may be the reason for why there is no coherent and general agreed upon formal model. The one used in this thesis will therefore be extrapolated from different theoretical and empirical papers. However, in order to structure this theoretical section, this part will in broad brushstrokes be structured as a model where the decision to become self-employed is dependent on the costs and alternative costs of being self-employed.

2.1 Entrepreneurial income and entry

How entrepreneurship is affected by taxes may be explained by looking at how the entrepreneurial income is affected by taxes. Entrepreneurs are often described and analyzed as rent seekers which imply an attempt to reap rewards in excess of the level determined in competitive equilibrium or rates exceeding the risk-adjusted market rate of return, i.e. they try to create or discover economic rents. This rent may be obtained through productive or unproductive entrepreneurial activities: introducing a new innovation in the form of a superior or more cost effective product, attaining a monopoly position when granted a government license or bribing an official to keep competitors out (Henrekson 2007). The specific compensation for effort of the self-employed usually comes in the forms of wages, dividends, sale of the business or a combination of these, which all may be taxed differently (Stenkula 2009) which give motives for study of several tax rates.

Gentry and Hubbarb (2000) model the entry decision for an entrepreneur as a comparison between the incomes the potential entrepreneur expects as self-employed compared with income from paid-employment. The decision to become self-employed is taken if:

$$\pi_s \theta_i k_{it}^{\alpha} - rk_{it} > w_{it}(e_{it}, x_{it}, z_{it}) \qquad (3)$$

This occurs when the potential entrepreneur receives a wage, *w*, when working for a firm that is lower than the expected return to entrepreneurship. The wage is based on education (*e),* experience (*x)* and other household characteristics (*z*). The left hand side is determined by the probability of success (π_s), entrepreneurial ability (θ_i), investment (*k)* and the risk-free interest rate *(r).* The gross return from entrepreneurship, if successful, yields $\theta_i k_{it}^{\alpha}$. This provides a picture of the theoretical micro-model of the entrepreneurial decision. Since this study investigates the aggregated determinants it will not be able to control for individual level factors such as ability, but the above model may illustrate the entry choice as determined by certain costs and alternative costs to self-employment. This gives structure for the next theoretical part. Further, this expression does not take into account the effects of taxes, but as will be discussed in the next section, taxes will affect both incentives and risks associated with the left hand side *relative* to the right hand side of this expression.

2.2 The effect of taxes

Income taxes consist of personal income taxes and corporate income taxes. Personal income taxes are taxes on wages and possibly business income from unincorporated businesses whereas corporate income taxes are levied on the net return on incorporated businesses (Stenkula 2009, Bruce and Moshin 2006). These taxes and their trends within the context of OECD countries will be commented on in section 2.6 and they may affect individual behavior and the employment decision in different ways, both negatively and positively. In short, personal income taxes affect the return to all forms of labour which includes self-employment (Bruce and Moshin 2006); hence an intuitively negative effect may be expected. As parts of the incorporated businesses earnings may be partly taxed as earned income, this may have a negative effect on these businesses as well (Stenkula 2009). This relationship has generated more interest in the literature compared with corporate income taxes, which nevertheless affect the earnings of incorporated business income. These taxes may also have a combined impact on how new businesses will be organized in terms of incorporation or not (Bruce and Moshin 2006).

The effect of taxes on entrepreneurship could therefore be a 'simple' one: since higher taxation reduces the level of profit opportunities, it reduces the expected income of the potentially self-employed and thus may reduce productive entrepreneurship. Following the previous vocabulary, taxes is then perceived as a cost of self-employment. Baliamoune-Lutz and Garello (2011) call this intuitive and straight forward negative labour supply effect the 'incentive' effect. However, as will be demonstrated, economic theory actually implies ambiguous effects on the level of entrepreneurship.

To start with, economic theory predicts different forms of impacts of marginal and average taxes. Marginal tax rates have an impact on the distortions introduced to individuals and firms choices, as they influence decisions concerning the amount of investments to undertake, additional income to earn, and entrepreneurial effort. It affects additional income on the margin; hence it may be argued that it will affect the "rate of success" for the self-employed. On the other hand, average taxes are predicted to influence the discrete choice to invest or supply effort at all (Hansson and Dackehag 2012).

Robson and Wren (1999) construct a model on marginal and average taxes where they focus on two aspects of self-employment, which are said to differ from regular paid employment. First, self-employed are argued to have a greater sensitivity of pre-tax income to effort, due to the risk and second, by self-employment there is a greater opportunity to evade taxes on income. Here marginal and average tax rates have opposing effects on the supply of effort and evasion and thus on the labour choice between self-employment and paid employment[6]. An increase in marginal tax rates is expected to have a negative effect on self-employment while an increase in average tax is expected to have the opposite effect, in this particular model.

Bruce and Schuetze (2004) further points out that higher tax rates not only affects the expected return, but also the risk associated with self-employment. Unlike wage work, entrepreneurial income offers an uncertain return. Thus, both rents and risks need to be

[6] The intuition is that an increase in the marginal tax rate decreases the optimal level of effort and thus causes the individual to reduce effort as well as the level of evasion. However, the relative costs of this, in terms of the loss of expected post-tax income, are greater for self-employed in this model. Pre-tax income is more responsive to the reduction in effort in self-employment, but the benefit to the self-employed of being able to evade taxes is also reduced, as the level of evasion is lower. This is in contrast to the relation between an increase in the average tax rate that increases optimal levels of effort and evasion. Within this framework they model the individual problem of optimal level of effort and tax evasions that maximize the level of utility in each employment model (Robson and Wren 1999).

taken into account in order to understand the self-employment decision. Given this, taxation can make investment in the risky sector more or less attractive.

The aspects of risk in the entrepreneurial income have been elaborated on in the context of the personal income tax as a progressive tax schedule. On the one hand, tax progressivity can lower entrepreneurship via the 'incentives' effect (reducing the size of profit opportunity). On the other hand, it can lead to an increase in entrepreneurship if risk adverse entrepreneurs perceive it as an insurance mechanism. The rational is that the progressive tax system, by providing loss offsets or because of lower marginal rate in case of entrepreneurial failure, may reduce the risk perceived by the potential self-employed, as a form of "risk-sharing" with the government. Positive effects of tax progressivity, as in the case with tax rates, may also be due to the previous mentioned effects of tax avoidance (Baliamoune-Lutz and Garello 2011, Gentry and Hubbard, 2000).

The negative incentive effect and the positive risk-sharing effect are theoretically applicable to the corporate income tax as well. Cullen and Gordon (2007) have made important contributions to this relatively late theoretical and empirical literature where both taxes are considered in a combined framework and actual incentives is thus affected by the relative rates between personal income taxes and corporate income taxes. Cullen and Gordon (2007) identify three channels by which the corporate income taxes affect self-employment, in relation to personal income tax. First, since the corporate income tax often is lower than the personal income tax, this encourages the self-employed to incorporate his business and is referred to as a form of "income shifting". Second, if higher risk-taking by the entrepreneur generates higher returns to entrepreneurship, the combination of progressive personal income tax and usually a flat corporate income tax creates a "risk-subsidy" for self-employed if incorporating the business. These two channels increase incorporated self-employment, if the corporate tax is reduced. The third channel refers to the already described "risk-sharing" channel.

The taxes above were analyzed with the main assumption that higher taxes raises the costs of the potential income of the self-employed but positive channels were also briefly mentioned. One potential positive effect arises from tax evasion possibilities through self-employment. As self-employed reports their own income, it is easier to avoid tax by underreporting income for self-employed compared to employed individuals (Hansson 2008). Both theoretical and empirical evidence indicate that self-employed have plenty of opportunities to evade taxes, a reason for why tax policies that encourages entrepreneurial activity can be a "double edged sword" (Bruce and Schuetze 2004). A tax system that generates a desired increase in entrepreneurial activity by an increase in taxes in the wage sector raises the marginal benefit to self-employment for those wishing to engage in entrepreneurial activity, but also increases the marginal benefit to self-employment for those whose intent is tax evasion (ibid). Further, even with neutral taxes in all employment modes, the effective tax rate may be lower for self-employed business owners, due to the possibility to avoid or evade taxes which may encourage entrepreneurial activity without motives of innovations (Gentry and Hubbarb 2000).

Given the above, it is clear that the expected sign of taxes are not given *a priori* to the empirical investigation in the next two chapters.

2.3 Institutions, taxes and self-employment

As institutions have shown to be important determinants for economic growth, it is most likely that they also have an important effect on entrepreneurship. Similar to taxes, institutions affect the potential return for the self-employed and thus the incentives for entrepreneurship (Hansson 2008). In this section I will describe the theoretical relationship further and it will

be concluded with a joint theoretical analysis of the effect of taxes on self-employment in the context of the institutional set-up.

When defining what institutions are, one often refers to Douglas C. North, one of the most important contributors to institutional economics. North (1993b, p. 20- 21) argues that institutions affect all individual choices and thereby governs the whole economy, by incentive structures ("rules of the game") that affect transaction costs for exchange and production. Thus, economic growth is not caused by factor accumulation, but initially by the incentives structure that encourage individuals to make productive investments. Similarly, where economic growth is absent or slow, North points to the explanation of destructive incentives rather than the lack of capital accumulation. The "rules of the game" are either formal or informal, where the former consists of political constitutions and regulation that enables financial freedom and ensures rule of law for example, while the latter consist of traditions, norms and codes of conduct that constitute cultural inheritance (North 1993a, p. 16).

Henrekson (2007) argues that entrepreneurship not per se is determined to be socially productive and stresses that institutions can determine how the entrepreneurial talent is channeled to productive, unproductive and even predatory activities. For example, the incentives structure may imply that the return to entrepreneurial effort is larger trying to circumvent institutions, by for example avoiding taxes, rather than benefiting from given institutions to reduce uncertainty and enhance contracts when doing business. An example of formal institutions that reduce uncertainty and risk would be the protection of property. However, there might be ambiguous effects of different institutions as, on the other hand, excessive protection of property rights also may hamper productive entrepreneurship. The logic could be analogous to a tax on innovation; both the risk and expected expense associated with entrepreneurship rises sharply (if an innovator faces a high risk of being sued for infringement by other patent holders). In this situation, incentives promote business through large firms with financial strength relative to small start-ups.

Lee and Gordon (2005) directly highlight the absence of rule of law, in terms of corruption, as a barrier to entry for small business owners. This is because the need to pay bribes to government officials in order to obtain needed licenses, imply a direct cost to self-employment.

As an extension, Torrini (2002) makes a connection between the effect of taxes and the presence of corruption. He argues that self-employed are more likely to be engaged in tax avoidance, the greater the toleration of irregular activities is, implying that positive effects of taxes are more likely to be found in countries with relatively high corruption. As a consequence, in countries where laws are more likely to be enforced, high taxation could instead discourage entrepreneurial activities.

Given this combined analysis, for the second research question in this study, the hypothesis is that as institutions shape incentives it is possible that the same tax policy will generate different results depending on institutional differences in different countries.

2.4 Other determinants of entrepreneurship

In the empirical literature on the determinants of self-employment a division may be made between *pull* and *push* factors. An individual can either be pulled into self-employment in order to pursue a promising business opportunity (rather than having a paid job or being unemployed) or the individual can be pushed into it because no other option to make a living is available (Henrekson 2007). These factors may also be modeled as costs of self-employment, but pull and push factors both refer to situations where the cost of self-employment relative to other employment modes are lower, however with this definitions,

two different *motives* crystallizes. These factors can usually be divided in labour market, macroeconomic and demographical categories.

Based on Evans and Jovanovic (1989), individuals are pulled into self-employment on the basis of the expected returns to entrepreneurship relative to those earned in paid- or unemployment. However, another assumption within this framework is that entry into self-employment requires some form of *wealth* of the potential self-employment. Thus, where there are incomplete capital markets, self-employment is restricted to the ones with the necessary funds (or collateral) for *start-up capital*. Factors that influence access to finance and potential start-up capital are therefore likely to be important. Empirically, it has been confirmed that entrepreneurs tend to have more savings than other households and that access to finance is important in order to start up a new business (Henrekson 2007). Fölster (2002) highlights further that a tax-financed welfare system tends to reduce household savings which may negatively affect access to finance.

A comprehensive welfare system is likely to be correlated with the *size of the public sector*. Davidsson and Henrekson (2002) have argued that a large size of the public sector also decreases the need for a private sector on the market where the public sector is offering services. This effect on the private market may hence have a negative effect of the number of self-employed. Hansson (2008) clarifies the analysis by pointing out the effect of taxes on self-employment most likely is dependent on what the tax revenues are spent on. This allocation can have impact on the incentives for the level of savings as well as on the incentives for the unemployed to find new employment opportunities. Large tax revenues may also be spent on strengthening institutions that foster productive self-employment and other business forms in general.

A key alternative cost and (positive) push factor to self-employment may be the risk of *unemployment*. When unemployment is high the decision to become self-employed may be "the last option".[7] On the other hand, high unemployment may be related to low levels of demand, indicating a negative relationship. Parker and Robson (2004) call these channels the "recession push" versus the "prosperity pull", explaining why unemployment in many studies is said to control for *business cycles.* Robson and Wren (1999) argue that a negative effect is plausible since the attractiveness of self-employment may rise when a well performing labour marked can act as a "cushion against failure" of the entrepreneur (i.e. if the entrepreneurial venture does not pay off the individual may be able to find a wage and salary employment). This corresponds to the potential "insurance" effect of taxes that was discussed in section 2.2, which stresses the need to incorporate the risk of entrepreneurship in the theoretical framework.

Related to unemployment is the *replacement rate*, i.e. the level of unemployment benefits. If this level of benefits is high, i.e. the alternative cost is high, it might discourage self-employment. Moreover, if self-employed do not receive the same benefit entitlements as those in wage employment, high benefits may discourage paid employers to become self-employed in the first place (Parker and Robson 2004). A corresponding alternative cost is the real value of the minimum *wage* (Bruce and Moshin 2006) or average wage as used in Stenkula (2009), which also may capture motives to become self-employed.

Industrial composition has been discussed as a determinant of self-employment as it may affect the available opportunities for the potential self-employed. Control for the size of the service sector and the share of the labour force that is employed in the services sector has been included as to control for different business structures (Stenkula 2009). Another labour market characteristic is the share of *female labour participation of the total*

[7] This is usually referred to "necessity entrepreneurship", where the prime motive to start a business is to make a living (Nyström 2008).

labour force that in many studies have been included, due to empirical evidence indicating that self-employment is more common among men than females (Parker and Robson 2004).

Other *demographical changes* have also been controlled for in many studies. Bruce and Moshin (2006) control for median age and Fölster (2002) include a measure of the dependency ratio (share of population below 18 or above 65).These measures may capture tendencies in age groups who affect the rate of self-employment or capture effects on welfare expenditures that previously were discussed and linked to self-employment.

Lastly, the *aggregated income level* in the economy may have an impact on self-employment by different channels. Theoretically, higher per capita GDP might be negatively related to aggregate self-employment rates if it is associated with more capital per worker which may imply larger average firms. If the measure of self-employment captures a large group of necessity entrepreneurship, it is also plausible that the effect is negative, as found by Bjørnskov and Foss (2008). However, higher per capita GDP might have a positive impact on the demand within countries, which might (disproportionately) benefit the self-employed or it may stimulate available capital for the entrepreneur (Nyström 2008). Fölster (2002) argue that lower income levels may cause higher self-employment, due to less efficient distribution systems in these countries, which enables smaller shops to be established, that are typically managed by self-employed.

2.5 Previous research

Not only does the theoretical literature predict somewhat diverse effects of taxes on self-employment, the empirical research has generated inconclusive result as well (Stenkula 2009). Many studies have found that higher tax rates lead to higher rates of entrepreneurial activity, while a number of more recent studies have questioned this general finding, finding both negative and non-significant effects (Hansson 2012). A short summary will follow of these results, with focus on studies that use similar methods or data.

The effect of taxes on self-employment has been studied on different levels, with some variations in methods. One straight forward way to divide them, as done by Stenkula (2009), is along a micro- and a macro-line. As has been mentioned earlier, there have been few macro studies that have investigated aggregated cross-country differences in self-employment using longitudinal data. The main reason is the low quality of data on self-employment (especially from OECD Labour Force Statistics) and problems with omitted variables and endogeneity (Hansson 2008). Lack of data has also been an issue (Parker and Robson 2004).

Fölster (2002) studied how tax revenues over GDP as a measure of general tax burden affects self-employment in OECD countries and found a negative and significant relation. He, however, points out the risk of using OECD data as it suffers from measurement errors. Parker and Robson (2004) also studied OECD countries, but included average rates of personal income tax and employers' social security contributions. They concluded that the emphasis on macroeconomic and demographic variables in previous studies appears to have been misplaced, as taxes seem to have the strongest impact. They claim that some of the dominant positive effect found in their and other studies may be partly explained by the fact that marginal income tax rates have been the same or similar for employees and the self-employed in many countries. Therefore, labor supply effects of changing tax rates are similar in both occupations, while the effect of tax avoidance that is especially applicable to the self-employed, generate the dominating positive effect.

The timing of the effects of taxes and the potential short-run and/or long-run effect of taxes has also been partly investigated in the aggregated literature (using error-correction models and cointegration techniques for example). Robson and Wren (1999) studied effective marginal and average taxes within OECD and found support for their

theoretical model that marginal tax negatively affect self-employment, while the average have a positive effect, as in Parker and Robson (2004), in both short run and in the long run. Stenkula (2009) used aggregated measures of different taxes and self-employment measures but in a 50 year long time series with Swedish data. He found that self-employment rates adapt slowly to changes in tax rates which he concludes is plausible since most people will not immediately start or stop being self-employed as a consequence to changes in the economic environment. He also found no significant effect of marginal income taxes and negative but small effects of payroll taxes and corporate taxes. His study was based on the analysis of Bruce and Moshin (2006) using US data. They were one of the first scholars to include corporate taxes in their analysis. They found that corporate income tax and personal income tax had a negative and a positive effect, respectively, though the significance depended on the measure of self-employment, and the effects were small. Only the corporate income tax rate was significant when more advanced statistical tests were performed.

Bruce and Moshin (2006) and Stenkula (2009) used corporate tax rates, but the vast majority of the research has focused on personal income tax. As Da Rin et al (2011, p. 1048) formulate it, corporate income tax is "a policy instrument that has received surprisingly little attention so far". Djankov et al (2010) use a large novel micro data set covering 85 countries in order to investigate the effect of corporate income tax on entrepreneurial activity, FDI and aggregated investment and find a relatively large negative effect of taxes on entrepreneurial entry (a 10 percentage point increase of the tax decreases entry with 1,4 percentage points). Nonetheless, this was a pure cross-country study for one year. Da Rin et al (2011) add to this literature by using firm level micro data in 17 countries over 8 years. They also find a negative effect, but below a certain tax level. Above a certain level the positive "risk-sharing" effect seem to dominate. Djankov et al (2010) and Da Rin et al (2011) also recognize the importance of regulation, security of property rights among other to control for the effects of taxes. These are examples of micro data studies has generated a lot of interest in recent years, when more detailed individual data has become available. Many micro data studies have questioned earlier positive effects captured by aggregated data but this data is still claimed to find less conclusive results, i.e. not as positive as earlier aggregated data, but still without possibility to generate consensus on a negative effect (Bruce and Moshin 2006).

While corporate income tax has been used in these micro data studies and top statutory corporate taxes have been used in single country studies (Bruce and Moshin 2006, Stenkula 2009) it has not, as far as the author has been able to found, been used in a cross-country study. This thesis adds to this literature.[8]

In addition to the timing of the tax effect, the linear form of the effect has also been studied. For example, Georgellis and Wall (2006) allowed for a non-linear effect of the marginal income tax on the basis that assuming a non-trivial cost of being caught evading taxes, the incentives to evade will be higher for higher tax rates. This may cause a non-linear effect where the negative labour supply effect (the "incentive effect", according to previously vocabulary) dominates at lower tax rate, while a positive effect may be found at higher rates. In a dataset on 50 American states for 8 years they found support for this. De Mooij and Nicodeme (2007) also found support for this with corporate income tax in a large micro data set and so did Da Rin et al (2011).

[8] It has been used, together with other top statutory tax rates, in studies on economic growth as the dependent variable. For example, Lee and Gordon (2005) investigate this relation with motivation on previous studies on taxes and entrepreneurship. If entrepreneurship generates growth, a similar relationship should prevail between taxes and growth. They conclude however, that the aggregated information in their study provides insufficient information on the precise links between taxes and growth. This thesis, using similar tax measures as Lee and Gordon (2005) might hence be seen as a further investigation of this link.

Related to this, is the research of Torrini (2002) who is one of few who has studied self-employment from a joint tax and institutions perspective on an aggregated level. Torrini (2002) argues that given the persistence of self-employment rates across countries, variables that show little variability over time are likely to be important explanatory factors. Hence, institutional characteristics become natural candidates as they are relatively stable over time, however, still different between countries. This is also the motivation to a recent study of self-employment with respect to institutions by Nyström (2009). Using panel data on OECD countries between1972-2002 she found that smaller government sector, better legal structure and security of property rights and less regulation of credit, labour and business seem to have a positive impact on self-employment, though this analysis was in isolation of taxes.

Da Rin et al (2011) pursue a very similar theoretical argument as Torrini (see section 2.2.3) when they investigate if the relationship between corporate income taxes and entry rates (on firm level) are constant or dependent on what they call "institutional infrastructure". They operationalize it by tax accounting standards and divide his sample with dummies for "god" and "bad" standards. They find that the negative effect of increased corporate income is smaller in countries with "bad" standards and conclude that this is an effect of greater possibility for tax evasion in these countries. Torrini (2002) investigated five institutional variables, including one tax measure: public sector size, labour and product market regulation indicators, tax and social contribution wedge and corruption, where the latter is said to capture "law enforcement and the general attitude towards illegal activities" (ibid, p. 15). He also found differences in the impact of the tax wedge on self-employment when he interacted this wedge with the corruption dummy. Nonetheless, as his data on self-employment rate also suffered from measurement errors and as a lack of data prevented him to use many control variables in his panel data specification, this makes it relevant to revisit this study in order to expand and possibly verify the result. For this purpose, the perceived levels of corruption in the sample are shown in figure 2.

Figure 2. Perception of corruption in OECD countries 1981-2008[9].

Source: CANA database 2011

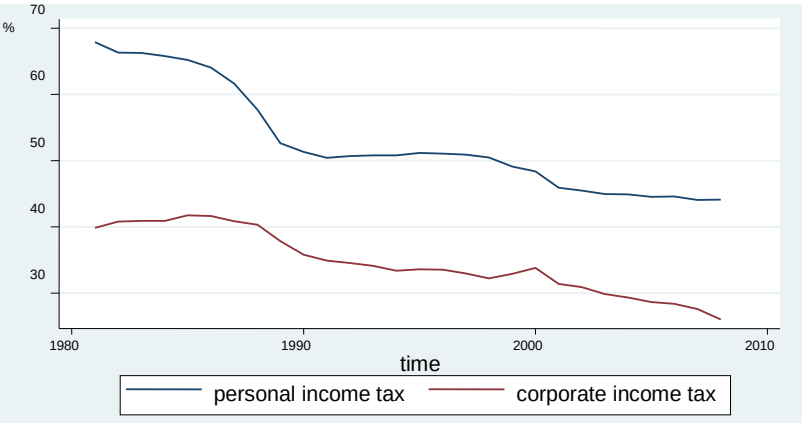
2.6 The context of OECD countries

For the business owner, globalization has made it possible to relocate the business in another country to take advantage of more favorable fiscal environments (Baliamoune-Lutz and Geller 2011). Overall, globalization has put pressure on the tax rates and system of developed

[9] The indicator consists of an index on a scale from 0 to 10 where 10 equal no perception of corruption. Further, the top rates refer to Denmark and New Zeeland, whereas the lower rates are found in Greece and Italy. Note that not all countries in the employed sample are represented in this figure (some countries with rates in the middle are omitted for clarity in presentation).

countries, when the international flow of capital and labour has increased. However, personal income taxes and corporate income taxes are together the largest source of taxes in OECD. Since they roughly represent 35 percent of the overall tax incomes and have done so during the last decades (OECD Tax Database, 2014), these taxes are likely to represent a substantive amount of the taxes affecting self-employed. The delimitation of the thesis is hence motivated. In the graph below the trend in the top statutory taxes since 1981 is shown.

Figure 3. Trend in average top statutory tax rates in OECD, 1981-2008

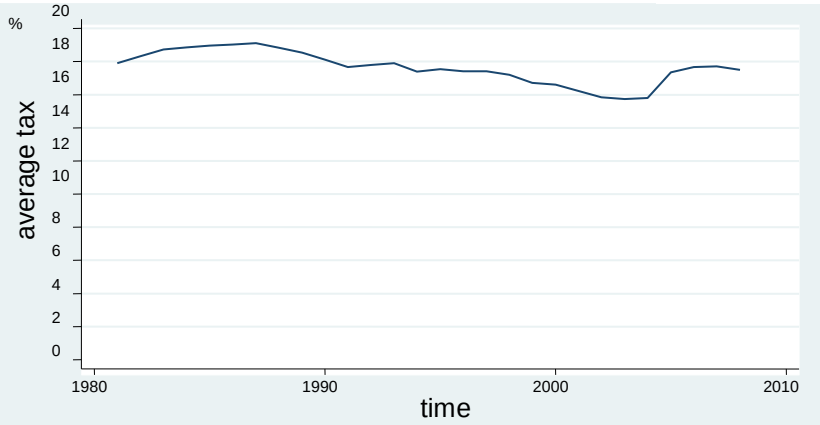


Source: OECD Tax Database (2014), own calculations on the employed sample

A wave of tax reforms took place in the 1980s and evened out many differences between the OECD countries in their tax schedules. Following these reforms, steep declines in top rates of personal income during the 1980s is one of the most notable changes over the last three decades.

When some convergence in different tax rates has taken place, self-employment rates have remained different between countries (as seen in figure 1 in chapter 1).The latter imply an argument for including slow-moving institutional variables into the analysis (Torrini 2002). Average income tax rates have also decreased on average among the OECD countries, but less dramatically compared with the top statutory income tax rates.

Figure 4: Trend in average income tax rates in OECD, 1981-2004



Source: OECD Tax Database (2014), own calculations on the employed sample

Chapter three

Empirical model

The aim of this study is to evaluate how taxes affect the level of entrepreneurial activity, while taking in to account institutional quality, in the OECD. This chapter gives a description of the chosen data, the empirical model and other methodological considerations.

3.1 Data

To assess the questions of the thesis I rely on a variety of data sources. The combination of these datasets is unique in the literature. The choice of the variables are based on what was discussed in the theoretical section and on what has been used in previous research as operationalization of theory. Detailed descriptions of the data and their sources can be found in Appendix A.

3.1.1 Dependent variable: self- employed business owners

Data on the dependent variable self-employment rate (*semp)* will be collected from the COMPENDIA database (EIM Business and Policy Research 2013) and refers to the number of self-employed in relation to the total labour force. The chosen measure will exclude self-employment in agriculture, hunting and forestry, as is usually done in the research. The rational is that these sectors have gone through significant structural changes throughout the end of the 20th century in most European countries, a development that has little to do with taxes (Stenkula 2009). These sectors are also argued to be influenced by historically and culturally determined traditions of family ownership and less by other factors that influence self-employment rates in the rest of the economy (Parker and Robson 2004).

The business ownership rates in the COMPENDIA database has been gathered with the specific purpose of constructing comparable rates within the OECD. The purpose is to address the measurement errors in the OECD Labour Force Statistic (LFS) database (van Stel 2003)[10]. There are other variables available that have been used as proxies for entrepreneurship in recent years, for example exit and entry rates of businesses and nascent businesses (in Gentry and Hubbarb 2000 and Baliamoune-Lutz and Geller 2011). Though, the main drawback of these measures is the limited availability over time (compared to the self-employment indicator).[11] Nonetheless, as was discussed previously, one should keep in mind that the COMPENDIA indicator may either under- or overestimate the full impact of entrepreneurship (see section 1.2).

Lastly, since previous research has shown that females are less likely to become self-employed (Hansson 2008), gender disaggregated data on the dependent variable would

[10] The main problem related to the OECD LFS indicator refers to variations across countries on how owners/managers of incorporated businesses are counted as self-employed or as employees. The latter case may prevail due to the fact that formally, owner/managers of incorporated businesses are employees of their own businesses. The Compendia database however addresses these differences by including owners/managers of both unincorporated and incorporated businesses in their measure. For countries where this data is lacking, an estimation of these numbers is done. Further, in contrast to the OECD LFS data, the Compendia database excludes unpaid family members working for the business, as well as people with self-employment as a secondary activity. See van Stel (2003) for a more thorough presentation of the database and on the draw backs of the original OECD statistics on self-employment.

[11] There have been initiatives to create better measures of entrepreneurship, such as the indicators from Global Entrepreneurship Monitor (GEM), however this data does not distinguish between formal and informal entrepreneurship, nor does it exclude the agriculture sectors (Nyström 2008).

have enabled interesting closer analysis of these findings. Unfortunately the COMPENDIA database does not provide this detailed data. Nonetheless, a control variable for the share of females in the labour force will be included in order to control for this tendency.

3.1.2 Independent variable: personal income tax and corporate income tax

As Stenkula (2009) has pointed out, several tax measures are preferred over only one when studying the effect of taxes on entrepreneurship.

The three independent variables that will be used in this study will measure the personal income tax and the corporate income tax. The first tax will be measured by two indicators: top statutory personal income tax (*pit)* as a measure of the overall marginal tax rate and an indicator of the average income tax rate (*av_tax*) measured as the average tax rate for a single person earning an average wage[12]. Similar or equal measures have been used in for example Stenkula (2009) and Parker and Robson (2004).

There are many considerations when choosing representing measures of the personal income tax. The reason behind including two different measures of personal income tax is the fact that previous research has found an opposite effect of these tax rates (Robson and Wren 1999), a result that could be evaluated by including these measures. Moreover, as incentives depend on marginal tax rates, an indicator of this is especially appealing. Even if many self-employed do not face all tax top rates, they still may be appropriate. In fact, Lee and Gordon (2005) argue that it is a suitable measure of incentives for potential entrepreneurs but it would rather poorly approximate the incentives faced by the general labour force. Bruce and Moshin (2006) further argue that it is a simple way to measure the tax level for the self-employed, and it may be seen as an acceptable proxy for the maximal tax on the marginal investment. Third, it might also act as a policy signal affecting the aggregate level of self-employment in a country (Stenkula 2009) and fourth, it is one of few available tax rates for all years and countries in this sample. As Lee and Gordon (2005) highlight, it should though be recognized as a "noisy" measure of the incentives for the whole tax schedule.

The data on average tax is taken from the OECD Tax database (2014)[13]. A typical source for the top statutory marginal tax rate in the literature is the World Tax Database from the Office of Tax Policy Research (OTPR) at the University of Michigan. However, when a closer analysis of this data was made in preparation for this study it was noted that only the central government statutory tax rates are included in this measure. Given that many tax systems in for example the Scandinavian countries are highly decentralized, this may be misleading when comparing different countries[14]. As no other readily available data was found on the combined tax rates for the whole period, data on separate central and sub-central rates from the OECD Tax database (2014)[15] was added to each other manually, in order to create data on the top marginal tax rates that are more suitable for cross-country analysis. However, to be noted is that even though countries may have the same top marginal

[12] There has been research indicating that married individuals have a higher probability of becoming self-employed compared to single households (Hansson 2008), which could motivate the use of average tax on earnings of a married individual instead. Unfortunately the data on the latter is more limited for the given sample. Moreover, using this particular measure enables comparison with previous research since this is the measure used in for example Parker and Robson (2004).

[13] Available on request from the OECD.

[14] For example, according to the OTPR database, the top statutory income tax in Sweden in 1991 was 20 percent, while it in fact was just above 50 percent taking into account the sub-central tax. Moreover, the sub-central tax seems to be included in the data for Sweden *before* 1991 which imply inconsistent methodology in the time series. Same inconsistency is found in the data on Denmark. Further examples of decentralized tax systems is found and applied to the rates in Canada and Japan.

[15] Currently, this dataset is also available on request from the OECD.

rate, the thresholds for this top rate may be at different levels. This implies that it may vary considerably how many people actually are subject to the top marginal tax rate. Moreover, as surtaxes and tax credits are not considered in the measure, it is still a "noisy" measure, but possibly less "noisy" than the data used in Lee and Gordon (2005).

Furthermore, the harmonized measure of self-employed used in this study entails both unincorporated and incorporated business-owners. Given that the profit of the latter group of self-employed is likely to be taxed under corporate income tax, inclusion of the corporate income tax (*cit)* is motivated. Moreover, as mentioned previously, the corporate tax has been argued to possibly affect the decision on incorporation. This choice is of discrete nature; hence the average effective rate is an appropriate measure. According to de Mooij and Nicodeme (2007), the statutory corporate income tax is an appropriate approximation of this measure. This data was taken from the World Tax Database from the Office of Tax Policy Research (OTPR).

3.1.3 Institutional quality and other control variables

Given that Nyström (2008) found empirical evidence of the significance impact of business, labour and credit regulation (*reg)* on self-employment as well as legal structure and security of property rights (*leg)* but did not investigate the impact of taxes in combination with these institutional variables, these will be included as control variables in this study. The variables are collected from the Fraser Institute on Economic Freedom in the World (2014) and are qualitative measures on a 0 to 10 point scale. The assessment is performed by the Frasier institute based on several aspects per indicator, where a high value is positive and implies well regulated markets and protection of property rights. Between 2000 and 2008 the indicators are available on annual basis while between 1980 and 2000 it is available on a five year basis. Following Nyström (20008) the data will be interpolated during this period in order to match our annual data on self-employment[16].

In line with above theoretical framework and following the empirical model of Torrini (2002), the indicator of corruption perceptions (*corrp)* will operationalize overall institutional quality in the economies, and will be used in order to calculate dummies for higher or lower than average perceived levels of corruption. In studies of self-employment, Robson and Wren (1999) argue that it is desirable to include variables that control for both the ease and the severity of the punishments when tax avoidance is detected. In their study such indicators were not available. Even though corruption perception is not a distinct measure of present punishments, it might be seen as an indicator of the ease or presence of non-tax compliance in a country. The indicator will be taken from the new data set CANA gathered from different sources and constructed by Fulvio Castellacci and Jose Miguel Natera (2011). This is a large dataset without missing values for 34 indicators for 134 countries between 1980 and 2008, including a measure of corruption perceptions.

Since a large welfare state have been argued to negatively affect entrepreneurship, an indicator government size (*gov_size),* also collected from the Fraser Institute, will be added as a control variable. This is also a qualitative measure based on several aspects, such as government expenditures and number of public companies.

All the below variables, controlling for economic and demographic factors, is taken from the COMPENDIA database by EIM Business and Policy Research (2013).

Based on the theoretical discussion on alternative costs of self-employment an indicator of gross replacement rate (*grr*) will be included. Another alternative cost that has

[16] This is of course not without risks and warrant cautious interpretation. This is why the results and their robustness will be briefly tested and commented on in section 4.2.3 in chapter four with the results.

been used in the literature is average wage, unfortunately this indicator was not available for the given sample.

As previously mentioned, one of the most important factors affecting entrepreneurship found in previous research, is the access to finance. Nonetheless, due to limited available data for longer time periods this aspect is often approximated by other indicators. As in Baliamoune-Lutz and Geller (2011) an indicator on domestic credit by banking sector as a share of GDP (*bank_credit*) will be included, in order to measure financial development and capturing the aspects of access to finance. It measures the domestic credit to various sectors on a gross basis, with the exception of credit to the central government. Another indicator of access to finance could be the level of household savings, as is included in Fölster (2002), unfortunately it was not available for the given period nor was an indicator of the start-up costs used in Baliamoune-Lutz and Geller (2011). Therefore, another variable that may partly approximate available finance is the level of GDP. As in previous research (Torrini 2002, Fölster 2002, Parker and Robson 2004, Nyström 2008), levels in GDP per capita *(lag_gdp*) and unemployment (*unemp*) will be included to control for economic development, demand and business cyclical fluctuations. Purchasing power parity as of 2000 is used to make the GDP measure comparable across countries. Theoretically, the sign of GDP per capita could be expected to be either positive or negative as was discussed previously. Given the previous described theoretical link between entrepreneurs and growth, including GDP growth would clearly introduce endogeneity bias to the model. However, Fölster (2002) argue that GDP per capita might be endogenous too and due to this, the variable will be lagged by one year in order to reduce the potential simultaneity problem. This reduces the sample by one year.

As has already been mentioned, an indicator of female labour as a share of total labour force (*femp*) will also be included. Industry composition is debated as a determinant, but self-employment in service sectors are likely to be larger, hence an indicator of share of services in total employment (*serv_emp*) will be included. To control for demographic changes during the period, dependency ratio (*dep_ratio*) is included as in Fölster (2002).

3.1.4 Time and country dimension

The above mentioned variables constitute a balanced panel data set consisting of yearly observations for 17 OECD countries over the years 1982-2008. The delimitations of the time period were chosen from an availability perspective. The data on corruption perception only stretches as far as 2008 and limitations in the available tax data set determined 1981 as the first year of observation, while lagging GDP per capita reduces the sample by one more year. Countries were gradually dropped as more variables were added to the data set. The time period naturally excludes new OECD members such as the Eastern European countries, but the final set of countries is however believed to give fairly well representation of the OECD countries as a whole.

Table 1. Sample of OECD countries

Austria	Germany	Portugal	United Kingdom
Belgium	Greece	Spain	USA
Canada	Italy	Sweden	
Denmark	Japan	Switzerland	
Finland	New Zeeland	The Netherlands	*17 countries in total*

3.2 Empirical strategy and model specification

A multiple linear regression will estimate the effect of taxes on the rate of self-employed, using the above described variables. Proportions and percentages are kept unlogged but

currency units such as GDP per capita are logged (as is usually done for the sake of normality and in line with for example Stenkula (2009) and Fölster (2002)).

Causal inference is dependent on how unobserved factors are taken into account (Angrist and Pischke 2009, p. 221). Panel data or time series cross sectional data (TSCS)[17] have the advantage of allowing the researcher to control for individual heterogeneity, i.e. that unobserved individual-specific effects may be taken into account for. These unobservable individual-specific effects, which in this case are country-specific effects, may be assumed to be either random or fixed and the choice between specifications depends on the nature of the data. If observations are randomly drawn from a large population, a random effects model is most suitable, but if observations correspond to a specific country where the inferences are conditional on the observed units, a fixed effects model is more appropriate (Veerbeck 2001, p. 385, Beck 2001). Naturally, the fixed effects model can be expected to be the appropriate choice in this study with TSCS data.

Formally the fixed effects model is specified as a regression model in deviations from individual means, where the transformation is called the within transformation. In the first stage, the mean of each variable, noted by a "~" in equation 4 below, is withdrawn. This produces equation 5 where the country specific effect, α_i, is dropped as we assume the effects are constant over time, hence $\alpha_i = \tilde{\alpha}_i$.

$$y_{it} - \tilde{y}_i = (x_{it} - \tilde{x}_i)'\beta + (\alpha_i - \tilde{\alpha}_i) + (u_{it} - \tilde{u}_i) \quad (4)$$

$$y_{it} - \tilde{y}_i = (x_{it} - \tilde{x}_i)'\beta + (u_{it} - \tilde{u}_i) \quad (5)$$

The y_{it} equals the rate of self-employment, x_{it} is a vector of different explanatory variables (including the tax variables), i and t are cross country and time-dimensions and u_{it} is the error term and α_i represents the unobserved country factors. Exactly the same estimator for β that is obtained by the within transformation is possible to obtain by including dummies for each country, represented by d in equation 6 (Verbeek 2009, p. 377). This generates the least squares dummy variable (LSDV) estimator and will be the model employed in the empirical investigation of this thesis:

$$y_{it} = x_{it}'\beta + \sum_{j=1}^{N} \alpha_j d_{ij} + u_{it}, \quad (6)$$

where d_{ij} = 1 if i=j and = 0 otherwise. The baseline model will be the LSDV model with country dummies (one-way fixed effects model) but since the time dimension is fairly large a two-way fixed effect model, i.e. a model that controls for both unobservable country specific effects and time-specific (common trend) effects, will also be explored, in order to fully control for possible omitted variables and homogenous shocks (captured by the time dummies). By doing this, the model will eliminate all endogeneity problems related to unobserved country and time specific factors, given that the error term is identically independently distributed (IID) (ibid, p. 376).

Lastly, the asymptotics of the TSCS data set is crucial to consider when dealing with various diagnostic tests. Potential heteroskedasticity, autocorrelation and contemporaneous correlation will bias the results of the within estimator (Podesta 2003). Several panel adjusted measures rely on asymptotics of fixed T and N$\rightarrow \infty$, which is not the

[17] So far the data has been defined as panel data, due to the combined dimensions of cross-section and time series. However, it might be more correctly to define it as time serie- cross sectional (TSCS) data or panel time series as the asymptotics for TSCS data are in *T,* while opposite is valid for traditional panel data, where *N* is typically large and *T* is typically short or moderate (Beck 2001).

case with our data. Beck (2001) who has focused specifically on appropriate methods for TSCS data, proposes that *panel corrected standard errors* (*pcse*) should be used in order to deal with heteroskedasticity and contemporaneous correlation. Since the data suffer from both problems, *pcse* will be employed. The data also suffers from first-order serial correlation and two alternative remedies are typically available: treat the potential serial correlation by using a first order autocorrelation, structure, AR (1), or incorporate over time persistence in the model specification and include a lagged dependent variable (Beck 2001). Both methods will be investigated, but the former will be the default in all regressions. Lastly, normality is assumed based on the large number of observations.

3.2.1 Exploring the model

In order to investigate the effects of taxes on self-employment it may be informative to allow for different forms of these effects, which will be done by the below extensions of the baseline model.

Lagged and non-linear effects

In the literature, lagged tax effects have been indirectly investigated in different dynamic models (for example, one year lagged effects are de facto estimated in the error correction model employed by Robson and Wren (1999) and in the generalized method of moment model used in Baliamoune-Lutz and Geller (2011)). Moreover, since Stenkula (2009) argues that self-employed adjust relatively slow to changes in the economic environment, investigation of lagged tax effects is motivated to shed light on a potential delayed relation. Given the possible slow adjustment, a lagged effect on up to three years will be investigated. Using lagged tax variables may also be seen as a remedy to avoid issues of simultaneity, as was discussed in relation to the lagged indicator on GDP per capita (the other explanatory variables are however assumed to be exogenous). [18]

In addition, as an advantage of the many observations of panel data, non-linear effects may be allowed for, by including squared terms of the tax rates. It may be important to control for, due to the possibly opposing effects (of labour effort and tax avoidance) of the tax variable that was described in chapter 2.

Tax-effects conditionally on institutional quality

In order to explore the second research question of this thesis, namely, does the effect of the taxes vary with the quality of institutions; additional interaction terms will be included. First and following Torrini (2002), a dummy is created for countries with higher or lower corruption than average (compared between the countries included in this sample). The dummy will be noted *cpi* (corruption perception index) and takes on the value of one when the corruption perception is higher than average. In a fixed effects model, the effect of a

[18] For the within-estimator to be consistent, the explanatory variables are required to be strictly exogenous (Verbeek 2009, p. 376-8). Given that panel data reduces the effects of omitted variable bias by the LSDV model, reverse causality may instead be the cause to problems of endogeneity. The tax variables can introduce simultaneity problems to the model as taxes may not only affect self-employment but self-employment may affect taxes. This is usually a problem in micro dataset where individuals are followed over time (where the taxes is a endogenous outcome of the occupational choice) but it also has been argued as a potential problem on aggregated level. The rational is that policymakers may change the taxes as a respond to shocks in levels of entrepreneurship or that a large and potential influential group of self-employed may lobby for change in certain forms of taxes (Baliamoune-Lutz and Geller 2012, Hansson 2011, Robsson and Wren 1999). As a robustness check for the estimates and to mitigate potential simultaneity, lagged tax variables will therefore be included. This is based on the argument that past levels of taxes may affect current self-employment rates, but changes to current self-employment rates is likely not to affect past tax rates.

constant dummy, would be washed out if included individually[19], however, as an interaction term it may be included. Interaction terms with all three different tax variables will be investigated, in order to evaluate the effect of taxes on self-employment where corruption is relatively higher.

Dynamic model

Including a lagged dependent variable, as was briefly mentioned before, is a possible statistical solution to issues of serial correlation. A dynamic model should nevertheless be based on theoretical arguments. First, it may be motivated by the statistical characteristics of the data, i.e. the large degrees of persistence in the self-employment rates (recall figure 1 in chapter 1) Furthermore, given that Torrini (2002) suggests that tax morale of the self-employed may be conditional on the institutional set up (in his case: the presence of corruption) one might argue that self-employment is partial endogenously determined in the sense that attitudes around potential entrepreneurs affect the decision to become one. Thus, given the persistence and possible habit formation, including a lagged dependent variable seems plausible[20]. However, within a dynamic framework, the within estimator is usually claimed to be both biased and inconsistent when including a lagged dependent variable, as it is correlated with the error term[21] and hence violates the required assumption of strict exogeneity. However this severe (downward) bias of the lagged dependent variable, known as Nickel bias or Hurwicz bias, is especially crucial in case of traditional panel data asymptotics with large N and small *T* (<10). The bias disappears when $T \rightarrow \infty$ and is hence a small sample bias that may be acceptable for larger TSCS data (around T>20) (Verbeek 2009, p. 396-7, Beck 2011). Nonetheless, we should be aware of the small bias that does remain.

Robust test – stationarity

The time length of the data employed in this study is of a rather moderate size compared to the longer time series data that have been used to study self-employment using cointegration techniques[22]. For example, Nyström (2008) studied a similar period but did not consider the issue of non-stationarity. Furthermore, Beck and Katz (2011) argue that with political economy data the data may be very persistent, but still stationary on a longer term on which data is not available for and this could naturally be the case here. They also stress the importance of not solely accepting statistical results in the search for unit roots, but to ask whether the characterization of the data having a unit root is plausible. They highlight, as an example, that if series have a unit root, they would tend to wander away from their means, however, as political economy data often is measured as proportions, the data (its variance) is bounded by definition. This is the case in most of the data employed in this study including the dependent variable and the tax variables in focus[23].

[19] This is because the mean of a constant value is the constant value it self. Recall that the fixed effects model imply that the mean of a variable is subtracted from its every observation (see equation 3), a constant variable would thus be dropped from the regression.

[20] A third argument would be that of Fölster (2002) who includes a lagged dependent variable based on lingering effects on inflow of self-employment due to lower taxes one period, which then result in higher rates the following years.

[21] This could be seen in a simple example. Note that a first differencing specification, instead of the transformations in equation 4 or 6, yield the same estimate if T=2 (the country specific effects,α_i, are "differenced out" in below equation). Given this, it is easy to see that a lagged dependent variable, y_{t-1} included in a differenced model will be correlated with the error term, since both will be a function of the error term u_{it-1} (Angrist and Pischke 2009, p. 245): $\Delta y_{it} = \emptyset \Delta y_{t-1} + \Delta x_{it}' \beta + \Delta u_{it}$.

[22] Typical examples are Stenkula (2009) and Bruce and Moshin (2006) who study self-employment during a 50 years period, in Sweden and US, respectively.

[23] Moreover, if either series were integrated of order one, *I*(1), the implication is that it would be equally likely to see an increase or decrease in either variable regardless of its present value. Hence one may ask if there really is

However, keeping the above in mind, as an additional robustness test one might control for potential unit-root processes. Two tests are used for this purpose, where one does not indicate problems of non-stationary data in any variable while the other does in some (see further description of the tests in section 4.2.5). In order to be strict in the investigation, I continue with the diagnostic tests and find that a log-log specification imply that the second test indicate more non-stationary data (the result of this test is found in appendix B). Given the relatively limited time dimension of this data one must however note that these tests also may lack power (i.e. have low probability of rejecting a false null hypothesis which in these tests are that all panels have unit roots) (Verbeek 2009, p. 414, Parker 2004). If this is the case, the data may still be stationary and the fixed effects specification is valid.

Beck and Katz (2011) discuss how to estimate slow-moving data, in the light of somewhat uncertain risks for unit roots, and they suggest the error correction (EC) model[24]: "…whether series are integrated or stationary but slowly moving they may well be modeled by EC specification" (ibid, p. 344), which they write as:

$$\Delta y_{it} = \sum \Delta x'_{it} ø + \gamma (y'_{it-1} + \sum x'_{it-1} к) + u_{it} \qquad (7)$$

The parameter γ in front of the parenthesis is the error-correction parameter and if it is significantly different from zero there is a long-run relationship between *x* and *y (*the model suggests that the integrated variables y and x are cointegrated) (Beck and Katz 2011). As this is the model that Robson and Wren (1999) use, the choice of this specification enable our results to be compared. Their model is an "unpacked" version of above, which Podesta (2003) also uses to estimate non-stationary data, [25] and this is the model that will be estimated:

$$\Delta y_{iy} = \sum \Delta x'_{it} ø + y'_{it-1} \gamma + \sum x'_{it-1} \beta + u_{it} \qquad (8)$$

The parameter for the lagged dependent variable, γ, is expected to be between 0 and -1 in order to satisfy equilibrium properties of the model, where shocks are reduced over time, causing self-employment to converge to a long-run equilibrium. The parameter for the differenced (change) variable, ø, capture the short-run (*passing*) effect while *lasting* effects is captured by the parameter, β, in front of the lagged independent variables (Podesta 2003). Therefore, as a robust test, I conclude the next chapter of results by being strict in the interpretation of non-stationarity and use this model, however still cautiously, given the limited sample size and the theoretical risks of incorrectly characterizing the data as non-stationary.

---

no tendency for variables in focus to be more likely to rise when they are low and fall when high (Beck and Katz 2011).

[24] Whether or not the persistence is due to high serial correlation or a true sign of unit roots, a first difference model would also be plausible. First, because first differencing usually is an effective remedy to serial correlation and second, because all variables in this sample are found to be stationary in first-differences (implying that they are integrated of order one). However, unlike the EC model this model only captures short-run effects and not maturation effects (Podesta 2003).

[25] Podesta (2003) highlights that this EC model allows for an asymmetric relationship to be modelled, unlike classical cointegration methods that implies a symmetric relation. In this case, a symmetric relation would mean that self-employment adjust to changes in tax rates and institutions in the long run (if they are cointegrated), but the opposite also holds and the latter might be theoretically questionable. The asymmetric assumption implies that self-employment adjusts to a long-run equilibrium but not any of the independent variables. As noted by Verbeek (2009, p. 347-8) this requires the assumption of strictly exogenous *x*, which is materialized in the formula of Podestra (and Robson and Wren 1999) by the inclusion of $\Delta x'_{it}$ witch hence is assumed not to affect the error-correction term.

Chapter four

Results

This chapter presents the results of the study. First, the baseline regression will be analyzed followed by some extensions of the initial regression. Except for the output of the baseline regression, only the results for the tax variables will be shown in this chapter. All other output will be found in Appendix B. As described earlier, all regressions, if not stated otherwise, is estimated using panel corrected standard errors and corrected for a common AR (1) process across all panels. Correlations between the explanatory variables were analyzed initially and due to high correlation (0.8) between GDP per capita and share of employment in the service sector, the latter were dropped from the sequent specifications (see correlation matrix, table A.3 Appendix A).

4.1 Baseline regressions

Table 2 shows the regression output for the one-way (column 1) and the two-way (column 2) fixed effect (LSDV) model for the sample with 17 countries between 1982 and 2008.[26] The fixed effects, i.e. the country and annual time dummies, will not be reported in any table for the sake of brevity but will occasionally be commented on in the following results.

In the one-way fixed effects model, five out of thirteen variables are significant on either a 1, 5 or 10 percent level. One of these has a somewhat unexpected sign. The degree of female participation in the labour market, *femp*, seems to have a positive impact on self-employment, opposite to the theoretical predictions. One possible explanation to this could be that the measure captures something else in the economy with a positive impact, which is confirmed by the fact that when controlling for time effects (column 2), the magnitude of the parameter decreases and the significance is lost. Also opposite to the theoretical predictions, *bank_credit* seems to have a negative effect of self-employment. However as in Baliamoune-Lutz (2011) it is not significant. Unemployment, *unemp*, is significant in all two specifications and indicate a negative effect on self-employment, implying that necessity entrepreneurship is not dominating the measure of self-employment, instead the "insurance" effect seem to be supported by the data (entrepreneurship is not mainly a "last option", instead, if there are salary jobs available (low unemployment) for the potential entrepreneur this may be perceived as an "insurance" and decreases the risk associated with an entrepreneurial venture). As an indicator controlling for business fluctuations it suggests that people tend to become self-employed during better economic times. However, the effect is rather moderate: a 10 percentage point decrease in unemployment is predicted to increase self-employment roughly by 0, 6 percentage points[27]. The institutional variables are also significant and also have the largest positive impact on self-employment in both specifications, except for the measure on legal structures and security of property rights *leg* that is insignificant with time effects. This seems to support theoretical predictions that especially business regulation as well the size of the public sector do seem to have an important impact on the incentive structure of self-employed. Recall the (bounded) scalar for these measures: the higher values of *reg*, the better conditions regarding business and labour regulation prevails in the country (according to the

[26] Note that all values within parenthesis in the tables are *p-values.*

[27] Evaluated at the sample mean of self-employment of around 11 percent, it roughly represents an increase of 5, 5 percent.

Table 2. Baseline regressions, with country (1) and annual time fixed effects (2).

Variable	(1)	(2)
LAG_GDP	0.370	-0.404
	(0.497)	(0.673)
UNEMP	-0.036*	-0.058***
	(0.059)	(0.005)
FEMP	0.106***	0.049
	(0.000)	(0.120)
DEP_RATIO	-0.010	-0.006
	(0.687)	(0.807)
GRR	-0.009	-0.015*
	(0.254)	(0.051)
BANK_CREDIT	-0.002	-0.002
	(0.171)	(0.128)
AV_TAX	-0.008	-0.003
	(0.631)	(0.853)
CIT	-0.006	-0.004)
	(0.159)	(0.416)
PIT	-0.002	0.002)
	(0.736)	(0.703)
REG	0.169**	0.385***
	(0.022)	(0.000)
LEG	0.231***	0.083
	(0.001)	(0.286)
GOVSIZE	0.150**	0.202***
	(0.017)	(0.009)
R-squared	0.919	0.927
Obs	458	458

subjective assessment by the Frasier Institute). Higher values of the measure of *gov_size* imply a smaller public sector (there may be relatively few public companies, as an example). The significant effect of the latter measure supports the hypothesis by Davidsson and Henrekson (2002) that a large public sector may crowd out entrepreneurial activity. Concerning the insignificant variable of legal protection and protection of property rights, it might be a consequence of the sample consisting of developed countries; hence, the variation in this variable may not be large enough to capture any influential effects. It is likely that this variable would show larger signs of significance in a larger and more economically diverse sample. When the time dummies are included, gross replacement rate *grr* also gains significance and show a small but expected negative effect. Lagged GDP per capita *lag_gdp* is not significant and this is in fact consistent with several other studies that did not find any robust significant relationship between self-employment and income levels (Nyström 2008, Stenkula 2009, Robson and Wren 1999) however opposite to Parker and Robson (2004) and Fölster (2002).

Lastly, the parameters of the tax variables are small, and as in previous literature, indicate both positive and negative effects, however and most importantly, none of the tax variables are significant in this baseline specification. As noted in chapter 2, previous

research has found both positive and negative effects from tax on entrepreneurship. However, insignificant effects have also been found by Gentry and Hubbarb (2004) and Baliamoune-Lutz and Geller (2011), using measures of marginal and average taxes on UK and OECD data respectively. A careful interpretation of these insignificant results may however be that the channels of impact are not correctly captured by this specification. Using this novel harmonized measure of cross-country rates on self-employment, the next section will therefore be devoted to further investigation of this relationship (as specified in section 3.2.1).

It should be noted though, that the introduction of country dummies increases the variance inflation factor in the model substantially. When analyzing the explanatory variables and different specifications, it seems that some of the explanatory variables are correlated with the country dummies (as the colinearity is not affected if one or several explanatory variables are dropped and is less driven by the time effects). If these country effects are excluded, keeping only the time effects, all variables but the tax variables and regulation are significant. However, this is neither surprising nor desirable as the advantage of panel data is the possibility to control for heterogeneous effects by the country fixed effects, in order to get unbiased estimates. The problem is not unusual; in fact, it is likely to be a problem when including country fixed effects in a model where the variables change slowly over time (Podesta 2003)[28]. In short multicollinearity implies reduced efficiency where true significant parameters might be rejected, while excluding dummies, as a remedy, might imply biased estimates. The main interest in this thesis is the tax variables and fortunately two of three tax variables indicate no problem of colinearity. Average tax does however indicate problems (VIF tolerance above 0.1), though, it is not a surprise that it is collinear with the country fixed effects as there are limited variance in the measure (recall the mean trend in figure 4, chapter 2). This might explain possible insignificant results of the variable in the subsequent results. Concerning the control variables, *lag_gdp, femp* and *leg* are some of the most affected and might explain their insignificant signs. As dropping variables or getting more data does not solve the problem or is available as an option, I therefore opt to proceed with the given specification, especially since unbiased estimates of the tax variables is prioritized to the option of significant control variables. To conclude, it should be noted though that all regressions presented in this chapter have a mean VIF below the tolerance level which somewhat strengthen the overall validity.

Lastly, computation of goodness-of-fit measures for panel data is "somewhat uncommon" (Verbeek 2009, p. 386), since usual R-squared and adjusted R-squared criteria are appropriate only if estimated by OLS. In this case the OLS is used to estimate the LSDV estimator; however, given the large amount of dummies in combination with the slow-moving data structure, the R-squared values are very high and somewhat misleading, [29] which is common in the context of panel data. In this case, the reported R-squared may cautiously be used to compare different specifications to each other but not to evaluate the estimation.[30]

[28] Statistically, colinearity problem implies that the X'X matrix is not invertible, leading to high standard errors and individual effects being hard to distinguish; i.e. colinearity reduces the efficiency of the estimates and increases the risk for type II errors. However, if the colinearity problem is affecting variables that we are not mainly interested in (control variables) this problem does not affect the standard errors of the variables we *are* interested in. Hence, the colinearity of the former may be "almost arbitrarily high" (Verbeek 2009, p 43-45). Nonetheless, there is a potential trade-off between including country effects or not: between a parsimonious model (as a multicollinearity remedy) and a fixed effect specification (as a remedy for endogeneity caused by omitted variables) (Podestra 2003).

[29] This is because the R-squared takes into account the variation explained by the fixed effects, i.e. the country dummies, and these effects in great extent perfectly fits the variation. This is somewhat misleading as they do not explain the variation; rather they cover up the model from omitted variables (Podesta 2003).

[30] Had the estimation been made by the within transformation, an alternative statistic to use would be the within R-squared (Veerbek 2009, p. 384-6). In the preferred estimation of the LSDV model with panel corrected standard errors, this statistic is however not reported in the statistical package, neither is the adjusted R-squared.

Table 3. Lagged effects of taxes with country and time fixed effects

Variable	(1)	(2)	(3)
AV_TAX L1.	-0.002		
	(0.891)		
L2.		-0.004	
		(0.810)	
L3.			0.025
			(0.101)
CIT L1.	-0.000		
	(0.950)		
L2.		-0.003	
		(0.464)	
L3.			-0.009**
			(0.040)
PIT L1.	0.003		
	(0.611)		
L2.		-0.004	
		(0.463)	
L3.			0.002
			(0.775)
R-squared	0.932	0.939	0.942
Obs	441	424	407

4.2 Results of exploring the model

In this section the baseline regression is explored in accordance with the empirical strategy that was previously described. As previously mentioned, only the tax variables will be presented, where the full output will be found in Appendix B.

4.2.1 Including lagged effects

Stenkula (2009) highlights the plausible slow adjustment of self-employment to surrounding economic conditions which warrant further study of potential lagged effects of taxes on self-employment. Seen in table 3 below, using the two-way fixed effects model (i.e. with country and annual time dummies), lagged effect of *cit* is indeed significant in the third lag (column 3). Notably is also that the *av_tax* is just almost significant on a 10 percent significance level in the third lag (column 3). The *cit* seem to be negative in all lags, while *av_tax* have a positive effect only in the almost significant third lag. As described earlier, including lags may be a remedy strategy to avoid simultaneity problems between our variables of interests, and these results suggest that there indeed is some one-way effect of taxes on self-employment. Cautiously looking at the R-squared values also indicate that the model with three lags improves the fitting. This positive effect of the average tax is in line with the theoretical and empirical findings in cross-country studies by Parker and Robson (2004) and Robson and Wren (1999), where the latter argue that an increase in average tax, holding the marginal tax rate constant, makes self-employment a relatively more attractive employment

However, using the formula for adjusted R-squared, $\bar{R}^2 = 1 - (1 - R^2)(\mathrm{n} - 1)/(\mathrm{n} - \mathrm{k} - 1)$ (Wooldridge 2012), and cross-checking some values by hand yield approximate the same values as the R-squared, only slightly smaller, therefore these will not be reported.

mode compared with paid employment. However, these scholars found statistically significant and also larger effects than these results show[31]. As Robson and Wren (1999) studied a shorter period (14 years) one reason might be that this large effect is bounded to their sample period. However, Parker and Robson got a somewhat similar magnitude of their effects and they studied a period between 1972 and 1996. One thing they do have in common is the use of the non-harmonized dataset from OECD LFS where incorporated business is not consequently included in all countries. As they are included in the measure in this study it might point to the fact that the unharmonized measure exaggerates the positive effect of average taxes. Nonetheless, lagged and hence delayed effects in one and almost two of the tax rates are supported by this data. One may also note that the magnitude of the effect also increases with lags, which also supports the delayed relation where the decision of starting or closing a business is not an immediate decision of changes in taxes.

Concerning the control variables, those who were significant in the previous two-way model (column 2, table 2) seem to be robust, while the effects seem to be slightly larger in this model.

4.2.2 Allowing for non-linear effects

When including squared terms in the two-way fixed effects model in table 4, both corporate and personal income tax turns significant in levels and as squared terms (seen in column 2 and 3, respectively) and in a combined specification (column 4). However, the effects of the squared terms are almost negligible. Nonetheless, the interpretation of the inclusion of these terms is that corporate tax up to a certain level seems to have a positive impact on self-employment while above this level it turns negative. The opposite relation is found for personal income tax. Again, average tax is not significant. The result suggests that while a linear effect on self-employment is not supported by the data in the baseline regression, a non-linear effect seems to be confirmed. Even though Robson and Wren (1999) received higher effects of their marginal tax measure, the negative effect of *pit* is in line (even slightly larger) than the effects found by Bruce and Moshin (2006) and Stenkula (2009) in US and Swedish data, respectively. This result suggests that a 10 percentage point decrease in *pit* would imply roughly a 0, 4 percentage point increase in self-employment. Evaluated at the sample mean of the self-employment rate of around 11 percent of the total labour force, this would increase self-employment by 3 percent. Given that top statutory income tax vary more between countries compared with the corporate income tax (there are several countries where the difference in marginal taxes is 10 percent or more), this explains the different self-employment rates to a greater extent[32]. In contrast to Robson and Wren (1999), Stenkula (2009) and Bruce and Moshin (2006) used the same measure of marginal taxes as is used in this study, which might also explain the different magnitudes in effect of taxes on self-employment.

The "U"-shaped effect of taxes on self-employment is explained by the possible counteracting effects of labour effort and tax avoidance (Garret and Wall 2006). Given this, the signs of *pit* might indicate that the negative labour supply effect dominates, but at higher rates this negative effect seem to be somewhat offset, possibly by increased behavior of tax avoidance at higher tax rates, according to this theory. Or in terms of the theory of Cullen and Gordon (2007) and Djankov et al (2010), at lower rates the negative effects of risk subsidies and income shifting is dominating, while being slightly offset at higher rates by the positive

[31] Robson and Wren (1999) found that self-employment would increase by 0.8 percentage points when average tax increased by one unit, compared with this value of almost 0,03 percentage points.
[32] In fact, Fölster (2002, p. 138) receive somewhat similar estimates which he consider is "quite a lot".

Table 4. Quadratic model with country and time fixed effects

Variable	(1)	(2)	(3)	(4)
AV_TAX	-0.018			-0.011
	0.585			0.730
AV_TAX squared	0.000			0.000
	0.597			0.745
CIT		0.029*		0.048***
		0.066		0.009
CIT squared		-0.000**		-0.001***
		0.043		0.006
PIT			-0.028*	-0.041**
			0.052	0.010
PIT squared			0.000**	0.000***
			0.045	0.008
R-squared	0.922	0.926	0.922	0.928
Obs	458	458	458	458

effect of risk-sharing with the government. However, the opposite relation is shown by the *cit,* indicating that tax avoidance (or risk-sharing) would be the dominating effect at first. As *cit* is levied on mainly the incorporated businesses, it might indicate a somewhat different behavior between incorporated and unincorporated businesses, but the "U"-shaped effect gets somewhat support from this data. Compared with the previously lagged and negative effect one might speculate if the immediate effect of a tax raise in corporate income tax spurs tax avoidance whereas the negative labour supply effect dominates later on.

Concerning the control variables, the signs and effects of *unemp, femp, grr* and the two institutional variables are robust to the earlier specifications (see table B2 in Appendix B).

4.2.3 Exploring tax effects conditionally on institutional quality

Given that the convex relationship between self-employment and taxes seem to get some support by this data on harmonized data on self-employment, it seems valid to pursue with the investigation of the second research question: does institutional quality have an impact on the effect of taxes on the self-employed? So far, the results seem to support the idea of the importance (and positive effect) of the quality on regulation on self-employment. What about when institutional quality overall may be relatively low in the economy due to for example corruption?

Following Torrini (2002), a dummy was created based on average perceptions on corruption in each country for the whole period. The countries that through out the sample period had higher rates than average were coded "1" and the countries with lower than average corruption perceptions were coded as "0". Of 17 countries 7 countries had higher than average corruption in this sample. In table 5, the results are found where interaction terms between the dummy and the respective tax measure have been included to the two-way fixed effects model.

Again, the previous significant control variables are robust in signs and magnitude while *av_tax* is still not significant, nor *pit*. However, turning to *cit*, both terms are significant and both have their expected signs (seen in column 3 and 4). These results indicate

Table 5. Interaction terms between taxes and corruption, with country and time fixed effects.

Variable	(1)	(2)	(3)	(4)
	(0.015)	(0.011)	(0.003)	(0.005)
AV_TAX	0.007			0.023
	(0.678)			(0.182)
*AVTAX*CPI*	-0.029			-0.036
	(0.392)			(0.299)
PIT		-0.005		-0.002
		(0.402)		(0.727)
*PIT*CPI*		0.012		0.008
		(0.179)		(0.386)
CIT			-0.012**	-0.011**
			(0.018)	(0.03)
*CIT*CPI*			0.031***	0.028***
			(0.001)	(0.006)
R-squared	0.922	0.924	0.936	0.933
Obs	458	458	458	458

that when corruption is lower than average, *cit* has a negative effect on self-employment, while in the presence of relatively high corruption, the effect is positive. Hence, the findings of Torrini (2002), seem to be robust in longer time series data and concerning corporate income tax, in fact, the magnitude seems to be slightly larger with this data. Note however, that when Da Rin et al (2011) studied the effect of corporate income tax interacted with "good" or "bad" tax accounting standards on firm level he does not find a positive effect in the latter countries, but rather a smaller negative effect. Hence, on an aggregated level the effect of low law enforcement seem to be larger.

So far some interesting patterns have emerged in this data. The effect of taxes on self-employment is not necessarily linear and counteracting effects demand careful modeling of the data generating process. It might suggest that there indeed is a negative effect of tax on self-employment, but given the institutional environment in a country, the impact may be altered. This warrants a contextual analysis before policymakers start tweaking in the respective tax schedules, in order to receive the desired results.

In addition to the interaction terms, the institutional indicator on regulations seems to have a robust impact, as well as the government size. However, the results should be interpreted carefully as some of the observations have been generated by interpolation were there were missing values (as done in Nyström 2009). The robustness of the results was however tested by rerunning the above regressions were the missing data instead were replaced by the previous known value in each indicator. This created small periods (5 years) of constant values in each indicator and hence less variance in the sample. Overall these alternative estimations indicated

almost the same robust results what regards the above significance and signs in the tax variables. The same was found with respect to the regulation indicator, however the magnitude were slightly smaller, however still large. The indicator on government size was also somewhat smaller and did turn up significant in much fewer specifications, however where significance were not robust, in most cases the p-value was just above the significance level of 10 percent. Hence, given the similar results and for the sake of brevity these results are not reported.

Table 6. Lagged dependent variable with country and time fixed effects.

Variable	(1)	(2)	(3)	(4)
SEMP L1.	0.883***	0.869***	0.887***	0.875***
	(0.000)	(0.000)	(0.000)	(0.000)
AV_TAXS	-0.000		0.008	0.001
	(0.955)		(0.656)	(0.946)
CIT	-0.003		0.025*	-0.008**
	(0.362)		(0.053)	(0.022)
PIT	-0.000		0.017**	0.003
	(0.994)		(0.019)	(0.373)
AV_TAX L3.		-0.003		
		(0.673)		
CIT L3.		-0.005*		
		(0.065)		
PIT L3.		0.006*		
		(0.066)		
AV_TAX squared			-0.000	
			(0.685)	
CIT sqaured			-0.000**	
			(0.036)	
PIT squared			-0.000***	
			(0.006)	
*AV_TAX*CPI*				-0.001
				(0.949)
*CIT*CPI*				0.014**
				(0.013)
*PIT*CPI*				-0.007
				(0.197)
R-squared	0.996	0.996	0.996	0.996
Obs	441	407	441	441

4.2.4 Lagged dependent variable

As was previously discussed, persistence in self-employment rates and the possibility of habit formation motivated the inclusion of a lagged dependent variable. Moreover, Beck (2001) suggests that serial correlation should be dealt with by including a lagged dependent variable.

The results are found in the above table 6. The lagged dependent variable turns up strongly significant which is not surprising given the persistence in the data. However, most of the other control variables loose their significance except for *unemp* and *reg*. Lagged GDP per capita also turns significant and has a negative impact on self-employment rate. This might suggests that higher income levels imply more capital per worker and hence larger firms in general and/or supports the argument of Fölster (2002) that countries with lower income levels have less efficient distribution system which creates job opportunities for self-employed. The variables *grr*, *gov_size* and *leg* are only partly significant in different specifications (compare column 2-4, table B.4, Appendix B). This might be a reason for not to prefer this specification as with lagged dependent variables it is common that the results are too strongly driven by this variable. Therefore it is interesting to note that *reg* is robust in its significant impact and moreover, that all the above results on taxes seem to be somewhat

robust in significance. The earlier significant lagged effect of *cit* is robust in this specification as seen in column 2, and *pit* is also significant and positive in this specification. Notable is that both *pit* and *cit* in the non-linear specification (column 3) seem to be dominant by possibly tax avoidance behavior at lower rates, while dominant by the negative labour supply factor at higher rates, which is opposite to what Georgellis and Wall (2006) found. When including a lagged dependent variable in the specification with interaction terms, the expected sign and significance of *cit* is robust. Even though this specification might not be our preferred one, it strengthens the hypothesis that *cit* does have a significant, although small effect on self-employment.

4.2.5 Stationarity

Beside the earlier discussed theoretical reasons and the forthcoming statistical reasons, test for non-stationarity may be motivated due to the relatively high significant parameter of the lagged dependent variable, in the previous section (seen in the first row of table 6). A parameter close to 1 may indicate non-stationarity. However, in light of the previously discussed theoretical and statistical reasons (see section 3.2.1) against the possibility of unit roots in this sample, the following analysis is pursued cautiously as a diagnostic test of the robustness of previously results.

The Levin-Li-Chu (LLC) test does not indicate unit root in the self-employment variable. However, since the Im-Pesaran-Shin (IPS) test is said to be a more powerful (Parker and Robson 2004) and less restrictive test[33], this test is also used and three different functional forms are tested. First, including a trend and subtracting cross-sectional averages from the series, secondly using only a trend and third including neither of these two factors. Lags are included based on the akaike information criteria (AIC). The null hypothesis is that all panels have unit roots and the alternative is that a fraction of the panels are stationary (Verbeek 2009, p. 413-4). The conclusion is that according to this test, self-employment and all tax variables seem to be non-stationary and some of the control variables (see results in table B5 in Appendix B).

Continuing by estimating the error correction model, it may be plausible to omit some of the variables as all parameters are entered twice in this model. Given that *dep_ratio, leg, grr* and *bank_credit* were rarely significant in previous specifications, they were omitted. This roughly corresponds, hence enable further comparison, to the specification employed by Robson and Wren (1999) (except for regulation and government size which are included based on previous significance).

Table 7 presents the results from the error-correction model earlier specified in section 3.2.1. Following Robson and Wren (1999), the estimation is done on the natural logarithm of all but the tax variables. Two-way fixed effects were at first modeled but since almost none of the time effects were significant in this specification (most likely due to the included first differenced variables that instead capture these annual effects), they were omitted in order to reduce the persisting colinearity in the model. This did not change any of the variables in a notable way. The drop in the value of R-squared is likely due to the dependent variable being measured as changes, rather than in levels.

[33] Assuming that a variable is generated from an autoregressive process, this process can be rewritten to: $\Delta y_{it} = \alpha_i + \pi_i y_{i,t-1} + u_{it}$, where $\pi_i = \gamma_i - 1$. The null hypothesis of both these tests corresponds to $\pi_i = 0$ for all i, whereas the LLC alternative hypothesis assume the same π across all countries. The IPS test allow for different π_i and hence a heterogeneous (less restrictive) alternative hypothesis (Verbeek 2009, p. 412-3).

Table 7. Error correction model			
Variable	(1)	(2)	(3)
SEMP L1.	-0.095***	-0.179***	-0.223***
	(0.000)	(0.003)	(0.000)
AV_TAX. L1.	-0.000	0.002	0.000
	(0.606)	(0.202)	(0.860)
CIT L1.	-0.001***	-0.001	-0.001**
	(0.005)	(0.364)	(0.016)
PIT L1.	0.000	-0.001**	-0.000
	(0.729)	(0.010)	(0.992)
AV_TAX D1.	-0.000	0.002	0.001
	(0.856)	(0.564)	(0.697)
CIT D1.	-0.001	-0.001*	-0.001*
	(0.191)	(0.076)	(0.054)
PIT D1.	-0.000	-0.001**	-0.001
	(0.298)	(0.028)	(0.253)
R-squared	0.179	0.336	0.429
Obs	459	187	187

In column 1, the results indicate that there is a significant error correction process that adjusts self-employment to an equilibrium level, as the lagged parameter in front
of the dependent variable is significant and between 0 and 1. However, according to this result, the adjustment seem to be rather fast where around 90 percent of a one unit shock will die out in a year in the rate of self-employment (calculated by taking 1 minus the parameter of the lagged dependent variable) (Beck 2011). Moreover there seems to be a lasting, negative effect of *lag_gdp* as well as a significant short-run effect of *lag_gdp* on self-employment (see table B.6 in the appendix). The only other significant lasting effect on self-employment seems to be *cit* (see column 1), an indication of this result being robust through specifications and possibly indicating a long run relationship between the variables. Given that this model is a log-linear model with respect to the tax rates, the interpretation of the effect is similar, only slightly smaller, than previously found: Evaluated at sample mean, a ten percentage point decrease in *cit*, would yield approximate a one percent increase in the self-employment rate. The interaction terms of taxes and corruption dummies was also included to the model as further test, however, without
significant result, why for the sake of brevity these results are not reported. This model is possible to augment in different ways to fit the data generating process (for example by adding other leads or lags of the explanatory variables); there is a possibility that such measures would have improved the significance of some variables and the model as a whole, but as the control variables are of secondary interest of the study it will not be investigated further.

In the second column an attempt to further compare the results with Robson and Wren (1999) is done, by decreasing the sample size to match the one used in their study[34]. The lasting effect of income level vanish, however, notably is that *pit* now show, similarly as in their study, a significant negative effect on both short and long run, while the effect of *cit* is no longer significant. Instead there is a significant short-term effect. Even though not significant (which again might be due to problems with multicolliniearity), it is also notable

[34] They study the period 1978-1992 (with three year gaps), but since the dataset in this thesis start from 1982 the period 1982-1992 on annual data is used.

that the tax variable *av_tax* also changes sign and becomes positive in this reduced sample, as is also found in the study by Robson and Wren (1999). What regards the pure effect, this reduced sample weakly confirms their findings of opposite effects of average and marginal tax, however this result is not robust in a longer setting. This might highlight the importance of the period you are studying when trying to determine influential variables on self-employment. Even though the variables *reg* and *gov_size* were not significant in the larger sample, they have in previous results been strongly significant, that is why they are included in column 3. In the reduced sample they indeed turn significant, similar to previous results, once again highlighting how different results may be obtained depending on the time dimension. Furthermore, including these imply that *cit* regain significance, which may highlight the need to jointly include institutions and taxes in order to accurately capture the effects of taxes on self-employment rates.

Chapter five

Discussion

The aim of this thesis was twofold: To evaluate how taxes may affect the rate of self-employment, as a proxy for entrepreneurs, and how these effects may be affected by the quality of institutions. The first part was investigated by using a relative novel harmonized data set on self-employment that so far has not been used in order to investigate the impact of taxes. Moreover, both measures of personal income taxes and corporate income taxes were included as independent variables, where the latter on an aggregated level in a cross-country setting seems not to have been investigated previously. As research has highlighted the importance of well-functioning institutions to foster growth, research on entrepreneurs has under the last couple of years taken interest in these aspects of an innovation-promoting environment. Against this background, variables capturing aspects of regulation and legal structure were included. Moreover, as the effect of taxes on self-employment has been found to generate counteracting effects between negative labour supply and tax avoidance, the effect of how institutional quality may have an impact of the effect on tax avoidance behavior was investigated. As the employed dependent variable is a relatively new measure and likely superior to the OECD LFS indicator, previously used, in combination with modeling several different channels of tax effects on self-employment, this study fills a gap in the literature.

To summarize the main results, average tax was never significant in any specification. However, the results of top statutory marginal income tax (*pit)* and top corporate income (*cit)* tax suggested that there is some relation between taxes and self-employment but the effects are very small. This is line with the results found by Bruce and Moshin (2006) among others, but smaller than the results found by Parker and Robson (2004) for example. It might mean that when longer time-series are studied, the effect of taxes is reduced. It might also indicate that the unharmonized measure from OECD LFS used in the latter (as well as in Robson and Wren 1999) does cause biased results. A third explanation to the smaller estimates compared to some (but not all) studies could be statistically. It is known that fixed effects estimates may suffer from attenuation bias due to enlarged measurement error in this model, hence it may account for smaller fixed effects estimates (Angrist and Pischke 2009, p. 225). In the case of *pit*, significance is not robust, but in general a negative effect is found, consistent with most previous research (on marginal rates). The effect of *cit* is robust through almost all specifications. Interestingly, is that these results of a small, negative effect of *cit*, but of no other tax rate, correspond to the study by Lee and Gordon (2005) where they use similar tax measures but investigate their effects on GDP growth, with the motivation that the linkage between taxes and growth could be through entrepreneurs.

Is there a specific reason to the fact that *cit* seems to have a more prominent effect on this particular self-employment measure? One explanation could be that the main improvement of this measure is the consistent inclusion of incorporated businesses, which in the unharmonized measure were included differently depending on the country. As pointed out by Steel (2003), the number of incorporated businesses have increased faster than unincorporated, which implies that this group of self-employed is relatively larger in this measure compared to the one employed by Parker and Robson (2004). It does not all in all explain the insignificance of *pit*, but it gives to reason to the significant *cit.*

However the expected negative sign of *cit* seems to be dominant where the channel of effect is allowed to be dynamic or lagged, which is one strong indication of the need to carefully model the determinants of self-employment (especially given the absence of

a general accepted theoretical model of entrepreneurship). This might be the reason for why the standard specification for the error-correction model indicated that only two variables were significant related to self-employment. Further investigation in the modeling might have generated somewhat other results. This might also be the reason for why studies have begun to look at the question of taxes and entrepreneurship using more complicated indicators of the tax system (Georgellis and Wall 2006). Nonetheless, as in Stenkula (2009), the effect of the significant corporate tax variable in the EC model was similar to the magnitude in the fixed effects model. Moreover, the data suggest that opposite effects of taxes on self-employment may exist and especially confirms the findings of Torrini (2002) indicating that a positive effect of corporate tax might be due to the current institutional context, rather than a general feature of self-employed. However, what Torrini (2002) does not discuss (nor Da Rin et al (2011)) is the fact that the construction of the dummy to investigate this relation, implies a subsample analysis. That is, the dummy takes on value of 1 for all observations of one country, if the corruption in general has been higher than the average corruption for the whole sample of countries. That is, caution when interpreting this result is crucial, as it in fact means that any potential common feature of these countries may cause the positive (or less negative) effect of taxes.

To summarize the effect of the control variables, only a few seemed to be robust in signs and magnitude. Unemployment seems to have a negative effect on self-employment rates indicating that there is no dominating recession push into self-employment. Instead, a well-functioning labour market might function as a risk-reducing effect. In a couple of specifications the gross replacement rate indicated signs of an expected negative effect on self-employment. This variable acted as an alternative cost to self-employment and further, if self-employed do not receive the same benefit entitlements as those in wage employment, high replacement rates might increase the perceived risk associated to self-employment. The causality is of course not clear cut, high replacement rates might also be an indication of a larger public sector, which according to Davidsson and Henrekson (2002) might crowd out entrepreneurial activity in several sectors. Given that the measure of government size also indicated somewhat robust significance this might be the case. Lastly, the qualitative institutional measure of business, labour and credit regulation seemed the most robust of all control variables, indicating that this might be perceived as a more crucial impediment by potential self-employed, compared with for example tax rates. Complicated business regulation may be a larger burden to manage for a small business owner compared to tax rates and a possible reason could be that the former is perceived as a less transparent "cost" compared with tax rates.

Since the data suffered partly of multicolliniearity when controlling for fixed effects, this might be a potential explanation for the overall insignificance. This is also a potential statistical explanation to why the variable on average tax showed robust insignificance, as the variance of this variable is less compared to the other two tax rates. However, it is interesting to note that no other study have commented on potential colliniearity issues between slow-moving variables, such as the average tax rate, and country fixed effects, which I find reason to believe must be somewhat present also in other studies.

Aside theoretical explanations, as well as methodological considerations, the results are a result of the data employed. This means that small effects of the tax variables may simply be a result of the chosen measures. As the measure in this study corresponds to those of Stenkula (2009) and Bruce and Moshin (2006) which uses the same tax measures and receive similar results, this is a plausible explanation. The small effect of *pit* in all these studies might suggest that the top marginal tax rate is not applicable to most of these entrepreneurs. Indeed, Robson and Wren (1999) used the effective marginal rate of average earnings, which is a somewhat different measure. Lastly, the fixed effects model assumes

poolability of the data, given that heterogeneous effects are allowed as constants, but the slopes (i.e. the β) are assumed to be homogenous in the sample. Including interaction terms as in Torrini (2002), indicated that the parameter of a subsample of countries, have a different sign. This might suggest that the effect of taxes is not mainly a discussion of the right indicators or model, but rather of the possibility to pool different countries with different tax system together.

5.1 Conclusions

As to answer the research questions in this thesis and to summarize the above results and discussion, based on the results of data on 17 countries over 26 years, taxes seem to have a small, if, any impact on self-employment. The negative sign of corporate tax is however robust and the magnitude consistent with other studies that have found small effects. Concerning quality of institutions, the hypothesis that presence of corruption may have an impact on the sign and effect of taxes on self-employment received some support, i.e. where corruption is relatively higher, the positive effect of taxes seem to dominate the negative labour supply effect. This warrants careful use of the tax schedule, when the desire is to stimulate entrepreneurship. Hence, one of the main conclusions of the results in this study is that using the tax system in order to have an impact on the rate of self-employment seems to be somewhat inefficient government policy. Enhancing regulation and cutting red-tape might be of greater importance.

Since the research area itself has generated very mixed results, in both controls and independent variables, and since the longer time dimension of this data nonetheless implied limited available variables, the expectations were not to produce hard evidence. Moreover, when evaluating the effect of taxes on self-employment, aggregated studies have some drawbacks in comparison to individual micro data studies and it is important to bear in mind that aggregated models provide only a general insight into the principal determinants of self-employment. Nonetheless, a general insight with regards to several countries also have some value in itself, as micro data usually is concentrated on a single country context. Further research using cross-country individual data with the possibility to evaluate different measures of taxes would most likely provide interesting contributions to the field.

Enhanced strategies and considerations regarding the choice of specific data for potential future research have already been discussed. In addition to the methodological aspects that have been analyzed, more prominent results in the impact of taxes on self-employment would also warrant a more close attention to possible simultaneity problems in future aggregated research. Except for the use of estimation methods such as generalized methods of moment (GMM), suitable for shorter dimensions of time, no other instrumental technique has been used in macro data research on self-employment as far as I have been able to find. One instrument of taxes that could be considered in future research is the instrument of taxes proposed by Lee and Gordon (2005), where country tax rates are instrumented by the distance weighted neighboring countries tax rates, which have been used in other fields of research.

Due to mixed results of the effect of different taxes on self-employment it is likely that this research will continue to grow. The development of different and hopefully enhanced measurements of entrepreneurship will probably enlighten the discussion. Moreover, the stickiness and persistence in different self-employment numbers within countries is something that still seem to be rather unexplained with a lot of potential in theory development. As traditional economic policy variables such as taxes and unemployment clearly are unable to explain the whole picture, there is probably much research to be done in the future with regards to attitudes and cultures, variables that traditionally share the same persistent patter as self-employment.

List of references

Aghion, P and Howitt, P (2009). The economics of growth. The MIT Press. Cambridge.

Agrist, J, D and Pischke, J-S (2009). Mostly harmless econometrics. Princeton University Press. Princeton.

Baliamoune-Lutz, M and Geller, P (2011). "Tax structure and entrepreneurship". *Institute for Research in Economic and Fiscal Issues working paper* , 1/2011.

Beck, N (2001). "Time-series–cross-section data: What have we learned in the past few years?". In *Annual Review Political Science*, 4, 271–93.

Beck, N and Katz, J,N (2011). " Modeling dynamics in time-series–cross-section political economy data". In *Annual Review Political Science*, 14, 331–52

Bjørnskov, C and Foss, N, J, (2008). Economic freedom and entrepreneurial activity: Some cross country evidence. In *Public Choice*, 134, 307– 328.

Bruce, D and Schuetze, H, J (2004). "Tax policy and entrepreneurship". In *Swedish economic policy review* , 11, 233-265.

Bruce, D and Moshin, M (2006). "Tax Policy and Entrepreneurship: New Time Series Evidence". In *Small Business Economics* , 26, 409–425.

Castellacci, F and Natera, J, M (2011). "A new panel dataset for cross-country analyses of national systems, growth and development (CANA)", In *Innovation and Development*, 1 (2).

Castellacci, F and Natera, J, M (2011). "CANA Database". Retrieved 2014-04-08. [http://english.nupi.no/Activities/Projects/CANA]

Cullen, J,B and Gordon, R,H (2007). "Taxes and entrepreneurial risk-taking: Theory and evidence for the U.S", In *Journal of Public Economics*, 91, 1479–1505

Da Rin, M, Giacomo di, M, Sembenelli, A (2011). "Entrepreneurship, firm entry, and the taxation of corporate income: Evidence from Europe ". In *Journal of Public Economics*, 95, 1048–1066.

Davidsson, P and Henrekson, M (2002). " Determinants of the Prevalence of Start-ups and High-Growth Firms". In *Small Business Economics*, 19:2, 81-104.

Djankov, S, Ganser, T, McLiesh, C, Ramalho, R and Schleifer, A (2010). "The effect of corporate taxes on investment and entrepreneurship". In *American Economic Journal: Macroeconomics* 2(3), 31-64.

EIM Business and Policy Research (2013). "COMPENDIA". Retrieved 2014-04-29. [http://data.ondernemerschap.nl/webintegraal/userif.aspx?SelectDataset=31&SelectSubset=113&Country=UK]

Evans, D, S and Jovanovic, B, (1989). "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints." In *Journal of Political Economy*, 97(4), 808-827.

Fölster, S (2002). "Do Lower Taxes Stimulate Self-Employment?". In *Small Business Economics*, 19, 135–145, 2002.

Garret, T, A and Wall, H, J (2006). "Creating a Policy Environment for Entrepreneurs ". *Federal reserve bank of St. Louis research division working paper*, 064B

Gentry, W, M and Hubbard, R, G (2000). "Tax policy and entrepreneurial entry". In *American Economic Review*, 90(2), 283-287.

Gentry, W, M and Hubbard, R, G (2004). "Tax policy and entry into entrepreneurship", Mimeo, Williams College.

Georgellis, Y and Wall, H,J (2006), "Entrepreneurship and the Policy Environment". In *Federal Reserve Bank of St. Louis Review*, 88:2, 95-111.

Hansson, Å (2008). "Skatter, entreprenörskap och nyföretagande". *Underlagsrapport nr 12 till Globaliseringsrådet.*

Hansson, Å (2012). “Tax policy and entrepreneurship: empirical evidence from Sweden”. In *Small Business Economics*, 38, 495–513.

Hansson, Å and Dackehag, M (2012). ”Taxation of Income and Economic Growth: An Empirical Analysis of 25 Rich OECD Countries”. *Lund University department of economics working paper* 2012:6.

Henrekson, M (2007). “Entrepreneurship and Institutions”. *Research institute of industrial economics (IFN) working paper*, 707.

Henrekson, M and Sanandaji, T (2008). “Entrepreneurship and the Theory of Taxation”. *Research institute of industrial economics (IFN) working paper*, 707

Lee, Y and Gordon, R, H (2005). “Tax structure and economic growth”. In *Journal of Public Economics* 89, 1027–1043.

Mooij de, R, A and Nicodème, G (2007). “Corporate tax policy, entrepreneurship and incorporation in the EU”. *European Commission economic papers*, 269

North, D, C (1993a). "Five Propositions about Institutional Change," in Knight J. and Sened I., (red), *Explaining Social Institutions*, University of Michigan Press, 15-26.

North, D (1993b). Institutionerna, tillväxten och välståndet. SNS Förlag, Stockholm.

Nyström, K (2008). “The institutions of economic freedom and entrepreneurship: evidence from panel data”. In *Public Choice*, 136, 269–282.

OECD (2014). “Tax database”.

OECD (2014). "1981-1999 Personal Income Tax Data ". Received 2013-06-11. [http://www.oecd.org/tax/tax-policy/tax-database.htm]

OECD (2014). "Comparative tables- historical model". Received 2013-05-14- [http://www.oecd.org/tax/tax-policy/tax-database.htm]

Office of Tax Policy Research, OTPR, University of Michigan. “World Tax Database”. Retrieved 2014-04-16. [http://www.bus.umich.edu/otpr/otpr/default.asp]

Parker, S, C, and Robson, M, T, (2004). “Explaining international variations in self-employment : evidence from a panel of OECD countries”. *Southern economic journal*, 71:2, 287-301.

Podesta, F (2003). “Econometric solutions vs. substantive results: a crucial trade-off in the time-series-cross-section analysis”. *ISER Working Paper Series*, 34.

Robson, T and Wren, C (1999). “Marginal and Average Tax Rates and the Incentive for Self-Employment”. In *Southern Economic Journal*, 65:4, 757-773

Stel, AJ van (2003). “COMPENDIA 2000.2: A Harmonized Data Set of Business Ownership Rates in 23 OECD Countries”. *EIM Research Report*, H200302.

Stenkula, M (2009). “ Taxation and Entrepreneurship in a Welfare State”. *Research institute of industrial economics (IFN) working paper*, 800.

The Fraser Institute (2014). “Economic Freedom of the World”. Retrieved 2014-04-22. [http://www.freetheworld.com/]

Torrini, R (2002). “Cross-country differences in self-employment rates: the role of institutions”. *Banca d’Italia temi di discussione del servizio Studi*, 459.

Veerbek, M (2009). A guide to modern econometrics. John Wiley & Sons. Ltd., Publication. West Sussex.

Wooldridge, J, M (2012). “Introductory econometrics: A modern approach”. South-Western College Publishers.

Appendix A

Data and descriptive statics

Table A.1 Variable description[35]

Variable	Obs	Mean	Std. Dev.	Min	Max
AV_TAX	476	17.639	9.232	0.399	44.390
BANK_CREDIT	475	123.794	54.897	23.181	312.784
CIT	476	34.880	9.595	8.500	56.000
CORRP	476	7.611	1.7621	2.99	10
DEP_RATIO	476	49.571	3.165	43.302	58.255
GDP	476	23483.26	5395.50	10007	38713.0
GOV_SIZE	476	4.857	1.319	1.800	8.000
GRR	476	28.073	13.595	0.300	64.900
LEG	476	7.997	0.923	5.600	9.600
PIT	476	53.105	13.126	11.500	93.000
REG	476	6.850	1.189	3.900	8.900
SEMP	476	10.931	3.930	5.600	21.000
UNEMP	476	7.092437	3.356136	0.2	19.2

[35] Since employment in the service sector (*serv_emp)* was omitted due to high multicolinearity in the empirical section of the thesis, it is not reported.

Table A.2 Variable description[36]

Variable	Description[37]	Source[38]
AV_TAX	Average income tax rate on 100% of average earnings for a single person.	5
BANK_CREDIT	Domestic credit by banking sector as a share of GDP.	1
CIT	Top statutory corporate tax rate.	4
CORRP	Corruption perception index, scale 0-10.	2
DEP_RATIO	Population aged 0-15 and >65 as share of total population.	1
GDP	GDP purchasing power parity (PPP) per capita (in year 2000 US dollars).	1
GOV_SIZE	Qualitative measure of government size, scale 0-10.	3
GRR	Gross replacement ratio.	1
LEG	Quality of legal structure and security of property rights, scale 0-10.	3
PIT	Top statutory marginal tax rate on labor income.	5[39]
REG	Quality of business, labor and credit regulation, scale 0-10.	3
SEMP	Non-agriculture self-employment as share of labor force.	1
UNEMP	Unemployment as share of labor force.	1

Table A.23 Correlation matrix. sample 1981-2008[40]

	AVTAX	*B_CRED*	*CIT*	*D_RAT*	*FEMP*	*G_SIZE*	*GRR*	*LEG*	*GDP*	*PIT*	*REG*	*S_EMP*	*UNEMP*
AVTAX	1.00												
B_CRED	-0.36	1.00											
CIT	0.13	-0.32	1.00										
D_RAT	-0.06	-0.03	-0.28	1.00									
FEMP	0.49	0.11	-0.24	-0.08	1.00								
G_SIZE	-0.28	0.30	0.09	-0.19	0.15	1.00							
GRR	0.39	-0.32	-0.01	-0.01	0.22	-0.11	1.00						
LEG	0.14	0.09	-0.21	-0.05	0.46	0.05	0.37	1.00					
GDP	0.20	0.47	-0.41	-0.13	0.48	0.07	0.09	0.28	1.00				
PIT	0.27	-0.19	0.51	-0.08	-0.23	-0.17	0.14	-0.22	-0.39	1.00			
REG	0.10	0.36	-0.18	0.03	0.62	0.48	0.04	0.51	0.39	-0.28	1.00		
S_EMP	0.36	0.33	-0.32	-0.16	0.65	0.30	0.20	0.33	***0.80***	-0.33	0.57	1.00	
UNEMP	0.14	-0.41	0.18	-0.11	-0.27	0.03	0.12	-0.13	-0.36	0.10	-0.25	-0.10	1

[36] Since employment in the service sector (*serv_emp)* was omitted due to high multicolinearity in the empirical section of the thesis, it is not reported.

[37] Where qualitative measures are used in a 0 to 10 point scale, high values indicate no corruption/ small government/ well protected property rights/ well regulated business, labour and credit markets.

[38] Source: Online Databases
1 - COMPENDIA: EIM Business and Policy Research (2013)
2 -CANA: Castellacci, F and Natera, J, M (2011)
3 -EFW: The Fraser Institute (2014)
4 -OTPR: Office of Tax Policy Research, University of Michigan
5- OECD: OECD, Tax database (available on request)

[39] The raw data from OECD consists of top statutory central and sub-central personal income tax rates that have been added together by the author, see further explanation in section 3.1.2.

[40] In order to fit the table appropriately, some variables have shorter names in this table. *AVTAX = AV_TAX; B_CRED=BANK_CREDIT; D_RAT= DEP_RATIO; G_SIZE=GOV_SIZE; S_EMP= SERV_EMP*

Appendix B

Regression outputs (Note that all values within parenthesis in the following tables are *p-values*).

Table B1. Lagged effects of taxes with country and annual time dummies.

Variable	(1)	(2)	(3)
LAG_GDP	-0.254	-0.385	-0.130
	(0.793)	(0.697)	(0.894)
UNEMP	-0.059***	-0.078***	-0.076***
	(0.004)	(0.000)	(0.000)
FEMP	0.053	0.033	0.016
	(0.108)	(0.280)	(0.611)
DEP_RATIO	-0.012	-0.013	-0.023
	(0.628)	(0.618)	(0.391)
GRR	-0.014*	-0.016**	-0.018**
	(0.068)	(0.046)	(0.031)
BANK_CREDIT	-0.002	-0.002	-0.001
	(0.162)	(0.199)	(0.289)
AV_TAX			
L1.	-0.002		
	(0.891)		
L2.		-0.004	
		(0.810)	
L3.			0.025
			(0.101)
CIT			
L1.	-0.000		
	(0.950)		
L2.		-0.003	
		(0.464)	
L3.			-0.009**
			(0.040)
PIT			
L1.	0.003		
	(0.611)		
L2.		-0.004	
		(0.463)	
L3.			0.002
			(0.775)
REG	0.421***	0.436***	0.474***
	(0.000)	(0.000)	(0.000)
LEG	0.058	0.056	0.059
	(0.436)	(0.441)	(0.432)
GOVSIZE	0.197**	0.201***	0.224***
	(0.011)	(0.007)	(0.004)
R-squared	0.932	0.939	0.942
Obs	441	424	407

Table B.2 Quadratic model with country and annual time dummies.

Variable	(1)	(2)	(3)	(4)
LAG_GDP	-0.378	-0.333	-0.121	-0.002
	0.694	0.727	0.901	0.999
UNEMP	-0.057***	-0.059***	-0.055***	-0.057***
	0.006	0.004	0.008	0.006
FEMP	0.041	0.047	0.034	0.036
	0.206	0.131	0.278	0.246
DEP_RATIO	-0.005	-0.007	-0.004	-0.008
	0.861	0.777	0.891	0.775
GRR	-0.016**	-0.015*	-0.015*	-0.015*
	0.047	0.052	0.053	0.061
BANK_CREDIT	-0.002	-0.002	-0.002	-0.002
	0.143	0.119	0.165	0.139
REG	0.368***	0.381***	0.377***	0.388***
	0.000	0.000	0.000	0.000
LEG	0.093	0.086	0.075	0.081
	0.229	0.263	0.330	0.292
GOVSIZE	0.190**	0.191**	0.183**	0.182**
	0.014	0.014	0.018	0.018
AV_TAX	-0.018			-0.011
	0.585			0.730
AV_TAX squared	0.000			0.000
	0.597			0.745
CIT		0.029*		0.048***
		0.066		0.009
CIT squared		-0.000**		-0.001***
		0.043		0.006
PIT			-0.028*	-0.041**
			0.052	0.010
PIT squared			0.000**	0.000***
			0.045	0.008
R-squared	0.922	0.926	0.922	0.928
Obs	458	458	458	458

Table B.3. Interaction terms between taxes and corruption, with country and annual time dummies.

Variable	(1)	(2)	(3)	(4)
LAG_GDP	-0.322	-0.171	-0.265	-0.042
	(0.737)	(0, 858)	(0.778)	(0.965)
UNEMP	-0.056***	-0.054***	-0.057***	-0.056***
	(0,007)	(0.008)	(0.004)	(0.005)
FEMP	0.042	0,041	0,057	0,047
	(0,191)	(0,182)	(0.06)	(0.124)
DEP_RATIO	-0.004	-0.001	-0.004	-0.005
	(0.879)	(0.961)	(0.872)	(0.838)
GRR	(-0.015*	-0.014*	-0.013*	-0.011
	(0.054)	(0.078)	(0.087)	(0.171)
BANK_CREDIT	-0.002	-0.002	-0.002	-0.002
	(0.146)	(0.138)	(0,.144)	(0.197)
REG	0.371***	0.387***	0.414***	0.424***
	(0,000)	(0.000)	(0.000)	(0.000)
LEG	0.096	0.079	0.07	0.078
	(0.215)	(0.306)	(0.364)	(0.319)
GOVSIZE	0.189**	0.197**	0.226***	0.218***
	(0.015)	(0.011)	(0.003)	(0.005)
AV_TAX	0,.007			0.023
	(0.678)			(0.182)
*AVTAX*CPI*	-0.029			-0.036
	(0.392)			(0.299)
PIT		-0.005		-0.002
		(0.402)		(0.727)
*PIT*CPI*		0.012		0.008
		(0.179)		(0.386)
CIT			-0.012**	-0.011**
			(0.018)	(0.03)
*CIT*CPI*			0.031***	0.028***
			(0.001)	(0.006)
R-squared	0.922	0.924	0.936	0.933
Obs	458	458	458	458

Table B.4. Lagged dependent variable with country and time annual dummies.

Variable	(1)	(2)	(3)	(4)
SEMP L1.	0.883***	0.869***	0.887***	0.875***
	(0.000)	(0.000)	(0.000)	(0.000)
LAG_GDP	-0.819**	-0.802*	-1.136**	-0.739
	(0.047)	(0.055)	(0.010)	(0.108)
UNEMP	-0.025**	-0.024**	-0.027***	- 0.021**
	(0.018)	(0.025)	(0.009)	(0.047)
FEMP	-0.023	-0.017	-0.014	-0.017
	(0.112)	(0.227)	(0.389)	(0.252)
DEP_RATIO	-0.010	-0.006	-0.014	-0.010
	(0.188)	(0.468)	(0.104)	(0.208)
GRR	-0.004	-0.007*	-0.003	-0.004
	(0.242)	(0.059)	(0.343)	(0.218)
BANK_CREDIT	-0.001	-0.001	-0.001	-0.001*
	(0.109)	(0.223)	(0.104)	(0.085)
REG	0.166***	0.162***	0.148***	0.168***
	(0.000)	(0.000)	(0.002)	(0.000)
LEG	0.052	0.043	0.069*	0.051
	(0.133)	(0.234)	(0.072)	(0.150)
GOVSIZE	0.053	0.070*	0.055	0.052
	(0.154)	(0.079)	(0.133)	(0.164)
AV_TAXS	-0.000		0.008	0.001
	(0.955)		(0.656)	(0.946)
CIT	-0.003		0.025*	-0.008**
	(0.362)		(0.053)	(0.022)
PIT	-0.000		0.017**	0.003
	(0.994)		(0.019)	(0.373)
AV_TAX L3.		-0.003		
		(0.673)		
CIT L3.		-0.005*		
		(0.065)		
PIT L3.		0.006*		
		(0.066)		
AV_TAX squared			-0.000	
			(0.685)	
CIT sqaured			-0.000**	
			(0.036)	
PIT squared			-0.000***	
			(0.006)	
*AV_TAX*CPI*				-0.001
				(0.949)
*CIT*CPI*				0.014**
				(0.013)
*PIT*CPI*				-0.007
				(0.197)
R-squared	0.996	0.996	0.996	0.996
Obs	441	407	441	441

Table B.5 Im-Pesaran-Shin unit root test.

Panel Unit Root Test in levels, with the natural logarithm.W-t-bar statistic. Asterix indicates stationarity on 10, 5 and 1 percent significance level.			
Variable	Trend and demeaned	Trend, not demeaned	No trend, nor demeaned
SEMP	-0.5594	0.4642	-0.5058
GDP	-0.2650	-1.2743*	1.2754
UNEMP	-2.0229**	-3.3208***	-2.8924***
FEMP	-1.3518*	-1.5419*	-5.0241***
DEP_RATIO	-3.0746***	-4.1275***	-4.5640***
GRR	-2.9106***	-7.0991***	-3.8876***
BANK_CREDIT	1.3453	1.4800	1.9143
REG	-0.6709	-1.6211*	1.3701
LEG	-2.2782**	5.4483	-1.6023*
GOVSIZE	1.5224	1.5440	-0.1642
CIT	-3.0736*	-1.1693	0.6297
PIT	-0.7491	0.1492	0.7744
AV_TAX	0.8492	-0.8134	-0.2749

Table B6. Error correction model

Variable	(1)	(2)	(3)
SEMP L1.	-0.095***	-0.179***	-0.223***
	(0.000)	(0.003)	(0.000)
GDP L1.	-0.037*	-0.020	-0.028
	(0.058)	(0.686)	(0.596)
UNEMP L1.	-0.003	-0.003	-0.004
	(0.552)	(0.744)	(0.676)
FEMP L1	0.006	0.008	-0.038
	(0.858)	(0.886)	(0.480)
REG L1	0.032		0.139***
	(0.132)		(0.001)
GOVSIZE L1.	0.016		0.110***
	(0.253)		(0.000)
AV_TAX L1.	-0.000	0.002	0.000
	(0.606)	(0.202)	(0.860)
CIT L1.	-0.001***	-0.001	-0.001**
	(0.005)	(0.364)	(0.016)
PIT L1.	0.000	-0.001**	-0.000
	(0.729)	(0.010)	(0.992)
GDP D1.	0.215***	0.267***	0.193***
	(0.004)	(0.001)	(0.007)
UNEMP D1.	-0.009	-0.013	-0.007
	(0.458)	(0.320)	(0.612)
FEMP D1.	-0.130	-0.133	-0.190
	(0.289)	(0.382)	(0.172)
REG D1.	0.069		0.192
	(0.109)		(0.126)
GOVSIZE D1.	0.027		0.270***
	(0.352)		(0.000)
AV_TAX D1.	-0.000	0.002	0.001
	(0.856)	(0.564)	(0.697)
CIT D1.	-0.001	-0.001*	-0.001*
	(0.191)	(0.076)	(0.054)
PIT D1.	-0.000	-0.001**	-0.001
	(0.298)	(0.028)	(0.253)
R-squared	0.179	0.336	0.429
Obs	459	187	187