Norwegian School of Economics

Bergen, Spring 2020

Inventors in Tax Havens

Helene Merrild Ro and Jørgen Langli

Supervisor: Maximilian Todtenhaupt

MSc in Economics and Business Administration,

Business Analysis and Performance Management & Economics

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

The aim of this thesis is to examine to which extent inventors engage in tax havens, as well as the characteristics of the inventors engaged in tax haven compared to the average inventor. Inventors are behind the innovations that drive economic growth and technological development. However, the patents they create can be used to shift income to tax havens. This study examines whether inventors are overrepresented in tax haven, compared to their representation in the general population. In addition, examining characteristics of inventors that use tax havens can enlighten which type of inventors seek engagement in tax havens.

Our empirical methodology consists of matching names in the PATSTAT register with names of shell company owners in the Offshore Leaks. Identification of inventors in the Offshore Leaks reveals that inventors are overrepresented in tax havens. This is especially prominent in East Asian countries. Inventors who engage in tax havens are more productive than the average inventor and there is a higher share of inventors in the field of electrical engineering in tax havens than on average. We find indications of inventors collaborating with their employer to engage in tax havens due to their higher share of corporate and public institution patents compare to independent individual patents compared to the average inventor.

Contents

1. Introduction

This thesis explores to what extent inventors engage in tax havens, as well as examining the characteristics of the inventors engaged in tax haven compared to the average inventor. There are several reasons that make inventors a particularly interesting to study in the context of international tax policy. First, inventors are behind the innovations that drive economic growth and technological development (Akcigit, Grigsby, Nicholas & Stantcheva, 2018; Wong, Ho & Autio, 2005). Hence, many governments like to attract them by implementing competitive tax systems. Inventors are often also top earners and thus likely to engage in offshore tax evasion (Alstadsæter, Johannesen & Zucman, 2019a). This is useful information when attempting to assess the impact of different tax policies. However, little is known about the extent to which inventors locate wealth in tax havens. Second, the patents they create can be used to shift income to tax havens (Böhm, Karkinsky & Riedel, 2012; Tørsløv, Wier & Zucman, 2018). While this is usually addressed as a corporate tax issue, the presence of network spillovers creates the opportunity for avoidance behaviour to spread throughout the organisation (Bohne & Nimczik, 2018; Paetzold & Winner, 2016). Examining characteristics of inventors that use tax havens can enlighten the presence of such effects. Third, inventors are subject to the income tax of the country they work, and corporate R&D is often allocated in high-tax countries that offer R&D tax credits (Böhm et al., 2012; Griffith, Miller & O'Connell, 2014). The literature shows that while inventors are mobile and attractive workers, few actually change country during their careers (Akcigit, Baslandze & Stantcheva, 2016; H. Kleven, Landais, Muñoz & Stantcheva, 2020). This might only be a reflection of their income level or lifestyle preferences but could also indicate that they avoid taxes and are thus less concerned with high tax rates.

To identify the extent to which inventors engage in tax havens, we examine whether inventors are overrepresented in the population of individuals owning shell companies in tax haven jurisdictions. To better understand the engagement of inventors in tax havens, we further investigate the characteristics of those inventors we find to be engaged in tax havens and compare them to the average inventor. Specifically, we examine if they are more successful than the average inventor and look at the patent technology groups to see if there are specific groups of inventors that engage in tax havens. We also look at whether tax-haven-involved inventors are more or less likely to cooperate with corporations, public institutions or operate independently than other inventors. And finally, we also look for indications that inventors in tax haven are cooperating with each other or their employer to examine the presence of network effects.

To identify inventors who engage in tax havens, we link the names of inventors in patent data with the names of shell company owners in the Offshore Leaks by name matching. Our source of patent data is the EPO Worldwide Patent Statistical Database (PATSTAT). It covers patent applications in more than 80 countries, providing name, addresses, and country of residence of inventors and patent applicants. PATSTAT has worldwide coverage and is the most extensive patent database available (Kang & Tarasconi, 2015). Our tax haven data is retrieved from the Offshore Leaks database released by The Consortium of Investigative Journalists (ICIJ), which contain leaked material of the Panama Papers (2016), Paradise Papers (2017), Bahamas Leaks (2016) and the Offshore leaks (2013). This includes name, addresses and country of origin of individuals owning a shell company in tax haven jurisdictions. Omartian (2017) found significant similarities across two sets of leaked documents in how individuals and their banks responded to external pressure in setting up or shutting down shell companies, which suggests that homogeneity of shell companies may be assumed across tax havens. The Offshore Leaks are thus likely to be a random sample of individuals engaged in tax havens and the distribution of inventors in this source would not be different had we picked another offshore source. This allows us to use this data to make inferences from the distribution of the information provided in the database to the whole population of tax haven investors.

We use the names of inventors from PATSTAT and match them to the names of individuals in the Offshore Leaks database using a “fuzzy matching” algorithm. The identification assumption is that two observations with the same name from the same country in both data sets refer to the same person. Lack of additional information and incomplete address data prevent us from using a narrower criterion. Both data sources contain unprocessed data, and we need to do a comprehensive cleaning and standardisation of the names in both datasets before the matching process. Fuzzy matching is intended to overcome remaining misspelling or separational errors in the name, by exploiting the available overlapping information and providing a score measuring the similarity between the names. We then manually filter the matching results provided by the algorithm, reducing the likelihood of false positive matches (including individuals who are in fact not the same person) and false negative matches (not including individuals who are the same person). After we identify the inventors who engage in a tax haven, we compare the ratio of inventors among individuals in tax haven to the ratio of inventors in the general population. If the former is larger, it implies an overrepresentation of inventors in tax havens. We do this on a country-to-country basis for the 55 countries with the highest number of patent applications. We find that inventors are overrepresented amongst individuals engaged in tax havens. For each country with at least one inventor engaged in tax havens, the share of inventors in tax havens are higher than the share of inventors in the general population of the country. This also holds true if we compare it to the share if inventors in the labour force of each country, except for Luxembourg. Japan, South Korea, China, United States and Germany are the countries with the highest share of inventors amongst individuals in tax haven. Hong Kong, Taiwan and China are the countries with the highest share of inventors engaged in tax haven of all inventors in the country.

To examine the characteristics of the inventors we identify in the Offshore Leaks, we use patent quality data from the OECD Patent Quality Indicators database. To examine if inventors in tax havens are more successful than the average inventor, we look at inventor quality, meaning we compare the average number of citations and patents for each inventor in tax haven to the average of all inventors in PATSTAT. Patent citations as an approximation of patent quality and economic value was proven robust by Hall, Jaffe, and Trajtenberg (2005). We also examine whether certain technological fields are overrepresented in the group of inventors in tax haven, by using technology classification provided by the OECD database. To find whether inventors in tax haven are more likely to be employed by a company or public institution, we connect each inventor to the applicants in their patent application. The inventors are usually employees of the patent applicant (OECD, 2009). We compare the share of company, public institutions and independent employment amongst inventors engaged in tax haven to the average inventor in PATSTAT. We also examine to what extent the inventors who are found in the same patent application, are also found in the same tax haven source, indicating cooperation in tax haven engagement between inventors or the inventors and their employer. We call these inventor collaborations. We find that inventors engaged in tax havens have, on average, more patents and more citations than all inventors in PATSTAT in average, indicating that they are more productive. They do not, however, have more citations per patent. There is a higher share of inventors within electrical engineering in tax havens than amongst inventors in general. Inventors in tax havens have higher share of public patents than the average inventor. We also find indication of collaboration between inventors, or inventors and their employer, to engage in tax havens; inventors who are registered in the same patent application, are also found in the same tax haven source. As well as the inventors who are collaborating with other inventors in tax haven have a higher share of corporate and public employment than the average inventor, and a lower share of individual independent patents.

This introduction section will be followed by a review of the relevant literature and hypothesis development. In Section 3 we explain the empirical methodology. This includes presentation of the data sources, description of the data cleaning and preparation, matching procedure, and a discussion about external validity. Section 4 contains the results of the empirical analysis, where we present and discuss our findings about inventor population in tax havens, including a case study on the three countries with the highest share of inventors in tax haven (China, South-Kora and Japan). We also present our results on inventor quality and technological field, inventor employment and inventor collaboration. In the end we discuss limitations of the study. Finally, in section 6 we conclude and provide suggestions for future research.

2. Literature

In this section we i) elaborate the framework that shaped our motivation to investigate the extent of inventors with shell companies ii) present theory to construct hypotheses iii) discuss how our data and research design allows us to answer hypothesis and iv) relate our findings to existing literature.

2.1 Framework

Knowledge of inventors and their engagement in tax havens is limited. We chose a descriptive research design to get an understanding of the phenomenon (Saunders et al, 2012). Descriptive research design is most appropriate when the purpose is to describe the characteristics of people, objects, organisations, environment, or groups (Zikmund, Babin, Carr & Griffin, 2012). We start by developing a framework in to develop our hypothesis of what we can expect to find using descriptive data on inventors engaged in tax haven.

2.1.1 Tax havens and offshore tax avoidance

Tax havens pose a challenge to governments' efforts to end corruption and tax avoidance (Johannesen & Zucman, 2014; Malan et al., 2017). There are different views to what constitutes a tax haven. We use the OECD definition, which is widely used and accepted by The Tax Justice Network (Malan et al., 2017). We continue with definitions from the OECD later on for consistency. The OECD defines a tax haven by the following criteria:

1. Applying no, or nominal taxes, such as inheritance, income, or corporate tax, to non-residents (individuals and corporations) primarily with a view to the avoidance of taxation in their home jurisdictions.

2. Having laws or measures which prevent the effective exchange of relevant information with other governments on taxpayers benefitting from the low or no tax jurisdiction.

3.Lacking in tax transparency, making it harder for home countries to take defensive measures which usually involves a favourable regulatory environment for tax evasion and avoidance.

The jurisdictions in the Offshore Leaks database meet these criteria when the Panama Papers were leaked (Remeur, 2018). Some countries define tax havens based on the relative tax difference between the two countries (Dischinger & Riedel, 2011). Tax havens allow non-residents to take advantage of the low tax by setting up companies, such as shell companies. The OECD provides a definition of shell companies that captures their main function:

> *A shell company is a company that is formally registered, incorporated or otherwise legally organised in an economy but which does not conduct any operations in that economy other than in pass-through capacity.*

Shell companies can be used to avoid and evade taxes, as well as funnel money to corruption and terrorism without the oversight of authorities (Malan et al., 2017; O'Donovan, Wagner & Zeume, 2019). The OECD defines tax avoidance as a term used to describe an arrangement of a taxpayer's affairs that is intended to reduce his tax liability, and although the arrangement could be strictly legal it is usually in contradiction with the intention of the laws it purports to follow (OECD, 2020a). Tax evasion is defined as illegal arrangements where tax liabilities are hidden or ignored. Tax evaders thus i) fail to declare all or part of their income ii) claim deductions from taxable income to which they are not entitled or iii) submit tax returns that appear to be legal but only because relevant facts are not disclosed to the authorities (Malan et al., 2017; OECD, 2020a). Multinationals can incorporate shell companies as subsidiaries to avoid taxes through income shifting to the subsidiary via intra-firm transactions or the booking of intangible income, such as patent income (Tørsløv et al., 2018). Individuals can use shell companies for tax evasion, by stowing away assets otherwise taxable in the home country.

The arrival of the Panama Papers (Harding, 2016) marked a new era for data availability to research the secret nature of tax havens and clientele of shell companies. Omartian (2017) investigates how the number of shell companies in the Panama Papers and Bahamas leaks changed in response to the enactment of the EU Savings Directive (EUSD) in 2003, and finds that a significant number of new shell companies were intermediated by banks shortly after the enactment. The EUSD required that banks in EU states and other participating countries reported bank information of non-resident individuals back to the client's home country. Banks were given the option to disclose the identity and bank statements of the client or levy a withholding tax on the interest income earned by the non-residents' account. Banks had incentives to circumvent the requirement to prevent clients from changing bank to one that was not affected by the directive. Not until 2013 did the EUSD require banks to report corporate accounts back to the non-residents home country, which represented a loophole for banks to circumvent the requirement between 2003 - 2013. Omartian (2017) provides evidence that banks took advantage of the loophole, and helped individuals switch from private to corporate ownership of their accounts by the use of shell companies. His results also showed that shell companies set up by different law firms in different tax havens were used for the same purposes, which clearly indicates some homogeneity in motives to set up shell companies: hide assets from government disclosure.

Consistent to the findings in Omartian (2017), a recent paper by O'Donovan et al. (2019) finds evidence that publicly listed firms implicated in the Panama Papers experienced significant drops in stock prices in response to the leak. They find the effect to be caused by the companies' lost opportunities to avoid or evade taxes and commit bribery, which reduces after tax profits, and consecutively shareholder value which stock markets react negatively to. Criminal investigations and several news reports have also revealed the use of shell companies by criminals, world leaders and rich business people (David, 2016; Pegg, 2016). Shell companies may also be used for legitimate purposes. For instance, companies from two or more different countries could conduct a merger under a shell company in a third tax haven jurisdiction to avoid preferential legal treatment from their home country towards either company (van der Does de Willebois, 2011).

Alstadsæter, Johannesen, and Zucman (2018) estimate that global offshore wealth held in tax havens equalled about 10% of global GDP in 2007, which they find by computing country specific discrepancies in bilateral banking statistics. They find that countries differ substantially in fractions of offshore wealth as share of own GDP, with East Asian and Scandinavian countries accounting for just a few percent except for Taiwan with about 20%; the U.S and continental Europe between 7%-15%, Greece and Argentina around 35%, and Venezuela and Gulf states closer to 60%. The offshore wealth distribution is similar to each country's ownership share of shell companies in the Panama Papers, except for China, which is over-represented in the Panama Papers (2016). Possible interpretations of this is that their use of shell companies has less to do with wealth than it has to do with circumventing investment regulations or protecting funds from expropriation (Wang, 2014). A new law in China also aims to reduce the use of shell companies to avoid taxes (KPMG, 2018), which suggests that some of the tax haven activity is motivated by tax avoidance.

2.1.2 Multinational companies and patent income

Tørsløv et al. (2018) compare the profits of multinationals to the profits of local firms in tax-havens and non-havens. They find that subsidiaries of multinationals systematically have lower profits than local firms in high-tax countries and higher profits than local firms in low-tax countries. The article estimates that 40% of multinationals profits are shifted to tax havens, through transfer pricing or booking of intangible income. Patents represent such a source of intangible income. A patent is a legal right granting its' proprietor the right to prevent third parties from commercially using an invention without authorisation in a designated period (EPO, 2020). A growing literature has studied firm specific data and found evidence that multinationals move legal ownership of patents to low-tax countries or tax havens (Alstadsæter, Barrios, Nicodème, Skonieczna & Vezzani, 2018; Böhm et al., 2012; Dischinger & Riedel, 2011; Griffith et al., 2014). Patent allocation in tax havens is most prominent in the case of higher quality patents with greater earnings potential, with CFC rules[1] found to reduce the extent of patent relocations (Böhm et al., 2012; Griffith et al., 2014). Except for where CFC rules are in place, the tax havens allow the multinational companies to receive low or no tax on patent income, while keeping the research and development (R&D) at a location more optimized for recruitment, knowledge flows and R&D tax credits (Akcigit & Stantcheva, 2020; Alcácer & Zhao, 2012; Bloom, Griffith & Van Reenen, 2002).

2.1.3 Inventors importance to multinationals

An inventor in this thesis is an individual categorised as the inventor in at least one patent application registered in PATSTAT. The patent applicant can also be the inventor, but most often the applicant is the employer of the inventor (OECD, 2009). According to PATSTAT patent applications, 80% of all applications are filed by corporations. A paper by Akcigit et al. (2016) find that 75% of inventors work for multinationals using U.S patent data. A survey of inventors from Germany, U.K, France, Spain, Netherlands and Italy shows that about 70% work for large firms with at least 250 employees (Giuri et al., 2007). From this we can derive that most inventors work for corporate R&D departments (keeping in mind the definition of inventor in this thesis: has at least one patent application). Multinational companies relocating patent ownership to a low tax country suggests that the R&D is conducted in a higher tax country

[1] Controlled-Foreign-Company rules are enacted in most states and many tax havens. CFC rules deny multinationals to apply local tax rates to subsidiaries if the income is of a certain nature, e.g. passive, and the local tax rate is below a certain threshold set by the multinationals' home country (Dueñas, 2019).

(Böhm et al., 2012; Dischinger & Riedel, 2011; Griffith et al., 2014). This suggests that inventors usually live in high tax countries, which for the case OECD implies personal income taxes that seldom are lower than 30%, but usually closer to 40% based on recent OECD tax wedges (OECD, 2020b).

The patents that inventors create for multinationals are not only instrumental to shifting income, but also strategic components to influence industry architecture[2] and retain market power. Big corporations like IBM, Huawei, and Samsung harvest patents to preserve licensing rights and block imitators or rivals on new and existing technologies (Merges & Nelson, 1990; Torrisi et al., 2016). This allows such firms to appropriate considerable fractions of the value chain through their market power and charge prices above marginal cost (Jacobides, Knudsen & Augier, 2006; Teece, 2014).

To the extent inventors or inventor collaborations have competence the firm cannot easily replace, the inventors also become of strategic importance to the firms (Jaravel, Petkova & Bell, 2018). Another concern (and opportunity) for firms is that of *knowledge spillovers*, defined as the external benefits from the creation of knowledge that accrue to parties other than the creator (Agarwal, Audretsch & Sarkar, 2010). Firms consider knowledge spillovers as channels to strategically gain information or influence industries (Yang & Steensma, 2014), but they also represent risks for firms by losing the best ideas to competitors or spin-off ventures (Agarwal, Audretsch & Sarkar, 2007; Agarwal, Ganco & Ziedonis, 2009). This could substantially reduce the value of the idea to the firm, and microeconomic evidence finds that knowledge spillovers occur within industries and between regions (Acs, Braunerhjelm, Audretsch & Carlsson, 2009; Audretsch & Lehmann, 2005). Firms thus have incentives to retain key inventors and information that can compromise the value of their own projects. Firms are found to disincentivize spillovers through higher wages and improved integration with the organization (Alcácer & Zhao, 2012; Møen, 2005). Empirical papers find that knowledge spillovers, firms act on incentives to absorb and protect knowledge from clusters, for instance by moving the most valuable R&D projects away from tech clusters, while staying in clusters when there is more information to gain (Alcácer & Zhao, 2012; Yang & Steensma, 2014).

[2] Industry architecture refers to how an industry is organized to allocate profit margins between suppliers competing in the value chain, based on the leverage that each actor has at each level of the architecture. For instance, a computer industry where different companies make hardware, software and chassis separately represents one architecture, while an industry where each company makes everything themselves represents another.

2.1.4 Inventor mobility and questions of tax avoidance

Papers by Akcigit et al. (2016) and Moretti and Wilson (2017) show that inventors to some degree make relocation decisions with respect to marginal tax rates. Akcigit et al. (2016) synthesize inventors' earnings potential by their quantity and quality of patents, and compute mobility relative to tax differences between countries. They find that between 3.7% - 4.6% of the superstar[3] top 5% with the highest earnings potential change country during their careers, and less than 0.7% of inventors below top 5% ranked inventors change country during their careers. Thus, earnings potential appears to be associated with mobility.

A European survey on inventors find that about 20% of inventors change employers after making an innovation (Giuri et al., 2007). Earlier studies on mobility find that about 30% of R&D workers at some point change jobs (Giuri et al., 2007; Trajtenberg, 2005; Trajtenberg, Shiff & Melamed, 2006). Trajtenberg (2005) also proves a positive relationship between mobility and citations on patents, implying that mobile inventors are more productive or impactful.

Based on the aforementioned studies, it appears that inventors in general are mobile, and the most productive ones even more. Yet, few appear to move with respect to taxes. Could this be affected by inventors using tax havens, thus reducing the number of inventors potentially responding to tax incentives? We pursue this question in the next section.

2.2 Theoretical considerations and hypothesis development

The objective of the thesis is to explore the extent and characteristics of inventors that use tax havens. We have developed a framework to interpret the data and derive hypotheses. We make theoretical considerations before we formulate hypotheses that can be addressed by descriptive data.

[3] Superstar inventors are inventors with disproportionate amount of technological impact, for instance measured by number of patents and forward citations. One example is an inventor at Johnson&Johnson with more than 700 granted patents.

2.2.1 Slippery Slope framework

Because offshore wealth is associated with non-compliant tax behaviour (Alstadsæter et al., 2019a; Alstadsæter, Johannesen & Zucman, 2019b), the "slippery slope" framework by Kirchler, Hoelzl, and Wahl (2008) can be used to discuss differences in offshore wealth distributions, also among inventors. The framework describes two dimensions that governments need to consider for achieving tax-compliant citizens: *trust* and *power*. The trust dimension refers to the legitimacy a government achieves by acting fairly and providing public functions to citizens. The power dimension refers to citizens' perception of the states' ability to pursue and punish non-compliance. The model assumes that when trust and power are high, compliance is high. If perceptions of power increase without raising trust, non-compliance increases. A recent cross-country study on a diverse sample of 44 countries found consistent support for the assumptions of the model, which added to existing support for the assumptions (Batrancea et al., 2019; Kastlunger, Lozza, Kirchler & Schabmann, 2013; Kogler et al., 2013). Thus, using the Perceived Corruption Index and Inclusive Development Index (Samans, Blanke, Corrigan & Hanouz, 2017; Transparency International, 2016), one might observe traces of non-compliance where trust is perceived to be low or disproportionate to power.

2.2.2 Network effects and tax evasion

Two recent studies show that tax avoidance behaviour has been learned by workers starting at new jobs. Paetzold and Winner (2016) find that new members of a firm learn that inflating commuter tax allowance goes undetected, and subsequently adopts this behaviour. Bohne and Nimczik (2018) find causal evidence of workers learning how aggressively they can exploit tax deduction schemes after changing jobs or networking with tax experts. Thus, both studies depict individuals learning what level of evasion that they can get away with without getting caught through. These outcomes seem consistent with the prediction of the seminal model on tax evasion by Allingham-Sandmo (Allingham & Sandmo, 1972; Sandmo, 2005). The model considers a taxpayer's gamble against the tax authorities' ability to detect and penalize evasion. The network effects found by the aforementioned studies could thus be interpreted to reduce the perceived risk of detection, resulting in greater non-compliance. H. J. Kleven, Knudsen, Kreiner, Pedersen, and Saez (2011) find evidence of self-reported income being more likely to be evaded when taxpayers know they will not be audited, which is consistent with the findings in the aforementioned studies of network effects and the Allingham-Sandmo model.

As accounted for in the framework, multinational firms exploit tax havens to shift patent income (Böhm et al., 2012; Tørsløv et al., 2018) or even corruption or tax evasion (O'Donovan et al., 2019). If individuals in a given multinational company know how to make use of tax havens, this information could hypothetically spread throughout the organization. While cheating on tax allowances and evading taxes through shell companies arguably represents different levels of tax evasion, the network principle could still apply to some extent.

2.2.3 A firms' incentive to retain key inventors

In Section 2.1.3 we establish that multinationals harvest patents to protect market power and employ measures to prevent knowledge spillovers detrimental to R&D prospects. Inventors do not appear very sensitive to changing country for taxes purposes (Akcigit et al., 2016), and evidence indicates that multinationals use tax havens to shift income (Böhm et al., 2012; O'Donovan et al., 2019; Tørsløv et al., 2018). Thus, with this in mind, we propose that multinationals could be forthcoming towards key inventors and provide opportunities to invest in tax havens, as a means to further integrate them in the organization and foster strategic behaviour.

2.2.4 Hypotheses

In the framework, we present research linking i) inventors to multinationals ii) multinationals to tax avoidance and iii) avoidance behaviour being adopted through networks. Based on this, we make the first hypothesis:

> **A)** Inventors are overrepresented in tax havens compared to the share they make up of the general population.

Patent portfolios indicate an inventor's quality, which should represent his or her bargaining power on the labour market. In a competitive labour market, higher quality should generate higher income to better inventors (Akcigit et al., 2016). In the framework, we link i) tax evasion to wealth and ii) wealth to inventor quality. Based on this we present the second hypothesis:

> **B)** Inventors in tax haven are more successful than the average inventors, measured by more patents and more citations than the average inventor in PATSTAT.

2.3 Research design

We will now explain how our data and methodological approach allow us to shed light on inventor engagement in tax haven and the hypotheses we have developed, as well as our underlying assumptions. We chose a descriptive research design. Descriptive research is aimed at casting light on current issues or problems through a process of data collection that enables them to describe the situation more completely than was possible without employing this method (Fox & Bayat, 2007). The disadvantage of this method is that it does not look at cause or effect.

A) To address if inventors are overrepresented in tax havens compared to the general population, we identify inventors in the Offshore Leaks database by the use of name matching. We then compare the share of inventors in the Offshore Leaks database to the share of inventors in the general population. If the former is larger, inventors are overrepresented in tax havens. For the validity of overrepresentation to hold, we assume that the Offshore Leaks data represent a random sample of individuals engaged in tax havens. Omartian (2017) found significant similarities in how shell companies respond to exogenous pressure, which suggests that homogeneity of shell companies may be assumed across tax havens. This allows us to use this data to make inferences from the distribution of the information provided in the database to the whole population of tax haven investors.

Furthermore, we examine the presence of network effects that can induce tax avoidance behaviour between inventors or inventors and their employer (Section 2.2.2). We therefore link tax haven inventors to each other and their employers through patent application ID's. We further examine whether inventors who collaborate, have a higher share of corporate or public institution patents than the average inventor. We assume that if tax haven inventors are found to collaborate with each other, it can be an indication of network effects. We also assume that a higher share of corporate and public patents than independent patents might indicate that inventors learn about tax aviation from their employer.

B) To examine the hypothesis of more successful inventors in tax haven, we compare the quantity of their patents and their forward citations to the average inventor. We assume that higher levels of citations and patents means higher quality of inventor, which leads to higher levels of earnings. Patent citations is used as an approximation of patent quality and economic value (Hall, Jaffe & Trajtenberg, 2005). Patent allocation in tax havens is most prominent in the case of higher quality patents with greater earnings potential (Böhm et al., 2012; Griffith et al., 2014). Inventors are also found to be top earners and thus likely to engage in offshore tax evasion (Alstadsæter et al., 2019a).

2.4 Related literature

2.4.1 Tax evasion in tax havens

The literature on tax evasion in tax havens has in recent years made empirical strides to quantify lost tax revenues and uncover the dynamics of the offshore wealth economy. Alstadsæter, Johannesen, et al. (2018) make several contributions to estimate how much household wealth is hidden in tax havens and how much belongs to each country by leveraging discrepancies in bilateral banking statistics (Alstadsæter et al., 2019a; Zucman, 2015). Studies on the effects of tax repatriation programs for offshore tax evaders include, but are not limited to Alstadsæter et al. (2019b) and Johannesen, Langetieg, Reck, Risch, and Slemrod (2018). While repatriation programs have positive effects on tax revenues, the article by Johannesen and Zucman (2014) finds that assets hidden in one tax haven are likely to slip to another tax haven if the first haven is pressured to comply with regulation. Some scholars also study the impact of institutional factors and flawed regulations facilitating the perseverance of offshore wealth (Andersen, Johannesen, Lassen & Paltseva, 2017; Roussille, 2015). Our thesis contributes by identifying a subgroup who accounts for an unknown share of the wealth estimated by Alstadsæter, Johannesen, et al. (2018) and Zucman (2015).

A recent literature is leveraging information of leaked bank accounts to expose activities in tax havens, wherein the data is limited and originating from leaks such as the Panama Papers. This literature makes contributions to the macro based literature on offshore wealth holdings. Omartian (2017) finds empirical evidence that the banking sector facilitates offshore tax evasion, and that different havens and law firms attract clients with the same motivation: asset concealment. O'Donovan et al. (2019) study the response of financial markets to publicly listed firms being exposed in the leak. They see that the firms' stock prices fall because they lose opportunities to evade taxes and bribe, which reduces after tax profits and subsequently shareholder value. Our thesis relates to this literature by exploring the engagement of inventors in tax havens.

2.4.2 Inventor mobility

A recent paper stresses the importance of governments to not make costly tax reductions to attract skilled migrants, because too little is known about the efficiency of tax as a migration stimulus (H. Kleven et al., 2020). Several studies have documented the domestic mobility of inventors (Akcigit et al., 2018; Giuri et al., 2007). Moretti and Wilson (2017) find that state taxes matter for inventor mobility within the U.S. Miguelez and Fink (2013) provide a database on the international mobility of inventors, which shows that 10% of inventors worldwide have immigrant background between 2001-2010, up from 7.5% between 1991 - 2000. Akcigit et al. (2016) are the first to provide empirical evidence on the international mobility of inventors with respect to taxation, using an eight-country sample with the bulk of patents in the U.S between 1977 - 2000. They estimate elasticities of international mobility relative to tax rates and the inventors' earnings potential. They show a migration rate of 4.6% among the top 1% inventors, and below 0.7% for the below top 5% inventors. Our thesis informs the efficiency question of tax as a migration stimulus to attract inventors, by providing data on inventors with offshore assets, who are less likely to react to tax incentives because part of their wealth is untaxed in tax havens.

3. Empirical methodology

To identify the presence of inventors and top scientists in tax havens, we use the names and country codes available from PATSTAT and the Offshore Leaks. Names in each dataset will be matched against each other, based on a string similarity calculation often referred to as "fuzzy" matching. We assume that the same name from the same country in both PATSTAT and the Offshore Leaks is the same individual, due to lack of additional information.

In this section, we describe the source of the data on which we base our analysis, followed by a description of how they were prepared and cleaned. Next, we explain our matching and filtering method. The analysis is based on secondary quantitative data from two main sources, as well as supplementary data from a third source.

3.1 Data sources

3.1.1 PATSTAT

Our source of patent data is the EPO Worldwide Patent Statistical Database (PATSTAT), which is prepared by the European Patent Office on behalf of the OECD Taskforce on Patent Statistics. It covers patent applications in more than 80 countries. The PATSTAT Register is a raw data product which is issued twice a year. We retrieved data from the 2017 Spring Edition. PATSTAT data is organized in subsets that cover specific details of patent applications. Information about the applicant and inventor is found in table "reg107". All patent applications have a unique numeric identifier called "ID". The information provided for each applicant and inventor is name, address, and nationality. 20.9 million observations containing names of applicants and inventors are registered in the PATSTAT 2017 dataset. Inventors are individuals, usually employees of the patent applicants. Applicants will have legal title to be the owners of the patent if it is granted and it is generally possible for the same person to be an inventor and an applicant (OECD, 2009). Therefore, we include both individuals classified as inventors and applicants in our analysis. The dataset was filtered to keep only observations for the 55 countries with the largest number of patent applications, which comprises 98.5% of the total observations in the PATSTAT database.

Reliability of the data source

PATSTAT are highly used among researchers on patent topics, and data description reports provides a good overview of the data. The main advantage of PATSTAT is its worldwide coverage and the inclusion of more information than other databases (Kang & Tarasconi, 2015), being a rich source of information about patents. We find the reliability of the data to be high.

External validity

The PATSTAT database gathers patent information from all over the world, but the database is biased toward European countries (Guerrero-Bote, Sánchez-Jiménez & De-Moya-Anegón, 2019). This can underestimate our results from non-European countries.

3.1.2 The Offshore Leaks

The Offshore Leaks is an online database containing a fraction of the aggregate leaked material of the Panama Papers (2016), Paradise Papers (2017), Bahamas Leaks (2016) and the Offshore leaks (2013). In addition to releasing the material, The Consortium of Investigative Journalists (ICIJ) decoded the leaks and built the open access database exposing the relationships between individuals, intermediaries, and shell companies. The database holds more than 785.000 offshore companies, and individual names are divided into categories called “officer”, “entity” and “intermediary”. Based on their investigations, ICIJ (2020b) offer the following understanding of the categories: Entities are the shell companies themselves. Officers are either beneficial owners in name or decision makers on behalf of the final owner. Intermediaries are law-firms or middlemen that creates entities on behalf of clients. Names of natural people from all categories are included in our analysis. While many of the high-profile leaks that circled the news wire were labelled as - or connected to - officers, ICIJ maintain that the ultimate beneficial owner may still be kept secret, partially because of the trust agents such as Mossack Fonseca placed in intermediaries to keep track of this information. Officers is however the category most likely to name individuals related to ownership. The Offshore Leaks are available for download at the ICIJ website as .csv files. Names and entities are connected by a variable NODE_ID. The information available are names, addresses and country of origin. Our dataset from all sources contained 1.5 million observations with names of entities, officers and intermediaries.

Reliability of the data source

Records of shell companies in tax havens are not intended for public disclosure, and the legitimacy of the records is demonstrated by prosecutions, stock markets and political shifts in the aftermath of the release. The Offshore Leaks database make up the largest compilation of exposed tax haven records to date (ICIJ, 2020a). The conveyor of the leak, ICIJ, won a Pulitzer for its' evidence-based work and is rated a low bias outlet by the News Fact Network (News Facts Network, 2020). Therefore, we consider the data source to be of high validity and the data to be suitable for our purpose.

External validity

Despite the inherent secrecy of tax havens, recent literature suggests that external validity can be attained. Omartian (2017) found significant similarities in how shell companies respond to exogenous pressure, which suggests that homogeneity of shell companies may be assumed across tax havens. The Offshore Leaks as a tax haven source is therefore likely to be random and the distribution of individuals in this source would not be different, had we picked another tax haven source.

3.1.3 OECD Patent Quality Indicators database

PATSTAT has become a standard among patent databases, and databases based on and linkable to PATSTAT have been produced by other institutions. For information about patent quality and technological fields, we used data from the OECD Patent Quality Indicators database from January 2020, which can be linked to the PATSTAT database through the common variable "APPLN_ID". The APPLN_ID was renewed and stabilized in 2011 (European Patent Office, 2016), which means that not all patent applications will have APPLN_ID and therefore miss quality and technological field information. The variables included in our analysis were FWD_CITS7, which is the number of patent citations received up to 7 years after publication, and TECH_FIELD showing the categorised technology field of the patent (Squicciarini, Dernis & Criscuolo, 2013). We found this to be relevant to our study on inventors in tax havens, to assess the characteristics of inventors with tax haven affiliation.

Reliability of the data source

The OECD Patent Quality Indicators are well developed and publicly available. Like the PATSTAT database, it has data description reports which provides a good overview of the data. We consider the data source to be reliable and suitable for our purpose.

3.2 Ethical considerations

One should always consider if using the data for other purposes than its initial intent could cause discontent with the subject of the data, and whether the data is suitable for the analysis in question. Because patents are public records and available for download by anyone, we did not consider the PATSTAT or OECD databases to be problematic to use in our analysis. Their size and coverage make them highly suitable. The Offshore Leaks are also publicly available for download, though the intended purpose of the information was for it to be kept confidential. We should therefore be cautious before drawing any conclusions about named individuals, as many people and entities have similar names. While we cannot conclude to what end shell corporations were created, several papers, investigative journalism and police investigations have revealed illicit use of these shell companies in tax havens[4], with anecdotal evidence suggesting that 95% of Mossack Fonseca’s work consisted of selling tax avoidance vehicles (Garside, Watt & Pegg, 2016; Omartian, 2017). Market and state level reactions to the leaks also suggest that tax avoidance and evasion were motivations behind using shell companies (Johannesen et al., 2018; O’Donovan et al., 2019).

3.3 Preparation of data

We identify inventors and patent applicants in the Offshore Leaks by matching the names in PATSTAT with the names of individuals in the Offshore Leaks database. Before the matching process, the names need to be cleaned and standardised in order to maximize the efficiency of the matching algorithm.

[4]There is considerable investigative journalism exposing high profile politicians and companies using the tax havens for personal gain or corruption(ICIJ, 2020c). Europol identified almost 3500 previously reported suspects of either money laundering, organized crime, terrorism, cyber criminals and VAT-fraud in the Panama Papers (Malan et al., 2017). In the framework, we explain how wilful concealment of assets was documented by Omartian (2017), corporate tax avoidance and corruption was documented by O’Donovan et al. (2019)

3.3.1 Problems with the datasets

PATSTAT and the Offshore Leaks data are close to its raw state, meaning the names have not been processed or standardised. Patent data is collected for various legal and administrative purposes, with no specific methodological requirements (OECD, 2009). Names and addresses found in the Offshore Leaks dataset are reproduced as they were leaked. Raw data needs to be cleaned and standardized in order to set individuals apart from companies, remove duplicates and be prepared for matching. The raw data imposes the following challenges: (i) noise within the name field; (ii) different formats of name; (iii) no unique identifier for one single individual; (iiii) typos in names and (v) missing information.

Noisy data can be manipulated with algorithms to increase its' applications. The fact that the Offshore Leaks and PATSTAT are rather noisy is not a big problem once appropriate procedures are implemented (Peruzzi, Zachmann & Veugelers, 2014). There are many available approaches to data cleaning and data harmonisation. However, many of the new and effective techniques from later years are often based on advanced techniques which apply complex algorithms or artificial intelligence to manage big data (Balsmeieri et al., 2018). Consequently, it can be difficult to reuse or replicate these approaches. The algorithms we perform are simple and easily replicable.

The objective of the cleaning and standardisation algorithms is to harmonise names in both datasets to improve the matching process. Our approach is based on the specific data content, and the same procedure is applied to both datasets. We remove all information (noise) in the name fields that is not part of the name, including addresses, company/university name and other miscellaneous information (id number, if deceased etc.). Due to the inconsistency in name formatting, we standardize all names to be in the same "first name surname"-format. We also perform character cleaning (e.g. removing ",", ".", double spaces etc).

Neither datasets have a stable identifier for each name. Address levels below country code (street, city, etc) are also missing for a lot of the observation. After name cleaning, we therefore remove duplicates in each dataset saying that the same name with the same country code is the same person. We remove all names that do not have a country code. Remaining typos or noise within the names are overcome using a "fuzzy" matching procedure to match names from PATSTAT to the Offshore leaks that exploits the available overlapping information in the two datasets and provides a similarity score for each matching name.

3.3.2 Coverage and accuracy

Coverage and accuracy are crucial when preparing and cleaning data (Eurostat, 2011). Coverage, or “completeness”, refers to the extent to which the cleaning procedure captures all name variations of the same person. Accuracy refers to the extent to which all name variations allocated to one person reflect one and the same person (Eurostat, 2011). Maximizing the completeness of the data requires automated procedures, though quality checks and validation are necessary to ensure accuracy of the outcome. Our methodology is based on an automated procedure, supplemented by manual quality control, to maximise both coverage and accuracy. It is in our interest to do so as the matching could only improve by having more candidates to match against.

Data cleaning is aimed to reduce noise without losing useful information, which is important for higher quality name matching. Raffo and Lhuillery (2009) showed the impact on different cleaning methods on matching results using the “bigram” algorithm[5]. One by one, each cleaning algorithm produces a small improvement. However, together, the gains in terms of precision and recall[6] are much greater than those provided by the simple addition of each technique`s marginal gain. The combination of cleaning techniques resulted in improvements in the precision rate from approximately 75% to 82%, and the recall rate from approximately 15% to 79% (Raffo & Lhuillery, 2009). Based on their result showing the importance of each step and their interaction, our cleaning stage comprises 17 steps. The effort of devoted cleaning is justified by that it allows for identification of more matches.

3.3.3 Cleaning procedures

Table 1 below shows an overview of our 17 cleaning steps. For steps that include dropping or splitting observations, the number of observations dropped or gained is shown for each dataset. An example is provided from the data showing how the names are cleaned. After the table follows a description of how each step were conducted. We start with 20,881,034 observations from PATSTAT and 1,549,731 observations from the Offshore Leaks.

[5] Matching algorithm using vectoral decomposition (Raffo, 2020)

[6] Precision rate = true positives / (true positives + false positives)
Recall rate = true positives / (true positives + false negatives)

Table 1. Overview of steps in the cleaning process with examples and number of observations dropped or gained

Step	Number of observations dropped or gained		Example		
	PATSTAT	Offshore Leaks	ORGNAME	NAME	
1. Pre-processing			MINORE, Jerry	minore, jerry	
2. Character and space cleaning			so chung keung, alfred ???	so chung keung, alfred	
3. Country code correction			CRISTOI, Florin / DE	cristoi, florin	
4. Drop missing country code	- 29,329	- 647,820			
5. Drop missing names and non-names	- 138	- 8,951	Apt. 102, 2-18-3 Shiboku	(deleted)	
			phd	(deleted)	
6. Country in parenthesis			raymond chan (china)	raymond chan	
7. Separate individuals from companies and universities	- 6,917,175	- 581,063	taipei medical university	(deleted)	
			samsung electronics ltd	(deleted)	
8. Remove titles and honorifics			MR. JUREIDINI, Michael	jureidini, michael	
			Heidrich, Adolf, Dipl.-Ing.	heidrich, adolf	
9. Remove end of string that is not part of name			SMITH, Adam Douglas 651 Franklin Street	smith, adam douglas	
			BENNETT, Alan, B., University of California	bennett, alan, b.	
10. Split aliases	+ 3,620	+ 1,977	john francis a.k.a. sean lee	john francis	sean lee
11. Split several names in name field	+ 58	+ 4,138	emmanuel ducrest and shahram diri	emmanuel ducrest	shahram diri
12. Special characters			HÃ¼bner, Heimo	hubner, heimo	
13. Remove miscellaneous information			tsakane mageza identity number: xxxxxxxxx	tsakane mageza	
14. Change name format			LAMAS, Carlos	carlos lamas	
15. Punctuation cleaning			BEHRENS, Timothy W.	timothy w behrens	
			Lay, Wai- Ming	wai-ming lay	
16. Drop one word names	- 1,176	- 737	GONZALEZ	(deleted)	
17. Duplicate names	- 9,909,262	- 44,776	Edlinger Alfred, Dipl.-Ing.	alfred edlinger	
			Edlinger, Alfred Dipl. Ing.		

1. Data pre-processing

Before starting name standardization, the input files have been checked to correct for any character encoding, normalize the format (to make sure data are in correct and comparable formats) and remove redundancies. These corrections are important to guarantee a proper application of the cleaning matching algorithms. After this preliminary data cleaning stage we executed manual inspection of a random sample of the data to better understand the characteristics of the PATSTAT and the Offshore Leaks dataset. For continuous validation throughout our cleaning process, we generated a variable that contains the original name - ORGNAME. The variable NAME contains the "cleaned" name. All names in NAME were made lower case.

2. Character and space cleaning

Double or several consecutive spaces are replaced with single spaces. Question marks are removed.

3. Country code correction

The name field in PATSTAT may contain the persons country code, indicating their country of origin. A forward slash followed by two letters corresponding with a valid country code were removed from the patentee names. The blank COUNTRY variable was replaced with the corresponding country code.

4. Drop if missing country code

As mentioned, one person can be entered in the PATSTAT database several times. Due to lack of additional information in the datasets, we assume that an individual with the same name from the same country, is the same individual both within the two datasets and between them when matching. We remove all names without country codes. Table 2 shows how many observations are removed in this step for each data source.

Table 2. Observations without country codes in each data source as share of total observations

Data source	*Total observations*	*Observations without country codes*	*Share of total observations*
The Panama Papers	484,146	95,915	19.8 %
Offshore Leaks	222,232	51,015	23.0 %
The Paradise Papers	641,662	299,813	46.7 %
Bahamas Leaks	201,691	201,077	99.7 %
PATSTAT	20,881,034	29,329	0.14 %

Table 2 shows that a lot of observatoins in the Offshore Leaks have missing country codes, and are therefore not included in the analysis. The high number of missing country codes in the Bahamas Leaks limits its application in our study.

5. Dropping missing names and those that are not names

As mentioned previously, both PATSTAT and the Panama Paper dataset can contain missing information. In addition, the name field can include words or characters that does not make up a name. The name field in the Offshore Leaks would also just refer to "the bearer" (see section 3.5). These are removed.

6. Remove name of country in parenthesis

The name field in the Offshore Leaks would sometimes include a country in parenthesis. These are removed. This help us in our next step, as these parentheses are often situated at the end of the name field with company names.

7. Separate names of people from companies and universities

The variable NAME in both PATSTAT and Offshore Leaks contain names of universities, companies, or natural people. There is also no variable indicating the classification. We are only interested in natural people.

To separate individuals from companies and public institutions, we wish to maximising the number of generic rules that can translate clues found in the name field into the proper classification (Eurostat, 2011). Such clues can be part of names, specific words (e.g. "government") and/or terms signalling legal forms (e.g. "Inc."). The starting point is an initial list of keywords/clues that are considered indicative of a certain category. These keywords/clues are applied to the full list of names. Case-based adaptions are introduced as needed, i.e. when too many false hits are generated by a particular rule. A case-based level increases quality levels – both in terms of completeness and accuracy (Eurostat, 2011).

To identify natural people in the datasets, we do the following: First, an inventor in PATSTAT will always be an individual, and are classified as such. Second, all observations in PATSTAT which included “c/o” in the name field, are categorised as individuals. We observe that names containing “c/o” had an individual name prior and a company name following. Third, all observations with titles such as “dr.”, “prof.”, “mr.”, “mrs.” are categorized as individuals.

For company categorization, we categorize every observation containing a legal identifier such as “ltd.”, “Inc.”, “s.a.”, “a.s.” at the end of the string as companies. In addition, if the name field contained specific words such as “company”, “enterprice”, “financial” or “ventures”, the observation are categorized as a company. Likewise, “university”, “faculty”, “department” etc. are used to categorize universities. An example of a case-based adaption to this method is if individuals are falsely classified as companies, such as "ruc, jacques a.g.", because “a.g.” also stands for Aktiengesellschaft (German public limited company). These individuals are reclassified as individuals at a case-based level.

8. Remove titles and honorifics

In both PATSTAT and Panama Papers, names will include titles such as “Mr.” and “Mrs.”, as well as honorifics such as “Dr.”, “Prof.”, “Dipl.” and name suffixes “Jnr.” “Esq” “III”. Honorifics are most prominent in the German PATSTAT names.

9. Remove end of string that is not part of the name

Upon visual inspection of the data we observed general problems with address information being added to the name field (e.g. road names and numbers). USA and Asian countries write their addresses with house number first, followed by street name. In these cases, any word that contain digits and all subsequent words are deleted from the name. In other cases, such as in Germany, house number are written after the name of the street, which require a case-based search and delete. Words such as “building”, “p.o. box” and “apartment” as well as the following words are also deleted.

In PATSTAT, company names or universities are sometimes included in the name field. We remove the company name from the name field. This is done by assuming that the inventors name will appear before that of the company, and then using the following steps: (1) identify a name containing a word associated with a company rather than a person, (2) read back from what word until either a punctuation or “c/o” is found, (3) delete everything from that punctuation mark onwards (Intellectual Property Office, 2019). Such punctuation mark are often not present, and personal name and company name will not have a separator (e.g. “holmes, elaine metabometrix ltd. rsm”). In case of the United Kingdom, the problem is extra prominent, and a separate file containing 19,949 first names are used to determine whether the word after the last punctuation are a first name or company name. Key words such as “University of”, “department of”, “dept. of” are used as separational words, meaning everything before is assumed to be a personal name, and everything after will be part of a university or company name. Specific company names are also used, as several company names occur with high frequency due to the high number of patent applications they file, e.g. “Glaxosmithkline”.

10. Split aliases and separate names

Names from certain countries (e.g. Germany and France), occasionally contain “geb.”, “nee”, “born” between two surnames. This indicates that the first surname is the persons (mainly female) married name, and the last surname is their given/birth name (e.g. “Ziegler, geb. Stadler, Elisabeth”). Patent attributes are usually a snapshot of data at the moment the dataset producer releases them (G Tarasconi, 2014). If the producer does not receive updates, such attributes are frozen at the moment of last update. To which extent Mossac Fonseca was updating their records is unknown. Therefore, to avoid the risk of one person having applied for a patent before they were married, using only their given name, but register in Mossac Fonsecas systems with their married name, we generate two separate variables containing each name (e.g. “ziegler, elisabeth” and “stadler, elisabeth”) as well as keeping both surnames as is in NAME.

The same division is done for aliases. Most prominent in the Offshore Leaks, names could contain both the individuals name, alias, or former name. The more alternative names we have, the more matches we generate (G Tarasconi, 2014). All name variations are used in the matching stage, avoiding risk of incompleteness.

11. Split several names in name field

The names of two or more different individuals might appear in the same name field. This is mostly the case in the Offshore Leaks if two or more people are joint shareholders. We include all names, as well as names of heirs and representatives of inventors (e.g. “DUBNOV, Halina, Heir Of Dubnov, Boris (deceased)”) for the case of completeness.

12. Special characters

Accented characters and characters with diacritic mark are shown as special characters (e.g. Ã©). When the set of special characters can be identified, the characters are replaced with their non-accented equivalents. This was done specifically in the German, Norwegian and French names where possible. When the special characters are not identified, they are left as is.

13. Miscellaneous information

In the Offshore Leaks, the names field will include other information on the client and general noise, e.g. id number, shareholder percentage, data of birth etc. This is removed.

14. Change in name format

In PATSTAT, generally the name format will be surname first followed by a comma and the first name - “Surname, First name”. In the Offshore Leaks, depending on the source, names are written in both “Surname, First name” and “First name Surname” format. Because the possible matches will be manually inspected, the ease of inspection will be higher if the names are formatted in the same way, as well as the similarity score the name matches get will be higher. Since 76% of the Panama paper names are written in “First name Surname”-format, and the first and last name in PATSTAT are usually separated by a comma, which makes an automatic approach easier, we change all names to the “First name Surname”-format.

15. Punctuation cleaning

Names may not only contain letters but also characters such as “,”, “;”, and ”-“ used to separate words or to indicate abbreviations and combinations. These characters might complicate or disturb the matching process, affecting the similarity score negatively. Period and commas are removed, and dashes followed by a space are replaced with dashes without a leading or trailing space.

16. Removing one word names

A name field in PATSTAT and Panama Papers will in a few cases only contain a surname, without first name or even an initial. In the Offshore Leaks, this is also often the case with only first names. For our matching purposes, it will not give us enough information to identify matches, and these observations are dropped.

17. Duplicate names

The same names can appear in different form in the databases for several reasons: titles, inclusion of company name, address, different character placement or order of first name and surname. There is a trade-off between accuracy and completeness, where this step favours accuracy. A transparent and accurate set of harmonized names in which completeness can be gradually improved, is considered far more appealing than a more complete set which contains the risk of not being accurate or being unsuited to specific analytical purposes (Eurostat, 2011).

Individuals in either dataset do not have a unique identifier. The address information is also missing and incomplete. Because of the lack of information, we have assumed that the same name from the same country (similar country code) are the same person. After undergoing the previous cleaning phases, we end by dropping duplicate names within the same country.

This strategy could incorrectly merge two people with similar name from the same country. This is a source for additional noise in the matching process, but we have too little information (in both PATSTAT and the Panama dataset) to adequately improve this part of the algorithm. A further improvement of this method could be to extract information from other sources.

3.4 Matching procedure

We now describe how we match the names of inventors and applicants present in the PATSTAT database to the names in the Offshore Leaks using a "fuzzy" matching process and how we filter the matching results. Obtaining the final results involved three stages: cleaning, matching, and filtering the results. Raffo and Lhuillery (2009) concluded that the interaction between the three stages is critical for achieving high precision and recall rates. We included the 55 largest countries from the PATSTAT register. This leaves us with 3,612,805 observations from PATSTAT and 206,982 observations from the Offshore Leaks to be matched against each other.

3.4.1 Matching algorithm

The matching is based on a "fuzzy" procedure that exploits the available overlapping information across the two databases: the name and country of origin (Gianluca Tarasconi & Menon, 2017). Although both datasets have undergone thorough cleaning procedures, there might still be spelling errors and other noise left in the data. For instance, Slavic names when translated from the Cyrillic to the Latin alphabet could be spelled in different ways. Due to the size of the data, it is not possible to correct this noise further within our timeframe. Fuzzy matching is intended to overcome these kinds of problems. We use a user developed algorithm in Stata called Matchit. Matchit is a tool to join observations from two datasets based on string similarity between the chosen variables (Raffo, 2020). In computer science, a string is a sequence of characters, such as letters, symbols, or numerals. Matchit returns a new variable called SIMILSCORE that ranges from 0 to 1 using the Jaccard distance[7]. A similarity score of 1 implies a perfect similarity according to the string matching technique chosen and decreases when the match is less similar (Raffo, 2020). This makes Matchit a convenient tool to join observations when the string variables are not necessarily exactly the same.

Matching algorithms can be categorized in three main families: Vectorial decomposition, Phonetic and Edit-distance algorithms (Raffo, 2020). Vectorial decomposition algorithms, such as N-Gram, Token, etc, compares the elements of two strings. The N-gram algorithm decomposes the text string into elements of N characters (grams) using a moving-window basis. Bigram (2-gram) splits text into grams of 2 moving chars. e.g. "John Smith" splits to “Jo oh hn n_ _S Sm mi it th” (Raffo, 2020).

However, vectoral decomposition algorithms do not need to have a moving-window structure (Raffo, 2020). For instance, the Token algorithm splits a text string simply by its blank spaces. In “John Smith” there are only two elements (or grams) “John” and “Smith”. These match perfectly with “Smith John”, but only one from either “Smith, John” or “Smit John”.

The N-gram algorithms work effectively on misspellings as well as large string permutations. Even though a clear hierarchy is hard to achieve for several reasons, 2-gram are found to be better performers than 3-gram, 4-gram, or Damerau-Levenshtein algorithms (Christen, 2006).

Although we standardised names as part of the cleaning process (step 14), first name and surname could still appear in the wrong order of each other when matching the two databases. For this reason, we first match using the Token-modification, to save time filtering the wrong-order-perfect matches. All names with a similarity score of one are kept. After this we perform a 2-gram algorithm on the same dataset. The matching results with a similarity score between one and 0.7 are manually inspected. The threshold of 0.7 was determined after trial and error, with using both testing data sets to find how differences in names would affect the similarity

[7] Calculated by the intersection between the two strings over the union of them: m/sqrt(s1*s2) (Raffo, 2020)

score as well as investigating the matching results of the main data sets. Visual check, based on both harmonized (NAME) and original names (ORGNAME), serves to minimize errors in approximate matching.

3.4.2 Filtering stage

Aside from name we use country code as a criterion for matching. Only names with similar country of origin are matched against each other. This can be considered a weakness in our process, because as mentioned not all names had country codes and the names are therefore left out. Other sources of errors can be the wrong country code assigned to a person, or people could have moved. In the Offshore Leaks 24,763 names have two or more country codes, e.g. “NOR;DEU”, which order were not found to be systematic. These names are matched with all corresponding countries in PATSTAT, e.g. both Norway and Germany. If a name have a match in both countries, the first country in the country code variable are chosen – in this case Norway. This is also considered a weakness.

For two names to be considered a “match”, first name and surname must be identical. Deviations from identical can occur when we find it highly likely that there has been a spelling mistake or separation error, taking a conservative approach. This filtering phase aims to delete false matches and add false negatives. Whenever trying to match two distinct data sources, an inevitable trade-off arises between flagging positive matches as false negatives and including false positives as true matches, measured by recall rate and precision rate. In both cases, the values should be as close to one as possible (G Tarasconi, 2014).

The reason for inspecting the names with similarity score one in bigram-matches, is that names such as “shih-chuan cheng” and “shih-chen chuang” will be considered a perfect match (similscore 1) although they are not the same name. This is because they are split into similar grams. These cases were few, and only apparent in Asian countries. Thus, false positive matches are removed.

For false negatives, a manual check is carried out on all matches below one and above 0.7 in similarity score. This entailed looking for possible spelling errors or separational errors in the name field. We also perform a brief online search such as looking for them on ESPACENET (online patent database provided by EPO), in order to determine whether or not a close match actually could be a match (G Tarasconi, 2014). Table 3 shows examples of names with similarity score below one, but which we consider to have cases of spelling and separational mistakes, and therefore to be a match after all.

Table 3. Examples of names with similscore below 1 considered a match

NAME (PATSTAT)	*NAME (OFFSHORE LEAKS)*	*SIMILSCORE*
martine-marcelle-maria campmas	martine marcelle maria campmas	0.8431
lars erik alm	larserik alm	0.8704
igor vorob'ev	igor vorobyev	0.8571
aleksandr vladimirovic kuzmin	alexandr vladimirovich kuzmin	0.8667
domenico di cesare	domenico de cesare	0.8824

Another criterion for matching is that a middle name occurring in one of the names, have to occur in the other as well. We accept a match where the middle name in one name corresponded to the initial in the other name. We also accept when an initial in both names matched, as well as the rest of the name. To which extent the registration of names in the Offshore Leaks are including middle names is unknown, and the exclusion of matches when a name is similar, but the middle name is missing (e.g. “Paul Smith” would not match “Paul George Smith” but might refer to the same person) might underestimate the total amount of matches. We accept only a few cases with missing middle name if the online search provided a bases for us to do so. Table 4 below shows examples of matches where the middle name is an initial in one or both names, but we still consider the names to be a match.

Table 4. Examples of matches with incomplete middle names

NAME (PATSTAT)	*NAME (OFFSHORE LEAKS)*	*SIMILSCORE*
albert eltjo doewe van capelleveen	albert e d van capelleveen	0.8451
daniel r williams	daniel r williams	1
neil r jones	neil richard jones	0.7635

This final filtering stage depends on the ability to obtain and implement complementary information in order to identify and reject false positives. This will also result in a higher likelihood of improved precision. However, we do not have additional information about the names in our databases. Address information is often incomplete or very noisy. We will only rely on name and country code, as well as results in online searches.

Our resulting dataset contain 15,896 matching names of inventors in PATSTAT and the Offshore Leaks database, from 53 of the 55 countries. 954 of these are patent applicants and 14,942 are inventors.

3.5 External validity

Type 1 and 2 errors

External validity refers to the degree to which the findings of the study can be generalized to other relevant scenarios (Saunders, Lewis, Thornhill & Bristow, 2012). The external validation of our findings can be limited due to what we have called “false positive matches” or “false negative matches”, known as type 1 or type 2 errors respectively. The number of inventors we find in the tax haven data will be overestimated if there are mostly type 1 errors or underestimated if there mostly type 2 errors. A type 2 error will occur if there are inventors in the tax haven data that we have not identified. This can be due to misspellings in the name which were not detected in the filtering stage, or the absence of a middle name in one of the databases which would not fulfil our criteria for a “match”.

An inventor could also have moved to another country. If the databases are not updated, this might lead to the inventor having the wrong country code and not be detected due to our “same name, same country”-criteria. Akcigit et al. (2016) show, however, that less than 5% top inventors actually change their residence country during their careers. Another reason for our results to be underestimated is the fact that not all names are given in the Offshore Database. A lot of the name fields only refer to “the bearer”, meaning shares that are considered to be owned by whoever physically holds a share certificate (ICIJ, 2020b). Bearer shares provide one of the deepest levels of secrecy. We will not be able to know how many of these are inventors.

A type 1 error will occur if the person we have identify in the Offshore Leaks as an inventor, is in fact not the same person as the name in the PATSTAT database. This will overestimate our findings. Generally, a cross-check against affirmative sources is needed to achieve absolute certainty of a name match, which is practiced in similar studies conducted by Europol and the European Anti-Fraud Office (European Parliament, 2016; Malan et al., 2017). Inconsistent address data prevent the use of a stricter criteria than name-country match, but name-country matches can be particularly weak. For instance, there were about 45 000 “John Smith”s in the United States in 2010 (Auron Technologies, 2009). There was no actual John Smith in our matches, but several names were more common than others. It is likely that commonness decreases as the number of words in a name increases and further that longer name matches are more reliable. Table 5 shows the distribution of name lengths of matches.

Table 5. Distribution of name length of matches

Number of words in names	*Freq.*	*Percent*
2	13,649	85.86
3	2,134	13.42
4	101	0.64
5	11	0.08
6	1	0.01
Total	15,896	100
Average number of words in names	2.15	

The high number of names with shorter name lengths, considered to be more likely a source of type 1 errors, might overestimate our result of the number of inventors with investments in tax haven location. Particularly Chinese names are prone to being overestimated, as most names consist of only two names (a single first name and surname).

In Section 4.4 we present what we call inventor collaborations, where several names of inventors from the same patent application (same ID) are found in the Offshore Leaks. The original name in the Offshore Leaks (ORGNAME) confirms that this is likely to be more than a coincidence. For example, we found two Swedish inventors from the same patent application, where both names of the inventors are in the same name field in the Offshore Leaks. We also found several Israeli inventor collaborations where all or several of the individuals` original names in the Offshore Leaks stated that an employee remuneration trust where nominee of the inventor (e.g. the Offshore Leaks name “employees remuneration trust company as nominee for samuel faran” where “Samuel Faran” is found to be an inventor in PATSTAT, and in inventors in the same patent application in PATSTAT also match with the same type of original name in the Offshore Leaks). The inventor collaborations were mostly Chinese, which mitigates the concerns about name disambiguation for Chinese inventors. In summary, some matches are subject to disambiguation, but observed inventor collaborations suggest a reasonable credibility of the matches.

To which degree a type 1 or type 2 error is most likely to occur is unknown. Steps to reduce both are implemented. The thorough cleaning of name and manual check in the filtering stage reduce the likelihood of type 2 mistakes. Findings such as inventor collaborations confirms the validity if the matches and reduces the likelihood of type 1 mistakes. There will be matches that are not in fact the same person, but there is also a number of inventors that we will not find because they use bearer shares. Overall, we still might underestimate the share of inventors in tax haven.

Generalisation of inventors in tax haven

The Offshore Leaks only comprises a fraction of the total concealed shell companies and general information about tax haven investors (Johannesen & Zucman, 2014). Omartian (2017) found evidence suggesting that homogeneity of shell companies may be assumed across tax havens. The distribution of inventors found in our four sources in the Offshore Leaks is showed in Table 6.

Table 6. Distribution of inventor matches across Offshore Leaks data sources

Offshore Leaks data source	*Individuals in tax haven*	*Individual from each source (%)*	*Inventors in tax haven*	*Inventors from each source (%)*
Bahamas	45	0.02	0	0.00
Offshore	50,330	25.87	3,529	22.20
Panama	60,322	30.99	5,767	36.29
Paradise	83,961	43.13	6,600	41.52
Total	194,658	100	15,896	100

Note: The individuals in tax haven show the number of individuals with country code information. The Bahamas leaks had very few observations with country information, which severely limited its' application for our study.

Table 6 shows that the distribution of inventors in tax havens between the different sources are similar to the general distribution of individuals between the sources. If we pick the Offshore Leaks as a tax haven source, it is likely to be random and the distribution of inventors in this source would not be different, had we picked another tax haven source. This suggests that inventors have no, or weak preferences of the tax haven used, and have similar behaviour to other people investing in tax haven shell companies. This strengthens the external validity of how we can generalize the behaviour of inventors in tax haven from our findings in the Offshore Database.

The use of name-country code information only

In step 4 in the cleaning process, we removed all names that do not have a corresponding country code. This amounted to 0.14 % of the PATSTAT database and 42 % of the total Offshore Database, most profoundly is the lack of country code in the Bahamas Leaks. We assume that missing country codes are random, and the relative findings from each country will still be valid, although not being able to take full advantage of our tax haven data might underestimate the absolute number of inventors engaging in tax haven activity.

Company names within the natural persons population

The case-based approach of separating natural persons from companies and universities in step 7 of the cleaning phase show that some individuals were wrongfully categorised as companies, and vice versa. In particular, it could be difficult to tell the difference between a law firm and two individuals being named together in the Offshore Leaks. Companies may also be named after individuals. Hence, there might still be companies within the final dataset of individual names. Specifically, this will affect the ratios where we use the total number of individuals in PATSTAT or the Offshore Leaks in the nominator or denominator. When inspecting the data, and applying case-based approach, we found that more companies were wrongfully categorized as individuals, than the other way around, ensuring completeness in our dataset of individuals. In the filtering stage we examine the matching result with similarity score above 0.7, including similarity score one. The presence of a company name within the names of inventors with investments in tax havens (our matching result) are therefore less likely.

4. Analysis

The objective of our analysis is to identify to what extent inventors engage in tax havens and their characteristics. We base our analysis on descriptive research. Descriptive research design is most appropriate when the purpose is to describe the characteristics of people, objects, organisations, environment, or groups (Zikmund et al., 2012). The research design addresses the what, who and where question (Wilson, 2010). However, descriptive research cannot identify any causal relationship (Saunders et al., 2012).

In this section, we present the inventors and patent applicants found in the Offshore Data base. To identify the extent to which inventors engage in tax havens, we compare the frequency of inventors in tax havens to the frequency of inventors in the general population. We do this on a country by country basis. This way, we find out if inventors are overrepresented in tax havens compared to the general population within the country and on average.

To better understand inventor's engagement in tax havens, we further investigate their characteristics compared to the average inventor. Specifically, we look at forward citations to see if inventors in tax havens are more successful than the average inventor and number of patents to see if they are more productive. In addition, we look at the patent technology groups to see if there are specific groups of inventors that seek tax haven activities. We also look at their connection to corporate or public institutions compared to the average inventor. In the end, we present a finding where groups of inventors from the same patent application are found in the Offshore Leaks.

4.1 Inventor population in tax havens

We start by looking at the geographic distribution of inventors identified in the Offshore Leaks. Figure 1 shows how many inventors and patent applicants we identified in the Offshore Database from each of the 55 countries in our patent data.

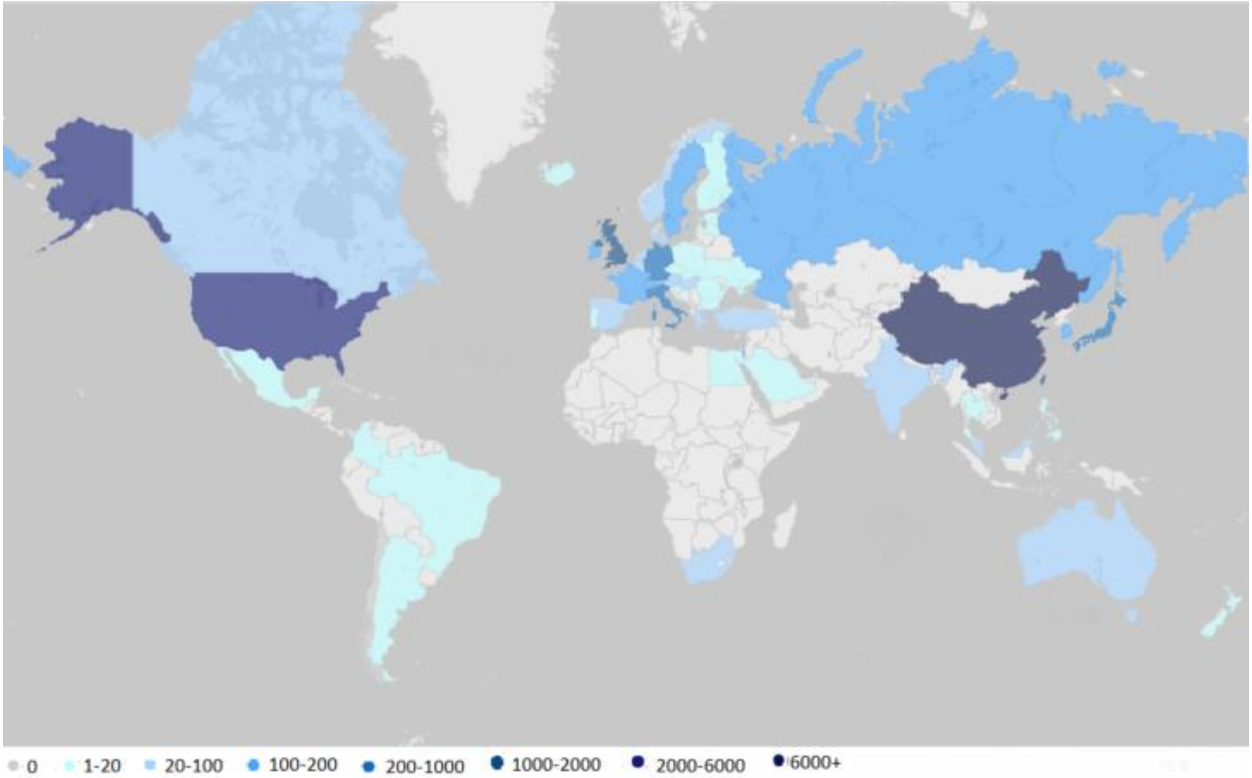
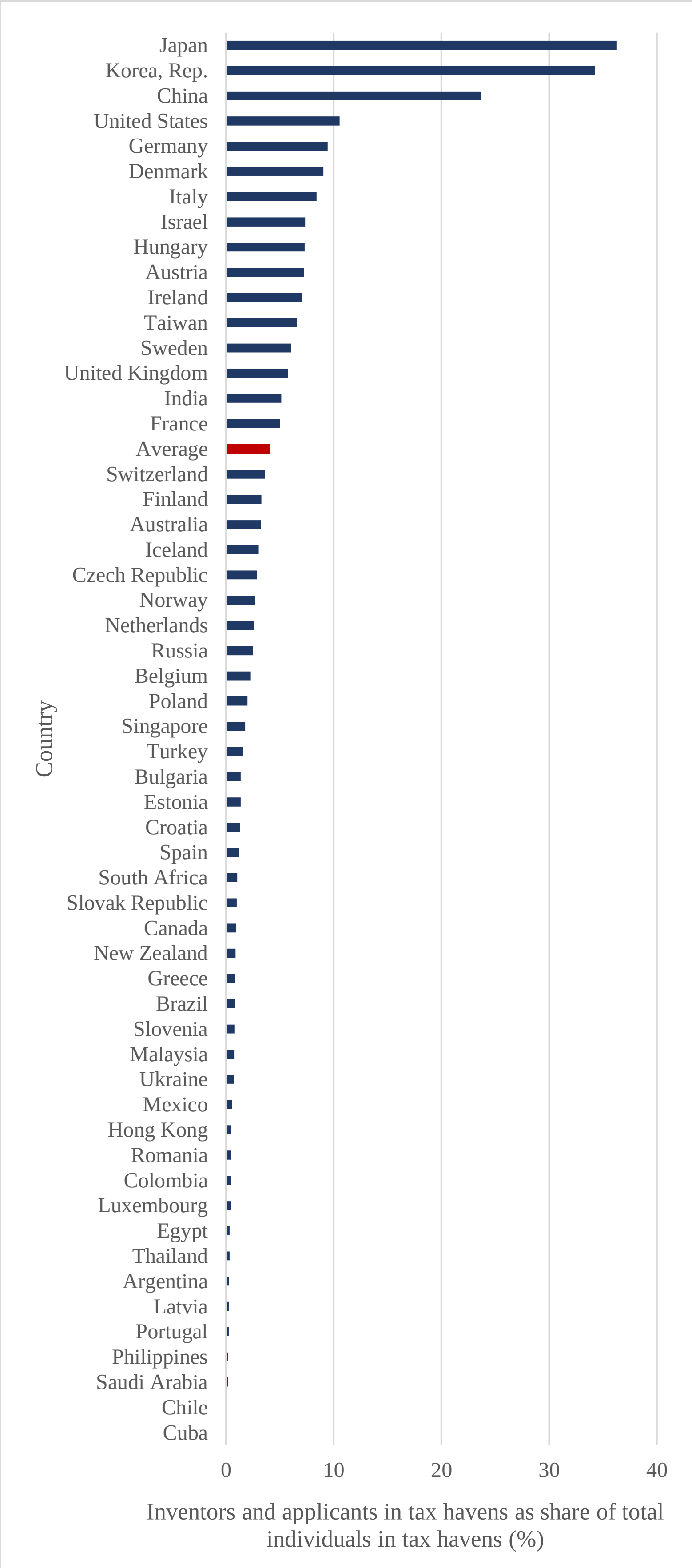


Figure 1. Absolute number of inventors found in the Offshore Leaks for each country

China had the highest absolute number of inventors and patent applicants invested in tax havens, with 6,961 individuals, followed by United States (2,765), Taiwan (1,205) and United Kingdom (1,144). Cuba and Chile are also among the 55 countries we examine, but we found zero inventors or patent applicants in the tax haven data.

4.1.1 Inventors in tax haven vs inventors in general population

We wish to see if inventors and patent applicants are overrepresented in tax havens compared to the general population. We do this by looking at two ratios: the share of inventors among the population of tax haven individuals (IT) and the share of inventors in the whole population (IP). If the former is larger, it implies an overrepresentation of inventors in tax havens. Figure 2 below show IT and Figure 3 shows IP for each of the 55 countries we are looking at. They are shown in two separate figures due to the large difference in distribution, IT ranges from zero to 36% while IP ranges from 0.0009% to 0.74%. Figure 4 shows IT over IP as a measure of overrepresentation.



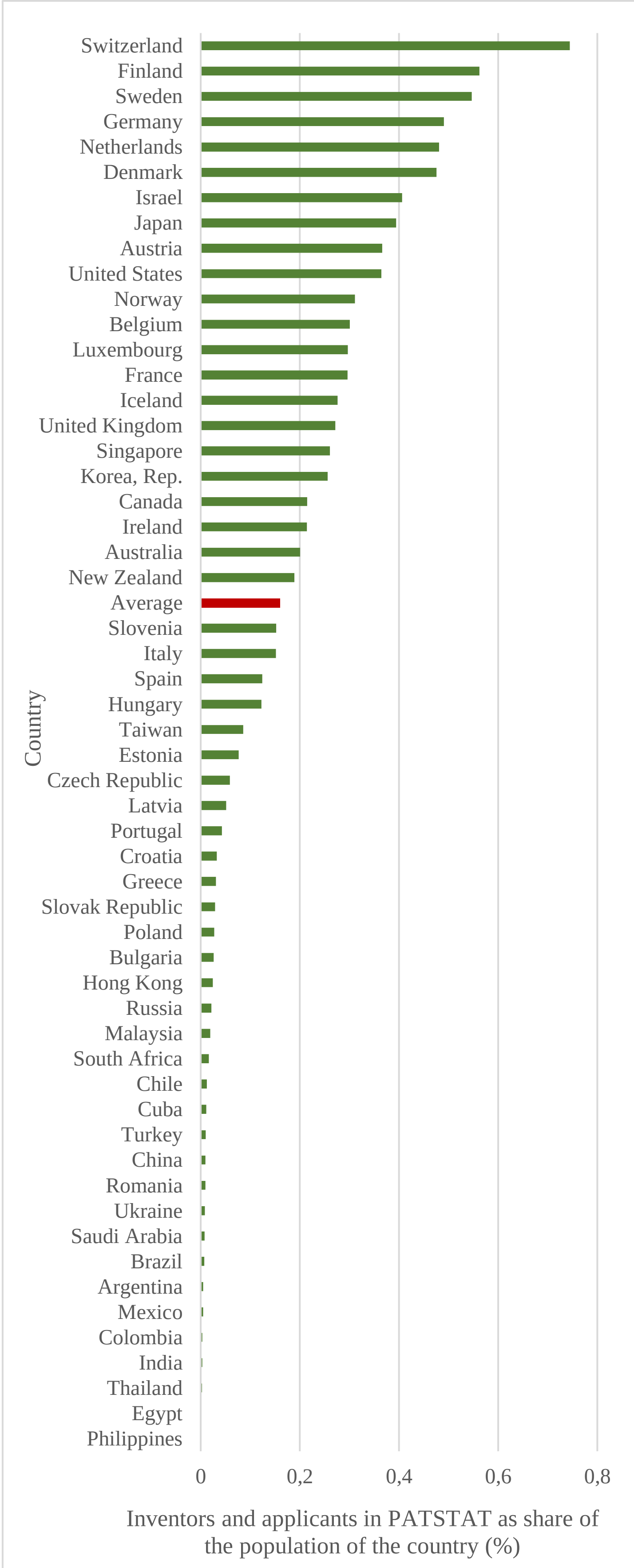


Figure 2. IT: Inventors and applicants in tax havens as share of total individuals in tax havens for each country

Figure 3. IP: Inventors and applicants in PATSTAT as share of the population of each country

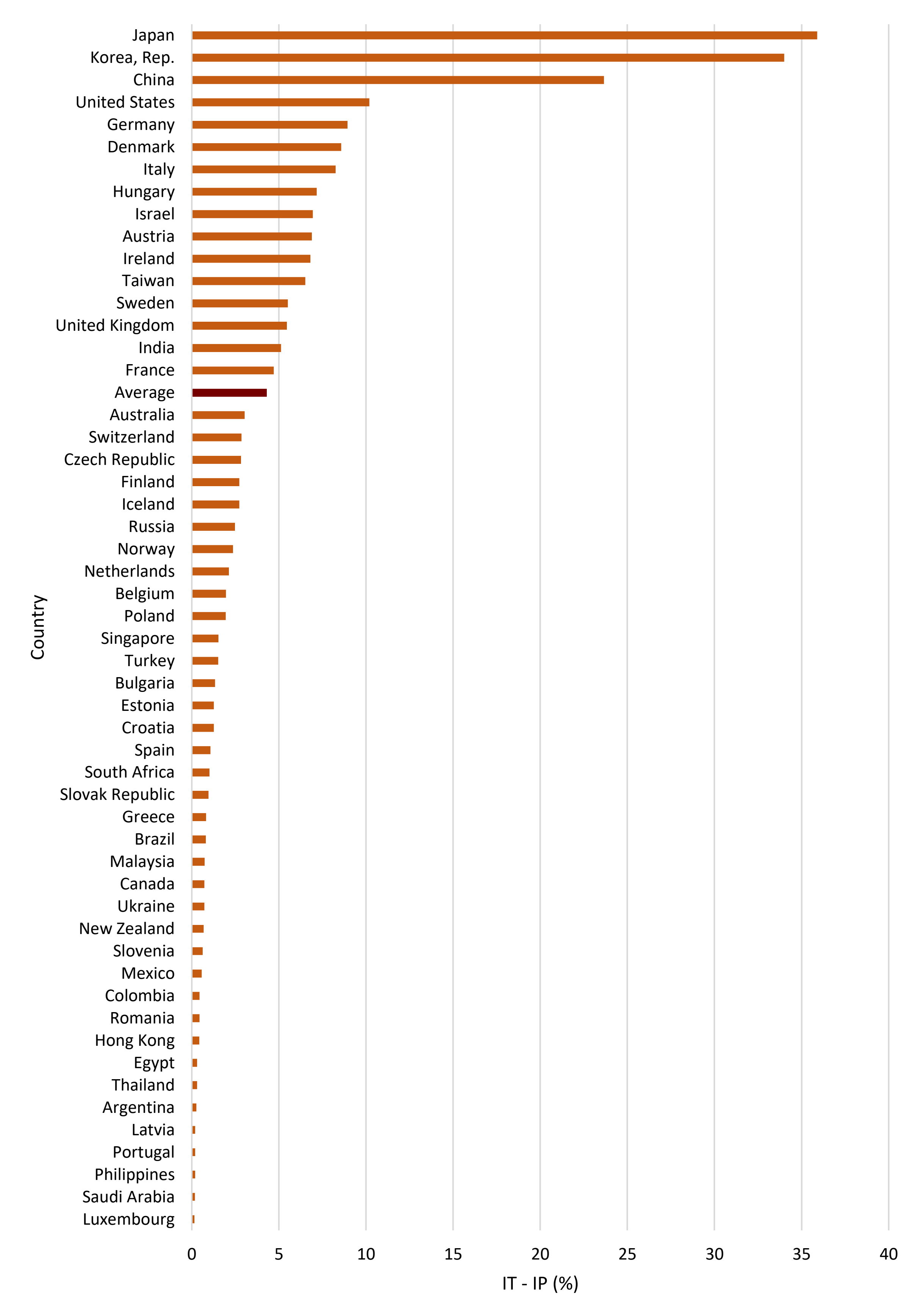


Figure 4 IT over IP as a measure of overrepresentation of inventors in tax havens for each country and the average

Figure 2 shows inventors and patent applicants as share of individuals engaged in tax haven per country, from the country with the highest share, to the country with the lowest. The top three countries are East Asian, showing that 36% of Japanese, 34% of South Korean and 24% of Chinese individuals engaged in tax havens are inventors. We address these countries in a case study later in this section. The United States is fourth with an IT of 11% and Germany is fifth with 9%, followed by mostly Western European countries in the top 10. There are stark differences between the top three countries as well as between them and the other countries. This is emphasized by the difference between China in third and the United States in forth at 10.5 percentage points, being the largest relative drop on the distribution. Below the United States, the differences in IT between countries are more gradual.

Figure 3 shows inventors as share of the general population, based on PATSTAT and population data[8] (The World Bank, 2020). We see that Switzerland has the largest inventor stock as share of the population with 0.74%. Finland, Sweden and Germany are the only other countries with inventor stock as share of total population above 0.50% followed by the Netherlands with 0.48%. The differences between countries in Figure 3 is more gradual than the distribution of Figure 2. Germany is the only country present within the top five countries in both Figure 2 and Figure 3.

Figure 4 shows the difference between IT and IP as a measure of overrepresentation of inventors in each country. For each country that had at least one inventor in tax haven, the ratio of inventors in tax haven is always greater than the ratio of inventors in the general population. This means that for each of the 53 out of 55 countries, inventors are overrepresented in tax haven. Figure 4 shows the degree of overrepresentation, from highest to lowest for each country and the average. Chile and Cuba had no inventors in tax haven. The mean IT of all the 53 countries is 3.96 % and the mean of IP is 0.15 %. We conducted a two sampled T-test of unequal variances and compared IT ($M = 0.0396$, $SD = 0.69$) to IP ($M = 0.0015$, $SD = 0.0018$) and see that $t(54) < 4.07$, $p = 0.001$. Overrepresentation is significant at the 0.01% level

The significance is subject to Type-I error because of skewness in the underlying distributions and non-random sampling of individuals in tax havens. We reject the hypothesis that inventors

[8] The population represent the population in each country in 2017, because our PATSTAT data is from 2017 and the latest addition to the Offshore Leaks is in 2017, making comparisons more valid. The data is gathered from the World Banks website 4. May 2020.

are equally represented in the tax havens as non-inventors. In Table 7 we show the number of inventors in tax haven, IT, IP, and IT-IP by geographical region

Table 7. Number of inventors in tax haven, inventors in tax havens as share of total individuals in tax havens (IT), inventors as share of general population (IP), and IT-TP as a measure of overrepresentation by geographical region

	Number of inventors in tax haven	*IT (%)*	*IP (%)*	*IT-IP*
Western Europe	3,228	5.57	0.31	5.26
Eastern Europe	286	2.37	0.02	2.35
East Asia	9,040	12.11	0.05	12.06
Indo-Pacific	331	1.88	0.01	1.87
North America	2,799	9.39	0.35	9.04
Latin-America	33	0.53	0.01	0.52
Africa	29	0.89	0.01	0.88
Middle East	150	3.21	0.04	3.17
Total	15,896			

Note: The number of inventors in tax haven in each country is found in Table A1 in Appendix. Western Europe includes United Kingdom, Italy, Germany, Switzerland, France, Ireland, Sweden, Austria, Denmark, Netherlands, Belgium, Spain, Norway, Greece, Iceland, Finland, Luxembourg, and Portugal. Eastern Europe includes Russia, Hungary, Poland, Czechia, Ukraine, Bulgaria, Estonia, Slovakia, Romania, Croatia, Slovenia, and Latvia. East Asia includes China, Japan, Taiwan, Hong Kong, and South Korea. Indo-Pacific includes Australia, New Zealand, India, Philippines, Thailand, Singapore, and Malaysia. Latin-America includes Mexico, Colombia, Argentina, Cuba, Chile, and Brazil. North America includes Canada and United States. Africa includes Egypt and South-Africa. Middle East includes Saudi-Arabia, Turkey, and Israel.

Table 7 shows that the highest IT ratios belong to East Asia, North America and Western Europe, and the lowest to Latin-America. Neither region in Table 7 represent the regions as a whole, and we caution that Africa, Asia, Indo-Pacific and the Middle East had very few countries among the 55 most patenting countries we looked at. In addition, the countries in Eastern Europe and Latin-America have very few identified inventors each in the Offshore Leaks. For example, the number of inventors identified from each country in Latin-America except Brasil (with 20) are between zero and four, and between one and 18 in Eastern Europe except Hungary (57) and Russia (168). These low numbers might suggest random matching due to name similarity, as discussed in section 3.5, and lower the validity of the results from these regions.

To pave the way for inventor-specific theories on tax haven investments, we first compare the distribution in Figure 2 to general patterns in offshore tax avoidance. Figure 2 shows that it does not conform to recent studies on the offshore wealth distribution of countries. (Alstadsæter, Johannesen, et al., 2018) found that a country's offshore wealth holdings were associated with proximity to Switzerland, political instability, and access to natural resources. The upper half of the distribution consists of more politically stable countries than the lower half (Samans et al., 2017; The Global Economy, 2018), with and without large natural resources and proximity to Switzerland. Top personal income taxes could explain some of the differences, but countries with income tax rates close to the OECD average of 40,2% (2017) are distributed both high (Denmark, Germany, Japan) and low (Belgium, Netherlands, Spain). In Figure 2, there are large countries (Germany, U.K) alongside smaller countries (Denmark, Sweden), and less corrupt (Japan, Austria) countries alongside more corrupt countries (South Korea., Hungary), as per the Corruption Perception Index (Transparency International, 2016). The top five countries conform to the top five countries in global patent applications, but the rest of the distribution does not conform any specific measure of innovative activity by comparison to statistics from WIPO (2017), although several countries with considerable innovative activity predominantly lie in the upper half of the distribution. Several countries in the bottom half have greater reliance on natural resources (Saudi-Arabia, Mexico), impose lower taxes than the OECD 2017 average of 40.2% (Ukraine, Brazil, Hong Kong) and score lower on the Inclusive Development Index by the World Economic Forum (Samans et al., 2017). These observations imply that the degree of innovation, or factors associated with the degree of innovation, could explain some of the variation.

The distribution of Figure 2 indicated that over-representation and country differences probably is explained by the interaction of several factors. A simple explanation for over-representation might be that on average, inventors are more likely to have shell companies than other citizens because they are wealthier than the average. This makes sense under the assumption that the shell companies they own are often used for tax evaison (Omartian, 2017), and offshore tax evaison is associated with wealth (Alstadsæter et al., 2019b; Zucman, 2013). We investigate the success of the inventors in tax haven in Section 4.2.1. A partial explanation could be that country differences are driven by how common it is to get rich from innovation compared with other activities.

Another explanation for over-representation could be that tax avoidance by multinationals lead to network effects. One line of thinking that could support this is that inventors live where the R&D takes place, which is often in non-havens where income taxes apply (Böhm et al., 2012; Karkinsky & Riedel, 2012). If the multinational is already using tax havens to avoid taxes, there could be a case for network effects as explained in Section 2.2.2, where communication between near affiliates lead to people learning how they can evade without getting caught.

Another explanation for over-representation could be that multinational corporations provide opportunities for inventors to invest in tax havens. We provide a recap of the incentives to provide this opportunity: Multinationals tend to allocate ownership of high value patents to low tax jurisdictions, to receive low or no taxes on patent income[9]. Multinational firms with significant R&D activities, like IBM and Huawei, harvest patents as a strategic means to influence industry architecture[10] and gain market power (Torrisi et al., 2016). Inventors are key for sustaining such competitive advantages through their continued innovation (Gambardella, Harhoff & Verspagen, 2005), but they also known to be mobile (Agarwal et al., 2009; Miguelez & Fink, 2013; Møen, 2005). Hence, to prevent the loss of key personnel or knowledge spillover to the competition, multinationals could provide offshore investments to attach key workers closer to the company. If a higher share of inventors in tax haven are found to work with corporations, compared to the general inventor, this could indicate a link between tax havens investments and certain employers. We investigate such links in Section 4.4, inventor collaborations.

This sub-section illustrated that inventors are over-represented for every region, but also that having one inventor in a tax haven was enough to be over-represented. We discussed possible explanations of differences in over-representation, such as wealth of the inventors, degree of innovative activity in the country and observed that countries with high income taxes were both

[9] Griffith et al. (2014) and Karkinsky and Riedel (2012) studied how multinationals allocate patents amongst subsidiaries with respect to corporate taxes, accounting for regulatory challenges and costs associated with shifting patents, still finding significant allocations towards low tax countries. Tørsløv et al. (2018) estimated that multinationals shift close to 40% of global profits to tax havens each year, with non-haven European countries being the main losers from the shifting.

[10] Industry architecture refers to product- or industry standards and how ownership of relevant technology determines which firm are positioned to extract the most value. One example is the Windows operating system, which is included as opposed to a purchase extra when buying PCs. This way, Microsoft prevents competition on operating systems and receives income from selling programs tailored to their operating systems. See Pisano and Teece (2007) and Jacobides et al. (2006) for detail.

high and low along the ratio of inventor in tax havens. We discussed how firms may have incentives to offer tax haven investments to retain inventors and avoid knowledge spillovers. Shell companies may be used for various legitimate purposes (Malan et al., 2017; van der Does de Willebois, 2011), but the majority of shell companies are used to conceal assets or obtain illicit advantages (Alstadsæter et al., 2019a, 2019b; David, 2016; Farrell, 2016; Johannesen & Zucman, 2014; O'Donovan et al., 2019; van der Does de Willebois, 2011). The implication of this is that we can assume inventors with tax haven investments to be wealthy, which should be related to their success.

Case study: Japan, South Korea, and China

As an extension of the discussion where we suggest a link between shell companies and returns from innovation, we consider how institutional trust and network effects might have led to the extra over-representation of (assumed) tax evasion by inventors in Japan, South Korea. and China. Top marginal income taxes are respectively 45%, 42% and 45% which suggests that tax rates themselves are not causing the extra evasion, because Germany (45%), Finland (55%) and France (45%) operate with comparable tax rates and had considerably fewer inventors in tax havens (PWC, 2020).

Institutional trust

Inventors in China may be discouraged to expose assets in financial institutions that could be under government influence, because of a weak history of enforcing intellectual property rights (Dimitrov, 2009; Du, Lu & Tao, 2008; Keupp, Beckenbauer & Gassmann, 2009; La Croix & Eby Konan, 2002). As documented by Wang (2014) and Xin and Pearce (1996), weak Chinese law enforcement encourage individuals to use Guanxi networks to obtain protection of property rights and fend off government extortion. Guanxi are networks of individual relationships cultivated through the exchange of gifts and favours to attain mutual benefits, with reciprocal obligations to respond upon requests (Alston, 1989; Xin & Pearce, 1996). A lack of asset protection may especially be a concern for firms and individuals in position to commercialize on intellectual property. This form of non-compliance can be related to the "slippery slope framework" explained in 2.3.1, in the sense that Chinese inventors could lose confidence in institutions and reduce compliance.

Japan abolished a policy of publicly disclosing tax payments above certain thresholds in 2004, which led to an increase in tax payments by people who previously evaded to stay below the threshold (Hasegawa, Hoopes, Ishida & Slemrod, 2012). This conforms to the “slippery slope” dynamic of non-compliance, in the sense that the extra degree of control evidently had a non-compliant effect on a share of the taxpayers.

Network spillovers – learning tax avoidance through organisations

Firms belonging to Keiretsu networks in Japan are empirically found to redistribute profits to each other through transfer pricing, effectively reducing taxable profits (Gramlich, Limpaphayom & Ghon Rhee, 2004; Lincoln, Gerlach & Ahmadjian, 1996). Keiretsu are horizontal or vertical networks of businesses that work closely together, while being operationally independent (Liberto, 2019). South Korea has a different form of networks called Chaebol, which is made up of single or cooperating conglomerates (Kenton, 2019). Similar to Keiretsu, Chaebol are also found to shift income than firms not in Chaebol (Jung, Kim & Kim, 2009).Thus, if inventors are associated with firms in Keiretsu or Chaebol networks, it could increase their exposure to tax avoidance, and subsequently lead to adoption of this behaviour through network effects.

4.1.2 Inventors in tax haven vs inventors in the labour force

Using the general population in the denominator of IP and comparing it to individuals in tax havens, the denominator in IT, can be problematic. The general population will have a different age-structure than individuals who are owning shell companies in tax havens. In addition, it can be problematic when comparing different countries, as different age structures and working-age populations affect the comparability of the numbers. As a test of robustness, we compared the ratio of inventors in tax haven (IT) with the ratio of inventors as share of labour force (IL) for each country shown in Figure 5. The age-structure and composition of the people in the labour force might be more comparable to the shell company owners in tax havens. The countries are ordered from lowest difference between IT and IL, to the highest.

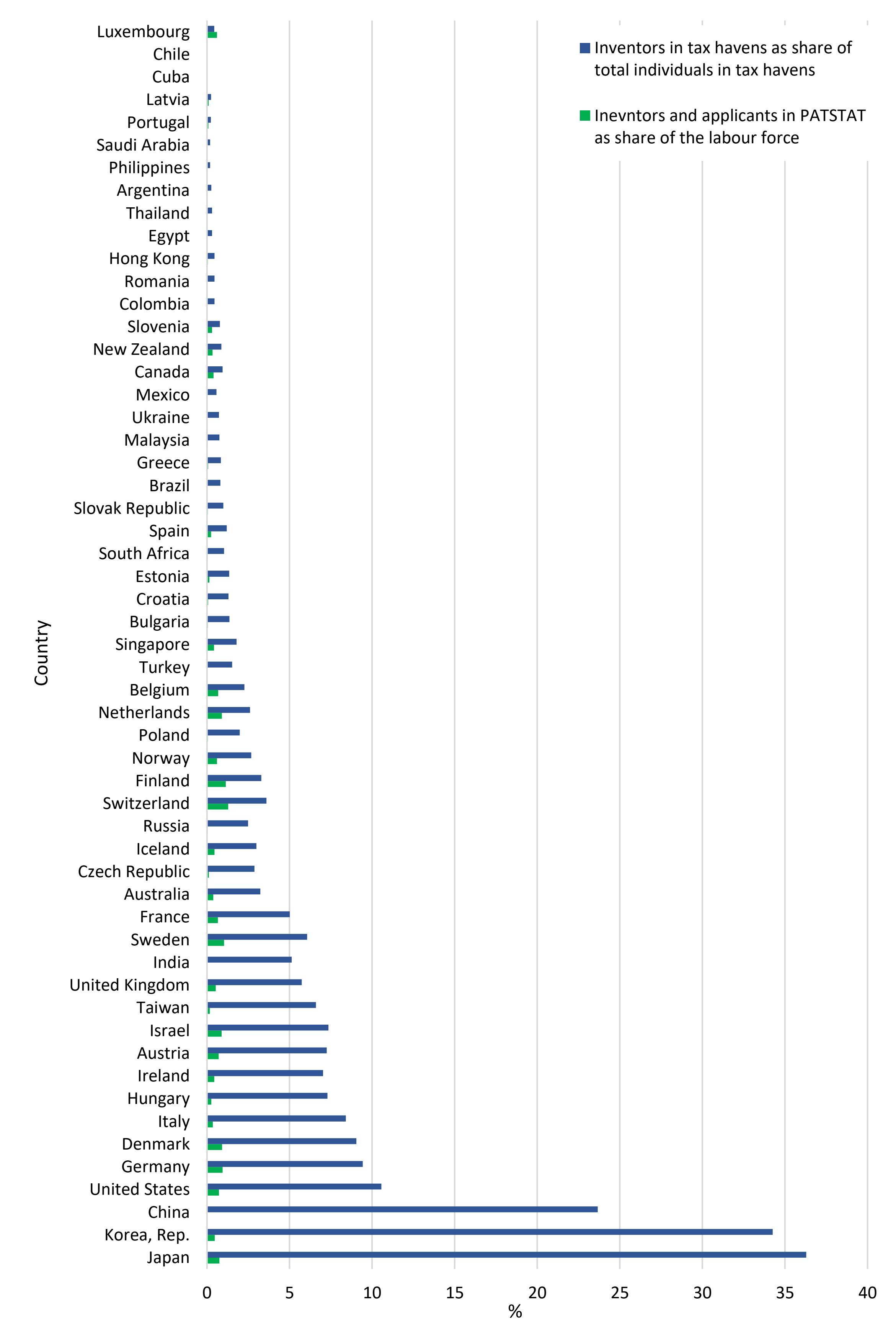


Figure 5. IT: Inventors in tax haven as share of all individuals in tax haven, and IL: inventors and applicants in PATSTAT as share of the labour force.

Figure 5 shows that for all countries with at least one inventor in tax haven, the ratio of IT is higher than IL, except for Luxembourg. This strengthens our results that inventors are overrepresented amongst individuals in tax havens. In Luxembourg, the ratio of inventors in tax haven are lower than the ratio of inventors in the labour force. However, Luxembourg only had three identified inventors in the Offshore Database, meaning this result need to be interpreted with caution.

4.1.3 Inventors in tax havens as share of all inventors

Figure 6 below shows inventors and applicants that engage in tax havens as a ratio of all inventors and patent applicants in each country (II). This is compared to all individuals that are engaged in tax havens as a share of the population of the country (TP).

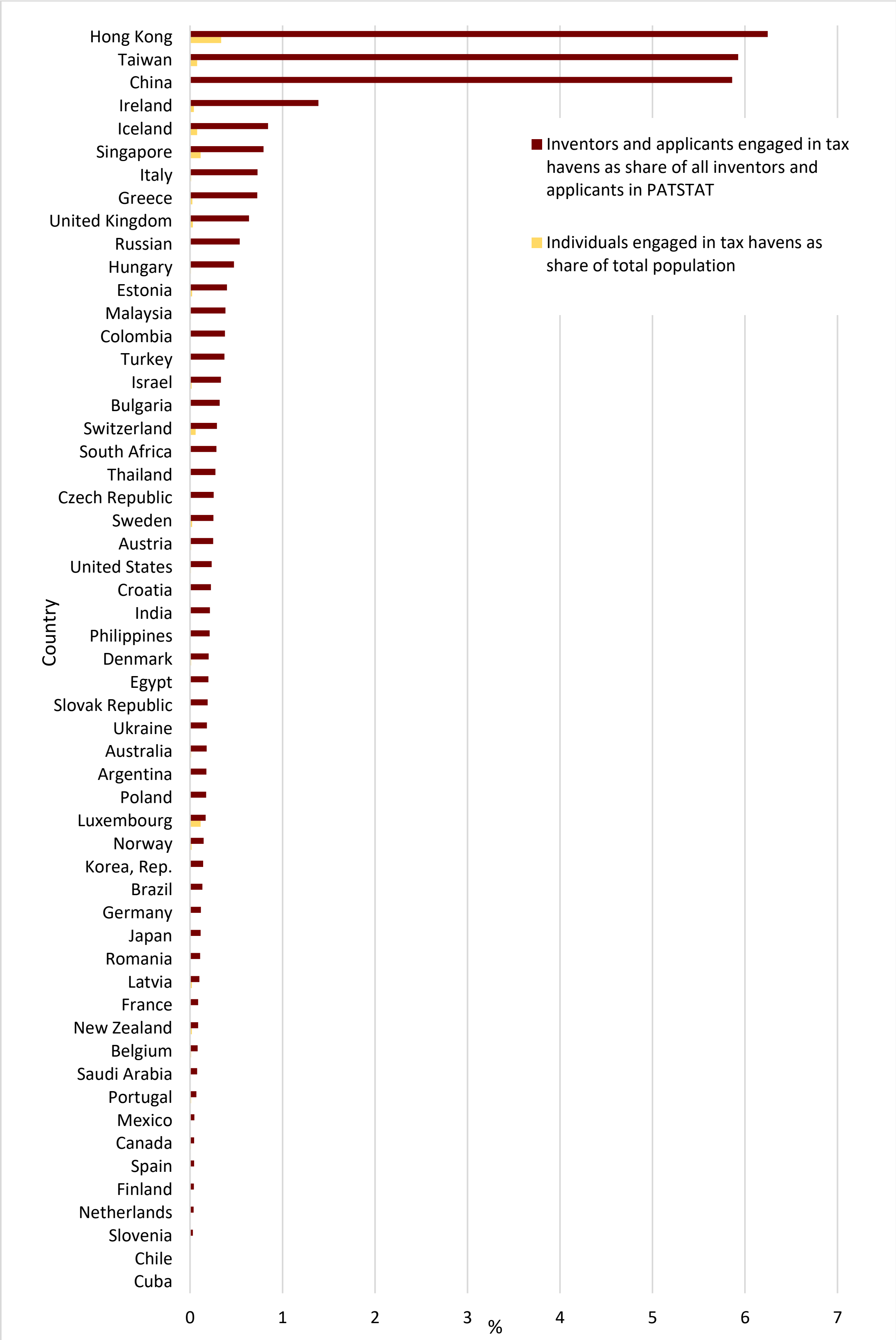


Figure 6. II:Inventors and applicants in tax haven as share of all inventors and applicants in PATSTAT and TP: Individuals engaged in tax haven as share of the general population

Hong Kong, Taiwan and China have the highest ratio of inventors engaged in tax haven as a share of all the inventors in the country with 6.25 %, 5.93% and 5.8% respectively. Hong Kong and Taiwan are also amongst the top five countries with the highest percentage of individuals in tax haven generally, together with Singapore, Luxembourg and Iceland. The difference in II between Hong Kong, Taiwan and China are less than one percentage point each, but the difference between China in third place and Ireland in fourth is 4.47 percentage points. After this, the difference in II between countries is gradual. This shows that the top three countries stand out.

In 2014, Hong Kong and Taiwan were no.1 and no.17 on the Index of Economic Freedom, which scores a country's commitment to rule of law, government interference, regulatory efficiency, and openness in markets (Samans et al., 2017). China was placed as no.137 (Samans et al., 2017). Hong Kong was until march 2019 considered a tax haven in its own right by the EU (Remeur, 2018; van der Does de Willebois, 2011), which makes the finding of inventors with offshore assets somewhat confusing. Perhaps the bar for offshore evasion is lower than other places because of the local financial industry already facilitating such asset concealment.

A potential explanation behind the stand-out rates for China, Hong Kong and Taiwan may be the lack of institutional trust or networking culture as discussed in the previous case study. Officials in pervasive networks of loyalty have the power to exert favourable treatment to friends through official bodies (Alston, 1989; Wang, 2014; Xin & Pearce, 1996). A perception that this can be expected from Chinese institutions may discourage inventors from allocating assets in domestic banks. When Hong Kong was handed over to mainland China in 1997, demand for asset protection among English speakers spiked (van der Does de Willebois, 2011). Similarly, inventors offshoring assets in Hong Kong may exhibit low confidence in the sovereignty of Hong Kong banks with respect to institutional influence from China. Taiwanese inventors may similarly exhibit low levels of trust in the sovereignty of financial institutions. In contrast to the abovementioned regional factors, general explanations of wealth or networking effects may still be better explanations.

The findings of Hong Kong, Taiwan and China having more inventors in tax havens than other countries, can also be due to their high name ambiguation. As mentioned, a lot of Asian names only consist of two short names, and we even accepted matches where these names were in different order. In section 3.5 we mentioned that individuals with shorter name lengths, considered as common names, might overestimate our result of the number of inventors with investments in tax haven jurisdiction. We found that the average name length of all inventors in tax haven were just over 2 names (2.15). However, the average name length of our matches for Hong Kong were 2.92 (113 observations). The average name length of matches from Taiwan were 2.01 (1,205 observations), but 91% of the names were hyphenated (e.g. kuang-cheng chao), strengthening the validity of the matches. In comparison, Japan had an average of 2.01 names (576 observations) and Ireland 2.12 names (143 observations), neither of them had any hyphenated names. In addition, the findings of numerous inventor collaborations from China support the validity of Chinese matches. This, as well as the belief that PATSTAT is biased towards Europe and might underestimate non-European results, supports the validity of our findings; of all the inventors in PATSTAT, China, Hong Kong and Taiwan have the highest share of inventors that engage in tax havens.

This section illustrated that inventors are over-represented for every region, but also that having one inventor in a tax haven was enough to be over-represented. We discussed possible explanations of differences in over-representation, such as wealth of the inventors, degree of innovative activity in the country, cultural factors and observed that countries with high income taxes were both high and low along the degree of over-representation. We discussed how firms may have incentives to offer tax haven investments to retain inventors and avoid knowledge spillovers. Shell companies may be used for legal purposes (Malan et al., 2017; van der Does de Willebois, 2011), but they rarely are. The majority of shell companies are used to conceal assets or obtain illicit advantages (Alstadsæter et al., 2019a, 2019b; David, 2016; Farrell, 2016; Johannesen & Zucman, 2014; O'Donovan et al., 2019; van der Does de Willebois, 2011). The implication of this is that we can assume inventors with tax haven investments to be wealthy, which should be related to their success. We will examine this in the following.

4.2 Inventor Quality and Technological field

To get an understanding of which type of inventor seek tax haven engagement, we examine the patent quality and technological field of the population of inventors that engage in tax havens and compared them to the average inventor. We will first look at the difference in inventor quality, then the difference in ratios within technological fields. We will also compare these measures between the top five countries with the highest ratio of inventors in tax havens; Japan, South-Korea, China, United States and Germany.

4.2.1 Inventor quality

By looking at the difference in number of patents and forward citations, we examine if inventors in tax havens are more successful than other inventors or vice versa. Patent citations as an approximation of patent quality and economic value was proven robust by Hall et al. (2005) and has later been used as a proxy by others (Akcigit et al., 2016; Bernstein, McQuade & Townsend, 2017; Kogan, Papanikolaou, Seru & Stoffman, 2017; Squicciarini et al., 2013). Citations are given to a patent A, when a future patent B, cites technology described in patent A. This means that patent B cannot take credit for the contribution of A, thus providing A with the opportunity to set a price on the authorized use of patent A. More citations indicate a broader technical applicability and commercial value of the technology, which reflects the scientific competence or ingenious entrepreneurship from the development process of the patent. Akcigit et al. (2016) use patent citations and patent quantity of inventors to benchmark their expected earnings.

The number of citations inventors can receive is a function of both how good or broad their patents are, but also how many patents they have. Lanjouw and Schankerman (2004) found that patent quality was inversely related to research productivity at the firm level, which suggests that patent quantity might come at the expense of patent quality. While patent citations may be a robust proxy for its' economic potential, the number of competing claims for the same patent, number of patent renewals and global filing for the same patents also indicate economic value. For instance, Griffith et al. (2014) identified high quality patent applications as those that were filed at patenting offices in the U.S., Europe (EPO) and Japan.

We first look at the average number of patents per inventors in PATSTAT in general compared to the average number of patents per inventor engaged in tax haven. We also look at the difference between these two measures for Japan, South Korea, China, United States and Germany shown in Figure 7.

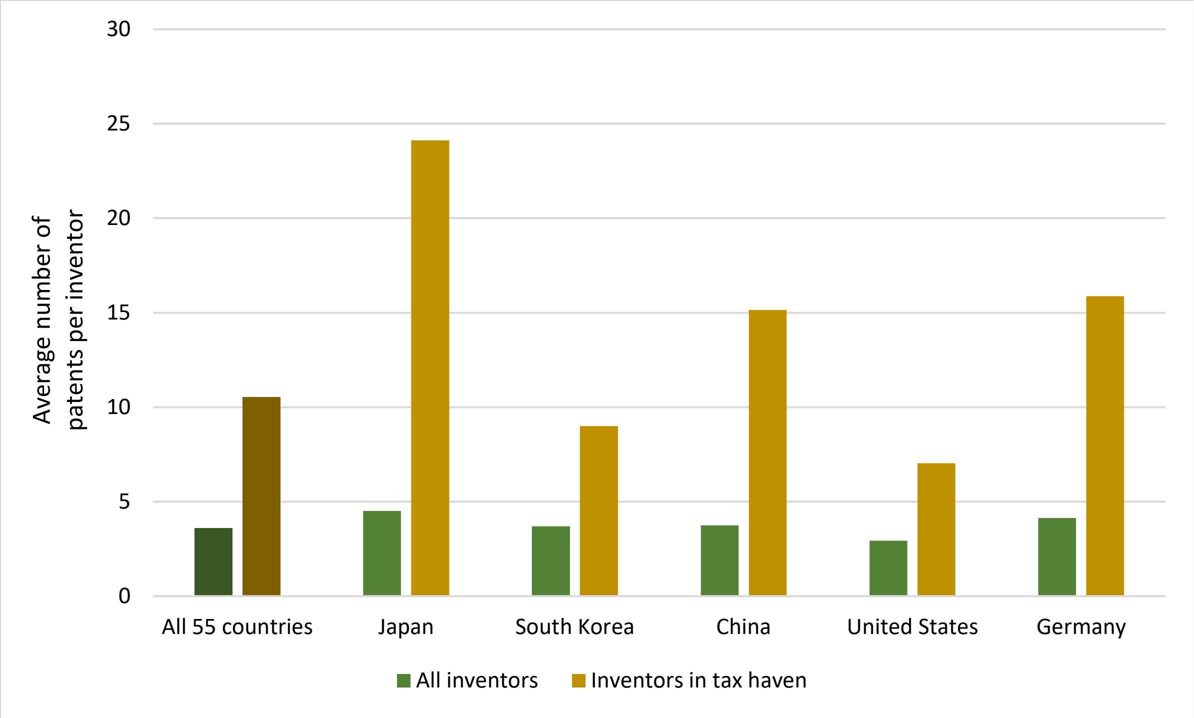


Figure 7 Average number of patents per inventor for all inventors in PATSTAT and all inventors in tax haven, and for Japan, South Korea, China, United States and Germany

Figure 7 show a higher average number of patents for inventors with tax haven investments. This shows that inventors with tax haven investments are more productive than average inventors in their region. An independent t-test shows that tax haven inventors (M = 10.54, SD = 29.95) have significantly more patents than the average inventor (M = 3.20, SD = 7.36) at the 0.01% level with t(15732) = -1.2e+2, $p < 0.001$. The difference is largest for Japan, China, and Germany, while more modest for the U.S and South Korea. To see whether these quantities could be affected by reduced quality, we look at average forward citations between all inventors and inventors in tax haven.

The number of patents is calculated by how many different patent applications one inventor is a part of, meaning how many times one inventor appears with a unique ID in the PATSTAT database. One unique inventor is identified by the same name and country, after the names have been cleaned and standardised as described in Section 3.3. This means that one inventor can be counted more than once if the name of the inventor is entered in the PATSTAT database differently each time (e.g. both John G. Smith and John George Smith). The average number of patents per inventor will be underestimated. It also means that two inventors with the same name from the same country can be merged together as one, and the average number will be overestimated.

We gather the number of forward citations from the OECD Patent Quality Indicators database. The OECD database is linked to the PATSTAT database through the variable APPLN_ID. This variable, however, is not stable in patent applications before 2011, meaning a lot of patent applications from before 2011 are missing this variable and cannot be linked to the OECD database (EPO, 2017). Therefore, we do not look at patent quality for the complete set of inventors in PATSTAT but assume randomness of missing APPLN_ID. Patent quality information for all 55 countries are found in the database.

We use the variable with the number of patent citations received up to 7 years after publication (FWD_CITS7) and find how many citations each inventor has for all their patents. Citation counts are inherently truncated (Hall et al., 2005), since patents continue to receive citations over long periods, but we receive only citations given up the last year of the available data in January 2020. We only count citation for patents that are registered in the PATSTAT Spring Edition 2017.

We look at the average number of citations per inventor for all inventors and the average number of citations per inventor for inventors found in tax havens, and for Japan, South Korea, China, United States and Germany in Figure 8.

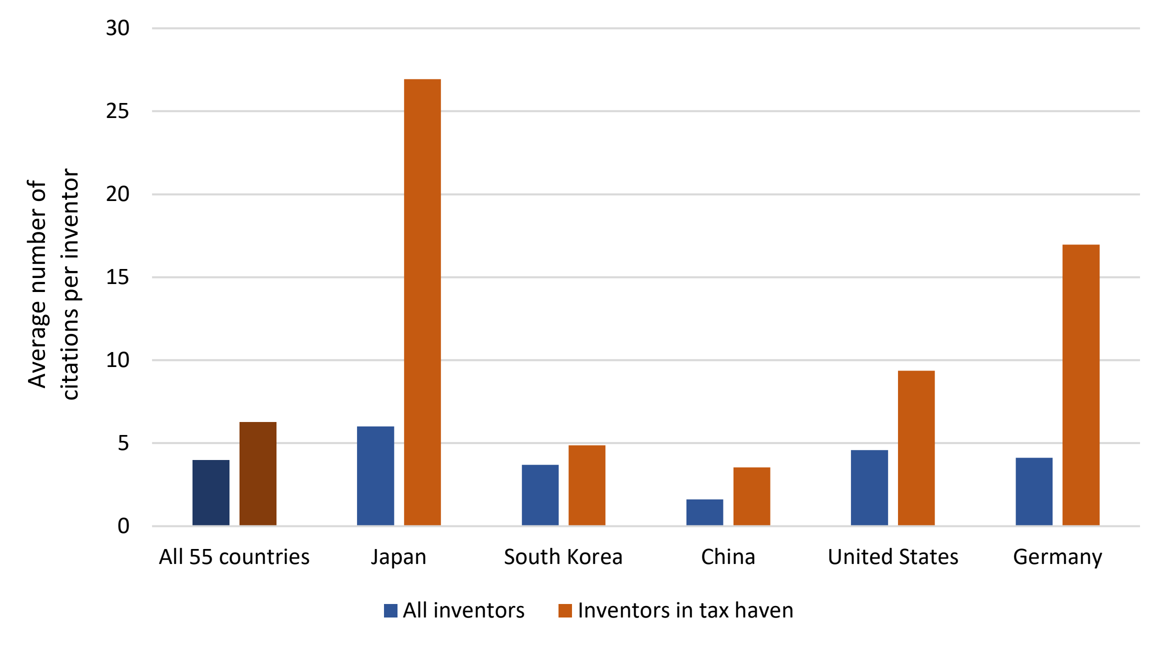


Figure 8. Average number of citations per inventor in PATSTAT and inventors in tax haven, and for Japan, South Korea, China, United States and Germany.

Note: Not all inventors contain citation information. The group "inventors in tax haven" consist of 11,363 out of 15,896 inventors from PATSTAT that are found in the Offshore Leaks. "All inventors" consists of 2,701,714 out if 3,656,208 cleaned names of individuals in PATSTAT with APPLN_ID, inventors from the 55 countries.

Figure 8 shows the average number of citations per inventor for all inventors and the average number of citations per inventor for inventors found in tax haven, as well each country. A t-test comparing the means shows that inventors in tax havens (M = 6.26, SD = 24.21) have significantly more forward citations than all inventors in PATSTAT (M = 3.98, SD 23.66), with t(2.7e+6) = -10.2, p < 0.001. The underlying distributions are however not normal but strongly skewed towards fewer citations. Test results are thus suggestive and imperfect. Figure 8 shows that in all five countries, the average number of citations are higher for inventors in tax haven than the average inventor, but the difference is highest for Japan, Germany and United States.

Both the quantity and citations of patents to an inventor are functions of time, which makes the inventor's age a variable of unknown effect on the results. A patent needs time to grow its impact, and research and development may also take years before it to materializes into tangible technologies. We therefore look at citations per patent for each inventor in tax haven and all inventors in PATSTAT in Figure 9.

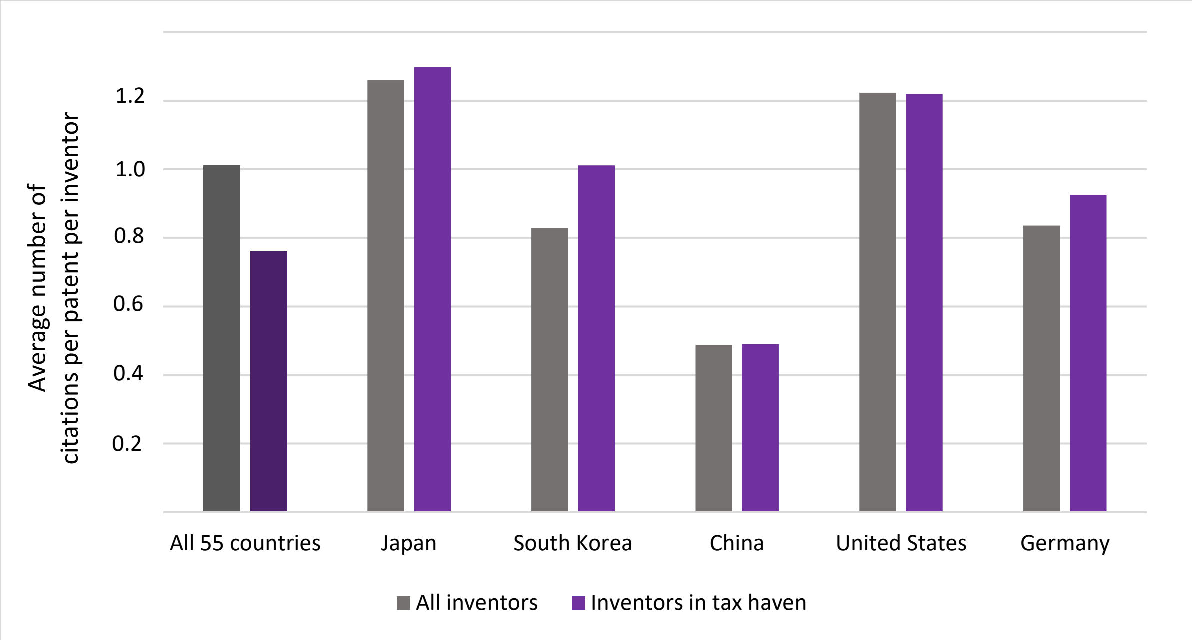


Figure 9. Average number of citations per patent for all inventors in PATSTAT and inventors in tax haven, and for Japan, South Korea, China, United States and Germany.

Figure 9 show that for the five countries, the average amount of citations per patent is either the same for tax haven inventors and all inventors, or only slightly higher. The average number of citations per patent for each inventor in tax haven are lower than average. An independent t-test for the all 55 countries, gives tax haven inventors on average ($M = 0.76$, $SD = 1.45$) significantly fewer citations than non-haven inventors ($M = 1.01$, $SD = 2.76$) with $t(2.3e+6) = 11.5$, $p < 0.001$. Thus forward citations of inventors with tax havens investments are influenced by quantity, and not quality. Lanjouw and Schankerman (2004) found that patent quality was inversely related to inventor productivity, which may support our findings. This higher productivity that tax haven inventors have, might be a source of wealth that have led them to tax haven engagement. As a reference to prior discussions of employer incentives (see section 2.2.3), if firms were involved in these inventors' engagement in tax havens, it could be because they are more instrumental in patent harvesting, than they are at making useful patents with technological influence on other inventions (Torrisi et al., 2016). Not knowing the ages between the samples also makes it challenging to assess actual productivity. A difference in average ages between inventors in tax haven and all inventors in PATSTAT may also have inflated the average patents held in each sample. However, citations are only counted 7 years after the grant, which reduces age bias in the way that it does not prevent non-haven inventors from having higher quality patents.

4.2.2 Technological field

We wish to examine if certain technological field(s) are overrepresented in the group of inventors and applicants engaged in tax haven activity. The OECD Patent Quality Indicators database groups patents by technology fields (TECH_FIELD). Technology fields are defined according to Schmoch (2008) classification (as updated in 2010 and 2011) which relies on the International Patent Classification (IPC) codes contained in the patent documents. This taxonomy features six main technology sectors, subdivided into 35 fields of balanced size, structured so as to maximise within-sector homogeneity and across-sector differences (Squicciarini et al., 2013). The OECD database merged two of Schmoch`s fields, leaving 5 main fields which we used: (1) electrical engineering (telecommunications, computer technology etc), (2) instruments (optics, medical technology etc) (3) chemistry (biotechnology, food chemistry etc), (4) mechanical engineering (machine tools, special machines etc) and (5) other fields (furniture, games, other consumer goods, civil engineering) (Squicciarini et al., 2013).

Figure 10 shows the total distribution of technological fields amongst the inventors engaged in tax haven activity on the left side, and the distribution of technological fields amongst all inventors on the right. As with patent quality, the variable APPLN_ID is used, meaning only the inventors with the available information are represented in both groups. We look at the categorization of one patent for each inventor, assuming that the inventor remain within the same field of research.

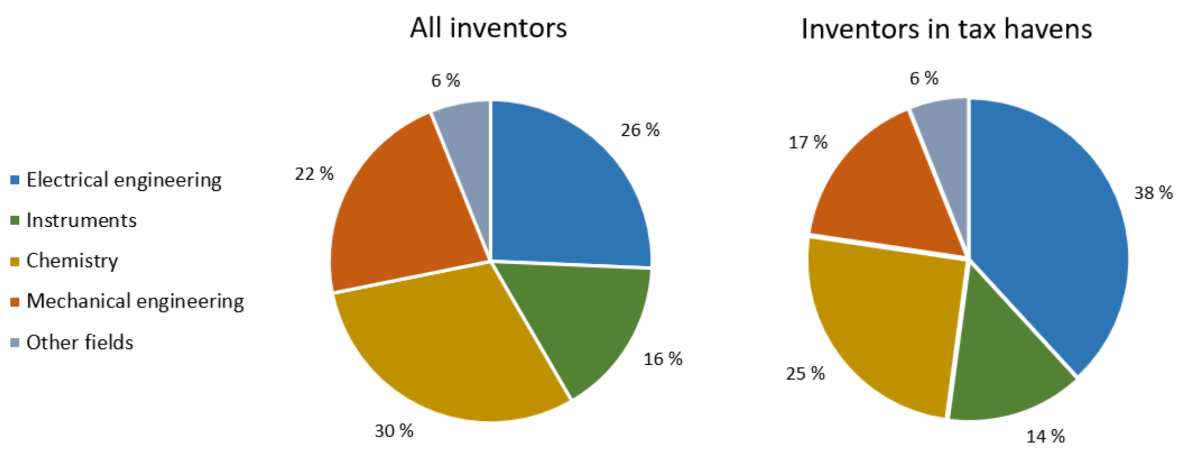


Figure 10. The share of technological field of inventors in tax havens and all inventors in PATSTAT

Note: Not all invenotrrs contain technology field information. The group "inventors in tax haven" consist of 11,358 out of 15,896 inventors from PATSTAT that are found in the Offshore Leaks and has APPLN_ID. "All inventors" consists of 2,712,450 out if 3,656,208 cleaned names of individuals in PATSTAT with APPLN_ID, inventors from the 55 countries.

In general, chemistry and electrical engineering are the largest technology fields within patenting in PATSTAT. From the pie charts in Figure 10 it seems that more inventors within the field of electrical engineering, and fewer inventors from mechanical engineering and chemistry, are engaged in tax havens. We perform logit regression[11] to test for significance in differences between proportions of tech fields, and find chi2 < 0.01 for each tech field comparison, except between instruments (2) – chemistry (3) and instruments (2) - other fields (5). We interpret this result as a significant difference in tech field proportions between tax haven inventors and all inventors. Electrical engineering represents the largest difference between the two groups.

Figure 11 shows the distribution of patents within technological field for Japan, Korea, China, United States and Germany, for both inventors engaged in tax havens and all inventors in each country.

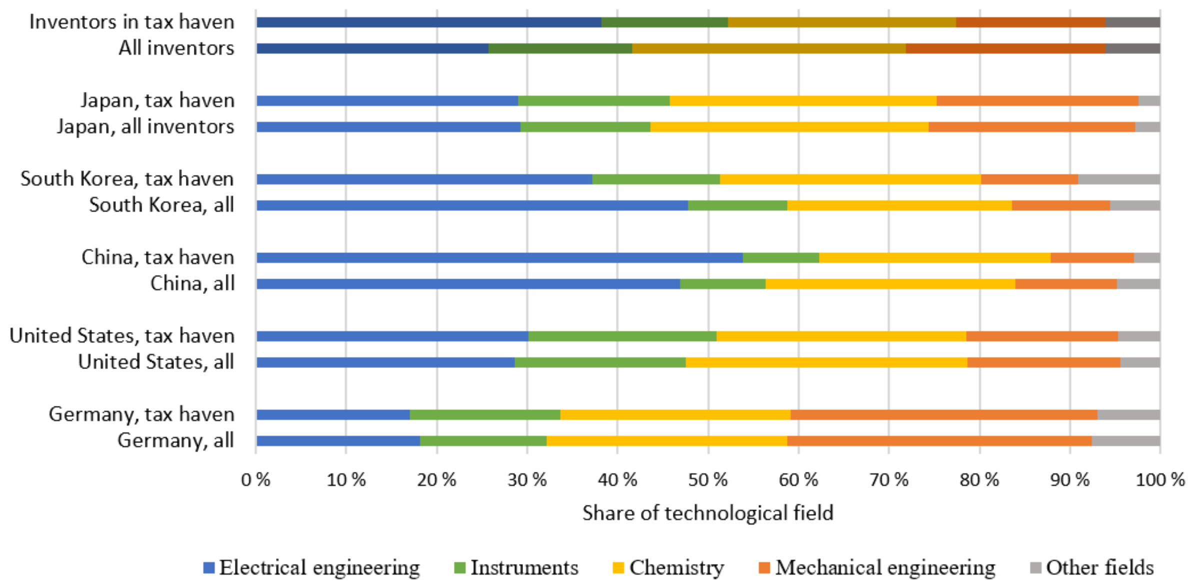


Figure 11. The average shares of technological field for inventors in tax haven and all inventors in PATSTAT for Japan, South Korea, China, United States and Germany

[11] For the logit regression we create indicator variables corresponding to each tech field from 1 - 5. (1) electrical engineering , (2) instruments (3) chemistry, (4) mechanical engineering and (5) other fields. We use these as independent variables in the logit regression and a tax haven dummy as the dependent variable. The regression tests for differences in the coefficients of the indicator variables, and if they are significantly different from each other.

Figure 11 shows that the distribution of patents is similar for each country, for both tax haven inventors and all inventors, except for China and South Korea. For China, seemingly more inventors from electrical engineering are engaged in tax haven, than the average and more inventors from chemistry are found in tax haven than average from South Korea.

We find that tax-haven inventors are more productive, as shown by the higher number of patents and citations per inventor. The productivity that separates them from the average inventor may be the source of wealth that motivated the acquisition of a shell company in the first place. We see that technology field, in particular electrical engineering, may also increase chances of inventors engaging in tax haven. However, if we interpret this as electrical engineering being where the most inventors can acquire the funds and networks to engage in tax havens, it is fair to consider that such overweight of a technology can be temporary. Technological fields sometimes have expansionary cycles, like electrical engineering in the era of digitalization and telecommunications, competition and product standards within each technology eventually decides how exclusively anyone may capture value from an invention over time (Jacobides et al., 2006; Pisano & Teece, 2007). Classical economic theory states that profitable markets will be targeted by new entrants, which would predict any technological field to be of similar profitability in the long run. Technical industries harvesting patents to create barriers of entry are however associated with more monopolistic or oligopolistic competition, which suggest long-term profitability above marginal cost. This could sustain differences in the attractiveness of industries to recruit inventors.

4.3 Stakeholders and patent owners

In the previous subsection we find that the average inventor engaging in tax havens is more productive and more connected to electrical engineering than the average inventor. We further want to look at whether the inventors in tax havens are employed or connected to companies, public institution or if they are independent. The sum of theoretical considerations under section 2.3 supports the hypothesis that employers behind inventors in tax haven are predominantly multinational companies or other institutions that facilitate research productivity.

We examine whether the inventors are employed by a corporation or public institution by looking at the patent applicant of each inventor. The inventors are usually employees of the patent applicant (OECD, 2009). Each patent application has its own patent id (ID), and each applicant and inventor within that application are assign this id. We connect inventors and their applicant together using ID. We use the same code as when separating individuals in step seven in the cleaning phase in Section 3.3.3, identifying if the applicants as either a corporation, public institution or an individual. We then calculated the share of corporate, public, individual, or any combination between the three, as share of all patents per inventor. Table 8 shows the mean share of each category per inventor, for inventors in tax havens and all inventors.

Table 8. Share of applicant classification for all inventors in PATSTAT and inventors in tax havens

Applicant classification	*All inventors (%)*	*Inventors in tax haven (%)*
Corporate	79.50	78.87
Public	4.26	5.14
Individual	11.22	11.37
Corporate and public	1.45	2.07
Corporate and indvividual	2.97	2.19
Public and individual	0.43	0.27
Corporate, public and individual	0.16	0.09

From Table 8 we find that the distribution of inventors employed by corporations are very similar between inventors engaging in tax havens and the average inventor. We conduct a t-test to compare the share of applicant types for all inventors and tax haven inventors and find significant differences, albeit very small. For the average inventor, the share of corporate patents ($M = 0.0795$, $SD = 0.39$) is higher than for inventors in tax havens ($M = 0.7887$, $SD = 0.34$), with $t(3.7e+6) = -2.045$, $p = 0.041$. For public institutions, tax haven inventors ($M = 0.0514$, $SD = .18$) have a higher share than all inventors ($M = 0.4263$, $SD = 0.19$), with $t(3.7e+6) = -5.71$, $p < 0.001$. The difference is insignificant for individual applicants, between the average inventor ($M = 0.1122$, $SD = .30$) and tax haven inventors ($M = 0.1137$, $SD = 0.27$), with $t(3.7e+6) = -0.614$, $p = 0.53$. The most significant difference was that of public institutions, which is interesting. If public institutions develop patents on public funding and inventors reap the benefits of this through tax havens, some of the desired externalities from public innovation could be lost through tax evasion.

Figure 12 shows the difference in ratios between Japan, South-Korea, China, United States and Germany.

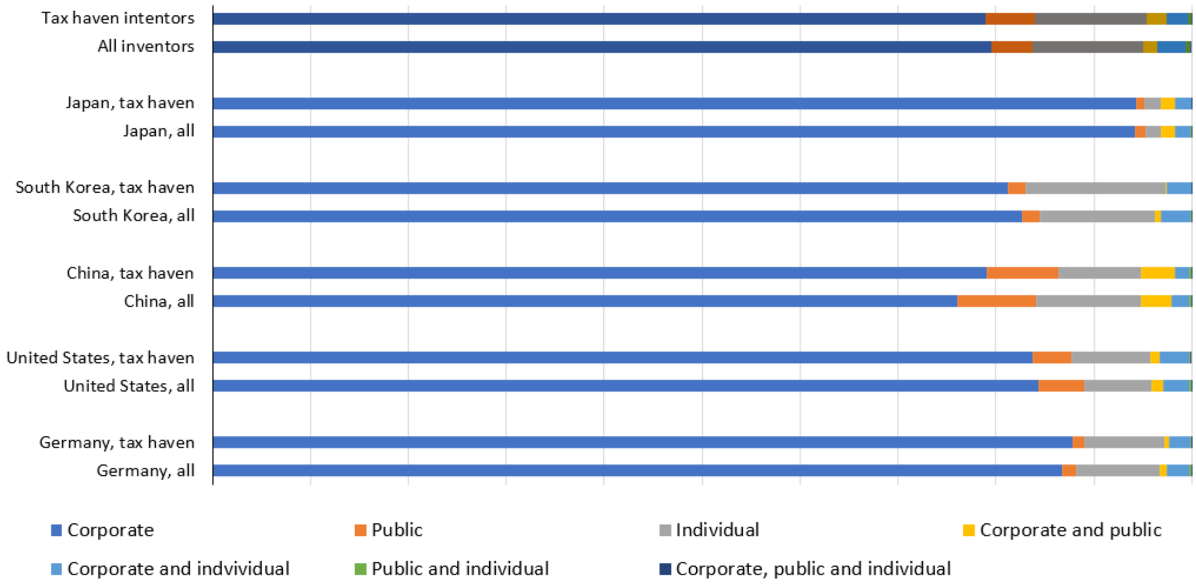


Figure 12. Share of applicant classification for all inventors in PATSTAT and inventors in inventors in tax haven, for Japan, South Korea, China, United States and Germany

The ratios in Figure 12 show that there are differences both between the countries and between the wo groups in each country. Japan has an overall higher share of corporate patents per inventors, but very little difference between the average inventor and inventor in tax haven. China and Germany have higher shares of corporate patents for inventors in tax havens compared to average, whilst South Korea has a lower share.

4.4 Inventor collaborations

We wish to examine the potential network effects among inventors. We find that several of the inventors in the Offshore Database have the same id, meaning they are registered on the same patent application in PATSTAT. Of the 15,896 inventors we found in the Offshore Leaks, 40.4% (6,360) of them had the same id as another inventor also found in the Offshore Leaks. We interpret this as inventors in the same patent application collaborating with either each other or their employer to engage in tax havens. We therefore call these “inventor collaborations”.

We find 23,432 inventor collaborations, with two to 13 inventors in each collaboration. One inventor might collaborate with several others, or the same inventor collaboration could have several patent applicants together, which explains the large number of collaborations compared to individual inventors. Table 9 shows how many inventor collaborations we find in the Offshore Leaks, as well as the number of inventors in each collaboration.

Table 9. Number of inventor collaborations and the number of inventors in each collaboration

Number of inventors in each collaboration (from same patent application)	*Number of collaborations*
2	17,350
3	4,322
4	1,209
5	385
6	102
7	32
8	17
9	2
10	2
11	2
12	2
13	1
Total	23 432

.

Of the 6,360 inventors who are part of an inventor collaboration, 83% of them are Chinese. All inventor collaborations with more than six inventors are Chinese, which strengthens the validity of the Chinese name matches. In comparison, 4% of all inventors in PATSTAT that are part of a patent application with more than six inventors are Chinese and 44 % of all inventors we find in the Offshore Leaks are Chinese. As mentioned in Section 3.5, we find several inventor collaborations where the inventors in PATSTAT were matched with names in the Offshore Leaks, which strengthens the validation of the name matches (such as two inventors in an inventor collaboration being in the same name field in the Offshore Leaks data or several inventors in one patent application having "employees remuneration trust company as nominee for [name]" in the Offshore Leaks name).

We discussed in Section 2.2.2 how avoidance behaviour may diffuse within and between networks, and we proposed that this effect on average would cause general overrepresentation of inventors, because tax avoidance is associated with multinationals. According to the sources we find, most inventors work for large firms or multinationals (Akcigit et al., 2016; Giuri et al., 2007). The finding of inventor collaborations makes a persuasive case for network effects, whether they are initiated by management or inventors.

To assess whether inventor collaborations are corporate affair, we examine the share of corporate, public, and individual patents for each inventor in an inventor collaboration. In Table 10, we compare the distribution of applicant types between collaborators, all inventors in PATSTAT and all inventors found in tax haven (data from **Feil! Fant ikke referansekilden.**). We use the same approach as in Section 4.3.

Table 10 Share of applicant classification for all inventors in PATSTAT, inventors in tax haven and inventors in inventor collaborations

Applicant classification	*All inventors (%)*	*Inventors in tax haven (%)*	*Inventors in inventor collaboration (%)*
Corporate	79.50	78.89	81.23
Public	4.26	5.14	7.16
Individual	11.22	11.37	6.22
Corporate and public	1.45	2.07	3.53
Corporate and indvividual	2.97	2.19	1.48
Public and individual	0.43	0.27	0.29
Corporate, public and individual	0.16	0.09	0.09

We perform two-tailed t-tests to compare the shares of corporate, public and individual applicant type for tax haven inventors and collaborators, and find significant differences at $p < 0.001$ for each applicant type using 95% confidence levels. We find the same significance for the difference in share between all inventors in PATSTAT and collaborators. In summation, inventors in inventor collaborations have a higher share of corporate and public patents and a lower share of individual patents than both the average inventor in PATSTAT and the average inventor in tax haven.

As most collaborations are Chinese, it is important to note that China also has the most shell companies out of all countries in the Offshore Database. The factors causing collaborations to conceal assets may thus not be inventor specific, but rather China specific. We discussed how low of faith in legal institutions may prompt employers or workers to protect assets abroad (Keupp et al., 2009; Wang, 2014). The sociocultural phenomenon Guanxi is found to distort official Chinese institutions to the point that outsiders invest in the exchange of favours with public officials to avoid extortion or asset appropriation (Wang, 2014; Xin & Pearce, 1996). Inventors with stakes in intellectual property may be particularly incentivized to not expose assets domestically, that may be extorted to make them share secrets of commercial or political value. We also theorized that firms, independent from Guanxi, may approach inventors with offshore investment opportunities to prevent them for changing jobs and leaving with irreplaceable competence. The shares presented in Table 10 may be affected by the high share of Chinese inventors in the inventor collaboration population, which is why we compare them to all Chinese inventors in PATSTAT and all Chinese inventors in tax haven in Figure 13.

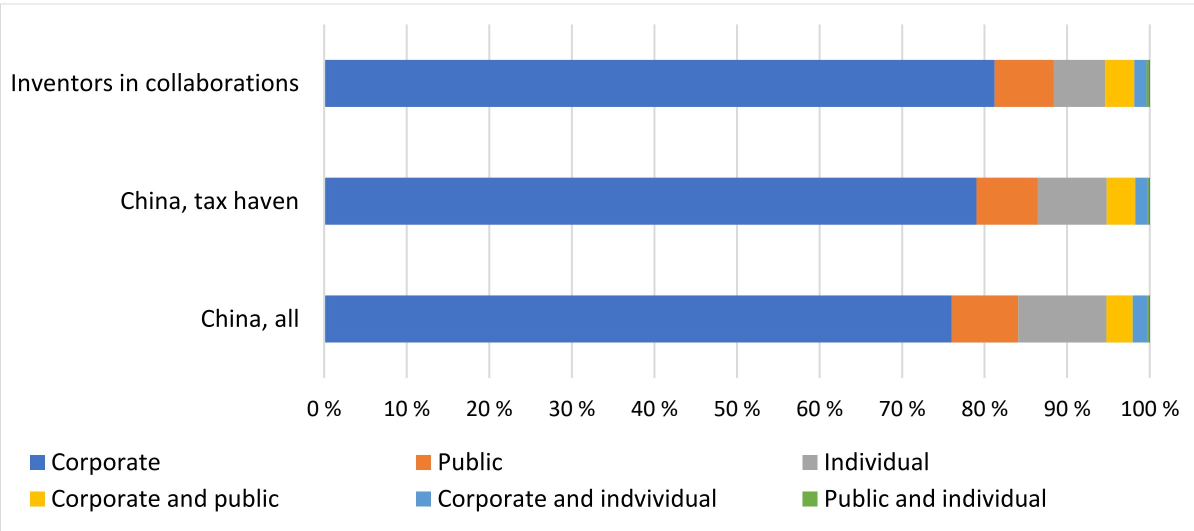


Figure 13. Share of applicant classification for all Chinese inventors in PATSTAT, Chinese inventors in tax haven, and inventors in inventor collaboration

The result in Table 10 coincides with Figure 13; there is a lower share of individual patents for inventors in collaborations than for inventors in PATSTAT. Also, inventors in tax havens have higher shares of corporate and public patents. This is could be related to earnings potential and patent quality, as well as the resources to commercialize the patent. Individuals are unlikely to have equal access to the level of facilities and knowledge networks that inventors at universities or corporate R&D departments do.

In general, the finding of inventor collaborations could be consistent with several rationales illuminated in the framework. If inventors working on the same patent for the same employer engages in tax havens, it indicates that this patent is important for their own earnings or their employer. The co-ordination between colleagues or employers would itself entail a network effect (see section 2.2.2). Underlying this effect could be motives both related to both wealth (Alstadsæter et al., 2019a), retention of strategically important inventors (see section 2.2.3) or generally low compliance triggered by low trust in government (see section 2.2.1). Low trust in institutional property rights is recognized as a legitimate concern in the case of China (Keupp et al., 2009; Wang, 2014). What is also interesting is the higher share of public patent applicants behind inventor collaborations, suggesting that either engaging in tax havens is not necessarily related to employment or connections in a multinational. This result is also impacted by China, which have several universities with inventor collaborations. In Chinese universities, state financed research may easily be appropriated by the state, which could make academic inventors more inclined to team up with private companies (Guo, 2007). Hence, university inventors may also be impacted by corporate ties.

4.5 Limitations of the study

Tax haven data is secretive in nature, and leaks such as the Offshore Leaks is the only source of information available to the public. Although Omartian (2017) suggests that where and by whom the shell company was set up does not matter for its' purpose, and we assume that the Offshore Leaks are a random sample of individuals who owns shell companies in tax havens, we do not have the full data or information on individuals engaged in tax haven activity. Within the data in the Offshore Leaks, there is another layer of secrecy by individuals owning "bearer shares" and therefore concealing their identity. We can therefore only make assumption about distributions based on our sample, but without knowledge of all tax haven individuals.

In addition, both PATSTAT and the Offshore Leaks data are lacking information such as addresses or even country of residence for many of the individuals, as well as the name field being very noisy. This makes it difficult to know for certain whether two people with the same name from the same country, are in fact the same person. Both in relation to removing duplicates within one database, but also when matching two names from the two different databases. Both data bases contain raw and unprocessed data and needed to be substantially cleaned to be usable for our purpose. Although both databases have undergone thorough cleaning procedures, there might still be spelling errors and other noise left in the data due to the size of the data and the limited time frame of this thesis. Fuzzy matching is intended to overcome these kinds of problems, but manual filtering of the results is also prone to human error. We relied on our own judgment whether a close name match could be the same person, as well as brief online search such as looking for them on ESPACENET. We also only had the capacity to filter name matches above a certain similarity score, leaving many matches unchecked.

In examining the extent to which inventors engage in tax haven activity, we compared the ratio of inventors in tax haven to the ratio of inventors in the population. One limitation to this is that the age and working structure of individuals who engage in tax haven are not the same as individuals owning shell companies. Correcting for working population was used as a robustness test but will not be a perfect comparison. The robustness test showed the same results of overrepresentation, except for Luxemburg. For some countries, we only had a very small sample of inventors to draw conclusions from. Our results showed that even one inventor in tax haven per country led to overrepresentation. Due to the limited information on the individuals, the presence of name ambiguity, as well as the limited data on tax haven individuals, the results from these countries should be interpreted with caution.

5. Conclusion

The purpose of this thesis was to examine to what extent inventors engage in tax havens. Inventors are important drivers of economic growth and technological development (Akcigit, Grigsby & Nicholas, 2017; Wong et al., 2005) and the patents they create can be used to shift income to tax havens (Griffith et al., 2014; Tørsløv et al., 2018). They represent a potential group of wealth owners that deserve a closer look to advance knowledge on tax avoidance behaviour. We examined to what extent they are engaging in tax havens and the characteristics of those who do. This will shed light on how widespread the occurrence of inventors in tax haven is and what type of inventors seek tax haven engagement.

We collected names and country codes of inventors and patent applicants from the PATSTAT database. We then collected names and country codes of individuals that own shell companies in tax havens from data provided by the Offshore Leaks database. Our purpose was to identify individuals found in the Offshore Leaks who are also inventors or patent owners. We did that by cleaning and standardising these names, followed by a matching and filtering process. We were able to identify 15,896 inventors in the Offshore Leaks, which we used to make population distributions and compare them to the general population. We also examined their characteristics and compared them to the average inventor in the PATSTAT database.

We found that inventors are overrepresented amongst individuals engaged in tax havens. For each country with at least one inventor engaged in tax haven, the share of inventors in tax haven is higher than the share of inventors in the general population of the country. This is also true if we correct the general population to the labour force, except for Luxembourg. Japan, South-Korea, China, United States and Germany are the countries with the highest share of inventors amongst individuals in tax haven. Hong Kong, Taiwan and China are the countries with the highest share of inventors engaged in tax haven of all inventors in the country. We discussed how the higher overrepresentation in Asia might be caused by amplified network effects or low faith in institutions' ability to protect property rights.

We further examined the characteristics of inventors that engage in tax havens, and found that they are significantly more productive and are more affiliated with electrical engineering than average inventors. However, they do not receive more citations per patents than the average inventor, which could associate tax haven inventors with patent harvesting. We found several “inventor collaborations”: inventors registered on the same patent application all found in the Offshore Leaks. These inventors also have a higher share of corporate and public patents, and a lower share of individual patents, than the average inventor in PATSTAT and the average inventor in tax haven. We interpret this as network effects recently studied by Bohne and Nimczik (2018). Most inventor teams came from China, which we relate to historically weak enforcement of property rights.

Looking forward, discovering inventors that engage in tax havens might have implications for tax policy aiming to attract foreign inventors. Inventors already evading taxes are less likely to respond to tax incentives as some of their wealth already is untaxed. This thesis may inform policy makers of which characteristics are associated with inventors less likely to respond to such incentives.

5.1 Future research

We encourage future research to determine the causal relationships behind the relative and absolute extents of inventors using tax havens. One study could for instance link the level of wealth, institutional trust, ongoing patent developments and employment situation of the inventor to the date which he or she acquires the shell company. A similar study could examine if acquiring a shell company was related to mobility in specific region or corporation. We also propose a cross-regional study of how tax avoidance behaviour appear to diffuse differently within and between East Asian and Western organization, as a continuation of the literature on network spillovers and learning dynamics in the context of tax behaviour (Bohne & Nimczik, 2018).

References

Acs, Z. J., Braunerhjelm, P., Audretsch, D. B. & Carlsson, B. (2009). The knowledge spillover theory of entrepreneurship. *Small business economics, 32*(1), 15-30.

Agarwal, R., Audretsch, D. & Sarkar, M. (2010). Knowledge spillovers and strategic entrepreneurship. *Strategic Entrepreneurship Journal, 4*(4), 271-283.

Agarwal, R., Audretsch, D. & Sarkar, M. B. (2007). The process of creative construction: knowledge spillovers, entrepreneurship, and economic growth. *Strategic Entrepreneurship Journal, 1*(3-4), 263-286. doi:10.1002/sej.36

Agarwal, R., Ganco, M. & Ziedonis, R. H. (2009). Reputations for toughness in patent enforcement: Implications for knowledge spillovers via inventor mobility. *Strategic Management Journal, 30*(13), 1349-1374.

Akcigit, U., Baslandze, S. & Stantcheva, S. (2016). Taxation and the International Mobility of Inventors. *American Economic Review, 106*(10), 2930-2981. doi:10.1257/aer.20150237

Akcigit, U., Grigsby, J. & Nicholas, T. (2017). *The rise of american ingenuity: Innovation and inventors of the golden age* (0898-2937). Retrieved from https://www.nber.org/papers/w23047.pdf

Akcigit, U., Grigsby, J., Nicholas, T. & Stantcheva, S. (2018). *Taxation and Innovation in the 20th Century* (0898-2937). Retrieved from http://home.uchicago.edu/jgrigsby/files/Research/AGNS_Taxation_Innovation.pdf

Akcigit, U. & Stantcheva, S. (2020). *Taxation and Innovation: What Do We Know?* (0898-2937). Retrieved from https://scholar.harvard.edu/files/stantcheva/files/w27109.pdf

Alcácer, J. & Zhao, M. (2012). Local R&D Strategies and Multilocation Firms: The Role of Internal Linkages. *Management Science, 58*(4), 734-753. doi:10.1287/mnsc.1110.1451

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

Alstadsæter, A., Barrios, S., Nicodème, G., Skonieczna, A. M. & Vezzani, A. (2018). Patent boxes design, patents location, and local R&D. *Economic Policy, 33*(93), 131-177.

Alstadsæter, A., Johannesen, N. & Zucman, G. (2018). Who owns the wealth in tax havens? Macro evidence and implications for global inequality. *Journal of Public Economics, 162*, 89-100.

Alstadsæter, A., Johannesen, N. & Zucman, G. (2019a). Tax Evasion and Inequality. *American Economic Review, 109*(6), 2073-2103. doi:10.1257/aer.20172043

Alstadsæter, A., Johannesen, N. & Zucman, G. (2019b). Tax Evasion and Tax Avoidance. Retrieved from https://www.nielsjohannesen.net/wp-content/uploads/Avoidance_FullPaper.pdf

Alston, J. P. (1989). Wa, guanxi, and inhwa: Managerial principles in Japan, China, and Korea. *Business Horizons, 32*(2), 26-32.

Andersen, J. J., Johannesen, N., Lassen, D. D. & Paltseva, E. (2017). Petro rents, political institutions, and hidden wealth: Evidence from offshore bank accounts. *Journal of the European Economic Association*.

Audretsch, D. B. & Lehmann, E. E. (2005). Does the knowledge spillover theory of entrepreneurship hold for regions? *Research policy, 34*(8), 1191-1202.

Auron Technologies, L. (2009). How Many Of Me. http://howmanyofme.com/

Balsmeieri, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., . . . Fleming, L. (2018). Machine learning and natural language processing on the patent corpus: Data, tools, and new measures. *JOURNAL OF ECONOMICS & MANAGEMENT STRATEGY, 27(3)*. doi:https://doi.org/10.1111/jems.12259

Batrancea, L., Nichita, A., Olsen, J., Kogler, C., Kirchler, E., Hoelzl, E., . . . Zukauskas, S. (2019). Trust and power as determinants of tax compliance across 44 nations. *Journal of Economic Psychology, 74*, 102191. doi:https://doi.org/10.1016/j.joep.2019.102191

Bernstein, S., McQuade, T. & Townsend, R. R. (2017). *Do household wealth shocks affect productivity? Evidence from innovative workers during the Great Recession* (0898-2937). Retrieved from https://www.nber.org/papers/w24011

Bloom, N., Griffith, R. & Van Reenen, J. (2002). Do R&D tax credits work? Evidence from a panel of countries 1979–1997. *Journal of Public Economics, 85*(1), 1-31.

Bohne, A. & Nimczik, J. S. (2018). Information frictions and learning dynamics: Evidence from tax avoidance in ecuador.

Böhm, T., Karkinsky, T. & Riedel, N. (2012). *The impact of corporate taxes on R&D and patent holdings.* Paper presented at the SBS Conference Paper.

Christen, P. (2006). A Comparison of Personal Name Matching: Techniques and Practical Issues. *in 'The Second International Workshop on Mining Complex Data (MCD'06)*. doi:10.1109/ICDMW.2006.2

David, G. J. W. H. P. (2016). The Panama Papers: how the world's rich and famous hide their money offshore. *The Guardian*. Retrieved from https://www.theguardian.com/news/2016/apr/03/the-panama-papers-how-the-worlds-rich-and-famous-hide-their-money-offshore

Dimitrov, M. (2009). *Piracy and the state: The politics of intellectual property rights in China*: Cambridge University Press.

Dischinger, M. & Riedel, N. (2011). Corporate taxes and the location of intangible assets within multinational firms. *Journal of Public Economics, 95*(7-8), 691-707. doi:10.1016/j.jpubeco.2010.12.002

Du, J., Lu, Y. & Tao, Z. (2008). Economic institutions and FDI location choice: Evidence from US multinationals in China. *Journal of Comparative Economics*, *36*(3), 412-429. doi:https://doi.org/10.1016/j.jce.2008.04.004

Dueñas, S. (2019). *CFC Rules Around the World*. Retrieved from https://taxfoundation.org/cfc-rules-around-the-world/

EPO. (2017). *PATSTAT*.

EPO. (2020). What is a patent? In.

Public Hearing: PANA Committee, EU. 3-5 (2016).

European Patent Office. (2016). DATA CATALOG for PATSTAT Biblio and PATSTAT Legal Status. In.

Eurostat. (2011). Patent Statistics at Eurostat: Methods for Regionalisation, Sector Allocation and Name Harmonisation. In: Publications Office of the European Union.

Farrell, H. (2016). A massive leak just revealed how the super-rich hide their money. Here's what you need to know. *Democracy Dies in Darkness.* Retrieved from https://www.washingtonpost.com/news/monkey-cage/wp/2016/04/05/a-massive-leak-just-revealed-how-the-super-rich-hide-their-money-heres-what-you-need-to-know/

Fox, W. & Bayat, M. S. (2007). *A Guide to Managing Research*: Juta Publications.

Gambardella, A., Harhoff, D. & Verspagen, B. (2005). The value of patents. *Universita Bocconi, Ludwig-Maximiliens Universitaet, and Eindhoven University, Working Paper: http://www. creiweb. org/activities/sc_conferences/23/papers/gambardella. pdf.*

Garside, J., Watt, H. & Pegg, D. (2016, 03.04.2016). The Panama Papers: how the world's rich and famous hide their money offshore. *The Guardian*. Retrieved from https://amp.theguardian.com/news/2016/apr/03/the-panama-papers-how-the-worlds-rich-and-famous-hide-their-money-offshore?fbclid=IwAR2NEVCCeLped60RsO3OEVd6glayU6M1VYlupj1ZnXp9ikxjSNsyLjKe8zs

Giuri, P., Mariani, M., Brusoni, S., Crespi, G., Francoz, D., Gambardella, A., . . . Verspagen, B. (2007). Inventors and invention processes in Europe: Results from the PatVal-EU survey. *Research policy*, *36*(8), 1107-1127. doi:https://doi.org/10.1016/j.respol.2007.07.008

Gramlich, J. D., Limpaphayom, P. & Ghon Rhee, S. (2004). Taxes, keiretsu affiliation, and income shifting. *Journal of Accounting and Economics, 37*(2), 203-228. doi:https://doi.org/10.1016/j.jacceco.2003.10.001

Griffith, R., Miller, H. & O'Connell, M. (2014). Ownership of intellectual property and corporate taxation. *Journal of Public Economics, 112,* 12-23. doi:10.1016/j.jpubeco.2014.01.009

Guerrero-Bote, V. P., Sánchez-Jiménez, R. & De-Moya-Anegón, F. (2019). The citation from patents to scientific output revisited: a new approach to Patstat / Scopus matching. *El profesional de la información, 28*(4). doi:https://doi.org/10.3145/epi.2019.jul.01

Guo, H. (2007). IP Management at Chinese Universities. In IPhandbook (Ed.), *Intellectual Property Management in Health and Agricultural Innovation: A Handbook of Best Practices*.

Hall, B. H., Jaffe, A. & Trajtenberg, M. (2005). Market Value and Patent Citations. *The RAND Journal of Economics, 36*(1), 16-38.

Harding, L. (2016). What are the Panama Papers? A guide to history's biggest data leak. *The Guardian*. Retrieved from https://www.theguardian.com/news/2016/apr/03/what-you-need-to-know-about-the-panama-papers

Hasegawa, M., Hoopes, J. L., Ishida, R. & Slemrod, J. B. (2012). The effect of public disclosure on reported taxable income: Evidence from individuals and corporations in Japan. *Available at SSRN 1653948*.

ICIJ. (2020a). About. Retrieved from https://offshoreleaks.icij.org/pages/about

ICIJ. (2020b). Frequently Asked Questions. Retrieved from https://offshoreleaks.icij.org/pages/faq

ICIJ. (2020c). The Panama Papers: Exposing the Rogue Offshore Finance Industry. Retrieved from https://www.icij.org/investigations/panama-papers/

Intellectual Property Office. (2019). *Gender profiles in worldwide patenting: an analysis of female inventorship*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/846363/Gender-profiles-in-worldwide-patenting-2019.pdf.

Jacobides, M. G., Knudsen, T. & Augier, M. (2006). Benefiting from innovation: Value creation, value appropriation and the role of industry architectures. *Research policy, 35*(8), 1200-1221. doi:https://doi.org/10.1016/j.respol.2006.09.005

Jaravel, X., Petkova, N. & Bell, A. (2018). Team-specific capital and innovation. *American Economic Review, 108*(4-5), 1034-1073.

Johannesen, N., Langetieg, P., Reck, D., Risch, M. & Slemrod, J. (2018). *Taxing hidden wealth: The consequences of US enforcement initiatives on evasive foreign accounts* (0898-2937). Retrieved from https://www.nber.org/papers/w24366.pdf

Johannesen, N. & Zucman, G. (2014). The End of Bank Secrecy? An Evaluation of the G20 Tax Haven Crackdown. *American Economic Journal: Economic Policy, 6*(1), 65-91. doi:10.1257/pol.6.1.65

Jung, K., Kim, B. & Kim, B. (2009). Tax Motivated Income Shifting and Korean Business Groups (Chaebol). *Journal of Business Finance & Accounting, 36*(5-6), 552-586. doi:10.1111/j.1468-5957.2009.02141.x

Kang, B. & Tarasconi, G. (2015). *PATSTAT revisited*.

Karkinsky, T. & Riedel, N. (2012). Corporate taxation and the choice of patent location within multinational firms. *Journal of International Economics, 88*(1), 176-185. doi:10.1016/j.jinteco.2012.04.002

Kastlunger, B., Lozza, E., Kirchler, E. & Schabmann, A. (2013). Powerful authorities and trusting citizens: The Slippery Slope Framework and tax compliance in Italy. *Journal of Economic Psychology, 34*, 36-45. doi:https://doi.org/10.1016/j.joep.2012.11.007

Kenton, Will (2019). Chaebol Structure. Investopedia url: https://www.investopedia.com/terms/c/chaebol-structure.asp

Keupp, M. M., Beckenbauer, A. & Gassmann, O. (2009). How managers protect intellectual property rights in China using de facto strategies. *R&D Management, 39*(2), 211-224. doi:10.1111/j.1467-9310.2009.00550.x

Kirchler, E., Hoelzl, E. & Wahl, I. (2008). Enforced versus voluntary tax compliance: The “slippery slope” framework. *Journal of Economic Psychology, 29*(2), 210-225. doi:https://doi.org/10.1016/j.joep.2007.05.004

Kleven, H., Landais, C., Muñoz, M. & Stantcheva, S. (2020). Taxation and Migration: Evidence and Policy Implications. *Journal of Economic Perspectives, 34*(2), 119-142. doi:10.1257/jep.34.2.119

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

Kogan, L., Papanikolaou, D., Seru, A. & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly journal of economics, 132*(2), 665-712.

Kogler, C., Batrancea, L., Nichita, A., Pantya, J., Belianin, A. & Kirchler, E. (2013). Trust and power as determinants of tax compliance: Testing the assumptions of the slippery slope framework in Austria, Hungary, Romania and Russia. *Journal of Economic Psychology, 34*, 169-180.

KPMG. (2018). PRC Individual Income Tax Reform. Retrieved from https://home.kpmg/cn/en/home/insights/2018/07/prc-individual-income-tax-reform.html

La Croix, S. J. & Eby Konan, D. (2002). Intellectual Property Rights in China: The Changing Political Economy of Chinese–American Interests. *The World Economy, 25*(6), 759-788. doi:10.1111/1467-9701.00462

Lanjouw, J. O. & Schankerman, M. (2004). Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators*. *The Economic Journal, 114*(495), 441-465. doi:10.1111/j.1468-0297.2004.00216.x

Liberto, D. (2019). Keiretsu. In *Investopedia*.

Lincoln, J. R., Gerlach, M. L. & Ahmadjian, C. L. (1996). Keiretsu Networks and Corporate Performance in Japan. *American Sociological Review, 61*(1), 67-88. doi:10.2307/2096407

Malan, J., Blomeyer, R., Smit, J., Vlcek, W., Jansky, P. & Nerudová's, D. (2017). *The Impact of Schemes revealed by the Panama Papers on the Economy and Finances of a Sample of Member States*. Retrieved from https://www.europarl.europa.eu/RegData/etudes/STUD/2017/572717/IPOL_STU(2017)572717_EN.pdf

Merges, R. P. & Nelson, R. R. (1990). On the complex economics of patent scope. *Columbia Law Review, 90*(4), 839-916.

Miguelez, E. & Fink, C. (2013). *Measuring the international mobility of inventors: A new database* (Vol. 8): WIPO.

Moretti, E. & Wilson, D. J. (2017). The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists. *American Economic Review, 107*(7), 1858-1903. doi:10.1257/aer.20150508

Møen, J. (2005). Is mobility of technical personnel a source of R&D spillovers? *Journal of labor economics, 23*(1), 81-114.

News Facts Network. (2020). International Consortium of Investigative Journalists. Retrieved from https://mediabiasfactcheck.com/international-consortium-of-investigative-journalists/

O'Donovan, J., Wagner, H. F. & Zeume, S. (2019). The Value of Offshore Secrets: Evidence from the Panama Papers. *The Review of Financial Studies*, *32*(11), 4117-4155. doi:10.1093/rfs/hhz017

OECD. (2009). *OECD Patent Statistics Manual*.

OECD. (Ed.) (2020a). OECD.

OECD. (2020b). *Taxing Wages*. Retrieved from https://www.oecd.org/tax/tax-policy/taxing-wages-brochure.pdf

Omartian, J. D. (2017). Do Banks Aid and Abet Asset Concealment: Evidence from the Panama Papers. *Available at SSRN 2836635*.

Paetzold, J. & Winner, H. (2016). Taking the high road? Compliance with commuter tax allowances and the role of evasion spillovers. *Journal of Public Economics, 143*, 1-14. doi:https://doi.org/10.1016/j.jpubeco.2016.08.001

Pegg, D. (2016). Panama Papers: Europol links 3500 names to suspected criminals. *The Guardian*. Retrieved from https://www.theguardian.com/news/2016/dec/01/panama-papers-europol-links-3500-names-to-suspected-criminals

Peruzzi, M., Zachmann, G. & Veugelers, R. (2014). Remerge- regression-based record linkage with an application to PATSTAT.

Pisano, G. P. & Teece, D. J. (2007). How to capture value from innovation: Shaping intellectual property and industry architecture. *California management review, 50*(1), 278-296.

PWC. (2020). Tax Summaries. https://taxsummaries.pwc.com/

Raffo, J. (2020). MATCHIT: Stata module to match two datasets based on similar text patterns. Retrieved from https://EconPapers.repec.org/RePEc:boc:bocode:s457992

Raffo, J. & Lhuillery, S. (2009). How to Play the 'Names Game': Patent Retrieval Comparing Different Heuristics. *SSRN Electronic Journal*. doi:10.2139/ssrn.1441172

Remeur, C. (2018). *Listing of Tax Havens by the EU*. EPRS Retrieved from https://www.europarl.europa.eu/cmsdata/147412/7%20-%2001%20EPRS-Briefing-621872-Listing-tax-havens-by-the-EU-FINAL.PDF.

Roussille, N. (2015). Tax evasion and the 'Swiss cheese'regulation. *Paris School of Economics and Ecole Polytechnique*.

Samans, R., Blanke, J., Corrigan, G. & Hanouz, M. D. (2017). *Inclusive Development Index 2017*. Retrieved from http://www3.weforum.org/docs/WEF_Forum_IncGrwth_2017.pdf

Sandmo, A. (2005). The theory of tax evasion: A retrospective view. *National tax journal*, 643-663.

Saunders, M., Lewis, P., Thornhill, A. & Bristow, A. (2012). Research Methods for Business Students. In: Pearson Education.

Schmoch, U. (2008). *Concept of a Technology Classification for Country Comparisons*. Retrieved from

Squicciarini, M., Dernis, H. & Criscuolo, C. (2013). Measuring Patent Quality. doi:https://doi.org/10.1787/5k4522wkw1r8-en

Tarasconi, G. (2014). Matching patent data with financial data. *OBSERVATOIRE DES SCIENCES ET DES TECHNIQUES: Les notes pratiques de l'OST*.

Tarasconi, G. & Menon, C. (2017). Matching Crunchbase with patent data. doi:https://doi.org/10.1787/15f967fa-en

Teece, D. J. (2014). A dynamic capabilities-based entrepreneurial theory of the multinational enterprise. *Journal of international business studies, 45*(1), 8-37.

The Global Economy. (2018). Political Stability - Country Rankings.

The World Bank. (2020). Populations. Retrieved from https://data.worldbank.org/indicator/sp.pop.totl

Torrisi, S., Gambardella, A., Giuri, P., Harhoff, D., Hoisl, K. & Mariani, M. (2016). Used, blocking and sleeping patents: Empirical evidence from a large-scale inventor survey. *Research policy, 45*(7), 1374-1385. doi:https://doi.org/10.1016/j.respol.2016.03.021

Trajtenberg, M. (2005). *Recombinant ideas: the mobility of inventors and the productivity of research.* Paper presented at the CEPR-Conference, Munich.

Trajtenberg, M., Shiff, G. & Melamed, R. (2006). *The" names game": Harnessing inventors' patent data for economic research* (0898-2937). Retrieved from https://www.nber.org/papers/w12479.pdf

Transparency International. (2016). *Corruptions Perception Index 2016*. Retrieved from https://www.transparency.org/en/publications/corruption-perceptions-index-2016

Tørsløv, T. R., Wier, L. S. & Zucman, G. (2018). *The missing profits of nations* (0898-2937). Retrieved from https://www.nber.org/papers/w24701.pdf

van der Does de Willebois, E. H., Emily M.; Harrison, Robert A.; Park, Ji Won; Sharman, J. C. (2011). *The Puppet Masters : How the Corrupt Use Legal Structures to Hide Stolen Assets and What to Do About It*. Retrieved from https://openknowledge.worldbank.org/handle/10986/2363

Wang, P. (2014). Extra-legal Protection in China: How Guanxi Distorts China's Legal System and Facilitates the Rise of Unlawful Protectors. *The British Journal of Criminology, 54*(5), 809-830. doi:10.1093/bjc/azu041

Wilson, J. (2010). *Essentials of business research : a guide to doing your research project*: London: SAGE.

WIPO. (2017). *WIPO IP Facts and Figures 2016*. Retrieved from WIPO: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_943_2016.pdf

Wong, P. K., Ho, Y. P. & Autio, E. (2005). Entrepreneurship, innovation and economic growth: Evidence from GEM data. *Small business economics, 24*(3), 335-350.

Xin, K. K. & Pearce, J. L. (1996). Guanxi: Connections as substitutes for formal institutional support. *Academy of management journal, 39*(6), 1641-1658.

Yang, H. & Steensma, H. K. (2014). When do firms rely on their knowledge spillover recipients for guidance in exploring unfamiliar knowledge? *Research policy, 43*(9), 1496-1507. doi:https://doi.org/10.1016/j.respol.2014.04.016

Zikmund, W., Babin, B., Carr, J. & Griffin, M. (2012). *Business research methods*: Cengage Learning.

Zucman, G. (2013). The missing wealth of nations: Are Europe and the US net debtors or net creditors? *The Quarterly journal of economics, 128*(3), 1321-1364.

Zucman, G. (2015). *The hidden wealth of nations: The scourge of tax havens*: University of Chicago Press.

Appendix

Table A1. Number of all inventors in PATSTAT, all individuals in the Offshore Leaks and all inventors found in the Offshore Leaks per country

Country	*Number of inventors in PATSTAT*	*Numbers of individuals in the Offshore Leaks*	*Inventors in the Offshore Leaks*
Argentina	2,248	1,525	4
Australia	49,358	2,751	89
Austria	32,231	1,117	81
Belgium	34,183	1,241	28
Brazil	15,088	2,458	20
Bulgaria	1,866	443	6
Canada	78,498	3,620	34
Chile	2,302	307	0
China	118,759	29,417	6,961
Colombia	1,594	1,342	6
Croatia	1,328	232	3
Cuba	1,268	55	0
Czech Republic	6,247	556	16
Denmark	27,415	608	55
Egypt, Arab Rep.	1,004	645	2
Estonia	1,005	297	4
Finland	30,958	396	13
France	197,867	3,516	176
Germany	405,505	5,060	477
Greece	3,304	2,837	24
Hong Kong SAR, China	1,809	24,822	113
Hungary	11,998	781	57
Iceland	948	267	8
India	40,958	1,716	88
Ireland	10,292	2,035	143
Israel	35,420	1,605	118
Italy	91,767	7,964	670
Japan	499,994	1,587	576
Korea, Rep.	131,376	540	185
Latvia	996	404	1
Luxembourg	1,770	680	3
Malaysia	6,005	3,071	23
Mexico	6,246	528	3
Netherlands	82,389	1,229	32
New Zealand	9,055	921	8
Norway	16,425	895	24
Philippines	937	1,040	2
Poland	10,273	911	18
Portugal	4,389	1,266	3
Romania	1,834	443	2
Russian Federation	31,204	6,740	168
Saudi Arabia	2,629	1,103	2
Singapore	14,614	6,505	116
Slovak Republic	1,582	304	3
Slovenia	3,147	129	1
South Africa	9,426	2,621	27
Spain	57,955	2,101	25
Sweden	54,989	2,292	139
Switzerland	62,901	5,082	183
Taiwan	20,331	18,276	1,205
Thailand	1,814	1,618	5
Turkey	8,070	1,962	30
Ukraine	3,858	965	7
United Kingdom	179,253	19,954	1,144
United States	1,184,123	26,202	2,765
Total	*3,612,805*	*206,982*	*15,896*

Note: The number unique of individuals in the Offshore Leaks is 194,658, but individuals with two country codes (e.g. "NOR;DEU") are accounted for in both countries and matched with both countries in PATSTAT, which is why the total number is higher in this table (see section 3.4.2 for more information)