Tax noncompliance and inequality - Three empirical essays for the case of Germany

Inaugural-Dissertation zur Erlangung des
akademischen Grades eines Doktors der
Wirtschaftswissenschaft

Fachbereich Wirtschaftswissenschaft
Freie Universität Berlin

vorgelegt von

Hannes Fauser, M.Sc.
aus Halberstadt

Berlin, 2021

Dekan:
Prof. Dr. Dr. Giacomo Corneo

Erstgutachter:
Prof. Dr. Dr. Giacomo Corneo, Freie Universität Berlin

Zweitgutachter:
PD Dr. Stefan Bach, DIW Berlin und Universität Potsdam

Tag der Disputation: 18.02.2022

Zusammenarbeit mit KoautorInnen und Vorveröffentlichungen

Kapitel 1: Germany's efforts to curb international tax evasion

- Kapitel 1 basiert auf einem bisher unveröffentlichten Papier, das zu gleichen Teilen mit Jakob Miethe verfasst wurde.

Kapitel 2: On income tax avoidance - a new micro data model for the German case

- Kapitel 2 wurde in Alleinautorenschaft verfasst und basiert auf einem Arbeitspapier, das zuvor unter anderem Titel als Beitrag für eine Konferenz öffentlich zugänglich wurde. Gegenüber dem zuvor veröffentlichten Papier ist das Kapitel inhaltlich und redaktionell verändert. Zudem wurde der Titel geändert.
- "On income tax avoidance - the case of Germany revisited", Beiträge zur Jahrestagung des Vereins für Socialpolitik 2019: 30 Jahre Mauerfall - Demokratie und Marktwirtschaft - Session: Taxation III No. E15-V2, Link.

Kapitel 3: Income tax noncompliance in Germany, 2001-2014

- Kapitel 3 basiert auf einem zum Zeitpunkt der Einreichung unveröffentlichten Papier, das zu gleichen Teilen mit Sarah Godar verfasst wurde.

Acknowledgments

First, I would like to express my sincere gratitude to Giacomo Corneo for supervising my thesis, providing excellent advice and staying patient. In particular, I am grateful that he directed me towards a very interesting research topic of high real-world relevance, and gave me the opportunity to start a PhD project based on an exposé that represented a bet on identifying and accessing suitable data. They were rarely liable to entertainment tax, but his seminars on theoretical topics in public economics and inequality have greatly improved my understanding of the larger field, as well as of the meaningfulness of neoclassical modelling under certain conditions. I would also like to thank Stefan Bach for co-reviewing this thesis and giving advice. Their contributions to the distributional analysis of income and wealth and its interlinkages with tax policy have been a great inspiration for my work.

I owe much to Jakob Miethe and Sarah Godar, both of whom are not only excellent co-authors but good friends that I could turn to when seeking advice on research matters and beyond. Without them, this thesis would not have materialized.

I also want to thank the (former) members of the PhD program on "Public economics and inequality", whose company I enjoyed very much and whose comments, suggestions and stimulating presentations have vastly improved my knowledge. Namely, I thank Mattis Beckmannshagen, Peihua Deng, Anja Gaentzsch, Rick Glaubitz, Christoph Halbmeier, Edison Jakurti, Robin Jessen, Markus Krecik, Maximilian Longmuir, Marc Lorbeer, Maria Metzing, Theresa Neef, Guido Neidhöfer, Birger Scholz, Maximilian Stockhausen, Anita Tiefensee, Marten von Werder and Christian Westermeier. As members of the faculty, I would like to thank Nelli Anan, Charlotte Bartels, Timm Bönke, Benjamin Fischer, Astrid Harnack, Dominik Hügele, Johannes König, Holger Lüthen, Astrid Pape, Merlin Penny and Miriam Wetter.

Moreover, thanks go to Natalia Danzer, Timm Bönke and Theresa Neef who were members of my PhD committee.

Coming to further researchers I am thankful to, Gabriel Zucman deserves a prominent place for inspiring much of my research and providing helpful advice in the earlier stages of my PhD project. Morevoer, comments and suggestions from numerous participants at various conferences, seminars and workshops are greatly appreciated. In particular, I would like to mention Irene Becker, Jana Emmenegger, Petr Janský, Tobias Hauck, Mehrene Larudee, Miroslav Palanský, Verena Pagels, Richard Murphy, Sara Torregrosa Hetland and Viktor Steiner. Nikolaus Wolf deserves thanks for guiding my research interests towards quantitative analysis during my Master's studies and

for supporting my applications for PhD funding.

The staff of the research data centres of the Federal Statistical Office and the Statistical Offices of the Länder has to be thanked for doing an excellent job under conditions of scarcity. Hence, I would like to thank Maurice Brandt, Tasso Brandt, Benjamin Gampffer, Melanie Heiliger, Steffen Lauf, Janina Loske, Stefanie Setzer, Stefanie Uhrich and Ulrike Türk-Schröter.

Without financial support, very little research is possible. I am therefore very grateful to the Rosa Luxemburg Foundation for granting me a PhD scholarship that made the largest part of this thesis possible. I would like to thank Cansel Kiziltepe and her team for the great time at Steuermythen. A big thank you also to Stefanie Knoth, for fruitful discussions on my research, sharing her tax policy experience and helping me to settle in at a new position. Many other colleagues deserve my gratitude. Representatively for the input received for my academic work, I would like to thank Klaus Wolter for advice on the tax statistics, and Karl-Heinz Nöhrbaß for discussions concerning my paper on income tax avoidance.

Many thanks also to friendly faces in the tax policy community, in particular to Lisa Großmann, Markus Henn, Stefan Herweg, Alexander Sacharow and Christoph Trautvetter. I furthermore owe a lot to my WiÖ friends Anja Breljak, Finn Müller-Hansen, Felix Kersting, Nikita Sorgatz and Iris Wohnsiedler. Our self-organized colloquium and a host of other activities were a central pillar for my persisting interest in economics. For their comradeship during my undergraduate and graduate studies, I want to thank Stephan Becker, János Jákli, Christopher Voss and André Schmelzer.

Let me wrap up by thanking Daniel Uhlig for stimulating discussions on political economy and law, Tobias Wolfram for his outstanding statistical and programming advice, and both for their commitment to open debate and intellectual honesty. Lots of other friends deserve my gratitude who I ask for leniency not to be named here.

I am deeply grateful to my parents Ellen and Frank for their unconditional moral and material support, and to my brother Henning for support in all aspects of life. See, in the end I did take longer despite the good wishes in your *remerciements*. Finally, my loving gratitude to Desa for sharing my life since more than five years.

Preface

When I started to work on my PhD project in 2015, I did so under the working title "Offshore wealth, tax evasion, 'shadow' capital markets – The German wealthy's money harbours". At the time, I had hoped to among others be able to determine the quantitative importance of tax evasion and avoidance, and to show how wealth is accumulated and protected from public scrutiny in Germany. In the last more than six years, I have followed this outline with some success. Nonetheless, it has to be stated that especially the latter target has been achieved only to a very limited degree. The accumulation and transfer of wealth inside the realms of financy secrecy as well as the tax shelters of the rich remain well protected from quantitative empirical scrutiny, at least in Germany. Thanks in particular to investigative journalists, we have learned more about the world of offshore finance and secrecy jurisdictions. In the past decade, leak after leak has torn holes into the layers of secrecy, yet for Germany there is still only very limited structural empirical evidence to the best of my knowledge.

The empirical identification ideas I had outlined in my PhD exposé, as well as some promising leads I followed over the course of my research work, have failed often times. Almost every time, the reasons included restricted access or limited availability of potentially interesting data sources. Most notably, a treasure of data sleeps in the hands of German federal states' fiscal authorities: Cases of tax evaders who self-reported their fraud in order to enjoy limited punishment. According to the former Minister of Finance of the most populous federal state of Northrhine-Westphalia, his state alone received 23,000 self-reports between 2010 and 2016, resulting in additional fiscal revenues of EUR 1.2 bn (WDR 2016). Back in 2016, I started to try to get access to such cases, attempting to convince fiscal authorities to build a micro dataset that could be used to study tax evasion directly and rigorously. In the last five years, studying direct evidence of tax evasion like leaks of evaders, amnesties or self-reports has been on the research frontier in the empirical literature on tax evasion. Following the seminal study of Alstadsæter et al. (2019), authors have been able to shed light on the relationship of tax noncompliance and inequality in particular. Unfortunately though, due to a mixture of legal concerns, ressource constraints and a lack of political will, such an endeavour has until now been futile for the German case. The tax statistics do not include cases that have not been closed 2.75 years after the tax assessment period. Therefore, most self-reports are neither included nor identifiable. Moreover, data on tax audits is not available for research.

Fortunately, these difficulties were overcome to a considerable degree, in the three self-standing papers that have been collected into this dissertation. By applying indirect estimation approaches, each one of them exploits one or more aspects of some dataset(s) to gain insights on tax noncompliance and its distributional implications: In chapter 1, we use monthly Bundesbank data on bank deposits to study reactions of tax evaders to policy measures and political events. Chapter 2 exploits the richness of the German Income and Consumption Survey in a novel way, by using information on incomes, expenditures and tax payments to derive a lower bound estimate on income tax avoidance. Lastly, chapter 3 uses well-established direct discrepancy and indirect expenditure-based approaches with survey and tax data to arrive at estimates of income tax underreporting for a longer period of time (2001-2014).

Setting aside the policy conclusion that can be drawn from the results on tax noncompliance and inequality that this PhD thesis collects, at this point I would like to stress a conclusion derived from the circumstances described above: Germany needs a better environment for research with administrative microdata. If the government strives to strenghten evidence-based policy making, it has to improve the competitiveness of using German data. Researchers interested in tax micro data for instance still lack remote access, having to rely on guest researcher workplaces or controlled remote data processing. More importantly, linking register data is prohibited in most cases, and statistical matching of these data with external sources is restricted and complicated. Therefore, I can subscribe to the demands of the scientific advisory council of the Federal Ministry of Finance, which published a report on the need, potential and approaches for improving the data infrastructure for tax policy in Germany(BMF 2020*b*). To end this preface on a positive note, it should be mentioned that in 2021, plans to improve the situation by establishing an institute for empirical tax research have been anounced by the Ministry of Finance (Dörrenberg & Peichl 2021). I am optimistic that its implementation will be a fruitful endeavour.

Contents

Chapter 1

Germany's efforts to curb international tax evasion

1.1 Introduction

An abundance of leaks of illegal tax noncompliance practices in offshore financial centers has revitalized attempts to curb international tax evasion. Tax evasion erodes both public finances and the legitimacy of governments around the world, both of which are reasons for concerted policy efforts in this area. A multitude of bilateral Tax and Information Exchange agreements has been signed over the last two decades and several countries have carried out unilateral measures. The implementation of the Common Reporting Standard (CRS) and the subsequent automatic exchange of tax information amongst a growing number of jurisdictions starting in 2017 was another such step.

We employ a monthly dataset of bilateral bank deposits provided by the German Bundesbank which offers a unique opportunity to evaluate the relative success of both national and international policy action in this area. This dataset provides deposits in German banks made by counterparties officially registered in tax havens. As we show in our analysis, these deposits include traces of evaded capital connected to its repatriation. In order to identify these traces, we exploit international tax information exchange (TIE) efforts, national regulatory changes, and direct leaks of information concerning German evaders. At the time of writing, we could not include the automatic TIE accords in our analysis, so the bulk of our TIE variation relies on the less effective TIE on request. Germany has been involved in all major regulation attempts and additionally carried out unilateral measures both in law and by exploiting leaks.

It is impossible to trace tax evaded deposits directly: they are hidden both from

authorities and researchers alike. The problem is - as Slemrod (2015, p. 10) succinctly put it - that we can neither observe the right hand side (the effect of policies targeting evasion) nor the left hand side (evaded capital) of a potential regression. Even in the detailed data we can employ by focusing on Germany, evaded capital is in all likelihood a small fraction of total deposits in German banks.

We therefore rely on an indirect identification strategy. The effect we expect if tax evasion is present, is a change of those deposits which are at some point connected to tax evasion. If deposits react to regulatory changes exclusively aimed at illegal behaviour or leaks which only concern evaded capital, this is interpreted as evidence of tax evasion. We follow Hanlon et al. (2015) by analysing positions inside a non-haven but attributed to counterparties in tax havens which can include German nationals under disguised identity. Our analysis hence hinges on two assumptions: i) Part of the offshore capital of German households or firms flows back to Germany, and ii) changes with respect to TIE or leaks of the names of evaders exclusively affect illegal capital related to tax evasion evaded by German nationals who worry about detection. We employ a narrative approach to identify regulatory changes in conditions affecting evasion for German nationals. This is based on an in depth analysis of the German tax code and changes concerning dividend payment, personal income taxes and the like. We hope that this database is a contribution useful for further research in its own right.

Germany has grown in popularity with researchers studying tax avoidance: Hebous & Johannesen (2015) use microdata for German multinational firms and their worldwide affiliates to estimate the extent of transfer mispricing in international trade with services. The same database is exploited by Gumpert et al. (2016) to study the incentives for multinationals to have affiliates in tax havens in the first place, finding that higher foreign taxes are associated with a higher probability of owning a tax haven affiliate. Langenmayr & Reiter (2017) and Reiter (2017) analze the tax avoidance of German banks using a dataset on banks individual positions. This research agenda on German tax noncompliance has not been extended to study tax evasion by German households, however. We aim at filling this gap by, to the best of our knowledge, providing the first study analysing tax evasion in general and repatriation in particular by using monthly bank deposits data.

On top of the availability of monthly deposit data, there are several reasons why Germany makes for a good case study to analyse household tax evasion. First, it is the largest economy in Europe which makes the effects we find economically meaningful.

Second, German residents hold about 16% of their wealth offshore, compared to a world average of only 9.8% (Alstadsæter et al. 2017). Moreover, they also hold a high net foreign wealth position in general (Piketty & Zucman 2014, p. 1279). Being a net exporter for more than a decade, Germany's high surpluses, accumulated predominantly with firms and wealthy households, are invested around the globe. Third, the use of tax havens has a long tradition in Germany. Particularly neighbouring jurisdictions like Switzerland, Liechtenstein or Luxembourg were prime destinations dating back to the First World War at least. Larudee (2015) mentions Keynes (1920) who already noted that much of the mobile German private capital had moved to Switzerland at the end of WWI, fearing reparations imposed on Germany. During the interwar period, funds sheltered in Swiss banks evaded the higher taxes imposed to cover the costs of the war. During WWII and thereafter, offshore money was protected for instance by the Swiss Bank secrecy law of 1934, anonymous accounts and secret Liechtenstein trusts. Starting in the 1980s, Luxembourg grew as an offshore financial center for German customers, as there are reports of massive capital outflows to the Grand Duchy during a 1989 withholding tax experiment in Germany (Nöhrbaß & Raab 1990), and after the introduction of the interest income tax in 1993 (Deutsche Bundesbank 1994). Journalistic sources also report that at the time, Germans were mostly unwilling to deposit their funds with non-German banks, and that most of the capital flowed back into German investment opportunities (Der Spiegel 1993).

During the timespan of our baseline sample (Jan 2003 - Dec 2016), offshore evasion tended to internationalise further, with growing complexity and specialization of individual tax havens. Our study adds to a growing literature analyzing offshore wealth, tax evasion and anti-evasion measures. We complement this literature by adding evidence of evasion effects in a developed economy which has so far only been provided for the US by Hanlon et al. (2015). We also provide for the first time in the literature a direct glimpse into bilateral bank deposits in Germany vis-a-vis a list of common tax havens and their reactions to information exchange and other policy measures.

The remainder of this chapter is organised as follows: Section 1.2 introduces the indirect identification strategy in more detail by outlining tax evasion strategies that produce the type of positions we use in our data. Section 1.3 outlines the policy measures we evaluate. In section 1.4, we describe the bank deposit data as well as our narrative database on Germany-specifc leaks, scandals and changes in the tax code. Results are presented in section 1.5 and section 1.6 concludes.

1.2 Identification

A simple scheme of offshore tax evasion during our sample period[1] may be sketched as follows: A German business owner who seeks to shelter wealth offshore incorporates a shell company or buys a shelved one from a law firm, for instance in the British Virgin Islands. This requires only little identity information, and the beneficial owner may be further disguised by using fiduciaries or additional layers of shell corporations. The firm then opens a bank account, say in the Cayman Islands. Next, the evader's German enterprise buys services that are difficult to monitor, like consulting services, from her new shell corporation and pays via bank transfer to the Cayman Island account. Since there are millions of international bank transfers between firms every day, these financial flows will hardly attract any attention (for a more detailed exposition, see Zucman 2014). There are two potential benefits of this construct: On one hand, the evader reduces the taxable income of her German company, on the other hand she may generate returns on the funds now taken offshore by investing them, typically in international financial markets (Zucman 2015). As long as neither her nor the Cayman Island bank report the account to German fiscal authorities, she may evade German capital income tax, too.

As soon as the tax evader wants to access her funds at home, she needs to repatriate. This potentially raises some issues for her, particularly by increasing the threat of detection. There are numerous ways in which to repatriate funds, only some of which will show up in bank data of the country to where the money flows.[2]

Undocumented cash transfers, either by the evader herself or a messenger, of course do not show up in bilateral positions. Simply wiring the funds to a domestic account without notifying domestic tax authorities is considered illegal, in the German-speaking context so-called "cold repatriation", which presumably was more common prior to extensive surveillance of domestic bank account transactions. The legal version is to

[1] Please note that conditions have changed since the implementation of automatic exchange of tax information, starting in 2017.

[2] A number of cases leading to convictions of tax evaders using offshore constructs is publicized each year by the US Internal Revenue Service. For the list of 2017 and links to earlier years: https://www.irs.gov/compliance/criminal-investigation/examples-of-abusive-tax-schemes-fiscal-year-2017. There are numerous cases where tax evaders and/or corrupt officials, outright criminals etc hold bank accounts in non-haven jurisdictions using a shell company in a haven jurisdiction. A World Bank report lists cases of Grand Corruption, one for example involving a corporate vehicle opening a USD bank account with UBS in London. Common characteristics of some 150 cases studied are inter alia that "a corporate vehicle was misused to hide the money trail [...] the proceeds and instruments of corruption consisted of funds in a bank account" (de Willebois et al. 2011, p. 2). Meinzer (2015) describes cases in the German context, where mostly foreign politically exposed persons were able to open accounts and hide/invest money in Germany using correspondent banking.

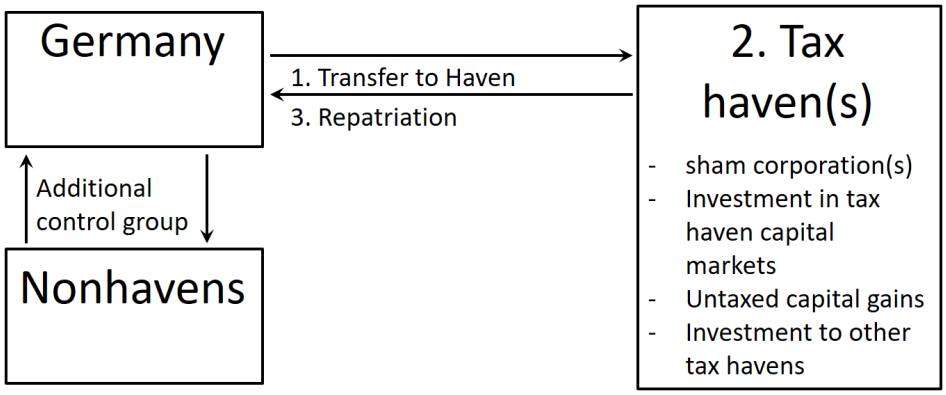
inform tax authorities about the funds, their origin and accepting tax due and possible back tax to be paid. Even if the repatriation move in itself is legal, it can betray earlier illegal activities.

An example that does show up in the international banking position, is when the sham corporation opens a bank account in Germany to either finance investment or ongoing consumption. Alternatively, the BVI shell may deposit funds in a trust held by fiduciaries of the evader. This would also be reported as a liability against a foreign counterparty. These funds show up in the data as deposits of a tax haven counterparty in Germany, as long as the entity is incorporated in one of the tax havens.

In line with this logic, part of the liabilities of German banks to non-residents, their subsidiaries and of foreign banks' subsidiaries in Germany actually belong to German nationals who reinvest tax evading funds previously hidden abroad. The funds would then be held by an offshore sham corporation, but the ultimate beneficial owner would be a German household. At some point, at least part of these funds flow back to Germany. Therefore, we expect deposit data to react to regulatory changes which affect the threat of detection or the benefits of evasion.

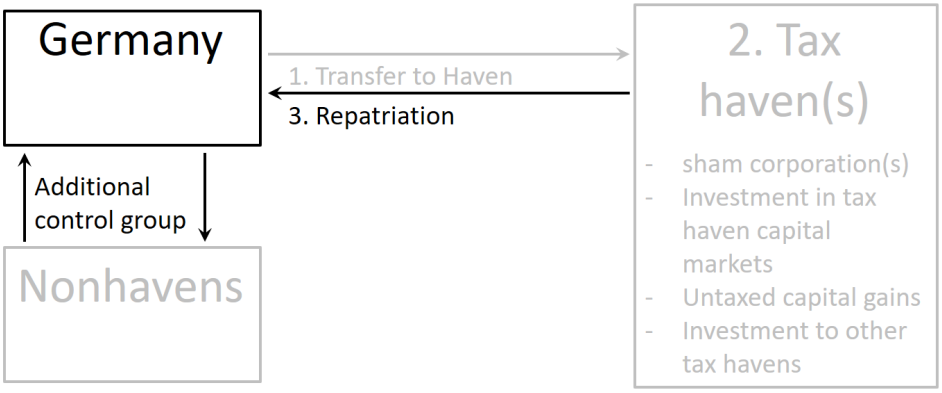
Figures 1.1 and 1.2 show a full international tax evasion scheme including repatriation. The most direct route evaded capital can take is round tripping, meaning offshoring (step 1), capital investment in tax havens (step 2) and direct repatriation (step 3). Such round tripping does not need to take place exclusively via one tax haven. It is therefore not necessarily (though possibly) the tax haven from which repatriation takes place that hides the illegal tax evasion scheme.

Figure 1.1: Schematic overview of flows connected to offshore tax evasion



The return of offshore funds may be due to several reasons: First, in line with Hanlon et al. (2015), investments that are financed by offshore funds can be driven by home bias effects. Like normal investors, evaders prefer to reduce uncertainty by using their funds in familiar environments. Even if some funds are kept within tax havens as

Figure 1.2: Capital moving out of other non-havens



savings for unforeseen circumstances[3], at least part of the money will likely be spent at home to raise the well-being of evaders where they live. If bank deposits made by tax haven counterparties in German banks thus include evaded capital, we expect reactions to international and national regulation attempts as well as data leaks. These policy measures are outlined in the following section.

1.3 Policy Measures

Starting with its report on harmful tax competition, the OECD committed to efforts aimed at combating tax avoidance and evasion. It identified tax havens as jurisdictions with the following traits (OECD 1998, p.23): Only nominal or no taxes, lack of exchange of information, lack of transparency, and no substantial economic activities. The OECD furthermore established the Global Forum on Transparency and Exchange of Information for Tax Purposes in 2000, which inter alia established standards for TIE and encouraged countries to sign new agreements or amend existing double taxation conventions (DTCs). During the financial crisis, the G20 started another crackdown on tax havens in 2009, mandating countries to sign at least 12 tax information exchange agreements (TIEAs) in order to avoid blacklisting.

These treaties specify quite restrictive exchange of information on request. The information has to be "foreseeably relevant" and has to be provided only if the requesting party "is unable to obtain the requested information by other means in its own territory, except where recourse to such means would give rise to disproportionate difficulty". Any request has to specify "with the greatest detail possible" both the "identity of the person under examination or investigation", "the period for which the informa-

[3]Anecdotal evidence has it that offshore bank accounts are in vogue with wealthy men in fear of payments due to divorce, or with business owners seeking to shelter funds from bankruptcy (Richter 2016).

tion is requested", "grounds for believing that the information requested is held in the requested Contracting Party", and several other conditions [4]. Only then is account information exchanged.

While these treaties only constitute a marginal threat of detection, shifting one's tax evasion setup to a different jurisdiction is cheap and simple (see Sharman 2010 for some real world examples) and such reactions to these treaties have been found in offshore deposits (Johannesen & Zucman 2014). Being an EU member state, Germany also implemented the EU Savings Tax Directive 2003/48/EC which seeks to establish information exchange about interest earned by EU residents.[5]

As part of the OECD and G20 initiatives, Germany signed 22 TIEAs and updated a number of DTCs to include information exchange. As of 2017, it could exchange bank account information in tax matters with 116 jurisdictions through 96 DTCs and 22 TIEAs. The regulatory foundations for information exchange between tax havens and Germany are summarized in table 1.1. The second column shows policy measures which, according to our identification strategy, are predicted to have negative impacts on deposits from the respective tax haven. We include TIEAs and DTCs (previously analysed by Johannesen & Zucman 2014 and Hanlon et al. 2015) as well as the EU Savings Directive (previously analyzed by Johannesen & Zucman 2014, Johannesen 2014). In October 2014, common agreement of 101 countries about automatic TIE was reached at the OECD Global Forum in Berlin. As of November 2016, 87 jurisdictions had joined the CRS MCAA, agreeing to over 1,300 bilateral exchange relationships which have started to be activated in 2017. However, at the time of writing it was still too early to study its effects empirically.

We also include a number of tax changes in order to capture increased or decreased benefits of tax evasion, e.g. via roundtripping as in Hanlon et al. (2015). Following this logic, evaders have a higher incentive for noncompliance when capital gains taxes or personal income taxes are high. They therefore evade higher sums and spend or repatriate (part of) them disguised as foreign residents. Only German citizens, who have to pay these taxes, should be affected, as foreign residents are exempt from them[6].

[4]The quotes here are taken from the TIEA signed between Germany and the Cayman Islands on the 27th of May 2010, available via the OECD exchange of information portal: http://eoi-tax.org/agreements/DE_KY_TIEA_18#default, last accessed 14.02.2018

[5]Both the 1998 OECD and the 2009 G20 measures have received a largely negative evaluation in the empirical literature (Kudrle 2008, Kudrle 2009, Johannesen & Zucman 2014). Similar conclusions are drawn by the majority of authors analysing the EUSTD (Johannesen 2014, Klautke & Weichenrieder 2010, contrasted by Hemmelgarn & Nicodeme 2009.

[6]This was not the case for the so-called "small tax on capital income" though, a short-lived source tax levied on interest generated from domestic debt securities in 1989.

Table 1.1: German treaties with tax havens in our sample period (2003-2016)

Counterparty	**Type**	**Signature**	**In force**
Andorra	TIEA	25 Nov 2010	20 Jan 2012
Anguilla	TIEA	19 Mar 2010	11 Apr 2011
Antigua and Barbuda	TIEA	19 Oct 2010	30 May 2012
Bahamas, The	TIEA	09 Apr 2010	12 Dec 2011
Belgium	DTC	21 Jan 2010*	not yet*
Bermuda	TIEA	03 Sep 2009	06 Jun 2012
British Virgin Islands	TIEA	05 Oct 2010	04 Dec 2011
Cayman Islands	TIEA	27 May 2010	20 Aug 2011
Costa Rica	DTC	13 Feb 2014	10 Aug 2016
Cyprus	DTC	18 Feb 2011	16 Dec 2011
Dominica	TIEA	21 Sep 2010	24 Nov 2014
Gibraltar	TIEA	13 Aug 2009	04 Nov 2010
Grenada	TIEA	03 Feb 2011	22 Nov 2013
Guernsey	TIEA	26 Mar 2009	22 Dec 2010
Ireland	DTC	30 May 2011	28 Nov 2012
Isle of Man	TIEA	02 Mar 2009	05 Nov 2010
Jersey	TIEA	04 Jul 2008	28 Aug 2009
Luxembourg	DTC	23 Apr 2012	30 Sep 2013
Malaysia	DTC	23 Feb 2010	21 Dec 2010
Malta	TIEA	17 Jun 2010	19 May 2011
Mauritius	DTC	07 Oct 2011	07 Dec 2012
Monaco	TIEA	27 Jul 2010	09 Dec 2011
Montserrat	TIEA	28 Oct 2011	31 Mar 2012
Saint Kitts and Nevis	TIEA	13 Oct 2010	19 Sep 2016
Saint Lucia	TIEA	07 Jun 2010	28 Feb 2013
Saint Vincent and the Grenadines	TIEA	29 Mar 2010	07 Jun 2011
San Marino	TIEA	21 Jun 2010	23 Dec 2011
Singapore	DTC	28 Jun 2004	12 Dec 2006
Switzerland	DTC	27 Oct 2010*	21 Dec 2011*
Turks and Caicos Islands	TIEA	04 Jun 2010	25 Nov 2011
Uruguay	DTC	09 Mar 2010	01 Jan 2012
2003/48/EC	EUSTD	03 Jun 2003	01 Jul 2005

Note: * indicates signature/enforcement dates of the latest TIE amendment to the DTC, rather than of the original DTC. Sources of data: OECD Exchange of Tax Information Portal and BMF (2017).

In the next section, we outline our data in detail.

1.4 Data

The most important data source for our analysis is monthly data on liabilities to non-residents of banks in Germany, and of German banks' foreign branches and sub-

sidiaries. Unfortunately, the non-resident category comprises the liabilities vis-à-vis foreign enterprises, persons, banks and governments. Data for banks as counterparty are published seperately for short term and long term loans and advances. However, excluding these series from the main series reduces the no. of useful observations by 2/3, pointing to underreporting issues for these interbank loans and deposits. Ideally, we would furthermore like to exclude liabilities to governments, but this is not possible, too.

The data are provided by the Bundesbank in the External Positions of Banks. They are available for different timespans for different countries, so we construct several balanced panels, the main one running from 2003 to 2016. Data coverage and quality are improved from January 2003 onward, because the statistics' scope was extended to all banks operating in Germany. Before, only institutions with external positions above EUR 10 million were required to report. Even though 90% of the volume had been covered before, the sample of reporting banks has broadened to the entire German banking population. This includes domestic banks, their foreign affiliates divided into branches and subsidiaries, and subsidiaries of foreign banks doing business in Germany. The main target of the data thus collected is to inform banking supervision and central banks including the ECB and the BIS, which uses the Bundesbank input to assemble its consolidated and locational banking statistics (Fiorentino et al. 2010).

Since this data is available in numerous sub-categories, it allows us to construct a bilateral control variables using different aggegates. Specifically, we construct two variables capturing the importance of bilateral bank integration based on bank claims in which we do not find tax evasion effects. For the tax havens in our sample, we construct a variable of their relative offshore weight. This variable divides the claims of German banks against counterparties in a specific tax haven by the sum of all claims on counterparty tax havens, or:

$$Offshoreweight_{it} = \frac{claims_{it}}{\sum_{i=1}^{I} claims_{it}} \tag{1.1}$$

where i is the tax haven in question, I is the total number of tax havens and t denotes the respective month. We create the same variable for the non-havens in our sample, that is:

$$Onshoreweight_{jt} = \frac{claims_{jt}}{\sum_{j=1}^{J} claims_{jt}} \tag{1.2}$$

where j denotes the non-haven in question, J the total number of non-havens and t again the respective month. These two variables are then combined so we are able

to use it in specifications including treated tax havens and non-havens. We call this variable 'relative weight' in the outputs.

Table 1.2: Tax haven descriptives, full list, sample period 2003-2016

country	pop pop	deposits mean	deposits st. dev.	ofc weight mean	imports mean	exports mean
Luxembourg	582291	118188	18033	37%	237	406
Switzerland	8179294	53585	9063	12%	2670	3335
Cayman Islands	57268	25592	12129	18%	8	16
Ireland	4952473	20293	6692	16%	1095	415
Hong Kong	7167403	7863	2677	2%	147	425
Singapore	5781728	7639	2237	4%	376	458
Guernsey	66297	6223	4859	1%	3	6
Jersey	98069	3016	2163	2%	5	9
Curacao	149035	2844	2567	0%	0	2
Liechtenstein	37937	2748	638	0%	32	44
Lebanon	6237738	2165	471	0%	3	56
Cyprus	1205575	1767	485	1%	17	58
Bermuda	70537	1278	438	1%	3	13
Bahamas	327316	938	642	0%	8	6
Virgin Islands, British	34232	828	193	0%	3	11
Panama	3705246	815	131	1%	10	23
Bahrain	1378904	790	303	0%	5	29
Jordan	8185384	664	177	0%	2	58
Malta	415196	518	278	1%	30	33
Marshall Islands	73376	487	301	1%	16	16
Liberia	4299944	421	114	1%	20	17
Isle of Man	88195	244	104	1%	4	8

Deposits are liabilities of German banks vis-à-vis non-residents. Deposits, imports and exports are given in millions EUR. Source: Own calculations, based on data from Deutsche Bundesbank, German Federal Statistical Office and CIA World Factbook.

Table 1.2 provides a descriptive overview of the dataset we use to evaluate the policy changes introduced below. The first column shows the population of the respective tax haven, taken from the CIA world factbook. Columns 2 and 3 show means (in million Euro) and standard deviations of bank deposits over the sample length. Column 4 shows the relative offshore weight of the respective tax haven which is highly collinear with but not identical to the deposits (as it is based on relative bank claims). Columns 5 and 6 show imports and exports (in million Euro) respectively. It is immediately obvious that there are some stark discrepancies with German banks reporting 57 billion Euros of liabilities against the Cayman Islands which have roughly 60.000 inhabitants. This bilateral integration, which on average accounts for 18% of the German offshore claims, only receives 16 million Euros worth of imports and provides 8 million Euros worth of imports to Germany. It is also visible, that the German offshore market is quite centralized. Only four jurisdictions (Luxembourg, Switzerland, the Cayman Islands, and Ireland), account for 84% of the relative offshore weight of Germany.

Concerning policy changes, the variables capturing TIEAs are constructed using the OECD database on Exchange of Tax Information Agreements, as well as information of the German Ministry of Finance. The latter also provides information on individual

agreements with tax havens that are related to the EU savings tax directive, on double tax agreements and on protocols or additional arrangements relating to tax information exchange which are included into our analysis. Table table 1.1 above provides an overview of the relevant treaties during the sample period analyzed. These treaties are exploited as dummies or treatment variables in our dataset. The list of tax havens is taken from Gravelle (2015) which is based on several sources and can be regarded as somewhat of a consensus list amongst academic economists.[7]

Our compilation of tax changes is based on summaries provided by the German Federal Ministry of Finance for 1964 until now (BMF 2014, 2015, 2016). We have extracted relevant changes for private households' asset allocation decisions. These regularly include changes of the income tax code, particularly with respect to capital income or income from entrepreneurial occupations that enjoy tax deductability for certain activities. We also consider measures that influence the threat of detection, e.g. via increased power of fiscal authorities to extract bank account information. Corporate tax code changes are usually excluded, except if they likely have a direct influence on income taxes, e.g. in case of dividend payments.

These changes in the tax code are quantified in the following way: We construct dummy variables for tax increases and decreases, and for changes related to the threat of detection, each for both the date they are enacted and the data they enter into force. In case the date of entering into force lies before the date of enactment, the latter is specified as enforcement date as well. This is done, because taxpayers may not adapt their behavior retrospectively.

First and foremost, we expect considerable effects on deposits by foreign counterparties for higher income households, so tax changes aiming for broader parts of society are less likely to be considered than those aiming at sources of income for the wealthy. The former include for instance child benefits or the tax exemption of the subsistence level, the latter the introduction of the withholding tax on interest income.[8] The dataset we have thus constructed runs from 1968 to 2016 and is available from the authors. In the balanced panel employed in the main analysis, we only use the years

[7]It should be noted that any binary measure of whether a jurisdictions qualifies as a tax haven is less than optimal, especially when it comes to havens within large economies that are hence left out of the analysis. More nuanced measures like the Financial Secrecy Index (Cobham et al. 2015) however are not available in sufficient frequency and would require a different econometric setup.

[8]This leaves of course some room for arbitrariness, but we think that it offers at least an approximation to taking into account the extent of tax changes beyond merely noting that some change occurred. Future research may produce more nuanced measurement, e.g. based on the - unfortunately also quite crude and ceteris paribus based - estimations of revenue effects done by the German Ministry of Finance, or by studying anecdotal evidence more profoundly.

in our baseline sample (2003-2016).

Table 1.3: Key tax policy changes in Germany over our sample period (2003-2016)

Policy type	**Enacted**	**In force**	**Effect**
income tax	14 Jul 2000	01 Jan 2003	minimum rate -2.9 %pts, top rate -1.5 %pts
income tax	14 Jul 2000	01 Jan 2005 (revised)	reduction across all income brackets, top rate -5.0 %pts
income tax and others	23 Dec 2003	01 Jan 2004	temporary tax amnesty at reduced rates
threat of detection	23 Dec 2003	01 Apr 2005	better access to domestic account information for fiscal authorities
income tax	29 Dec 2003	1 Jan 2004	early implementation of tax rates planned for 1 Jan 2005
income tax	24 Jul 2006	01 Jan 2007	top rate +3 %pts
local/federal corporate tax, income tax	06 Jul 2007	01 Jan 2008	diverse, in total decrease (e.g. corporate tax rate by 10 %pts)
income tax	06 Jul 2007	01 Jan 2009	new 25 % flat rate withholding tax on capital income
income tax	02 Mar 2009	01 Jan 2009, 01 Jan 2010	rightward shift of tax brackets, i.e. reduction
threat of detection	22 Dec 2014	01 Jan 2015	tightening of rules for self-disclosure of evasion

Note: Own compilation, based on BMF (2014, 2015, 2016).

Alternatively, we use the top statutory personal income tax rate on dividend income as calculated by the OECD. This rate for proceeds from equity securities should be the same for those from debt securities, i.e. interest income, if taxpayers were to abide the law during the baseline sample period. The advantage of taking the tax rate is that there is one simple variable instead of many dummies for different changes, but it comes at the cost of neglecting changes that are difficult to quantify, e.g. closing of tax loopholes or modifications to deductibilities. From the OECD database, we also include the combined and the simple (federal) corporate income tax rates.

A last "policy" change we include are the recent public scandals of prominent Germans evading taxes, which were part of a wave of foreign account information disclosed to German fiscal authorities. Most prominent examples are former Deutsche Post CEO Klaus Zumwinkel and Uli Hoeneß, president of football club FC Bayern Munich, who were caught evading taxes in 2008 and 2013, respectively. After so-called "tax-CDs" with account information of German customers in neighbouring tax havens were bought by German authorities, many tax evaders self-disclosed their misbehaviour out of fear of being detected, with numbers much higher than those included in the original data

files. We identify major events in this regard and use them as dummies in our regression analysis (see table 1.4).

Table 1.4: Major tax evasion scandals over our sample period (2003-2016)

Event	**Publication**	**Effects**
CD with account information from Liechtenstein uncovers tax evasion of wealthy individuals, prominent: Klaus Zumwinkel, former CEO of Deutsche Post	Feb 2008	increased no. of self-disclosures, investigations, fiscal surplus proceeds of around EUR 0.8 bn
fiscal authorities buy CD with Credit Suisse account data, followed by several purchases the same year	Feb 2010	investigations and self-disclosures
Federal Constitutional Court decides that acquiring account information is legal	Nov 2010	authorities can continue paying informants for account data
purchase of another three CDs with Swiss account data (Coutts, UBS)	Jul 2012	investigations and self-disclosures
Uli Hoeneß' evasion is publicised after faulty self-disclosure, another CD is bought, 'Offshore Leaks' (with relatively minor impact on German customers, though)	Apr 2013	Hoeneß resigns and is sentenced for prison, major public discussion about tax evasion of the wealthy, high no. of self-disclosures
'Swiss Leaks', HSBC branch in Geneva	Feb 2015	more than 2,000 accounts with German bearing
'Panama Papers'	Apr 2016	another major public discussion about offshore financial centers

Note: The date refers to when the general public was informed about the event, assuming that potential evaders thinking about self-disclosure did not know earlier. Source: Own compilation.

We include monthly goods trade data as a control variable, provided by the German Statistical Office for the years 2000-2016. Of course, a broader range of monthly controls would be desirable, but it is difficult to obtain data for small countries like island tax havens. For instance, we would like to use services trade data of the Bundesbank as well, but only a limited sample of quarterly bilateral data for too few countries were available. For three British crown dependencies (Guernsey, Isle of Man and Jersey), no data were available even for goods trade. A simple approximation based on population figures of these islands compared to the UK was used to fill the gap[9]. To deal with the unavailability of monthly data, we intrapolate yearly variables from the World Development Indicators of the World Bank to obtain a monthly frequency. These include the birth rate, access to electricity and land lines per 100 persons. Unsurprisingly, the monthly bilatal control variables (exports, imports, and the relative weight) perform much bettter in the estimations than these yearly control variables.

[9]We take the population numbers for each island for the respective years between 2000 and 2015, divide by the respective UK population, and multiply the result with the UK trade figures.

1.5 Results and discussion

As is convention in the literature, we show results based on a fixed effects panel of the form

$$y_{it} = \alpha_i + \gamma_t + \beta_1 Threat * Haven + \beta_2 x_i t + u_i t \tag{1.3}$$

where y_{it} denotes international bank deposits in German banks by country i in year-month t. Year-month (γ_t) and country (α_i) fixed effects are included and x_{it} collects control variables such as bilateral trade, the relative financial weight introduced above and intrapolates macroeconomic control variables. Finally, The $Threat * Haven$ variable is the treatment variable which captures the threat of detection of evasion strategies that depend on a tax haven in which the threat rises due to new tax information exchange arrangements (TIEAs, DTCs, the EUSTD). Reductions of bank deposits in tax havens after TIEAs for examples have been estimated at around 11-16% in Johannesen & Zucman (2014). Hanlon et al. (2015) find much stronger effects, of around 38-45%, in US debt and equity foreign portfolio investment inflows. Also, the effect of the implementation of the EU Savings Tax Directive, which represents a threat of detection in the likely case that tax evasion leads to interest earnings, has been estimated to lead to a roughly 40% reduction in deposits by foreigners in Switzerland (Johannesen 2014).

In line with our identification strategy, we argue that (parts of) the funds or the capital gains that German tax evaders store in tax havens end up back German bank accounts. Since the evasion scheme can still be discovered via the ownership strucure in the tax haven, a threat of detection always remains even if the repatriation move itself is legal. Thus, we expect β_1 to show a strong negative effect of information exchange on bilateral bank deposits in Germany from tax havens. Building on the aforementioned studies, we show effects for TIEAs/DTCs combined with the EU Savings Tax Directive both of which represent danger for evaders.

1.5.1 Baseline

Table 1.5 provides baseline results where y_{it} is set to total bank deposits of German reporting banks versus international counterparties. The first column shows the effect of bank deposits from tax haven counterparties to the introduction (using the signature date) of bilateral information exchange. This effect is strongly negative and highly significant. However, it decreases to roughly 32% when controling for bilateral trade

(column 2) and to 34% when adding the relative offshore weight (column 3). Economically, this result is in line with the effect Hanlon et al. (2015) find for foreign portfolio investment. It is, however much larger than the results in Johannesen & Zucman (2014) who find decreases between 10% and 20% of deposits in tax havens themselves. This is most likely due to our larger sample span: Johannesen & Zucman (2014) carried out their estimation at the height of international signature momentum and did not capture its entire full effects. Both exports and imports are significant and have the expected sign. They show sensible effects: a one percent increase in exports leads to a 0.19% increase in bank deposits in Germany mirroring closer ties between Germany and the trade partner. The relative weight variable, which collects the offshore weight of a tax haven and the onshore weight of a non-haven is also highly significant and strongly positive. Since it is based on bank claims while the outcome variable is based on bank liabilities, a strong positive correlation is to be expected. We hope to capture the integration of the bank sector here that is not connected to tax evasion.

Table 1.5: Baseline results

	Dependent variable:					
	log(deposits)					
	(1)	(2)	(3)	(4)	(5)	(6)
signed TIE with:	tax havens	tax havens	tax havens	non-havens	non-havens	non-havens
	−0.416***	−0.319***	−0.341***	−0.153	−0.171	−0.177
	(0.131)	(0.116)	(0.115)	(0.186)	(0.146)	(0.146)
log(imports)		0.088*	0.087*		0.091*	0.091*
		(0.052)	(0.051)		(0.050)	(0.050)
log(exports)		0.185***	0.185***		0.186***	0.187***
		(0.060)	(0.060)		(0.059)	(0.058)
relative weight			4.337***			4.188***
			(0.887)			(0.913)
country f.e.	Yes	Yes	Yes	Yes	Yes	Yes
year-month f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,952	16,849	16,849	17,952	16,849	16,849
R^2	0.050	0.103	0.112	0.043	0.100	0.108
Adjusted R^2	0.035	0.088	0.097	0.028	0.086	0.094

*p<0.1; **p<0.05; ***p<0.01

Note: Autocorrelation and heteroskedasticity robust standard errors in parenthesis. Log(deposits) refers to liabilities of German banks vis-à-vis non-residents. Column 1 shows reactions to information exchange with tax haven countries. Column 2 adds the goods trade statistics to control for real bilateral integration. Column 3 adds the relative financial trade variable introduced in the data section. Columns 4-6 repeat the same exercise for a threat-of-detection dummy interacted with non-haven countries.

1.5.2 Robustness

To show that we indeed establish a tax evasion effect here, we run a number of falsification exercises. Columns 4-6 repeat the same specifications but interact the treaty

variable with non-haven countries instead of tax havens. Non-haven economies, especially large ones, typically have double taxation conventions in place or other means to exchange information. They also do not specialize in providing financial services for tax evaders. For these reasons, we expect no effect to TIEAs and new DTCs including information exchange here and, indeed, all columns show economically and statistically insignificant results. This exercise also shows that we are not picking up particularities of the German banking sector in our tax evasion estimates. Instead, the fact that we exclusively find significant results for tax haven deposits reacting to information exchange strongly suggests that these positions include evaded capital. As outlined above, a depositor who does not fear detection should not react and, as columns 4-6 show, does not.

Table 1.6: Falsification using short-term claims

	Dependent variable:					
	log(short term claims on banks)					
	(1)	(2)	(3)	(4)	(5)	(6)
signed TIE with:	tax havens	tax havens	tax havens	non-havens	non-havens	non-havens
	−0.125	−0.050	−0.137	0.378	0.324	0.307
	(0.360)	(0.356)	(0.350)	(0.266)	(0.234)	(0.231)
log(imports)		0.111***	0.105**		0.111**	0.107**
		(0.043)	(0.043)		(0.047)	(0.045)
log(exports)		0.576***	0.585***		0.567***	0.576***
		(0.160)	(0.148)		(0.155)	(0.144)
relative weight			11.680***			11.510***
			(1.620)			(1.602)
country f.e.	Yes	Yes	Yes	Yes	Yes	Yes
year-month f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,767	14,952	14,952	15,767	14,952	14,952
R^2	0.047	0.097	0.125	0.052	0.101	0.128
Adjusted R^2	0.029	0.081	0.109	0.035	0.085	0.112

*p<0.1; **p<0.05; ***p<0.01

Note: Autocorrelation and heteroskedasticity robust standard errors in parenthesis. Log(short term claims on banks) indicates claims held by German banks against non-German bank counterparties. Column 2 adds the goods trade statistics to control for real bilateral integration. Column 3 again adds the relative financial trade variable introduced in the data section. Columns 4-6 repeat the same exercise for a threat of detection dummy interacted with non-haven countries.

There is a concern that an indirect identification strategy as the one employed here captures other movements in bank positions that are not connected to tax evasion. In order to rule out this possibility, table 1.6 shows results for running our estimation on short term claims against bank counterparties as a measure of the interbank market. We are not aware of individual tax evasion schemes which would show up in this data. Indeed, although the model is still able to explain these positions with the same precision and signs and significances of the bilateral control variables do not change,

our results on information exchange disappear entirely when looking at short term claims against bank counterparties.

In order to test our results for sample effects, we try to eliminate potential effects of the financial crisis which wrecked havoc not only on the international economy but on most financial statistics as well. Appendix table 1.9 reports the results for both tax havens (to show the robustness of the evasion effect) and non-havens (to show the robustness of the non-effect in the control group). Columns 1 and 3 drop the 18 months of the financial crisis from mid 2007 to end of 2008 from the sample. Our results are virtually unchanged by this modification: the evasion effect is stronger if anything and still significant. Columns 2 and 4 cut off all data before 2009 and test our effects exclusively in a post crisis sample. We observe a decrease in both economic magnitude and statistical significance but can still confirm the results. For the non-haven control group, there is again no effect which makes us confident that our identification strategy is not affected by the financial crisis or our sample length specification[10].

Results on information exchange are thus quite robust and falsification exercises suggest that we have indeed identified a tax evasion effect and not a more general effect in the bank system. Still, these effects could be driven by macroeconomic trends in tax havens that affect individual bank deposits but do not effect the intra-bank market or bank claims. In order to control for such effects, appendix table 1.10 inlcudes a number of yearly macroeconomic variables that have been employed in the literature (see for example Hanlon et al. (2015)). We sucessively control for the percent of the population that has electricity access, the birthrate, the population, the growth rate of GDP/capita and the number of landlines per 100 inhabitants. All resulting specifications show almost identical effects of international information exchange. The slight changes that are visible are due to the sample reduction that comes with the limited availability of such data. To show this, we provide results in a sample reduced to those observations where all data is available but use an estimation excluding the macroeconomic variables. Results are almost identical which indicates either that there is no hidden bias due to macroeconomic conditions or that the data we have access to is to broad to capture it. Therefore, we use the model including exports, imports and the relative weights as the baseline.

[10]Further robustness is provided by checking whether other flows like short-term claims react to changes in TIE as well. Figure 1.3 and table 1.6 show that they do not, confirming our analysis.

1.5.3 Further analysis

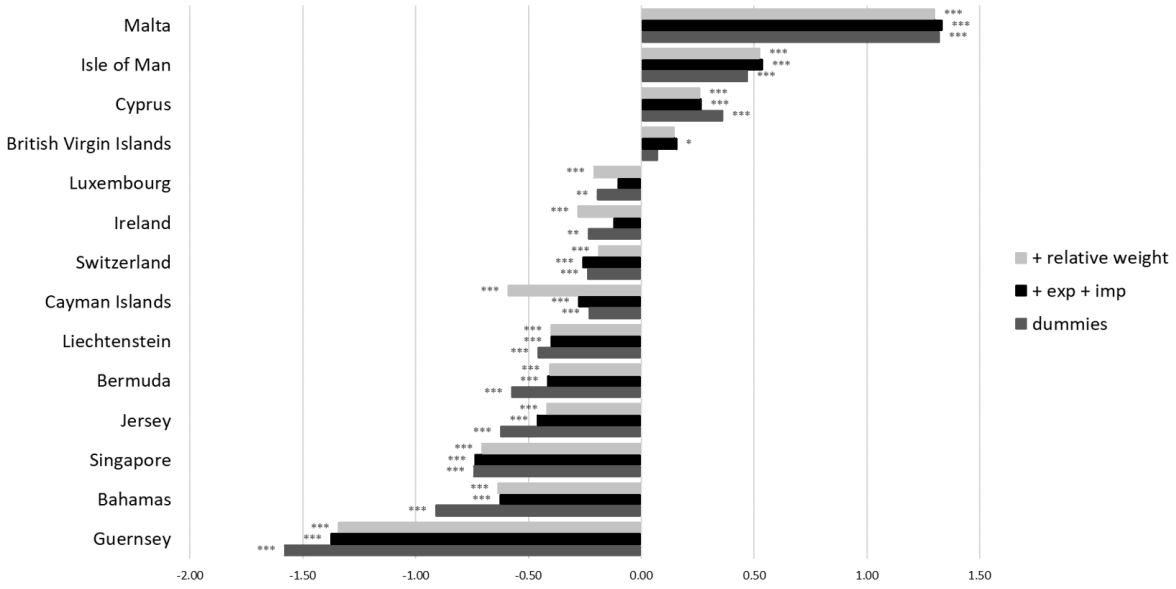
The bilateral nature of the deposit data allows us to take a detailed look at the composition of the effect of information exchange over the tax havens which have signed a treaty and for which we have data. Appendix table 1.11 and Figure 1.3 provide results for each of these tax havens seperately. This allows us to answer the question which tax haven is most implicated in evasion schemes of tax evaders responding to a threat of detection in Germany. Results (green bars) show first that deposits in Germany are reduced for the majority of tax havens. Guernsey, the Bahamas, Singapore, Jersey, Bermuda, Liechtenstein, the Cayman Island, Switzerland, Ireland and Luxembourg all show a large and highly significant effects[11] The red and blue bars add bilateral control variables as before, confirming results. Also, the relative impact of information exchange varies heavily. Guernsey, the Bahamas, Singapore and Jersey stand out with particularly strong reactions. However, these effects are to be interpreted in percent of deposits from that respective tax haven. The 26% decrease for Switzerland with mean deposits of EUR 53,720 million is economically much more meaningful (for Germany) than the 62.8% for the Bahamas with only EUR 962 million of mean deposits in Germany.

Interestingly, Cyprus, the Isle of Man and Malta show a positive effect while the effect for the British Virgin Islands is insignificant. The positive effect is puzzling since both Malta and the Isle of Man are widely regarded to be exemplary tax havens. Our aggregated data makes a further analysis of these cases difficult.A diseggregated analysis showing which sector(s) drive(s) these results could help to explain this development and is a promising venue for future research. Our best guess is that these jurisdictions are specialized in different aspects of the offshore industry and focus legal avoidance schemes rather than on the illegal evasion cases that we analyse. Malta's business model for instance includes attracting wealthy citizens to change their domicile, generous tax refunds for multinational enterprises or tax savings for yacht owners using a Maltese corporations.

As evasion becomes more and more risky, more evaders might opt for avoidance instead, which may explain the increase in Malta in spite of information exchange. A similar story might hold for Cyprus, which also seeks to legally attract wealthy

[11]Note that the OECD applies a peer review procedure to monitor whether treaties are properly drafted and enforced. In case of the DTC signed between Germany and Singapore in 2004, the assessment is negative, so results that show a strong significant effect for treaty signature could pick up other developments and should be treated with caution. The baseline results are robust to excluding Singapore.

Figure 1.3: Differential effects over tax havens



Source: Own calculations. For details, see Table 1.11 in the appendix.

individuals but was struck by a financial crisis in 2012, hence the positive effect is smaller. Lastly, the Isle of Man signed its TIEA around the same time it offered a new structure to be in compliance with the EU Savings Tax Directives. According to anecdotal evidence, this secured a competitive advantage over other jurisdictions, particularly Guernsey and Jersey, potentially explaining the rise of liabilities towards the Isle of Man. With the sectorally aggregated data available, we cannot confirm these conjectures empirically, however.

Having shown an effect of increased threat of detection, we now turn to the payoff side: the benefits of evasion. Theoretically, we should expect higher deposits of tax havens in Germany when taxes in Germany are raised as the incentive to evade increases. Some of that evaded capital will, as we argue in the identification section, end up back in Germany. Lower taxes should instead lead to lower deposits as the incentive to evade decreases. Moreover, the domestic threat of detection captures laws that increase penalties for evasion or change the probability to get caught, like expanded competences of German fiscal authorities to access German account information. However, table table 1.7 shows that neither effect is visible in the data. Column 1 shows that changes in the statutory top personal income tax rate on capital returns have no statistically significant effect on liabilities towards tax havens, and neither do the events we think change the domestic threat of detection (column 2).

We therefore are unable to confirm the results of Hanlon et al. (2015) concerning

Table 1.7: Results of tax law changes

	Dependent variable:	
	log(deposits)	
	tax havens	tax havens
	(1)	(2)
Top PIT rate on capital income	0.595	
	(0.725)	
Better access to domestic account information		−0.076
		(0.138)
EU Directive Administrative assistance		−0.092
		(0.072)
Domestic adaptation to FATCA and CRS		0.021
		(0.049)
Tightening of self-disclosure		−0.108
		(0.094)
log(imports)	0.088*	0.087*
	(0.051)	(0.051)
log(exports)	0.187***	0.188***
	(0.060)	(0.060)
relative weight	4.128***	4.093***
	(0.872)	(0.937)
country f.e.	Yes	Yes
year-month f.e.	Yes	Yes
Observations	16,849	16,849
R^2	0.107	0.109
Adjusted R^2	0.092	0.094
	*p<0.1; **p<0.05; ***p<0.01	

Note: Autocorrelation and heteroskedasticity robust standard errors in parenthesis. Log(deposits) refers to liabilities of German banks vis-à-vis non-residents. Column 1 shows reactions to the top personal income tax rate on capital income, including controls for goods trade and the relative financial trade variable introduced in the data section. Column 2 does the same for a range of domestic policy changes implemented as temporary dummy treatment variables.

tax effects. This might be due to relatively few tax changes in our sample period. Alternatively, it may indicate that changes of income taxes by a couple of percentage points do not sufficiently influence the decision whether to evade taxes to establish empirical effects. The difference is still small when compared to zero taxes paid in the stereotypical tax haven. Unfortunately, the biggest change in our sample period, the introduction of the flat withholding tax on capital income, was legislated in July 2007 and came into effect in January 2009, i.e. during the financial crisis which may bias results. Moreover, it should be noted that Hanlon et al. (2015) analyse a somewhat different evasion model using foreign portfolio investments.

We then test variables capturing major events of the recently bought "tax-CDs" during a 12 month window after the respective scandal went public. The coefficients are insignificant, but show the expected sign, i.e. the liabilities towards tax havens decrease. It may be worthwhile to study effects for a subsample of havens that were

particularly affected by these scandals, i.e. neighbouring offshore destinations like Switzerland and Luxembourg. In future work we hope to study these effects in more detail and test these marginal threats of detection in subsamples of tax havens.

1.6 Conclusion

Using a monthly dataset of bilateral deposit data in German banks, we show that positions of tax haven counterparties in Germany decrease by roughly 32-34% if bilateral information exchange becomes possible. We rely on an identification strategy that builds on the stylized fact of tax evaders shifting capital into tax haven jurisdictions. This capital returns to Germany and does so at a reduced rate when information exchange is enabled. Thus, we interpret these effects as reactions of tax evaders to a threat of detection.

A number of falsification exercises confirm that this effect is not driven by omitted factors. First, we find no effects for the same deposits originating from non-haven countries. We also find no effects whatsoever in bank data that can not be connected to tax evasion: short term bank to bank claims from Germany to offshore centers do not react to information exchange. Our results confirm those of previous studies in different samples and for different countries both in direction and in magnitude (Hanlon et al. 2015, Johannesen 2014, Menkhoff & Miethe 2017) which makes us confident in our choice of Germany as a valid case study to study international tax evasion in more depth. Consequently, we exploit a rich dataset of policy changes, leaks and regulations in Germany to take full advantage of the monthly bilateral deposit data available for this country.

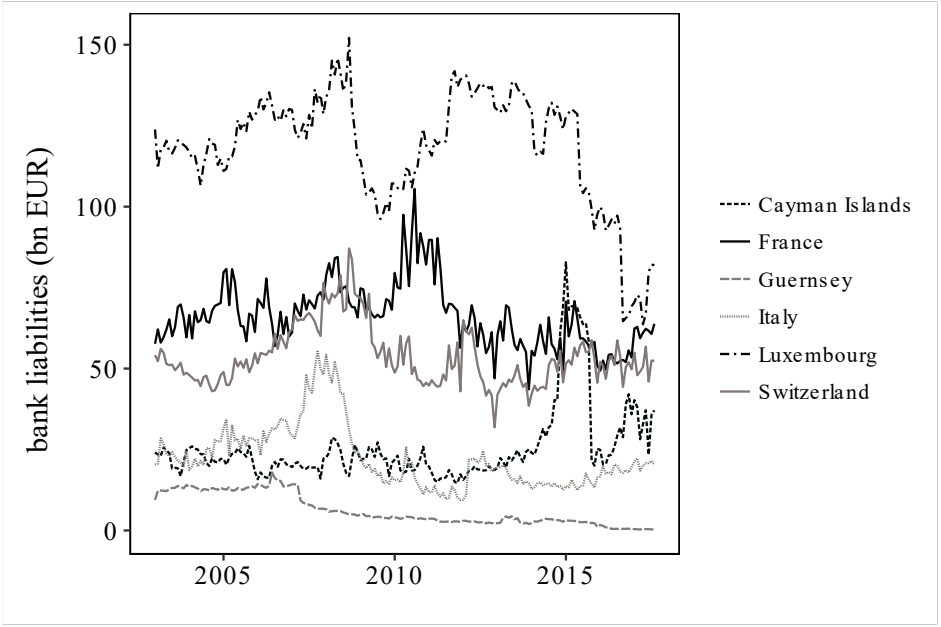
We first provide results for disaggregated reactions for a list of tax havens which have signed information exchange agreements. Amongst the tax havens which reduce deposits, Guernsey, the Bahamas and Jersey stand out with particularly strong effects. Counterintuitively, we find positive results for three tax havens that are of minor quantitative importance for German offshore money, though: Malta, Cyprus and the Isle of Man. We provide some anecdotal evidence to explain these effects but further research is warranted to support such evidence. Additional insights may be drawn from studying a longer time horizon, which unfortunately comes along with a smaller sample of tax havens with available data.

When investigating German policy changes, we are not able to confirm the results in Hanlon et al. (2015) on tax policy: all measures of tax changes and tightening of

rules employed by us remain insignificant. We interpret this as evidence for a very low elasticity of deposit shifting in response to tax changes. Additionally, we test a number of recent tax evasion scandals in Germany. While all of them show negative reactions, significance is at best marginal. This confirms information sharing as the main lever of regulatory attempt to curb tax evasion. Future research may improve on this analysis by disaggregating the analysis for the most relevant tax havens impacted by the respective "tax-CD" revealing account information of banks in its jurisdiction.

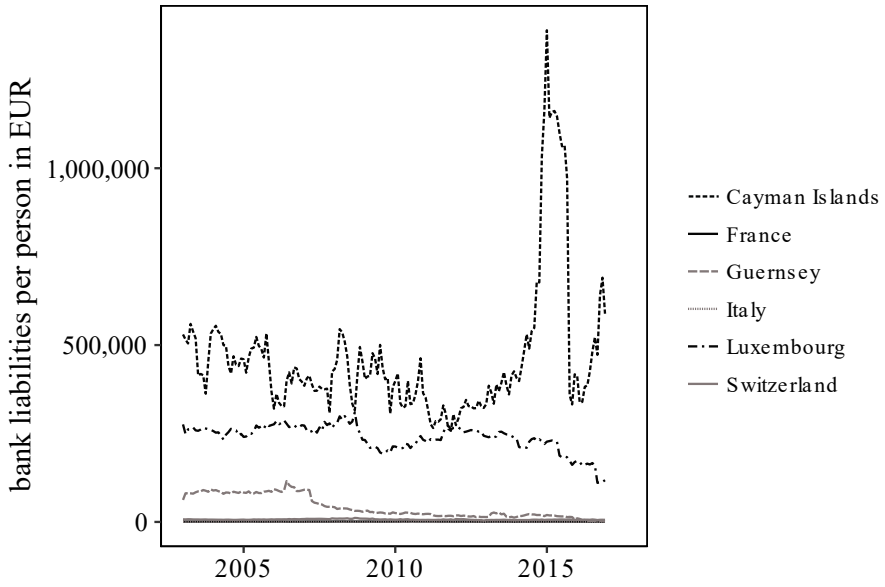
1.A Appendix

Figure 1.4: Liabilities, selected countries (2003:1-2016:8)



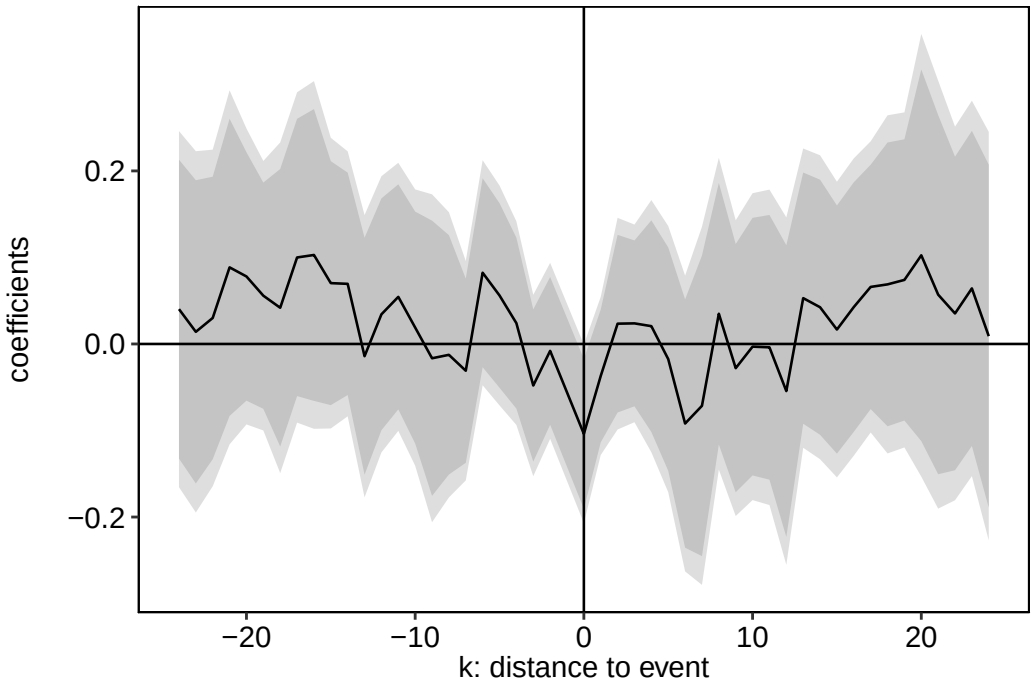
Source: Own representation, based on Bundesbank data.

Figure 1.5: Liabilities/population, selected countries (2003:1-2016:8)



Source: Own representation, based on Bundesbank data.

Figure 1.6: Treatment analysis: Hoeneß 2013



Source: Own representation.

Table 1.8: Results of public scandals ("tax-CDs")

	Dependent variable:		
	log(deposits)		
	tax havens	tax havens	tax havens
	(1)	(2)	(3)
Zumwinkel 2008		−0.053	−0.044
		(0.074)	(0.080)
1st CD wave 2010		−0.029	0.078
		(0.086)	(0.096)
Fed. Const. Court verdict 2010		−0.114	−0.020
		(0.094)	(0.092)
2nd CD wave 2012		−0.117	0.010
		(0.102)	(0.093)
Hoeneß2013		−0.163*	−0.062
		(0.099)	(0.077)
Swiss Leaks 2015		−0.196	−0.079
		(0.160)	(0.130)
Panama Papers 2016		−0.324	−0.237
		(0.209)	(0.190)
signed TIE with tax haven	−0.341***		−0.310***
	(0.115)		(0.103)
log(imports)	0.087*		0.087*
	(0.051)		(0.051)
log(exports)	0.185***		0.186***
	(0.060)		(0.060)
relative weight	4.337***		4.283***
	(0.887)		(0.923)
country fixed effect	Yes	Yes	Yes
year-month fixed effects	Yes	Yes	Yes
Observations	16,849	17,952	16,849
R^2	0.112	0.046	0.113
Adjusted R^2	0.097	0.030	0.098

*p<0.1; **p<0.05; ***p<0.01

Note: Autocorrelation and heteroskedasticity robust standard errors in parenthesis. Log(deposits) refers to liabilities of German banks vis-à-vis non-residents. Column 1 shows our baseline specification, column 2 shows results on dummies taking value 1 for the first 12 months starting with the date of the leak. Column 3 adds the baseline specification to these dummy variables.

Table 1.9: Robustness to sample specification

	Dependent variable:			
	log(deposits)			
	(no crisis)	(post crisis)	(no crisis)	(post crisis)
	(1)	(2)	(3)	(4)
signed TIE with:	tax havens	tax havens	non-havens	non-havens
	−0.352***	−0.254***	−0.174	0.0002
	(0.126)	(0.095)	(0.147)	(0.170)
log(imports)	0.094*	0.070**	0.097*	0.072**
	(0.055)	(0.033)	(0.054)	(0.032)
log(exports)	0.194***	0.144**	0.197***	0.144**
	(0.062)	(0.061)	(0.060)	(0.061)
relative weight	4.262***	3.599***	4.141***	3.427***
	(0.858)	(0.580)	(0.905)	(0.576)
country f.e.	Yes	Yes	Yes	Yes
year-month f.e.	Yes	Yes	Yes	Yes
Observations	15,045	10,737	15,045	10,737
R^2	0.114	0.074	0.111	0.072
Adjusted R^2	0.099	0.056	0.095	0.053

*p<0.1; **p<0.05; ***p<0.01

Note: Autocorrelation and heteroskedasticity robust standard errors in parenthesis. Log(deposits) refers to liabilities of German banks vis-à-vis non-residents. Column 1-2 shows reactions to information exchange with tax haven countries in our preferred specification. This includes controls for goods trade and the relative weight variable introduced in the data section. Columns 3-4 repeat the same exercise for a threat of detection dummy interacted with non-haven countries.

Table 1.10: Infeasibility of control variables

	Dependent variable:				
	tax-havens	tax-havens	log(deposits) tax-havens	tax-havens	tax-havens
	(1)	(2)	(3)	(4)	(5)
TIE signed with haven	−0.341***	−0.316***	−0.299***	−0.319***	−0.305***
	(0.115)	(0.114)	(0.116)	(0.122)	(0.118)
log(imports)	0.087*	0.079	0.081*	0.065	0.073
	(0.051)	(0.050)	(0.047)	(0.049)	(0.053)
log(exports)	0.185***	0.159***	0.143**	0.109**	0.115**
	(0.060)	(0.062)	(0.063)	(0.055)	(0.056)
relative weight	4.337***	3.715***	3.547***	3.005***	3.255***
	(0.887)	(1.117)	(1.140)	(1.004)	(0.989)
% electr. access		−0.005	−0.006	−0.010	
		(0.015)	(0.015)	(0.016)	
log(birthrate)			0.414	0.167	
			(0.602)	(0.614)	
log(population)			0.923**	0.914**	
			(0.430)	(0.436)	
growh GDP/capita				−0.004	
				(0.003)	
log(landlines per 100)				0.334**	
				(0.150)	
country f.e.	Yes	Yes	Yes	Yes	Yes
year-month f.e.	Yes	Yes	Yes	Yes	Yes
Observations	16,849	14,160	14,021	13,375	13,375
R^2	0.112	0.103	0.114	0.120	0.096
Adjusted R^2	0.097	0.087	0.098	0.103	0.079

*p<0.1; **p<0.05; ***p<0.01

Note: Autocorrelation and heteroskedasticity robust standard errors in parenthesis. The dependent variable log(deposits) refers to liabilities of German banks vis-à-vis non-residents. Column 1 repeats the baseline, columns 2 - 4 adds the controlvariables indicated thus decreasing the sample size as visible in the number of observations. Column 5 takes the smalles sample thus reached and estimates the baseline without macroeconomic control variables.

Table 1.11: Disaggregation of tax havens

	Dependent variable:		
	log(deposits) signatures of specific havens		
	(1)	(2)	(3)
Guernsey	-1.580^{***}	-1.378^{***}	-1.343^{***}
	(0.082)	(0.078)	(0.076)
Bahamas	-0.911^{***}	-0.628^{***}	-0.637^{***}
	(0.074)	(0.189)	(0.188)
Singapore	-0.744^{***}	-0.738^{***}	-0.705^{***}
	(0.082)	(0.073)	(0.072)
Jersey	-0.623^{***}	-0.463^{***}	-0.419^{***}
	(0.082)	(0.076)	(0.074)
Bermuda	-0.575^{***}	-0.416^{***}	-0.408^{***}
	(0.075)	(0.083)	(0.082)
Liechtenstein	-0.458^{***}	-0.401^{***}	-0.399^{***}
	(0.075)	(0.070)	(0.069)
Cayman Islands	-0.232^{***}	-0.279^{***}	-0.590^{***}
	(0.082)	(0.077)	(0.111)
Switzerland	-0.237^{***}	-0.260^{***}	-0.189^{***}
	(0.073)	(0.067)	(0.068)
Ireland	-0.235^{***}	-0.122	-0.280^{***}
	(0.083)	(0.077)	(0.089)
Luxembourg	-0.194^{***}	-0.104	-0.211^{***}
	(0.070)	(0.064)	(0.068)
British Virgin Islands	0.073	0.158^{*}	0.147
	(0.082)	(0.093)	(0.092)
Cyprus	0.361^{***}	0.267^{***}	0.260^{***}
	(0.083)	(0.090)	(0.090)
Isle of Man	0.470^{***}	0.539^{***}	0.527^{***}
	(0.082)	(0.076)	(0.077)
Malta	1.323^{***}	1.334^{***}	1.301^{***}
	(0.083)	(0.072)	(0.073)
log(imports)		0.085	0.084
		(0.054)	(0.054)
log(exports)		0.184^{***}	0.186^{***}
		(0.060)	(0.060)
relative weight			4.350^{***}
			(0.915)
country f.e.	Yes	Yes	Yes
year-month f.e.	Yes	Yes	Yes
Observations	17,952	16,849	16,849
R^2	0.063	0.114	0.123
Adjusted R^2	0.048	0.099	0.108

*p<0.1; **p<0.05; ***p<0.01

Note: Autocorrelation and heteroskedasticity robust standard errors in parenthesis. The dependent variable log(deposits) refers to liabilities of German banks vis-à-vis non-residents. The country names indicate a country dummy multiplied by the threat-of-detection variable used in the rest of the text. Column 1 shows results for only the resulting dummy variables. Column 2 adds the log of exports and imports of goods and columnd 3 adds the relative financial weights as discussed in the main text.

Table 1.12: Tax Havens Germany can exchange information with

Counterparty	**Type**	**Signature**	**In force**
Andorra	TIEA	25 Nov 2010	20 Jan 2012
Anguilla	TIEA	19 Mar 2010	11 Apr 2011
Antigua and Barbuda	TIEA	19 Oct 2010	30 May 2012
Austria	DTC	10 Mar 2000	31 Oct 2002
Bahamas, The	TIEA	09 Apr 2010	12 Dec 2011
Belgium	DTC	11 Apr 1967	30 Jul 1969
Belgium	DTC	21 Jan 2010*	not yet*
Bermuda	TIEA	03 Sep 2009	06 Jun 2012
British Virgin Islands	TIEA	05 Oct 2010	04 Dec 2011
Cayman Islands	TIEA	27 May 2010	20 Aug 2011
Costa Rica	DTC	13 Feb 2014	10 Aug 2016
Cyprus	DTC	18 Feb 2011	16 Dec 2011
Dominica	TIEA	21 Sep 2010	24 Nov 2014
Gibraltar	TIEA	13 Aug 2009	04 Nov 2010
Grenada	TIEA	03 Feb 2011	22 Nov 2013
Guernsey	TIEA	26 Mar 2009	22 Dec 2010
Ireland	DTC	30 May 2011	28 Nov 2012
Isle of Man	TIEA	02 Mar 2009	05 Nov 2010
Jersey	TIEA	04 Jul 2008	28 Aug 2009
Liberia	DTC	25 Nov 1970	01 Jan 1975
Luxembourg	DTC	23 Apr 2012	30 Sep 2013
Malaysia	DTC	23 Feb 2010	21 Dec 2010
Malta	DTC	08 Mar 2000	27 Dec 2001
Malta	TIEA	17 Jun 2010	19 May 2011
Mauritius	DTC	07 Oct 2011	07 Dec 2012
Monaco	TIEA	27 Jul 2010	09 Dec 2011
Montserrat	TIEA	28 Oct 2011	31 Mar 2012
Saint Kitts and Nevis	TIEA	13 Oct 2010	19 Sep 2016
Saint Lucia	TIEA	07 Jun 2010	28 Feb 2013
Saint Vincent and the Grenadines	TIEA	29 Mar 2010	07 Jun 2011
San Marino	TIEA	21 Jun 2010	23 Dec 2011
Singapore	DTC	28 Jun 2004	12 Dec 2006
Switzerland	DTC	11 Aug 1971	29 Dec 1972
Switzerland	DTC	27 Oct 2010*	21 Dec 2011*
Trinidad and Tobago	DTC	4 Apr 1973	28 Jan 1977
Turks and Caicos Islands	TIEA	04 Jun 2010	25 Nov 2011
Uruguay	DTC	09 Mar 2010	01 Jan 2012
2003/48/EC	EUSTD	03 Jun 2003	01 Jul 2005

Note: * indicates signature/enforcement dates of the latest TIE amendment to the DTC rather than of the original DTC. Source of data: OECD Exchange of Tax Information Portal and BMF (2017).

Table 1.13: Tax policy changes in Germany over our sample period (2003-2016)

Policy type	**Enacted**	**In force**	**Effect**
income tax	14 Jul 2000	01 Jan 2003	minimum rate -2.9 %pts, top rate -1.5 %pts
income tax	14 Jul 2000	01 Jan 2005 (revised)	reduction across all income brackets, top rate -5.0 %pts
income tax, corporate tax	16 May 2003	16 May 2003	restriction of loss reduction for dormant partners, other small increases concerning CFC tax rules
income tax, investment tax	15 Dec 2003	15 Dec 2003	diverse decreases, modernisation
income tax	22 Dec 2003	22 Dec 2003	closing of loopholes
income tax and others	23 Dec 2003	01 Jan 2004	temporary tax amnesty at reduced rates
threat of detection	23 Dec 2003	01 Apr 2005	better access to domestic account information for fiscal authorities
income tax	29 Dec 2003	1 Jan 2004	early implementation of tax rates planned for 1 Jan 2005
income tax	30 Dec 2005	30 Dec 2005	limitation of loss adjustments for tax deferral models
income tax	28 Apr 2006	28 Apr 2006	closing of loopholes
income tax	24 Jul 2006	01 Jan 2007	top rate +3 %pts
local/federal corporate tax, income tax	06 Jul 2007	01 Jan 2008	diverse, in total decrease (e.g. corporate tax rate by 10 %pts)
income tax	06 Jul 2007	01 Jan 2009	new 25 % flat rate withholding tax on capital income
income tax	21 Dec 2008	01 Jan 2009	temporary reintroduction of degressive depreciation
income tax	02 Mar 2009	01 Jan 2009, 01 Jan 2010	rightward shift of tax brackets, i.e. reduction
income tax, inheritance tax, land transfer tax	26 Jun 2013	30 Jun 2013	closing of loopholes
threat of detection	26 Jun 2013	01 Jan 2015	national rules for automatic TIE within the EU
threat of detection	28 Dec 2013	29 Dec 2013	adaptation to upcoming international automatic TIE
threat of detection	22 Dec 2014	01 Jan 2015	tightening of rules for self-disclosure of evasion

Note: The table includes all changes we incorporated into the respective variables for major/minor increases/decreases and the domestic threat of detection. Source: Own compilation, based on BMF (2014, 2015, 2016).

Chapter 2

On income tax avoidance - a new micro data model for the German case

2.1 Introduction

As inequality has risen in many countries worldwide[1], and as governments have faced the need to raise revenues during and after the Financial Crisis of 2008/2009, tax evasion and avoidance have become topics of rising interest for researchers and policy-makers alike. In particular, empirical studies by Kleven et al. (2011), Zucman (2013), Johannesen & Zucman (2014), Hanlon et al. (2015) or Alstadsaeter et al. (2018*a*), to mention just a few, have contributed to increased academic notice of these phenomena. On the policy side, the negotiation and implementation of the international Common Reporting Standards for automatic exchange of tax information is regarded as a milestone in the combat against illegal tax shelters.

Studying tax avoidance and evasion is fraught with difficulties. The quantity of interest is hidden by definition and agents go to great lengths to conceal it. In order to make progress, indirect estimation techniques have been applied widely in the literature. These include approaches that compare survey with tax data using national accounts, as implemented for instance by Matsaganis et al. (2010), Leventi et al. (2013) or Torregrosa (2015). Another way of indirectly estimating tax avoidance and evasion was established by Pissarides & Weber (1989), who exploit differences between dependly-employed and self-employed tax payers reporting similar consumption expenses but different taxable incomes. This was further refined by Lyssiotou et al. (2004), who estimate a demand-system approach that reduces biases arising from differential consumption patterns of self-employed, and applied to US tax data by Feldman

[1] See wid.world for an overview.

& Slemrod (2007) who make use of deductions for charities. In contrast, a contribution in the German context relies on a Tobit regression to estimate the amount and distribution of tax savings along the income distribution (Lang et al. 1997). Using micro data from the Income and Consumption Survey (EVS) of 1983, they estimate that the effective marginal tax rate for the highest income groups was 16 percentage points lower than legislated. Aggregate income tax losses amounted to 34% of taxes paid.

In contrast to these indirect approaches, the first best option to directly study tax evasion on the micro level are randomised tax audit programmes, at least for the bulk of the income distribution. At the top, it has proven beneficial to rely on leaks of data from offshore banks and other service providers. Examples are the 2013 Offshore leaks, the 2015 Swiss leaks, the 2016 Panama Papers or the 2017 Paradise Papers. Furthermore, valuable insights can be derived from cases of tax fraudsters caught by fiscal authorities or evaders who self-reported their abuse to decrease penalties imposed. The first major study that is able to exploit all types of sources mentioned above, including information on top income earners and wealth owners, is Alstadsaeter et al. (2018*a*), who link these incidents with administrative data for three Scandinavian countries. So far, yet highly desirable, implementing such an approach for Germany is not possible due to data availability.

Instead, inspired by the article of Lang et al. (1997), I use the most recent wave of EVS data from 2013 to estimate tax avoidance along the income distribution. Beyond their Tobit estimation approach, I have programmed a microdata model that seeks to exploit the survey information to the greatest degree feasible. To my knowledge, this estimation method for tax avoidance is new to the literature, made possible due to the rich combination of variables within the EVS. The main contribution consists of empirically modelling provisions of the tax code for 2013 as precisely as possible, given constraints of the data available. I provide the first micro data tax model tailored to the EVS, and hitherto the most-detailed empirical model of the German income tax code in the literature[2]. Hence, I am able to estimate the tax due of individual tax payers within a reasonable degree of certainty, and compare it with the tax actually paid. For this purpose, the quarterly values provided in the survey are adjusted to yearly level, taking into account the frequency of items to mitigate seasonal biases.

My results confirm findings in the literature claiming that tax avoidance and evasion

[2]The main reference is the German section of the EUROMOD tax-benefit model (Gallego-Granados & Harnisch 2017), which however is calibrated for the use with the Socio-Economic Panel SOEP. Therefore, it does not include reported tax payments and is less detailed concerning expenditures. It comes with the advantage of intertemporal variation within a panel dataset, and the ability to study reactions to policy measures, though. The same holds for the Tax-Benefit Microsimulation Model STSM (Steiner et al. 2012).

increase with rising income. My preferred estimate places a lower bound on the amount of tax avoided at 2.8% of taxable income before deductions for the richest decile, which equates to 15.8% of their taxes paid. Unfortunately, due to top-censoring of the data, households with a monthly net income above EUR 18,000 are excluded from the analysis. Nevertheless, my findings are in line with results of Alstadsaeter et al. (2018*a*) which show that wealthy individuals are much more likely to use illegitimate tax planning methods.

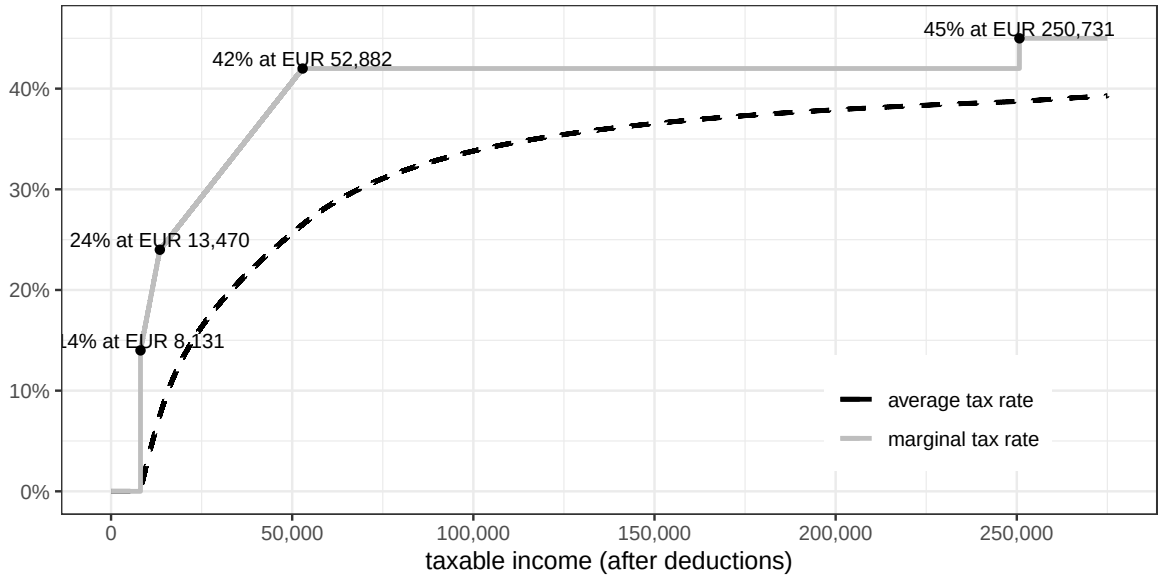
Germany is an interesting case study for tax avoidance for several reasons: It is the largest Euro area economy, its tax code allows for substantial deductions that aim to ensure individual fairness but leave room for semi-legal practices, and it has a long history of tax evasion and avoidance. To my knowledge, for Germany no recent empirical estimates, which are based on verifiable quantitative evidence, of the amount and distribution of legal and illegal income tax savings are available in the literature. The figures of Zucman (2013, 2015) and Alstadsaeter et al. (2018*b*, 2018*a*) pertain to offshore wealth and include estimates of its share for Germany. However, they exclusively capture offshore evasion and no other, less obviously illegal, tax shelters. On the other hand, estimating avoidance using the EVS comes with the grain of salt that the data are top-censored, and that the statements made by individuals in the survey may be biased downwards because people underreport their true income and wealth. Even though one may argue that the underreporting is smaller than for tax data, it is reasonable to expect some degree of it (Moore et al. 2000, Korinek et al. 2006). Hence, much of the (especially offshore) tax evasion probably goes unnoticed by my approach, as it presumably takes place outside of the confines of a household survey conducted by the Federal Statistical Office. As a consequence, my estimates provide for a lower bound of income tax avoidance in Germany in 2013.

The remainder of this chapter is organised as follows: Section 2.2 gives an overview about the institutional setting of the German income tax. Section 2.3 introduces the dataset, while the methodology of my estimation of the compliance gap is presented in section 2.4. Results are given and their robustness is discussed thereafter in section 2.5, followed by the concluding sixth section.

2.2 The German income tax

While its roots date back to Prussian and other German States' tax systems, the modern centralised income tax system was introduced during the Weimar Republic in 1920 (Bach & Buggeln 2020). Compared to pre-WWI levels, the tax rates were increased

Figure 2.1: Statutory German income tax schedule 2013, base rate (single filing)



sharply to cover the costs of war and its aftermath. Throughout the interbellum, and continuing after 1945, the federal income tax remained the backbone of German government finances, even though a decline of its importance can be observed since the 1980s (Corneo 2005). In 2013, 38.9% of tax revenues were generated by the income tax (including flat rate withholding tax on capital income and solidarity surcharge), followed by 31.7% of tax revenues from VAT (Destatis 2014*a*, p. 268).

The formula of the tax, its base and rates are specified in the tax law ("Einkommensteuergesetz"), which is updated continuously. The current system of family taxation with income splitting was introduced in 1958, the last change relevant for my analysis went into effect in 2013: The number of types of tax assessment ("Veranlagungsarten") was reduced from seven to four. Tax payers are assessed each year, based on their nominal taxable income less deductions. In 2013, tax payers enjoyed a basic allowance of 8,130 EUR, up until income went untaxed. Between the basic allowance and the top statutory tax rate, two tax brackets are defined with tax rates increasing linearly within each bracket. The top income threshold for the regular schedule was 52,882 EUR, however taxpayers are charged an additional rate on income exceeding 250,730 EUR, since the so-called "rich tax" was introduced in 2007 (see figure 2.1).

For married couples who are assessed together, the relevant amounts are doubled. They enjoy a lower tax burden if their individual incomes differ substantially, especially in case of the traditional sole breadwinner model. This is because the total tax due is calculated by doubling the tax payable on half of the added up taxable incomes of both partners. Hence, the couple is taxed at a lower point at the progressive tax schedule, paying less than when taxed individually.

The taxable income is determined from seven different types of earnings, which are sorted into the two categories of profit income and surplus income. The former is generated from self-employed activities, the latter from dependent employment, rent, capital and other sources. A plethora of deductions are substracted to get the taxable income, see table 2.1 for an overview. There are very detailed provisions concerning what is deductible by whom under which circumstances[3], which are described in more detail in the technical appendix.Once the final taxable income ("zu versteuerndes Einkommen") is determined and the tax due is calculated, where applicable direct tax reliefs ("Steuerermäßigungen") are substracted from that value to get the final amount of income tax to be paid.

Table 2.1: Calculation of taxable income according to the German tax code, 2013

Profit income	**Surplus income**
Income from self-employment	Income from dep. employment
+ Income from agriculture and forestry	+ Capital income
+ Business income	+ Income from rent and lease
	+ Other income
− related business expenses	− related professional expenses
= **Sum of Revenues**	
− Proportional relief for elderly retired persons	
− Relief for single parents	
= **Total amount of Revenues**	
− Special expenses	
− Extraordinary expenses	
= **Income**	
− Children's allowance	
= **Taxable income** (zvE)	

Source: Own table, based on Dittmann et al. (2014, p. 22).

[3]In short, the bulk of professional expenses are the commuter allowance, travel and educational expenses, expenses for work equipment, the additional meal allowance and expenses for double households. Special expenses are mostly precautionary expenses like contributions for insurances of old age, health and nursing. Moreover, among others they comprise of expenses for eduction of oneself, the spouse or children living away from home, childcare expenses, alimonies, paid church tax, or contributions to political parties and some clubs. Extraordinary expenses are typically case-specific, however some are categorized as well, for instance for people with disabilities, surviving dependants and non-remunerated care providers. Other expenses that fall under the extraordinary ones are sickness costs, costs of nursing homes and services, alimonies or expenses for modifications of buildings due to health condition. Direct tax reliefs are granted for a rather small range of expenses: Craftsmen and domestic work services, nursing homes and services, and membership fees and donations to political parties and independent voter groups.

2.3 Data

My analysis is based on the Income and Consumption Survey (EVS), which is conducted by the Federal Statistical Office of Germany every five years. The cross-sectional quota sample is representative for German private households, because it is adjusted to the sample census (Mikrozensus). However, households with a monthly net income of more than 18,000 EUR are excluded, so are people living in institutions and homeless people. The EVS is the largest household survey of its kind in Europe, multiple quality checks on multiple levels assure consistency (Destatis 2016). Especially, the plausibility of results is checked through a budgeting process that compares income and expenditures of the household, triggering further investigations, if necessary contacting the household again, to resolve differences that are too high (Destatis 2017*b*). However, no information is given concerning the treatment of missing values, which typically are imputed in household surveys during the data processing.

Importantly for my analysis, the survey provides quite detailed information on paid taxes, earnings and expenses, but also on financial assets and wealth. On average, the Federal Statistical Office selects a sample of roughly 80,000 households who answer voluntarily. The drop-out rate including those that stopped participating without even having answered the paper questionnaire was 27%, the drop-out rate during the sampling period was 10.5%. The Scientific-Use-File that is applied in my analysis consists of a subsample of 42,792 household observations.

Nevertheless, the EVS comes with a range of disadvantages as well. First of all, the top-censoring leads to a low coverage of high incomes by construction, and of high wealth as a coincidence. On top of that, the figures may be understated by households below the cut-off as well, who might feel uncomfortable to report their true income and wealth. Particularly problematic when studying tax avoidance are survey issues with self-employed individuals. Because they have more leverage about underreporting (part of) their income to tax authorities, it is important to gather or estimate their actual income as precisely as possible. Unfortunately personal draws, which are a source of income that should be stated, may be diffcult to ascertain, e.g. when private and business wealth cannot be separated easily. Generally, the volatility of self-employed incomes poses a challenge for proper data gathering in surveys, as households can cope more easily with the reporting of steady streams of income (Destatis 2016, p. 4, Becker 2014). As Becker (2014) furthermore notes, due to the voluntary and exclusively German-language sampling procedure the representativeness of the EVS is limited insofar as it inherits a middle-class bias and an undercoverage of households with

foreigners.

Moreover, a major obstacle arises from the surveying period: Each household is interviewed only for one quarter of the year, hence quarterly data is provided to the researcher. This approach was introduced with the 1998 wave of the EVS to decrease the drop-out rate of participants over the year. This is problematic for my analysis for several reasons: To begin with, I need yearly data to estimate the taxable income, because the tax code refers to yearly values. This requires a transformation of the data, which is rather straightforward for items with a high frequency, say a salary or food expenses, that may simply be quadrupled. However, it is much more problematic when thinking about items with a low frequency, like lump-sum payments of bonuses or purchases of durables. Biases may arise when simply quadrupling these values where they occur. Adding to that, the quarterly values also introduce seasonal biases: Single payments, e.g. at the end of the year due to the so-called "christmas bonus", are presumably higher and more frequent for households surveyed in the fourth, than in the first or second quarter. As a consequence, quadrupling these values to get yearly data possibly distorts results, especially as the sample sizes differ by quarter (Q1: n=11,134, Q2: n=11,665, Q3: n=10,379, Q4: n=9,614).

To address these issues, some adjustments have to be made when transforming the data from quarterly to yearly values. This problem was tackled by Bönke et al. (2013) before, but they do the transformation in reverse order: Using the EVS samples of 1978-2003, they establish a harmonised database (pooled cross-sections) at the quarterly level. I.e., they reduce the information of the 1978-1993 cross sections, when households were still surveyed over a full year, to get quarterly data for all years.

Adding to that, some adjustments for particular items were proposed by Becker (2014): For income from self-employment, the previous year's income which fortunately is recorded, is used as a plausibility check to keep the current year's values within reasonable bounds. A similar procedure is applied to capital income.[4]

To get yearly data, I first identify items that are sensitive to biases arising from multiplying them by four (for more details, see the technical appendix). These are items with low frequency that are relevant in the context of my analysis, i.e. for the

[4]If the current year's self-employment income is stated as zero, it is replaced by the previous year's value if that was at least EUR 12,000. If the quadrupled quarterly value deviates by more than 30% from the previous year, 70% of the preceeding year's self-employed income is set for low incomes, 130% for high incomes exceeding EUR 9,000. Also if the current year's capital income is stated as zero, she replaces it by the previous year's value. If the previous year's proceeds from capital are zero, current year receipts are set to the quarterly value stated, assuming no other capital income was generated that year. If both t and $t-1$ show positive amounts, the quadrupled value is restricted to 125% of the previous year's proceeds

tax code. The problem is addressed through two measures.

First, one has to cope with a bias for the number of positive observations in each quarter. For example, the main income earner of the household receives a positive one-time payment in only 6.2% of households surveyed in Q1, but in 37.7% of the households surveyed in Q4. Hence, for the relevant items the households are split by quarter. Next, values for the missing three quarters of each household are imputed using predictive mean matching within a multiple imputation with chained equations [5]. From the implicates thus generated, one is drawn and the subsequent calculations are applied to this dataset[6]. Finally, all quarters are added up to receive yearly values.

Second, the yearly values are weighted by the quarter's mean divided by the mean of the whole sample, to correct for higher or lower levels of payments in some quarters. This seasonal adjustment is done for all items, regardless of frequency.

Additionally, I partially follow Becker's ad-hoc approach by adjusting the specified variables in a similar way. Quadrupled self-employed incomes are restricted within bounds of 50% when deviating positively or negatively from previous year's incomes. Bonuses, lump-sum payments and indemnities are not quadrupled, but replaced by previous year's value when stated as zero. Also if they are more than 10% lower in t compared to $t-1$, the preceeding year's value is pasted. Considering a margin of 20%, the same is done with capital income. Moreover, it is restricted to 150% of the previous year's capital income, as proposed by Becker (2014).

2.4 Modelling the German income tax code

The basic idea of this study is to calculate two income aggregates for each taxpayer i. The taxed income Y_i^{tax} is based on the tax paid T_i, the taxable income Y_i^{all} is estimated using earnings and expenses as stated in the survey. Comparing the two provides may indicate tax savings that may be due to avoidance behaviour. As a next step, comparing paid taxes T_i with the hypothetical estimate of taxes to be paid based on taxable income T_i^{yall}, can be formalised as follows:

$$T_i^{loss} = T_i^{yall} - T_i$$

[5] As my problem is quite unique, suitable approaches are difficult to ascertain given that on average 3/4 of the observations have to be imputed for the affected variables. The latter entails a rather high computational load for many algorithms. Any ideas for further improvement are highly welcome.

[6] For robustness, I have applied a couple of different implicates to the subsequent modelling exercise, and the results changed only slightly. A small potential increase in robustness may be possible by generating more implicates, but does not outweigh the cost of the high computational and programming load this would entail.

To this aim, first the tax function is inverted to calculate the taxable income from taxes paid. Refunds are taken into account by using the previous year's value, corrected for the average growth of refunds between 2012 and 2013[7].

Second and more complicatedly, an estimate for the taxable income is generated from a variety of items. This follows the tax code as described briefly in table 2.1. On the income side, some taxable items are available on the household level only. For households with multiple tax payers, these are allocated among the individuals according to their share of tax payments of that particular household.

It is noteworthy that some deductions are quantitatively much more important than others, but still the highest degree of precision is aspired for all items. According to the tax statistics, for instance the commuter allowance makes up almost 60% of professional expenses for an average employed income earner (Destatis 2015*a*). As the EVS items differ from the tax concepts for many variables, a proper estimation requires additional information to supplement the relevant EVS items. This information is taken mostly from macro statistics, which subsequently contribute to the definition and selection of items for the estimation of deductions.

To illustrate, consider once more the case of the commuter allowance: Households that receive income from employment are entitled to deductions for the commuting distance they have travelled to work. In 2013, the allowance was EUR 0.30 for every completed kilometre one way, compensating the full round trip. In case public transport was used, the tax payer was allowed to claim the full cost as deductions, up to a limit of EUR 4,500 per year. Unfortunately, the EVS does neither contain the distance to work, nor expenses for commuting directly as separate items. As a consequence, one has to estimate them for all groups that may claim the deduction: car and motorcycle drivers, public transport users, bicyclists, pedestrians and carpool users. To get the deductible allowance, several steps are taken:

1. The relevant items in the EVS are identified. These are expenses for car fuels and lubricants (EF383) and third-person transportation services (EF386).

2. A macro estimate for the shares of employed persons commuting with any of the means of transportation is taken from the sample census of 2012, then these shares are applied to the relevant group within the EVS (employed persons) to get absolute numbers. For example, of the 9,256 single household heads with

[7]The same is unfortunately not possible for arrears, which are not given separately but included in the tax payment item. However, these are quantitatively less important because only 1.5 Mio taxpayers had to pay arrears while 11.5 Mio received refunds in 2013 (Destatis 2017*d*), the averages differing not much (EUR 988 and 935, respectively).

dependent employment, 4,397 are estimated to commute using their own car, while 951 use public transport and 616 a bicycle.

3. In the absence of more precise information concerning the distribution of commuters along other covariates (sex, age, income, etc), these absolute numbers of commuters of each type are drawn randomly from the relevant items in case of motorvehicle drivers, public transport users and carpool users. For pedestrians and bicyclists, the kilometres driven are estimated ad-hoc using normal distributions. Without more precise information concerning their travel distance, this approach is justifiable on the grounds of the small ranges covered by typical commuters of this kind.

4. For motorvehicle and public transport commuters, the share of the relevant items which are due to commuting have to be estimated. This is done using macro figures about the shares of purposes of kilometres travelled for different types of transport (Source: Radke 2014). Moreover, for motorized commuters the federal state is taken into account, considering the average journey time to the nearest large city. Also, the size of the community is used to weight the share.

5. Motorvehicle users' kilometres driven are estimated from their gasoline expenses. Therefore, the average price is used to estimate the amount of fuel consumed, taking into account the average kilometres driven by different types of car engines (petrol vs. diesel). From the average fuel consumption per 100km, one can finally get the kilometres driven commuting.

6. Now, the commuter allowance can be computed for all relevant means of transport: The kilometres travelled by motorvehicles, pedestrians and bicycles are simply muliplied by EUR 0.30. The amounts spent by public transport users are claimed under consideration of the maximum limit of EUR 4,500.

In a similar way, items are used to estimate deductible travel expenses for business trips or training courses. Several times, it is necessary to randomly assign some tax payers who report positive expenses for some item to a certain group (e.g., party members), because the items contain mixed information (e.g. party member fees, but also fees for union or club members).

Joint income tax filers, i.e. married couples who opt to be assessed together, are treated as follows: Since the EVS of course contains no information on the type of tax filing, they are first identified by a profitability calculation.[8] Next, the taxable

[8] Joint assessment is the default option for married couples and profitable in most cases, especially

income is estimated as if everyone was single filer. Finally, joint filing is accounted for by applying the same procedures as the tax authority, i.e. assuming that spouse's incomes are added up, then split by half to get virtual individual taxable incomes, which are finally subjected to the standard tax scale (figure 2.1). Adding the two amounts of tax up then provides the couple's common tax due.

Essentially, I estimate a full micro data model of the German Tax Code for 2013 on the individual taxpayer level, accounting for more details than anywhere in the literature to my knowledge. Detailed accounts of both the data preparation and the calculations performed to get the taxable income can be found in the technical appendix. An overview of the provisions that are taken care of can be found in table 2.2.

Table 2.2: Inclusion of tax code provisions in the micro data model

Tax code rule	**Included in the model?**
Taxable income components:	
Income from self-employment (private and material withdrawals, including from agriculture)	yes
Income from sale of solar power	yes, less estimated deductible expenses
Income from dependent employment: salaries, one-time payments, indemnities, bonuses, other employer benefits	yes
Capital-forming benefits for employees (if not used for private/occupational pension)	yes
Base salary when in part-time retirement	yes
Non-cash benefits for employees (if above allowance)	yes
Capital income: interest, dividends, payouts (subject to withholding tax)	yes

when the two incomes differ by a large margin, and the smaller amount is non-negative and not subject to progressivity proviso (Dittmann et al. 2014, p. 185). For the sake of simplicity, I assume that partners file their tax return jointly if one differs by more than 20% from the common total Sum of Revenues (i.e., profit income plus surplus income minus related expenses, see table 2.1). Everyone else is considered a single tax filer.

Capital gains: Sale of real estate, other economic goods, financial assets and company shares	yes, but only for financial assets
Income from rent and lease	yes
Income from subletting, if > EUR 520	yes
Other income	yes
Income from public, civil servant and occupational pensions: only partly taxable, depending on year of entry	yes
Income from private pensions: taxable share depends on type and contract signature year	yes
Professional and business expenses:	
Business expenses (of self-employed)	no, because they only report private expenses[9]
Allowance for honorary work	yes
Capital income allowance	yes
Special allowance for income from agriculture and forestry	yes
Fully deductible expenses of landlords: residential home apportionment for condos, interest on building loans and mortgages, residential building and landowner insurance, other operating expenses	yes
Partly deductible expenses of landlords: acquisition costs, expenses for wear and tear, maintenance and construction costs (multiple conditions apply)	yes
Special depreciation rules for victims of natural disasters	no, not observable
Side costs for sublet rooms, according to m^2 used by subtenant	yes
Expenses allowance for pension income	yes
Professional expenses of employees:	
Flat-rate allowance of EUR 1,000, applied if not claiming higher expenses	yes

[9] As the self-employed are asked to only report private withdrawals (including in-kind) in the survey, I assume that they basically directly report their Sum of Revenues for self-employed income. The direct tax relief for paid local business tax cannot be accounted for, see below.

Costs of a home office (some conditions apply)	no, cannot properly check eligibility
Costs of work equipment: can be deducted fully if < EUR 410, otherwise subject to depreciation over some years	yes, for most relevant items
Commuter allowance	yes, see above
Costs of further eduction, if related to work	yes
Occupational travel expenses, either actually incurred costs or allowance per km	yes, assuming allowance is used
Travel and accomodation costs wrt professional education	yes
Food expenses during business and educational trips, allowance graded by duration of trip	yes
Membership fees for unions and employer (and similar professional) organisations	yes
Motor vehicle repair if work-related	yes
Professional share of costs for tax lawyer services	no, not identifiable
Professional share of costs for accident insurance	yes
Professional share of costs for liability insurance	no, not observable
Costs of moving due to professional reasons	yes
Medical expenses related to occupational diseases and work accidents	yes
Costs of work-related telecommunications	no, not identifiable
Professional hospitality costs	yes
Allowance for professional account fees	yes, for some occupations
Special expenses:	
Standard EUR 36 allowance if not claiming higher expenses	yes
Retirement provision expenses (pension contributions): several conditions apply, e.g. a maximum amount that is cut for some occupations like civil servants	yes
Expenses for health and nursing care insurance, several conditions apply	yes
Unemployment insurance contributions	yes
Private accident insurance contributions	yes

Private share of costs for car liability insurance, fully deductible if commuter allowance is claimed	yes
Contributions for private liability insurance, disability insurance and term life insurance	yes
Contributions for some other capital/pension insurances, some conditions apply	no, not identifiable
Maximum amount for provisional expenses, depending on personal status	yes
Contributions to state-subsidised private pension scheme (Riesterrente)	yes
Contributions to some building societies (Wohnriester)	no, not identifiable
Payments based on pension rights adjustment of split couples, if contractual agreement	no, not observable
Alimony payments when divorced or living permanently seperately, up to EUR 13,805	yes
Expenses for own/spouse's education (some conditions apply, e.g. maximum amount of EUR 6,000): tuition and exam fees, equipment, transportation, interest on student loans	yes
Childcare expenses, up to EUR 4,000 per child	yes
Paid church tax	yes
Membership fees and donations to associations that are churchly, charitable or serving the public good	yes
Membership fees and donations to political parties and registered electoral groups, amount that exceeds the possible tax relief, up to 20% of the Total Sum of Revenues	yes, but only dues
Extraordinary burden:	
Itemised cases for costs of disabled people, nursing of other people; survivor allowance	no, not identifiable
Children's education, when grown up and staying outside of parent HH	no, not identifiable
Home reconstruction cost in special cases (disability, sickness, danger, etc)	no, not identifiable
Health insurance of other supported persons	no, not identifiable

Expenses for nursing homes, if not only there for ageing reasons	no, very few cases[10]
Expenses for ambulant nursing care	yes
Sickness costs	yes
Obligatory supportive payments that are not deductible as special expenses	yes
Voluntary supportive payments	no
Sacrifice restriction for supportive payments (Opfergrenze)	yes
Reasonable own burden for all expenses in this category, depending on Total Amount of Revenues and no. of children	yes
Direct tax reliefs:	
Paid local business tax up to a maximum reduction amount	no[11]
Craftsmen services at home of tax payer, 20% of the costs w/o material and transportation, up to EUR 1,200	yes
Services close to home, 20% of expenses (further conditions apply if only "minor employment")	yes
Expenses for ambulant or stationary nursing, some conditions apply (e.g., deduction as extraordinary burden comes first)	yes
Expenses for moving for private reasons, 20% thereof	yes
Maximum amount of EUR 4,000 direct relief for HH services	yes
Membership dues and donations to political parties and registered voter groups, up to EUR 825 per spouse	yes, but only dues for political parties
Further rules:	
Joint filing of married couples	yes
Proportional tax allowance for elderly retired persons	yes
Tax relief for single parents	yes

[10]This is due to people living in institutions not being captured by the EVS. Also, further conditions cannot be checked with the data.

[11]Unfortunately, local business tax payments are not observed in the survey. Moreover, it is debatable whether this deduction should be substracted here, because businesswomen and self-employed are asked to report private withdrawals only.

Progressivity proviso: some forms of income are tax exempt, but increase the tax rate payed on taxable income	yes
Fifth rule: mitigation of tax progression for high one-time incomes	yes
Check whether child allowance is more profitable than child benefits	yes
Check whether capital income rather be taxed under PIT schedule than with flat-rate withholding tax	yes
Check whether tax deduction rather be used instead of private pension (Riester) state benefit	no
Solidarity surcharge	yes

Own table, for more details see the technical appendix.

2.5 Results and Discussion

Before delving into the details, it should be noted that all results are weighted using the extrapolation factors supplied by the Federal Statistical Office.[12]

2.5.1 Descriptive overview

A descriptive overview about taxable income before deductions and taxed income is presented in table 2.3, further visualized in figure 2.3 in the appendix. Overall, the progressivity of the tax schedule is visible, albeit less so when comparing average tax rates based on taxable income before deductions (column 9) to taxed income (column 10). Clearly, tax payers are able to substantially reduce their taxed income (column 7), but the ratio of taxed to taxable income increases with income. The maximum for the average tax rate is reached at only 17.9% when based on taxable income (24.8% for taxed income). The coefficient of variation for taxed income is much higher than for taxable income, which is to some extent expectable because the data are ordered by taxable income. So far, the results are broadly in line with the findings of Lang et al. (1997) who report a little bit more progressivity. This fits to changes in the German

[12]These weights are adjusted by the Statistical Office using the sample census (Mikrozensus), a yearly 1% cross section sample of all German households. It is very reliable, because participation is obligatory. The adjustment mitigates representativeness issues of the EVS (see section 3).

tax code since 1983, which have mostly reduced the burden for higher incomes, but also increased the basic allowance [13].

However, a peculiarity shows up: Households in the lowest three deciles seem to experience a higher tax burden than expected, which could point to measurement or data preparation errors. For several reasons though, I am convinced that these effects derive partially from the composition of the sample: I have here included all individuals (not households!) that have *some* amount of income that is theoretically subject to tax, no matter how small. Especially, pensioners with only a small taxable component of their income are present in these deciles. Clearly, the incomes therefore recorded as belonging to the first three deciles are even below the basic allowance of EUR 8,130 which tax payers were granted in 2013. The nevertheless slightly positive mean tax payments, respectively, are driven mostly by outliers. They are in part probably artifacts from the transformation from quarterly to yearly variables. Especially the self-employed have to pay taxes based on previous year's income, so they might record tax payments while not having (as high a) taxable income in the same period[14]. The resulting measurement problem is exacerbated by the mere three month observation period during the year.

Concluding, I hold this supposedly too high average income tax burden of the poorest deciles to be largely a statistical artifact rather than a real finding.

Adding to that, table 2.4 shows the percentage of individuals of some taxable income (before deductions) decile with respect to the taxed income decile.[15] Again, it is noteworthy that the first two deciles end up being taxed a lot more than would be expected (see previous paragraph). In principle, as Lang et al. (1997, p. 330) put it, "households in low gross income deciles cannot conceivably reside in much higher taxed income deciles". Even with the spread of more precarious and hence volatile self-employment since the 1983 wave of the survey, this should still hold to some extent.

[13] See Bach et al. (2017) for an estimate of the distribution of the tax burden in Germany.

[14] Indeed, the self-employed account for many of the biggest outliers in the two deciles, the maximum being one tax payer with more than EUR 60,000 in tax payments while showing only some EUR 2,000 of taxable income. That person alone increases the mean tax payments of the decile by EUR 9.

[15] It should be noted that in this table, in contrast to my other results, only individuals with positive taxed income are reported. This is done, because due to many individuals with small amounts of income subject to tax, the first two deciles of the taxable income distribution show little to no tax payments. This is to be expected, as the taxable incomes are lower than the basic allowance of EUR 8,130. In turn, this generates the problem that no boundaries between the first two deciles can be determined, which precludes preparing the desired distribution table.

Table 2.3: Individuals by taxable income decile

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Taxable income decile	Taxable income (Y^{all})			Taxed income (Y^{tax})			Tax payments (T)		
	Decile range (EUR)	Mean (EUR)	Coeff. of var.	Mean (EUR)	Coeff. of var.	Y^{tax}/Y^{all} (%)	Mean (EUR)	T/Y^{all} (%)	T/Y^{tax} (%)
1	[1 – 2819]	1499	0.54	929	0.54	61.94	91	6.07	9.80
2	[2819 – 5637]	4286	0.19	1574	0.19	36.71	156	3.64	9.91
3	[5637 – 8132]	6851	0.10	2235	0.10	32.63	241	3.52	10.78
4	[8132 – 11633]	9780	0.10	4244	0.10	43.39	422	4.31	9.94
5	[11633 – 15683]	13610	0.08	7885	0.08	57.93	878	6.45	11.13
6	[15683 – 20921]	18187	0.08	12584	0.08	69.19	1575	8.66	12.51
7	[20921 – 28022]	24289	0.08	18165	0.08	74.79	2587	10.65	14.24
8	[28022 – 37367]	32433	0.08	24648	0.08	76.00	3963	12.22	16.08
9	[37367 – 52548]	44134	0.10	33432	0.10	75.75	6096	13.81	18.23
10	[52548 –]	85321	0.63	61540	0.63	72.13	15255	17.88	24.79

Note: Table includes only individuals with positive taxable income before deductions, which is calculated as outlined in table 2.1. Source: EVS 2013.

Table 2.4: Percentage of individuals of taxed income decile within taxable income decile

Taxable income	Taxed income										Total
	1	2	3	4	5	6	7	8	9	10	
1	86.62	7.26	1.90	0.68	0.77	0.50	0.41	0.50	0.50	0.91	100%
2	12.92	65.28	13.52	3.35	1.30	0.91	0.75	0.81	0.41	0.75	100%
3	0.37	21.63	53.30	16.08	4.49	1.57	0.58	0.79	0.48	0.66	100%
4	0.08	4.12	23.02	46.88	17.45	3.79	2.36	0.91	0.77	0.70	100%
5	0.06	1.18	4.88	23.29	44.31	18.44	4.14	1.70	1.06	0.91	100%
6	0.00	0.31	1.95	5.34	22.04	43.57	19.31	3.91	1.55	1.97	100%
7	0.00	0.06	0.95	2.73	5.90	22.77	43.67	18.65	3.44	1.84	100%
8	0.00	0.10	0.37	1.20	2.38	5.22	21.96	47.92	17.88	2.96	100%
9	0.00	0.00	0.04	0.31	0.99	2.92	5.32	21.09	55.26	14.05	100%
10	0.00	0.04	0.04	0.12	0.35	0.41	1.43	3.70	18.65	75.26	100%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	

Note: Table includes only individuals with positive taxed income, which is calculated using reported tax payments. Source: EVS 2013.

Apart from that though, the picture is quite comparable to that of the 30-year old data. Individuals in the upper income deciles are able to reduce their tax burden, the richest 10% slightly more than in 1983. Hence, they end up in taxed income deciles lower than their taxable income decile, as can be witnessed in the lower off-diagonal triangular. This indicates substantial horizontal inequalities, as some people sharing similar incomes end up paying different rates of tax. Whether or not these inequities are justified when investigating the deduction possibilities, is analysed in further detail in the upcoming paragraphs.

2.5.2 Outcomes of the tax model

The main estimation sets out to measure tax avoidance by comparing a hypothetical estimate of taxable income with taxed income. On the way, the components of taxable income before deductions are calculated, as are deductions and sub aggregates (as defined in table 2.1). Results are given in tables 2.5 and 2.6, and tables 2.7 and 2.8 in the appendix.

Starting with the income components, it is obvious that income from dependent employment is the major income source for most households across most deciles. Please note that the lower shares in the 1st - 5th deciles are due to retired people, whose pensions are partially liable to tax. Income from self-employment oscillates around 2-6% of taxable income, slowly increasing with income. Expectedly, the richest decile shows a larger share of 16.5%. The rich naturally also stick out with higher shares of income from rent and lease, and from capital income.

Coming to the descriptive statistics for the sub aggregates of taxable income (table 2.7 in the appendix), it can be seen that going down the calculations in the tax code overview (table 2.1), taxable income decreases as expected. In relative terms, the reduction attained by claiming professional and business expenses is slightly decreasing from 7 to 4% along the 2nd to 10th deciles. A different picture arises for the finally assessed taxable income: Relative to taxable income before deductions, the share goes up for every decile. Hence, when applied correctly the tax code shows a considerable degree of progressivity, as richer individuals are estimated to reduce their tax burden less.

Next, the different types of deductions are discussed (table 2.6). Obviously, on average special expenses are the most important deduction category, which is in line with the tax statistics (Destatis 2017*c*): The average ratio of special expenses, including precautionary expenses, to the Sum of Revenues[16] was 13.3% according to tax statistics,

[16] A comparable aggregate to my estimate of taxable income before deductions is not readily available

Table 2.5: Estimated taxable income, surplus and profit income components

(1) Taxable income decile	(2) Dep. employment	(3)	(4) Self-employment	(5)	(6) Rent and lease	(7)	(8) Capital	(9)
	Mean (EUR)	% of Y^{all}	Mean (EUR)	% of Y^{all}	Mean (EUR)	% of Y^{all}	Mean (EUR)	% of Y^{all}
1	857	57.17	43	2.90	16	1.07	7	0.46
2	1612	37.62	106	2.46	34	0.79	17	0.39
3	2238	32.67	148	2.16	44	0.65	21	0.31
4	4593	46.96	247	2.53	82	0.84	40	0.41
5	8222	60.41	434	3.19	165	1.21	77	0.57
6	13605	74.81	685	3.77	283	1.55	119	0.65
7	20288	83.53	1084	4.46	449	1.85	188	0.77
8	28262	87.14	1599	4.93	643	1.98	202	0.62
9	38389	86.98	2741	6.21	918	2.08	327	0.74
10	64123	75.15	14063	16.48	3501	4.10	1309	1.53

Note: Table includes only individuals with positive taxable income before deductions, which is calculated as outlined in table 2.1. Source: EVS 2013.

while my estimate based on the EVS puts it at 15.5%. In contrast, professional expenses reduce taxable income by roughly 2-6%, decreasing with income starting at the 2nd decile. These deductibilities are underestimated in my analysis, since on average they are put at 3.9% of the Sum of Revenues compared to 5.0% in the tax statistics. Extraordinary burdens, which as the title says are the most difficult to ascertain category, are generally less important. I estimate this deductions category to make up 1.4% of the Sum of Revenues, while the tax statistics report 0.9%. Direct tax reliefs are only marginally important across all deciles. There might be some underestimation here, because the tax statistics set them in the same range as the extraordinary burden, however I only estimate a share in the Sum of Revenues of less than 0.1%.

As a whole, these figures lend credibility to my claim of having programmed a rather benevolent tax authority, in ordert to arrive at a conservative estimate of tax underpayment. On average, my model seems to overestimate deductions by about 0.8% of the Sum of Revenues.

The main results of my analysis are shown in Figure 2.2 (and table 2.8 in the appendix), which gives my estimates of tax due and the resulting tax loss (or gain) from under(over)payment of tax. Keep in mind that at least the first three deciles are affected by the caveats made above. Namely, the average amount of tax paid in these deciles is probably largely a statistical artifact. At the aggregate level, assumed average tax overpayments of EUR 28 do not matter much, anyway. These figures pale when

from tax statistics publications, so I compare my figures to the Sum of Revenues which is reported.

Table 2.6: Estimated taxable income, deductions

(1) Taxable income decile	(2) Prof. expenses	(3)	(4) Special expenses	(5)	(6) Extraord. burden	(7)	(8) Direct reliefs	(9)
	Mean (EUR)	% of Y^{all}	Mean (EUR)	% of Y^{all}	Mean (EUR)	% of Y^{all}	Mean (EUR)	% of Y^{all}
1	422	28.13	498	33.23	32	2.14	1	0.08
2	277	6.45	864	20.15	64	1.48	2	0.04
3	326	4.76	1392	20.32	82	1.19	3	0.04
4	496	5.07	1807	18.47	192	1.97	5	0.05
5	641	4.71	2255	16.57	363	2.67	9	0.07
6	829	4.56	2784	15.31	483	2.66	15	0.08
7	1051	4.33	3599	14.82	500	2.06	19	0.08
8	1264	3.90	4800	14.80	607	1.87	23	0.07
9	1492	3.38	6437	14.59	785	1.78	28	0.06
10	2064	2.42	10878	12.75	1054	1.24	57	0.07

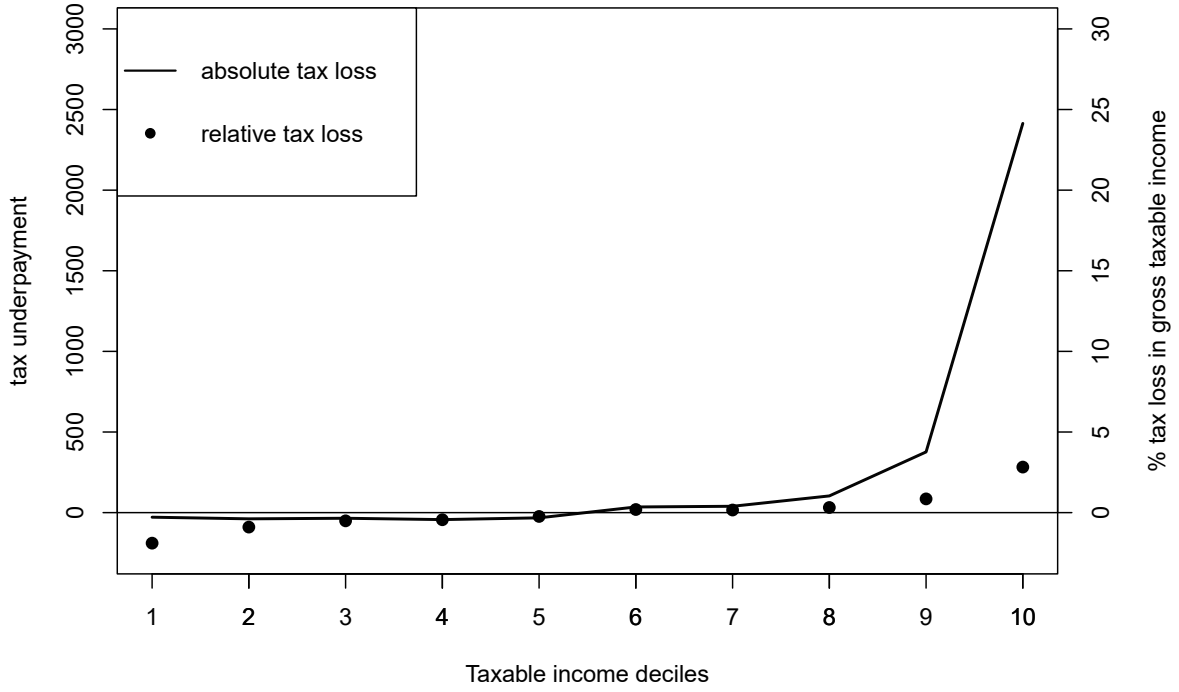
Note: Table includes only individuals with positive taxable income before deductions, which is calculated as outlined in table 2.1. Source: EVS 2013.

comparing them to the average underpayment of roughly EUR 2,400 for the richest decile. In relative terms, it is clearly visible that the amount of underpayment of tax rises with income when comparing it to taxable income before deductions. Amounting to 2.8% of taxable income before deductions, it is largest for the 10th decile, which equates to 15.8% of taxes paid. A graphic depiction of the estimated vs. statutory average tax rate can be found in figure 2.5 in the appendix.

Under the assumption that my EVS-based sample of tax-paying individuals is representative for German tax payers, which is reasonable after adjusting with the sample-census-based weights, some calculations out of the envolope may derive an aggregate estimate of income tax thus lost: Multiplying the number of tax payers (with a positive Sum of Revenues) captured by the tax statistics with the average tax loss for the same group in my sample gives an amount of EUR 10.73 bn of avoided tax. Of those, EUR 9.21 bn can be attributed to the richest decile, i.e. 86% of the avoided amount. Compared to the assessed amount of income tax, the tax loss is 4.4%. Obviously, this is considerably lower than the 34% estimated by Lang et al. (1997) for 1983.

Compared to a net tax gap of 13.7% (relative to the amount that should have been paid, i.e. 15.9% relative to taxes paid) in the U.S. in 2001 or an overall tax gap of 8% in Sweden in 2000 (Slemrod 2007), which were estimated from random tax audits, my estimate seems to be rather low. This strengthens my argument that the estimate is by any means a lower bound, even when taking into account that tax morale in Germany

Figure 2.2: Estimated tax loss (individual taxpayer level)



Note: A negative value for the tax loss implies that individuals have paid more tax than what is estimated as due. The graph includes only individuals with positive taxable income before deductions, which is calculated as outlined in table 2.1. Source: EVS 2013.

is rather high[17].

The availability of individual taxpayer level items in the household survey, allows me to compare the tax loss on the individual taxpayer level with the household level (table 2.9 in the appendix). As can be expected, the distributional impact of the studied phenomen is reduced by shifting the focus on the household (table 2.9). Nevertheless though, the general direction of the findings remains clear: Richer households are most able to reduce their tax burden, by about 1.82% of taxable income before deductions or 9.87% of taxes paid for the richest decile.

2.5.3 Robustness of the tax loss estimates

Of course, one may object that tax avoidance is notoriously difficult to measure, and therefore my results may be biased. For several reasons though, I am convinced that my estimate is by any means a lower bound. I shall discuss factors leading to a possible overestimation or underestimation in turn.

[17]Dörrenberg & Peichl (2018) report a high overall tax morale in a survey experiment with German participants. 89% of them answered that "evasion is 'not at all justifiable', 'not justifiable' or 'rather not justifiable' " (Dörrenberg & Peichl 2018, p. 25).

Starting with factors contribution to a possible overestimation, I cannot observe whether an individual carries over losses from previous years into the 2013 tax declaration. This possibility presents itself especially for business owners, even though it is subject to some limitations. Up to EUR 1m for single filers (doubled for joint filers), losses can be carried forward if they cannot be balanced within the current year or with the previous year. Further restrictions apply to certain types of losses, for partners in a limited partnership for instance it is restricted to her contribution in capital in the company. This could reduce taxable income, and therefore lead to some overestimation of tax avoidance. However, there were only EUR 4.9 bn losses carried over in 2013 according to the tax statistics, and this number includes not only losses from the previous year, but also losses of 2013 and of 2014 carried back. This figure equates to about 0.35% of the Sum of Revenues. Hence, the degree of overestimation resulting from this is presumably very limited, but could reduce the tax avoidance amount for richer individuals in particular.

Moreover, it was mentioned already that my estimate of direct tax reliefs seems to be too low, by about 0.8% points as a share of the Sum of Revenues when checked with the tax statistics. Moreoever, there are some deduction possibilities that cannot be accounted for because relevant information is unobservable in the data. For instance, this is the case for some extraordinary expense allowances: I cannot check for people with disabilities, non-remunerated care providers, grown-up children's eduction when staying outside of parent household or health insurance of other supported persons. Another example is due to the limited surveying period of the cross-sectional survey: If someone bought a computer the year before and deducted it over three years, I am unable to identify this deduction possibility[18].

Some general factors that work to the contrary have been mentioned before during the description of the data. First of all, the EVS does not contain the very top households or individuals of the income distribution. Roughly speaking, the top 0.5% income earners are excluded from the sample. By the general logic of my findings and of the literature, one would expect a higher degree of tax avoidance and evasion among them. Moreover, nonresponse problems and underreporting in surveys are typically rising with income, especially for wealth and asset income (Moore et al. 2000, Korinek et al. 2006). This makes underestimation of tax avoidance and evasion more likely when using the EVS. Furthermore, the distributional effect of tax avoidance and evasion is therefore presumably underestimated.

Adding to that, due to the construction of the EVS dataset, it was not possible

[18]See the technical appendix for a further discussion of the issue. For instance for landlords, this problems was mitigated by some simplified estimations.

to estimate deductions for business expenses for self-employed, freelancers and self-employed farmers. This is because in the survey, they are asked to exclude those and only report their private expenses. Hence, they basically report their Sum of Revenues and I am unable to check how identifiable business expenses reduce their taxable income.

Moreover, the modelling of the deduction possibilities in the German tax code is done in such a way that the tax authority is benevolent. I.e., when there is room for judgement on how some deduction item is distributed, I follow a conservative approach that increases deductions and therefore reduces assessed taxable income and tax due. For example, in the case of commuting expenses, I assume actual commuters to experience twice the average share of commuting in their total car use (in contrast to uses like shopping, holidays, etc.), thereby increasing the kilometres driven that are eligible for commuter allowance.

Furthermore, as shown when discussing table 2.6, I overestimate special and extraordinary expenses by about 2.2 and 0.5% points of the Sum of Revenues, respectively, when compared to the tax statistics. This seems to override concerns of underestimation due to unobservables that were mentioned before. At least, no inequality-reducing bias is to be expected, if these unobservable deduction possibilites are more or less evenly distributed along the income distribution. Given better health conditions and longer life expectancy of richer individuals, the allowances available for people with disabilities or non-remunerated care providers are unlikely to occur more frequently in the upper deciles of the income distribution for example.

Lastly, the EVS is unlikely to include a large fraction of black market incomes which are going untaxed. According to Schneider & Boockmann (2018), the size of the black market economy in Germany is estimated at 12.1% of GDP in 2013. Since these incomes are presumably distributed more in favor of lower income groups, being able to include them would increase the tax loss for the lower deciles. Note that this possibly reduces the distribution of the compliance gap towards high-income earners.

Concluding, it seems apparent that factors contributing to an underestimation of tax avoidance and evasion in my analysis are quantitatively much more important than those that work to the opposite.

It is important to note that I do not claim all of my tax loss estimate to reflect illegal tax evasion. On the contrary, as the most lucrative methods of tax evasion take place offshore, I assume that most of the difference I observe constitutes avoidance rather than evasion. This is because offshore tax evasion is more prevalent among the very top income individuals (Alstadsaeter et al. 2018*b*) who are excluded from the EVS. Rather, I interpret most of the tax loss estimated here to be the result of

(illegal) underreporting of income sources where the tax payer has some descretion over how much to report, combined with overreporting of deductibles and "semi-legal tax write-off opportunities", as Lang et al. (1997) put it. Even the standard work for tax advisors that I have used to grasp the plethora of deduction possibilities, is full of legal tax planning advice that one might evaluate as more or less illegitimate.

2.6 Conclusions

This chapter has set out to analyse income tax avoidance along the income distribution in Germany, using data from the 2013 Income and Consumption Survey (EVS). An estimate of taxable income is derived from a detailed micro data model of the German income tax code that exploits the richness of the survey items in terms of income, expenditures and taxes. This estimate of tax due is then compared to reported tax payments to get taxes lost due to tax avoidance.

Results confirm findings in the literature that detect more tax avoidance with rising income. The estimated amount of tax underpayment for the richest decile is 2.8% of taxable income before deductions, or 15.8% of taxes paid (at the individual taxpayer level). At the household level, these figures decrease to 1.8% and 9.8%, respectively. For German public coffers, taking these results to the tax statistics implies a loss of fiscal revenue of at least EUR 10.7bn, which equates to 4.3% of income tax proceeds. The richest decile accounts for roughly 86% of this figure. This confirms that inequality measures based on tax statistics underestimate income concentration by a considerable degree.

The results moreover underline the importance of fighting tax evasion and avoidance. Closing tax loopholes seems to be a cat-and-mouse-game where regulators react rather sluggishly, as demonstrated for instance by the the so-called "CumEx-Files" published by the "Correctiv" journalist team[19]. From an economic viewpoint, providing a cost-benefit analysis of measures to reduce the tax compliance gap would be desirable. Keen & Slemrod (2017) therefore propose to evaluate the "enforcement elasticity of tax revenue". Unfortunately, providing such an estimate is beyond the possibilities of this study, as it requires intertemporal variation that is unlikely to get from the database used here. Even when pooling the EVS waves, the fact that the survey is conducted only every five years presumably prevents to neatly identify sufficient variation.

Nevertheless, it seems safe to assume that some effective measures could be implemented rather cheaply. Using German data, Dörrenberg & Peichl (2018) identify a positive reciprocity treatment effect for a simple intervention: Participants who were

[19] See https://cumex-files.com/

told that spending on public goods could be increased if the revenue lost to tax avoidance and evasion was recovered, showed a significantly higher tax morale. When held against a share of merely 11% of participants that find cheating on tax acceptable in their study, increasing tax morale by 2-3 percentage points could be economically significant. By sweeping generalization to my results, tax avoidance could be reduced by an amount of EUR 1.9 - 2.9 bn. As tax payers receive mail from the fiscal authorities anyway, telling them about the drawbacks of tax avoidance and evasion while stressing the benefits of public goods is unlikely to be costly. Another possibility presents itself when assessing the performance of the fiscal authorities: For instance, the Federal Court of Auditors criticizes substantial amounts of revenue losses due to underinvestment in the number and training of tax inspectors (Schäfers 2018).

If access is granted and data protection regulation allows for it, future work could benefit from statistical matching of additional micro data. Particularly useful in the context of the commuter allowance and travel expenses could be the Sample Census (Mikrozensus) or the Employee History Statistics (IAB-Beschäftigtenhistorie), as they contain data on commuting distances. But also for other items they might prove useful. For obvious reasons, matching micro data from the tax statistics could also benefit the precision of the analysis, as it would allow for estimated and actually claimed deductions to be compared. A possible result could be a better calibration of some of the modeling concerning the hypothetically possible deductions.

The higher degree of avoidance among the richest decile also warrants to study actually claimed deductions more closely using the tax statistics. To my knowledge, no further analysis concerning the distributional implications of the actual use of deductions is yet available for Germany.

2.A Appendix

Figure 2.3: Tax payments and taxed vs. taxable income

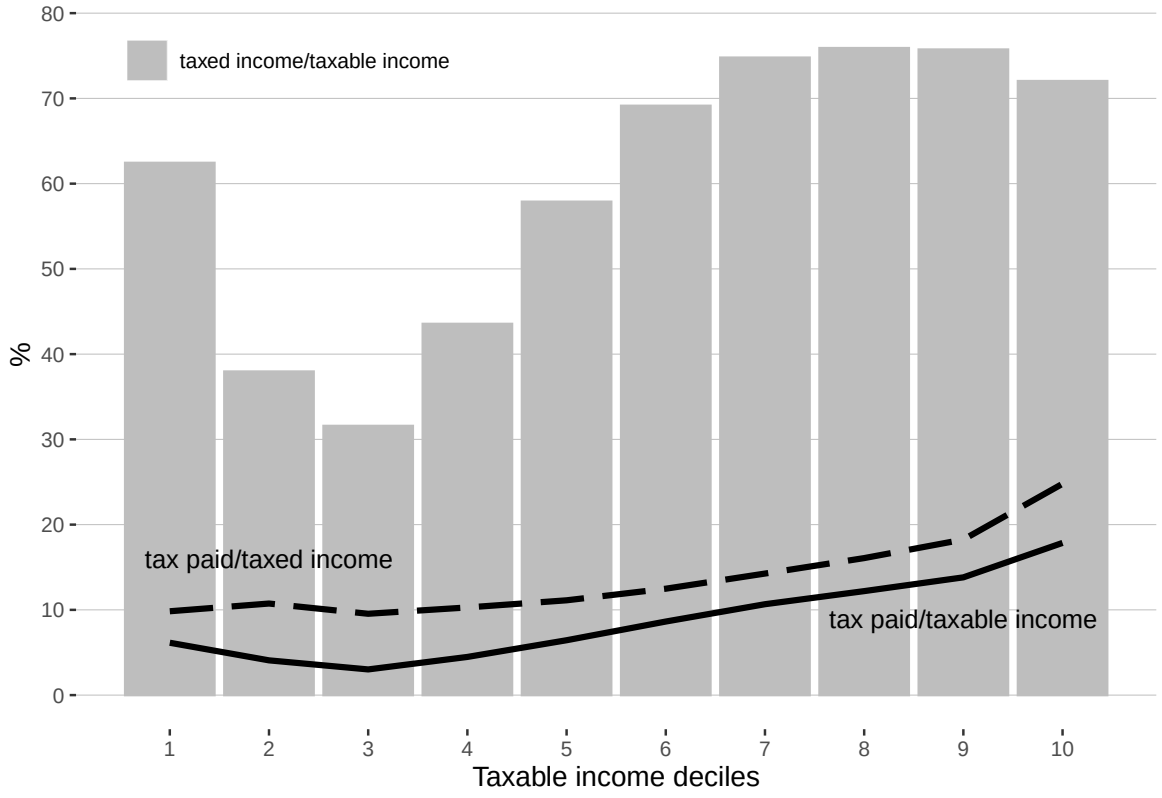


Table 2.7: Estimated taxable income, sub aggregates

(1) Taxable income decile	(2) Taxable income (Y^{all})	(3)	(4) Sum of Revenues	(5)	(6) Assessed taxable income	(7)
	Decile range (EUR)	Mean (EUR)	Mean (EUR)	% of Y^{all}	Mean (EUR)	% of Y^{all}
1	[1 – 2819]	1499	1167	77.80	606	40.42
2	[2819 – 5637]	4286	3959	92.36	3012	70.27
3	[5637 – 8132]	6851	6445	94.07	4949	72.24
4	[8132 – 11633]	9780	9153	93.59	7120	72.80
5	[11633 – 15683]	13610	12768	93.81	10097	74.19
6	[15683 – 20921]	18187	17187	94.50	13847	76.13
7	[20921 – 28022]	24289	22991	94.66	18763	77.25
8	[28022 – 37367]	32433	30809	95.00	25277	77.94
9	[37367 – 52548]	44134	42140	95.48	34778	78.80
10	[52548 –]	85321	81539	95.57	69448	81.40

Note: Table includes only individuals with positive taxable income before deductions, which is calculated as outlined in table 2.1. Source: EVS 2013.

Table 2.8: Estimated taxable income (individual taxpayer level), tax loss

(1) Taxable income decile	(2) tax due	(3)	(4) tax due – tax paid	(5)	(6)
	Mean (EUR)	% of Y^{all}	Mean (EUR)	% of Y^{all}	% of tax paid
1	64	4.25	-28	-1.89	-30.85
2	136	3.18	-38	-0.90	-22.01
3	171	2.50	-35	-0.51	-16.92
4	395	4.04	-43	-0.44	-9.87
5	844	6.20	-32	-0.24	-3.69
6	1605	8.82	35	0.19	2.23
7	2631	10.82	39	0.16	1.51
8	4063	12.52	104	0.32	2.62
9	6479	14.67	376	0.85	6.17
10	17669	20.68	2414	2.83	15.82

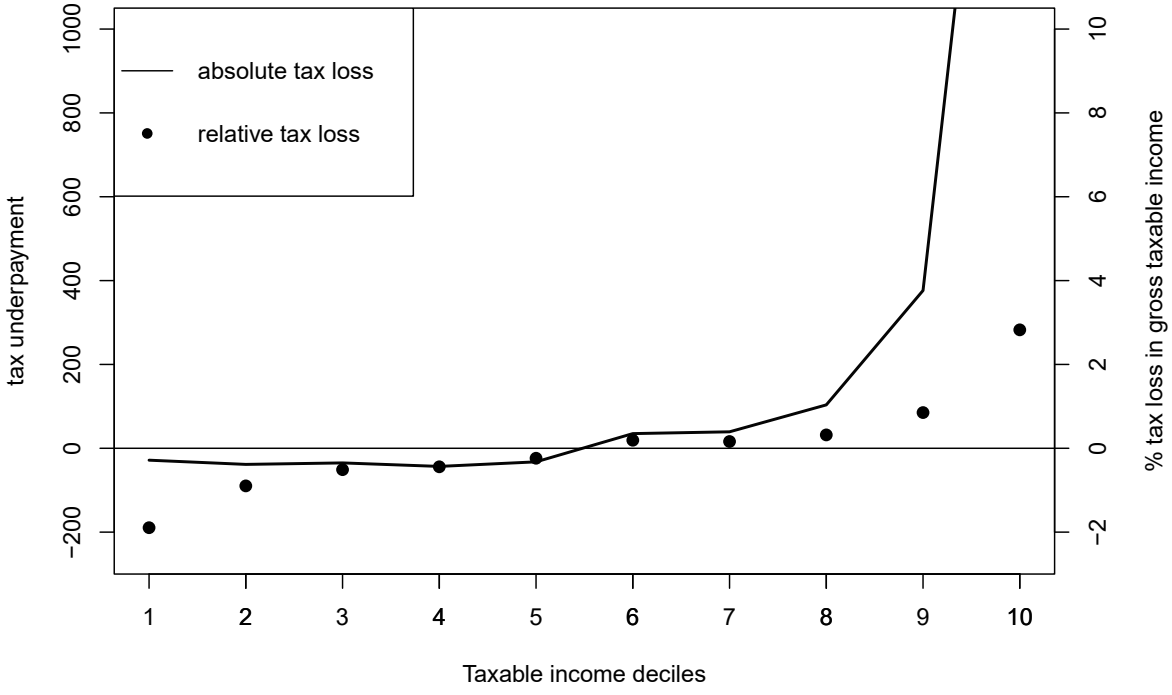
Note: A negative value for the tax loss (columns 4-6) implies that individuals have paid more tax than what is estimated as due. Table includes only individuals with positive taxable income before deductions, which is calculated as outlined in table 2.1. Source: EVS 2013.

Table 2.9: Estimated taxable income (household level), tax loss

(1) Taxable income decile	(2) tax due	(3)	(4)	(5) tax due – tax paid	(6)
	Mean (EUR)	% of Y^{all}	Mean (EUR)	% of Y^{all}	% of tax paid
1	199	4.97	-107	-2.68	-35.04
2	290	2.81	-116	-1.12	-28.57
3	483	3.06	-79	-0.50	-14.00
4	1027	4.57	-17	-0.08	-1.65
5	2117	7.02	-117	-0.39	-5.25
6	3981	9.90	52	0.13	1.33
7	6031	11.31	-55	-0.10	-0.91
8	9064	12.97	85	0.12	0.95
9	13719	14.74	144	0.15	1.06
10	35887	20.25	3198	1.80	9.78

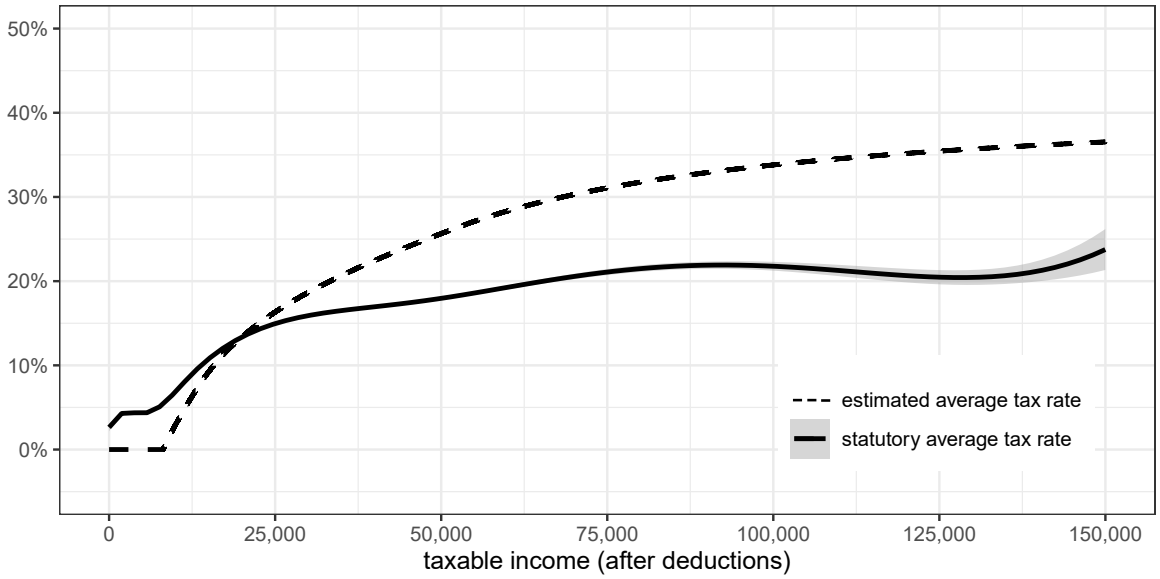
Note: A negative value for the tax loss (column 5) implies that households have paid more tax than what is estimated as due. Table includes only households with positive taxable income before deductions, which is calculated as outlined in table 2.1. Source: EVS 2013.

Figure 2.4: Estimated tax loss (individual taxpayer level), zoomed in



Note: A negative value for the tax loss implies that individuals have paid more tax than what is estimated as due. The graph includes only individuals with positive taxable income before deductions, which is calculated as outlined in table 2.1. Source: EVS 2013.

Figure 2.5: Statutory vs. estimated average tax rates



Please note: The estimated average tax rate is based on the reported tax payments and my estimate of taxable income after deductions (zvE). It is smoothed using a spline with four knots at the kinks of the tax schedule. The graph is cut at EUR 150,000 taxable income, because there are too few observations beyond this to give an acceptably precise estimate. The calculation takes into account that flat rate withholding taxes on capital income, which are part of the taxes paid reported in the EVS, have to be excluded to make the estimated ATR comparable to the statutory rate.

Chapter 3

Income tax noncompliance in Germany, 2001-2014

3.1 Introduction

Following the Financial Crisis of 2008, tax noncompliance has become a growing concern of policymakers and researchers alike. International public scandals like the Offshore Leaks, LuxLeaks, Panama Papers or Paradise Papers have contributed to rising awareness. In Germany, particularly the cases of celebrities have received substantial media attention (Garz & Pagels 2018), as well as widespread "Cum Ex" fraud perpetrated by banks on behalf of wealthy investors (Spengel 2017).

Accordingly, tax evasion is perceived as a more serious offence than previously: During the 1995-98 wave, only 46.3% of German respondends in the World Values Survey deemed cheating on taxes "never justifiable" (extreme value on a 1 to 10 scale). In the 2017-2020 wave, this number has risen to 75.2%. Dörrenberg & Peichl (2018) equally document a relatively high tax morale of German respondents in the 2014 German Internet Panel.

To some extent, also policy action in the field of personal income taxation[1] has been a result of the aformentioned growing concerns: In particular, the Financial Crisis of 2008 was followed by G20 measures in 2009 that increased the number of international tax information exchange agreements (for analyses of these TIEAs, see Johannesen & Zucman 2014 and Menkhoff & Miethe 2019). More importantly, in 2013 agreement was reached to introduce automatic information exchange of financial account information, which became effective in the beginning of 2017 (for early evaluations, see Menkhoff &

[1] Furthermore, measures were taken to fight base erosion and profit shifting (BEPS) in the corporate income tax.

Miethe 2019 and Casi et al. 2020).

A growing literature documents the extent of income tax avoidance and evasion, increasingly shedding light on the link with income and wealth inequalities. Most notably, Alstadsæter et al. (2019) have exploited the richness of Scandinavian administrative data by linking it with cases of caught or self-reported evaders, for instance from some of the leaks mentioned earlier. A similar analysis has been conducted with Dutch (Lejour et al. 2020) as well as US tax data (Guyton et al. 2021), likely with more to follow in upcoming years. Unfortunately, a likewise approach cannot be followed for the German case, due to legal restraints on the use of tax micro data.

Fortunately though, different strands of the literature have long established indirect approaches to measure tax avoidance and evasion. Firstly, some authors have tried to exploit differences between responses in survey and tax data. This may be called the direct discrepancy method, which seeks to compare samples of populations made comparable through weighting and is based on the debatable assumption that taxpayers report their incomes more honestly in an anonymous survey than in their tax declaration. The methodology has been applied to a range of European countries, e.g. by Fiorio & D'Amuri (2005), Matsaganis et al. (2010), Benedek & Lelkes (2011) and Leventi et al. (2013), but not yet to Germany.

Secondly, following the seminal paper of Pissarides & Weber (1989, henceforth PW), researchers have relied on differences in reported income compared to certain expenditures to detect underreporting of income. Typically, food is used for survey data and donations in case of tax data. This approach has proven valuable to study the underreporting behaviour of self-employed individuals in particular. As different income sources can be observed in tax data in more detail, studies following the paper of Feldman & Slemrod (2007, henceforth FS) are typically able to also estimate underreporting for other income types. For a literature review, see Albarea et al. (2020, Section 2.2). Estimates in this literature suggest that the self-employed underreport on average 15-40% of their income.

Recently, Albarea et al. (2020) have combined the two major methodologies, by enhancing a micro-simulation based discrepancy estimate with underreporting figures obtained from survey data with the PW methodology. Bazzoli et al. (2020) are moreover able to directly link tax data with household budget survey data at the micro level for seven years, thereby improving the distributional analysis when estimating self-employed income underreporting. For Germany, Bittschi et al. (2016) have applied the FS-approach to the Taxpayer Panel, finding rather small effects for the different income categories in a fixed-effects Poisson specification for 2001-2006. Fauser (2019) has estimated income tax avoidance along the income distribution, applying a micro

data model of the German income tax to the Income and Consumption Survey of 2013. The papers that go beyond constant shares of underreporting by income category show that tax noncompliance tends to be concentrated at the top of the income distribution (among the recent contributions, see e.g. Alstadsæter et al. 2019, Albarea et al. 2020 and Bazzoli et al. 2020 or Guyton et al. 2021).

In our study, we test both the discrepancy method and the PW/FS-approaches for Germany. We combine different types of micro data, but we are unable to directly match them for confidentiality reasons. Our main datasets for the full period 2001-2014 are the German Socioeconomic Panel (SOEP), a survey panel provided by the German Institute for Economic Research (DIW Berlin) that is widely used in the social sciences, and the Taxpayer Panel (TPP) which consists of tax records that were linked by the German Federal Statistical Office (Destatis).

Our descriptive analysis shows that SOEP and TPP samples differ somewhat with regards to demographic characteristics and the income categories included. As income variables in the tax data are generally more detailed than in the survey, we adjust the former to the latter by combining relevant variables. We then weight the SOEP figures using distributions drawn from the limited set of sociodemographic variables included in the TPP. The remaining differences between reported mean incomes in the SOEP and TPP are broadly in line with the patterns observed in earlier papers. We find higher average income from self-employment and income from renting and leasing in the survey than in the tax data and higher average discrepancies of self-employment income in the higher income quintiles. Tax evasion by income underreporting may, however, be only one of several possible explanations for the observed discrepancies.

In order to further investigate the potential underreporting of non-wage incomes for tax purposes, our analysis thus focuses on the regression analysis based on the PW/FS-approaches. We test different specifications on several datasets: Following PW and using the SOEP, food expenditure is regressed on different income measures and a host of control variables. We do not find indication of income underreporting by the self-employed in the SOEP based on the food-expenditure regressions. However, self-employment is associated with higher average expenditures on electricity, heating, and warm water and with higher total housing cost. Assuming that unobserved heterogeneity, e.g. with respect to working from home, does not fully explain these differences, this might indicate underreporting of self-employment income even in the SOEP. However, the coefficients are relatively small and the food regressions would not support such an interpretation.

Moreover, using the TPP we estimate donation regressions both on cross-sections and the whole panel. Results indicate that self-employment and business incomes are

significantly associated with higher donations, which can be interpreted as evidence of income underreporting. This would once more call into question the equality of tax collection by income source and hence the progressivity of the tax schedule, because self-employment and business incomes are more concentrated at the top of the income distribution. However, we cannot rule out that unobserved heterogeneity is responsible for some part of the effect we find. Unfortunately, the scarce sociodemographics do not allow to control for possibly relevant factors such as a presumably more frequent solicitation for donations or higher charitable giving for marketing reasons. This could lead to an overestimation of underreporting.

Finally, we use the our results from the FS-type regressions based on TPP data to derive estimates of aggregate underreporting and resulting tax revenue losses. Back-of-the-envelope calculations suggest a tax gap of EUR 21.3 bn in 2001 and 15.8 bn in 2014, when underreporting from all income categories is considered . Relative to "true" income tax due, this amounts to 10.7% in 2001 and 5.7% in 2014.

In a more detailed approach, we also take into account the progressivity of the income tax schedule by applying the estimated underreporting coefficients to the individual tax units observed in the panel. Furthermore, we assume that not the whole "underreported amount" would actually be taxed, e.g. due to eligible deductions. For all income categories, the estimated tax loss in the more detailed estimations is considerably larger: It ranges from EUR 70.2 bn in 2001 to EUR 32.4 bn in 2014, implying a tax gap relative to "proper" tax due of 28.4% in 2001 and 11.1% in 2014 (or of 39.6% and 12.5% relative to the assessed income tax).

The remainder of this chapter is organized as follows: Section 3.2 describes the datasets that were used. Section 3.3 covers the discrepancy method, our approach and results, while section 3.4 repeats the same exercise for the indirect regression-based approaches. Section 3.5 concludes.

3.2 Data

Our two main data sources are the German Socioeconomic Panel (SOEP) version 35 and the Taxpayer Panel (TPP). The Taxpayer Panel (TPP) includes annual data on German taxpayers since 2001. Until 2012, the data only included the whole population of taxpayers filing tax returns (around 28 Mio. filers per year). Since 2013[2], the data includes also information on the pay-as-you-earn cases which are usually not required

[2]In theory since 2012. In practice, data delivery delays have caused the TPP to include the additional cases only starting in 2013.

to file a tax return (some additional 12 Mio. cases). Therefore, the data are not representative for the whole population of taxpayers until 2013, but are biased against the income-poor wage earners who are less likely to file a tax return. Adding to that, the construction of the Taxpayer Panel changed in 2010. Before, the cases were linked as a panel based on several characteristics (for details, see Vorgrimler et al. 2006), since 2010 the newly introduced unique personal tax identifier is used to build the panel.

Due to the strict confidentiality requirements, researchers can have access to a random sample from the TPP only at regional statistical offices or via controlled remote data processing. For our analysis we thus rely on a 5-percent stratified random sample drawn from the TPP which includes about 840,000 wage earners, 540,000 earners of self-employment or business income and 380,000 earners of income from rent and leasing. Extrapolation factors included in the dataset allow us to correct for the oversampling of richer households compared to the whole population in the complete dataset. The TPP features only a limited set of demographic variables such as sex, age and religion. Number of children and marital status can be inferred from the tax allowances and tax classes. The TPP includes income variables in accordance with the different income categories on the tax return which are not necessarily consistent over time as the tax law changes.

The SOEP is a representative survey of private households in Germany and available as a panel data set since 1984. Between the relevant years (2001-2014), it includes income information on 12,000 to 16,000 households. In the questionnaire, respondents are asked to estimate their monthly earnings from dependent and independent employment, as well as from secondary jobs. In contrast to the TPP, negative self-employment income is not included. The annual labour income is imputed by multiplying the monthly earnings with the months of employment of the previous year. In addition, respondents are asked to estimate their annual income and losses from rent and leasing and from investment. The SOEP covers a broad range of demographic variables, among which detailed housing and education related variables. The data is made available to researchers by the German Institute for Economic Research (DIW Berlin).

3.3 Discrepancy approach

3.3.1 Building comparable samples and income categories

As mentioned before, not all income categories are easily comparable between TPP and SOEP as the TPP variables correspond to the legal definitions of the income tax

base while the SOEP has broader but stable income categories. For example, the TPP includes several detailed categories of business, freelance and other self-employment incomes but the SOEP only includes one broad category of self-employment income where respondents are asked to estimate their overall positive income from all types of self-employment. Due to tax reforms, the parts of passive capital income included in the TPP have changed several times over the sample period[3] and are thus not comparable to the capital income in the SOEP. The same applies for the pensions. We thus limit our analysis to the three broad income categories which can be defined consistently in the two databases: wage income, business income including all types of self-employment income, and income from rent and leasing. The samples are compared at the individual level.

In the SOEP, wage income is included in the variables ijob1 and ijob2. Unfortunately, ijob2 includes all sorts of secondary income which can stem from a second job but also from secondary self-employment, honorary work or family workers. Moreover, the source of the secondary income is only given for the years 2017 and 2018. We use this information to impute the shares of secondary income attributable to dependent and independent work for our sample years 2001 to 2014 and add this to wage and self-employment income from the main occupation. We further add extra payments, such as Christmas and holiday bonuses. In the SOEP, income information is "What did you earn from your work last month?". If people are self-employed they are asked to estimate their monthly income before and after tax. The annual income is then extrapolated by multiplying the monthly income by the number of months worked in the previous year. It is very likely that self-employed respondents report their income less cost, i.e. their profit which should make it conceptually comparable to the positive self-employment incomes in the TPP. However, it is not likely that business owners in the SOEP report retained earnings as part of their income even though these would be considered as taxable income in the TPP. For the income from rent and leasing we subtract the losses from renting and leasing and only include the net income if it is positive. To make the SOEP sample comparable to the TPP, we drop all individuals without positive income in any of those three categories. We also drop individuals that report positive income only once during the sample period, as these would not appear in the TPP.

[3]Most notably, since the switch to a dual income tax system in 2009, capital income is mostly not included in the income tax statistics anymore. Most passive returns are withheld at source, and no information concerning the taxpayer is transmitted to the tax authority. For details and some approaches to estimate capital incomes in the context of top wealth and income shares, see Bartels & Jenderny (2015).

Variables from the TPP sample are chosen accordingly: We take gross wages from dependent employment, which are gross of costs of obtainment and therefore match the SOEP ijob concept. For self-employment income, the matter is more complicated: We take incomes net of costs of obtainment[4], but substract capital gains related to self-employment to enable comparability to the SOEP. The same procedure is followed for incomes from business and agriculture and forestry. Moreover, income from sales of shares in unincorporated businesses is substracted, because these capital gains are likely included in a different variable in the SOEP (capital income). Incomes from these three revenue categories are then added up. For income from rent and lease, we take the revenues net of related costs claimed for tax purposes.

3.3.2 Adjusting the samples

It is not surprising that our two samples differ with regard to certain key characteristics as they represent two different populations. While the SOEP should be representative of the whole population, the TPP represents the population of taxpayers. We thus drop all individuals without income from our SOEP sample. Before 2013, the TPP only includes taxpayers who filed a tax return. For the years before 2013, we thus drop individuals from the SOEP who earned wage income only and whose income was below the income tax allowance as these are very unlikely to have filed a tax return. After these adjustments, the samples still differ with regard to the age structure, marital status, number of children and region of residence. The SOEP includes a slightly higher share of individuals at both tails of the working age distribution between 16 and 25 and 56 and 64, a lower share of married individuals, a slightly higher share of individuals in East Germany in some years and a slightly lower share in the income-rich region 1 (Hamburg, Bremen, Bayern, Baden-Württemberg and Hessen). As these sample characteristics correlate with average incomes, we reweight the SOEP sample to better match the TPP. For this purpose, we calculate post-stratification weights, treating the TPP as the population and the SOEP as the sample whose distribution needs to be adjusted to fit the characteristics of the TPP "population". As cross-tabulation of frequencies might produce unstable weights for rare combinations of characteristics, we iteratively fit the weights to reflect differences in the single variable frequency tables. After three iterations, our TPP and SOEP samples closely resemble each other in terms of age structure, marital status, regional distribution and number of children.

[4]These costs are not seperately available in the tax data. Self-employment incomes are only reported after related business expenses are substracted (yielding the revenue, or "Einkünfte" in German Income Tax law). On the contrary for dependent employment, we can observe the related professional expenses, i.e. costs of obtainment.

Table 3.1: TPP and SOEP samples

	TPP	SOEP	SOEP reweighted
Region			
Region 1	0.41	0.40	0.41
Region 2	0.43	0.44	0.43
Region 3	0.16	0.16	0.16
Marital status			
Married	0.67	0.53	0.67
Unmarried	0.33	0.47	0.33
Age class			
16-25	0.07	0.12	0.07
26-35	0.21	0.21	0.21
36-45	0.31	0.28	0.31
46-55	0.30	0.26	0.30
56-64	0.12	0.13	0.12
Number of children			
0	0.50	0.65	0.51
1	0.22	0.19	0.22
2	0.21	0.13	0.21
3 or more	0.07	0.03	0.07

Note: Table includes only individuals with positive taxable income.
Region 1: Baden-Württemberg, Bavaria, Bremen, Hamburg, Hesse; Region 2: Berlin, Northrhine-Westphalia, Lower Saxony, Rhineland-Palatinate, Saarland, Schleswig-Holstein; Region 3: Brandenburg, Mecklenburg-West Pomerania, Saxony, Saxony-Anhalt, Thuringia.

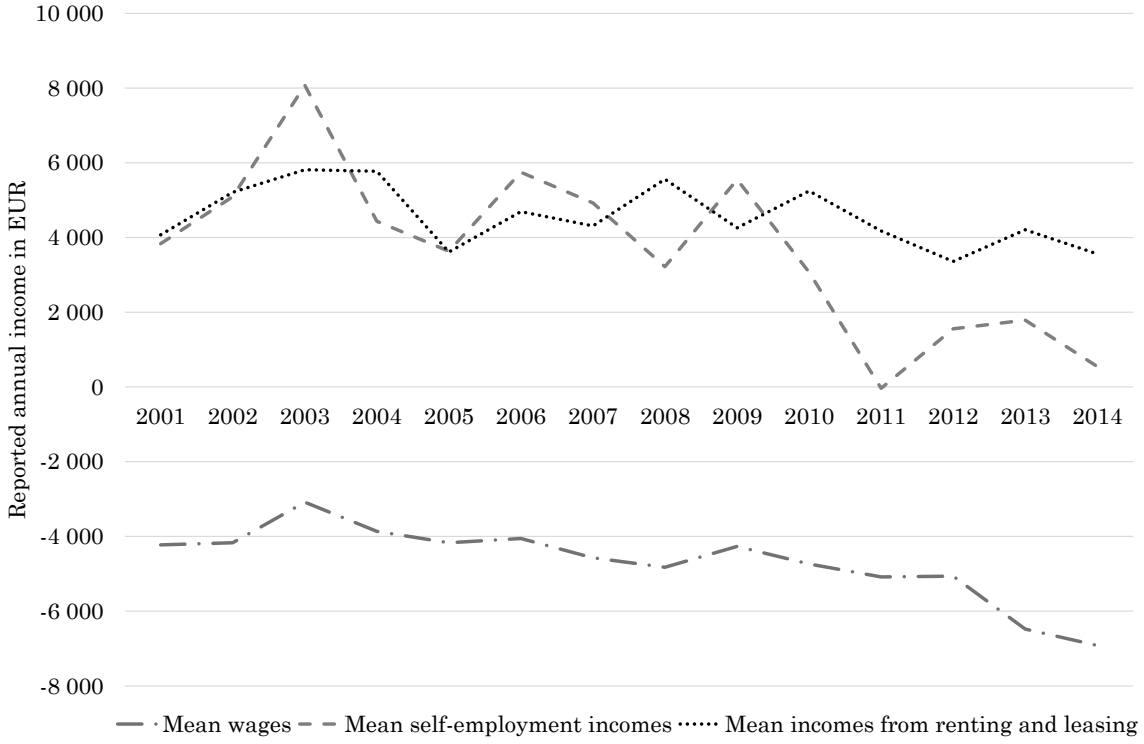
3.3.3 Results

Comparing the means of the different income categories between the adjusted SOEP sample and the TPP, we find that reported self-employment income and income from rent and lease are on average higher in the SOEP than in the TPP. Average wages, in contrast, are lower in the SOEP than in the TPP.[5]

Between 2001 and 2009, the discrepancies are broadly stable over time with the wage income being on average about 4,000 EUR lower in the SOEP than in the TPP,

[5]One might suggest that the lower average wages in the TPP stem from the fact that until 2012 only wage earners filing a tax return were included in the TPP and that these are more likely to be high-wage earners. However, after the inclusion of all wage earners in 2012/2013, the negative discrepancy in wages between SOEP and TPP even increases. Another possible explanatory factor could be the construction of the TPP: Only taxpayers who are observed at least twice over time, are taken into the panel. Therefore, a cross section of the TPP is likely biased downwards for wage incomes, compared with the (full sample) cross section of the wage and income tax statistics (LESt).

Figure 3.1: Mean discrepancies between SOEP and TPP samples



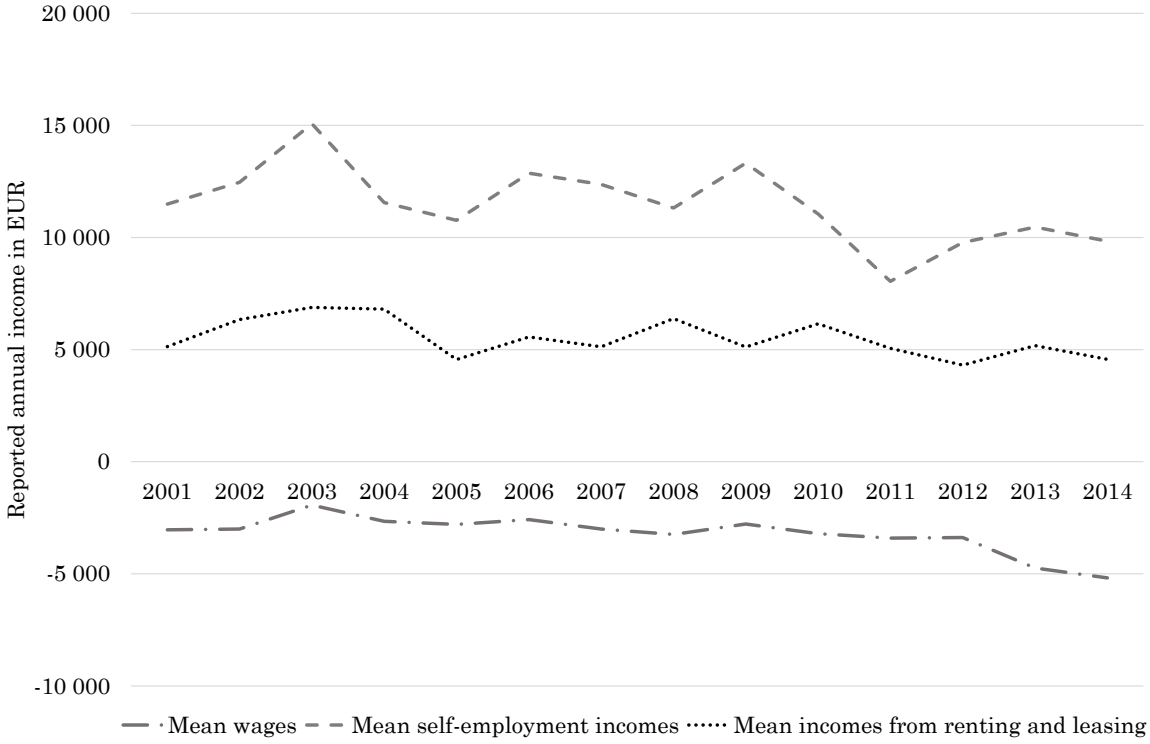
Source: Own calculations, based on SOEP and TPP data.

and self-employment and rent income being about 4,000 EUR higher (Figure 3.1). After 2009, the discrepancy between SOEP and TPP starts to widen for wages and to narrow for self-employment incomes. This might indicate that the two samples underlie different trends and are only of limited comparability. It is noteworthy, however, that those incomes which are self-reported for tax purposes are higher in the SOEP than in the TPP which would be in line with the underreporting hypotheses. However, if the discrepancy was only due to reporting behaviour, reported wage incomes should be the same in the SOEP and TPP, at least after 2012, as wage incomes are subject to the pay-as-you earn tax scheme.

As the top-income percentile is known to be underrepresented in the SOEP, we exclude the top one-percentile from our TPP sample and repeat the analysis. Comparing the SOEP to the top-censored TPP sample, the negative discrepancy between the average reported wage incomes decreases somewhat to approximately 3,000 EUR on average while the positive discrepancy for the self-employment incomes is much higher with approximately 10,000 EUR on average. The discrepancy of income from renting and leasing increases to 5,000 EUR on average (Figure 3.2).

In order to examine the size of the discrepancy along the income distribution, we

Figure 3.2: Mean discrepancies between SOEP and TPP top-censored sample



Source: Own calculations, based on SOEP and TPP data.

compare mean incomes by income quintile between the SOEP and the two TPP samples (Figure 3.3). We build the quintiles on the sum of wages, income from self-employment and income from renting and leasing.

We find that the negative discrepancy between reported wage incomes is broadly constant across income quintiles. As expected, the discrepancy narrows significantly for the top quintile, when comparing the SOEP to the top-censored TPP sample and remains the same for the other quintiles. For the self-employment incomes in contrast, we find slightly negative discrepancies for the first two income quintiles and positive discrepancies for the third and fourth income quintile. The top-censoring of the TPP affects the discrepancy of the top quintile very strongly as it switches from a slightly negative discrepancy to one of about 28,000 EUR. The average discrepancy of the self-employment incomes seems thus to be caused mainly by the two top income quintiles. For the income from rent and lease, we find a positive and significant discrepancy for all income quintiles. Similarly to the self-employment income, the discrepancy for the top quintile is negative in the non-censored sample but turns out positive and large for the top-censored sample.

In conclusion, we find relatively high and positive discrepancies when comparing

self-employment incomes and incomes from renting and leasing between the SOEP and the top-censored TPP. This would be in line with under-reporting to the tax authorities as these types of income are self-reported and thus leave more scope for tax avoidance and evasion. For wage income, in contrast, we find a negative and relatively small discrepancy. Further, the discrepancy for self-employment income seems to increase along the income distribution. For income of rent and lease and wages we cannot observe such a tendency. The observed discrepancies might indicate under-reporting by taxpayers to the tax authorities while revealing their true amount of income in an anonymous survey. They might, however, also be explained by other factors, such as a more accurate consideration of expenses and losses for tax purposes or unobserved differences in the underlying sample distributions.

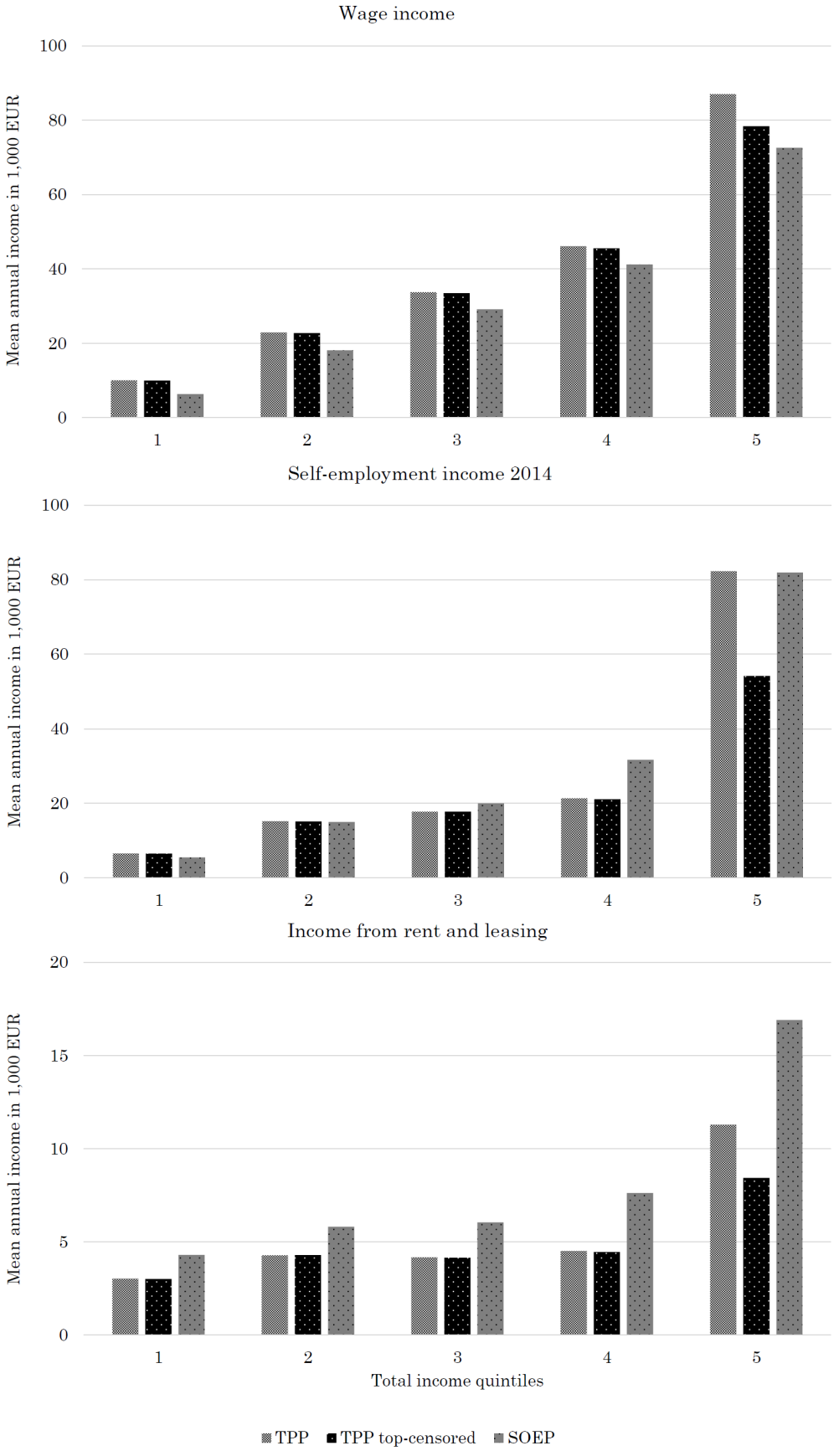
3.4 Regression-based approach

We test the classical PW approach (Pissarides & Weber 1989) to estimate Engel Curves in order to detect income underreporting for some income categories, using the SOEP. Moreover, for the same purpose we also follow FS by estimating regressions of donations on income types and a range of controls, using the TPP.

3.4.1 Food regressions using the SOEP

The PW approach is based on the idea that – in contrast to wage earners - the self-employed might underreport their income also in anonymous surveys but correctly report their expenditures for food consumption. As food is a basic necessity, the interpersonal variation of food expenditures in relation to income might be lower than that of other consumption categories and less affected by personal taste and status considerations. The authors thus assume that wage earners and self-employed having the same level of income and similar personal or household characteristics should spend the same amount of income on food consumption. However, regressing the logarithm of food expenditures on the logarithm of disposable income and a set of control variables, they find that self-employed report significantly higher food expenditures than wage earners which they attribute to underreporting of income by the self-employed. Variations of this approach have been used by several authors among which Engström & Hagen (2017) and Engström & Holmlund (2009) for Sweden, Kukk & Staehr (2017) for Estonia, and Kim et al. (2017) for Russia and Korea. A key challenge identified by most authors is that the underreporting of self-employed might be overestimated if based on current income instead of permanent income. Some studies use instrumen-

Figure 3.3: Mean incomes by income quintile, 2014



Source: Own calculations, based on SOEP and TPP data.

tal variable techniques to address this problem, others proxy permanent income by multiple-year averages of current income if panel data is available.

In this section, we apply the PW approach to the SOEP to test whether we find indication of income underreporting by the self-employed in Germany. The SOEP contains detailed consumption information only for the year 2010, which limits our analysis to this probably untypical year. However, the panel data allows us to calculate multiple-year averages of income around that year in order to proxy permanent household income. Similar to Engström & Hagen (2017) but limited to a single cross-section, we regress the logarithm of food expenditures of households on logarithmized 3-year and 5-year averages of their disposable income and a self-employment dummy. As additional control variables, we include the age, sex and education of the household head, marital status, number of children and adults living in the household, three regional dummies, a dummy variable indicating whether the household is a renter or house owner, and a dummy variable indicating whether the household is paying off a loan. We use different operationalisations of the self-employment dummy as suggested in the literature. A household may either be defined as self-employed when any of the household members reports being self-employed (A), when the household head reports being self-employed (B), or when the share of self-employment income in total household is more than 25 percent (C). We limit the sample to those households which do not switch between the categories for three or five years.

Hence, we estimate the following Engel curve equation:

$$\ln(C_i) = \alpha X_i + \beta \ln(Y_i) + \gamma SE_i + e_i \tag{3.1}$$

where subscript i denotes the household, X_i a vector of control variables, Y_i permanent household income and SE_i a dummy variable for self-employed households.

Surprisingly, in our data, wage earners and self-employed seem to spend the same share of their income on food consumption on average: 16 percent (see table 3.7 in the appendix). Using a simple OLS regression to control for additional household characteristics, we do not find any significant correlation of the self-employment dummy and food consumption in none of the specifications. Most of the control variables are significant with the expected signs. A higher household income, the age of the household head, the number of children and adults in the household and being married are associated with higher food expenditures. Being based in East Germany and being widowed is associated with lower food expenditures (see detailed results in the appendix table 3.9). The self-employment dummy is very small and negative but never

significant. These results would be consistent with self-employed reporting their income accurately in the SOEP. It should be noted however, that - depending on the definition of self-employed - only 431-600 households fall into this category as compared to about 4,000 households defined as wage earner households. This relatively low number of self-employed and the limited availability of consumption data shed doubts on the representativeness of results.

3.4.2 Housing-cost regressions using the SOEP

As an alternative to the food regressions, we estimate similar equations for expenditures on housing, an approach also taken by Albarea et al. (2020). The SOEP includes several housing-related variables which - in contrast to the food expenditures - are available for a greater number of households and years. These include expenditures on electricity, heating and hot water, additional cost, rent payments, amortization, and maintenance cost. Based on these variables, we build two alternative housing-cost variables, the first (EHW) including only electricity, heating and hot water, and the second (total housing cost) including all available housing-cost variables as in Albarea et al. (2020). For 2013, we obtain a sample of about 10,000 households of which we consider 500-800 as self-employed. A first look at the descriptive statistics suggests that the self-employed spend a little less on electricity, heating, and hot water, and on total housing cost on average as compared to the wage-earner households (appendix table 3.8). As in the previous section, we regress the logarithm of the dependent variable on a self-employment dummy, the logarithm of household income and the same set of control variables. The regression results suggest that being self-employed is associated with higher expenditures for electricity, heating, and hot water which exceed those of wage-earner households by approximately 10 percent on average. Total housing expenditures are higher by approximately 3-5 percent (see tables 3.10 and 3.11 in the appendix). The positive coefficients of the self-employment dummies are significant for all three definitions of self-employment and also for the other available years 2010-2012.

Under the assumption that - everything else being equal - self-employed and wage-earner households have the same preferences with regard to housing, the positive coefficient of the self-employment dummy variable could be interpreted as indication of underreporting of self-employment income even in the SOEP. This would suggest that our previous assumption of correct reporting in the survey - on which we base the discrepancy approach in section 3 - is not fulfilled. We would argue that underreporting of self-employment income in the SOEP might lead to even higher discrepancies between the TPP and the SOEP and would therefore not put our previous results into ques-

tion. However, the assumption that, in the absence of underreporting, self-employed have the same housing-related expenditures as wage earners might also be problematic. Some unobserved household characteristics might correlate both with the likelihood of being self-employed and the housing cost, e.g. a preference for spacious or prestigious appartments or a less economical consumption behaviour. Importantly, we cannot control whether the self-employed are working from home. We would expect that the self-employed are more likely to work from home which might partly explain higher expenditures for electricity and heating, and even higher total housing cost, if more space is needed. We would therefore interpret our results with caution and even more so, as the results from the food regressions did not seem to be line with underreporting.

3.4.3 Donation equations using the TPP

Most of the literature estimates underreporting of income for single years using cross-sectional data. With the panel structure of the TPP on the contrary, we are able to identify effects of changes in income variables over time. The disadvantage of standard panel data models in this context is, however, that fixed effects cancel out a big portion of the underreporting effect across income categories which is found using cross-sections. Therefore, with the TPP we estimate both single year (cross-sectional) levels of underreporting, and the effect of changes over time.

In contrast to the direct comparison performed in section 3, we employ different variables because there is no more need to adjust the samples to match the SOEP figures. Hence, to explain donation behaviour we take the household as the level of analysis here. Accordingly, monetary variables are aggregated at the household level. We use the seven different income categories of German income tax law, which are net of costs of obtainment but before other deductions: Income from agriculture and forestry, self-employment, business, dependent employment, capital, rent and lease, and other sources. Total income is summed up over all these categories.

As is standard in the literature, we construct a variable that measures the tax price of giving. The general idea is that due to the progressivity of the income tax, a donation is cheaper for richer households. Reducing their tax base yields higher tax savings at the margin. Therefore, the tax price of giving is defined as $1 - m$, where m is the marginal tax rate. Because the tax rates changed regularly over the observation period[6], there is sufficient intertemporal variation. In contrast to Bittschi

[6]Inter alia, the rates were adjusted to keep the minimum subsistence level tax-free and to account for the so-called "cold progression" through a rightward shift of the schedule. Moreover, there were substantial tax cuts at the beginning of the millenium and the introduction of the so-called "tax on the rich" in 2007, a three percentage point higher rate for taxable incomes exceeding EUR 250,000 for

et al. (2016), we do not assign a value of 1 to non-itemizers who exhibit donations below the standard deduction for special expenses for two reasons. First, as special expenses are the quantitatively most important deduction category in the German income tax, it is fairly unlikely that the standard deduction (which was set at merely EUR 36 for single filers during most of the observation period) would be exceeded only as a result of the donation.Second, assigning the full price of 1 would pertain to all observations that have zero donations (again, implicitly assuming that the standard deduction is not exceeded by other non-donation items), which biases results in case that these observations are kept and not discarded from regressions as missing values.

Furthermore, we construct several dummies for self-employment to test for differential effects of these operationalizations (which were shown to matter in Estonian survey data by Kukk & Staehr 2017). We differentiate by 25% and 50% thresholds of income (as a share of total income) derived from self-employment, business or agriculture and forestry, both seperately by income source and jointly. The remainder of control variables are standard demographics, their choice is mostly dictated by availability. It should be noted that the gender variable in the TPP is flagged as unreliable by Destatis, because values are missing in most cases if a couple is jointly assessed for tax.

A problem that arises when estimating donation regressions, is that many households report not having donated at all. Only using the observations with positive donations may then lead to biased estimates, because people who choose to donate may systematically differ from those who do not. One possibility to account for this possible selection bias is the use of Tobit models. Unfortunately, a consistent estimation of Tobit regressions requires assumptions that are unlikely to be fulfilled in the case at hand. Bittschi et al. (2016) discuss three points: Error terms that are neither normally distributed nor homoscedastic, differential effects of explanatory variables along the intensive and extensive margin (i.e., they may affect the decision whether to donate differently than the decision how much to donate, which a Tobit model assumes to be the same) and the infeasibility of estimating a Tobit panel model with fixed effects because of the incidental parameters problem (i.e., when the length of a panel is small and fixed, the MLE of nonlinear panel models is biased and inconsistent). Therefore, they resort to using fixed effects Poisson models (FEPM), which are borrowed from the trade literature (Silva & Tenreyro 2006). Some of these challenges can alternatively be met by using fixed effects with log-linearized OLS models, or with nonlinear least squares estimation. However, the former requires adjustments to the dependent

singles (this threshold increased over time). For details, see BMF (2020*a*).

variable and the latter is only feasible for cross-sections.

We tackle these points along different avenues. First, for cross-sections of the TPP we control for the selection problem by using a two-step Heckman approach following Torregrosa-Hetland (2020). From a 1st stage probit estimation, the inverse Mills ratio is derived and then included in the 2nd stage OLS and NLS estimations to account for the probability of selection into the positive donations sample. Second, we use both log-linearized OLS with fixed effects and FEPM for the panel dimension of the data. The former is quite robust, and the latter even accounts for nonlinearity and other factors (Silva & Tenreyro 2006 and Bittschi et al. 2016). FEPM was developed for count data, but it works well for continuous data as long as strict exogeneity of the conditional mean is given (Wooldridge 2010). With these specifications, we avoid both the incidental parameters problems and the computational complexity of fitting nonlinear least squares with panel data.

3.4.3.1 Single-year estimations

As in the SOEP case, due to the longitudinal structure of the data we are able to test different measures of income to proxy for (unobservable) permanent income. Following Engström & Hagen (2017), we use the mean for different ranges around the respective year, namely three, five and seven years. These may be interpreted as a medium choice between current yearly income and long-term permanent income.[7] To control for the large share of zero observations for the donations, we estimate a Heckman specification that uses a probit regression as the first stage selection equation. Following Torregrosa-Hetland (2020), we construct a wealth dummy that measures whether a household receives capital gains. It can be argued that this dummy is associated with status considerations of households and thereby only affects the decision whether to donate, but not the amount donated once income is controlled for. It may therefore satisfy the exclusion restriction required for at least one variable in the selection equation.

The first-stage Probit estimation seeks to explain who donates:

$$Prob(s_i = 1|\ln(Y_i), Z_i) = \Phi(\alpha + \beta \ln(Y_i) + \gamma X_i + \delta WD_i + e_i) \tag{3.2}$$

where Y_i are total revenues from all income categories , X_i are controls (included also in the second stage) and WD_i the wealth dummy that indicates capital gains in the

[7] Additionally, we have also tested instruments that are applied to control for permanent income when only current income is available. Using capital income as identified to be the best IV by Engström & Hagen (2017) however did not improve the precision or efficiency of estimations and was therefore discarded.

tax return. From the estimation, the inverse Mills ratio λ is calculated.

For the second stage main equation, we use both ordinary and nonlinear least squares to test different approaches that are common in the literature. While the OLS specification has the advantage of simplicity by requiring merely a dummy for the desired income category, the NLS allows to estimate the underreporting for all non-wage incomes relative to wage income in a single specification.

We apply an OLS specification that includes all income in one variable and identifies differences between households with a self-employment dummy, similarly to the food equation:

$$\ln(don_i) = \alpha + \beta \ln(Y_i) + \gamma SE_i + \delta X_i + \lambda_i + e_i \tag{3.3}$$

where the self-employment dummy SE_i is again operationalized in different ways: 25 vs. 50% share of income from self-employment, business and agriculture and forestry, and all three income categories separately or jointly in a composite dummy. Additionally, for total income Y_i the 7-year-average is applied to approximate permanent income. As a default, we only use balanced sample-observations that are available for 3 years before and after the current year. Moreover, the inverse Mills ratio λ_i from the first stage is included to account for the selection bias. X_i includes all available demographics (i.a. age, no. of children, religion and gender) and the tax price of giving $1 - m$, m being the marginal tax rate.

A second specification is run for all income types seperately using NLS:

$$\ln(don_i) = \alpha + \beta \ln \left[L_i + \sum_{j=1}^{6} k_j y_{ij} + k_7 N_i \right] + \gamma X_i + \lambda_i + e_i \tag{3.4}$$

where L_i is positive income from dependent employment, k_j are coefficients for positive revenues from the other j income categories y_{ij} (self-emplyoment, business, agriculture and forestry, rent and lease, capital and other income) and the absolute value of the sum of all negative incomes N_i, and X_i are controls. $1/k_j$ can be interpreted as the compliance ratio of an income category y_{ij} relative to labour income L_i.

To moreover test for distributional effects, we add an interacted term with a dummy for the top decile in a similar way like Torregrosa-Hetland (2020):

$$\ln(don_i) = \alpha + \beta \ln \left[L_i + \sum_{j=1}^{6} (k_j y_{ij} + k_j^{top} y_{ij} * top10_i) + k_7 N_i \right] + \gamma X_i + \lambda_i + e_i \tag{3.5}$$

where k_j^{top} denotes the coefficient for the *top*10-interacted income categories y_{ij}, and X_i includes a dummy for the *top*10.

Descriptives and results

In the TPP sample that is used for the donation regressions, on average about 36% of tax units have made a donation. This share increased over time, from 34% in 2001 to 42% in 2014. [8] Moreover, the average amount donated and the total income of the respective tax payers have increased as well (see table 3.2). COnsistently and expectedly, donors earn higher incomes than non-donors on average.

Table 3.2: Selected descriptive statistics for the TPP regressions sample

	donation				no donation	
Year	Donation	Total income	No. of obs.	% of all obs.	Total Income	No. of obs.
2001	347	53,890	6,805,963	34	31,117	13,502,416
2002	360	52,017	7,416,485	35	30,398	13,842,170
2003	351	51,786	7,402,453	33	30,301	14,891,212
2004	368	52,974	8,022,086	35	30,208	15,184,054
2005	392	54,903	8,483,991	35	30,296	15,704,795
2006	404	57,335	7,947,386	33	31,605	16,146,839
2007	479	60,165	8,038,604	33	32,788	16,265,129
2008	482	62,013	8,309,649	34	33,730	15,851,930
2009	455	58,026	8,493,536	35	33,035	15,872,698
2010	483	59,546	8,945,669	36	33,627	15,696,125
2011	498	62,218	8,929,930	37	35,961	15,121,196
2012	511	71,761	7,044,483	40	42,668	10,758,719
2013	550	73,552	7,185,375	41	44,200	10,141,568
2014	563	76,171	7,142,893	42	46,418	9,804,254

Note: Donations and total incomes are provided as the average for the respective groups, i.e. for people who donated in column 2 and 3 and for people with zero donations in column 6.

Results of the baseline OLS regression indicate underreporting of a magnitude that has been found for other countries in the literature as well. For the composite dummy that aggregates over the three self-employed categories in German income tax law[9], donations are increased by a range of 20 to 27 % over the years 2004 to 2011. Over time, the effect decreases somewhat from 27 % in 2004 to 21 % in 2011.

These results are affirmed by the nonlinear least squares baseline estimation (see table 3.4 and figure 3.4). In contrast to the previous OLS regression, current revenues

[8] The marked increase after 2011 is probably due to dataset issues, as the sample size drops at the same time. As was mentioned in section 2, due to data delivery issues apparently not all tax units that could be are already included in the TPP sample.

[9] These three are called "profit incomes", in distinction to the remaining "surplus incomes" (dependent employment, rent and lease, capital and other).

Table 3.3: Coefficients of self-employment dummy in OLS baseline

	2004	2005	2006	2007	2008	2009	2010	2011
self-employment	0.37	0.37	0.33	0.31	0.27	0.28	0.29	0.29
business	0.27	0.28	0.29	0.25	0.21	0.23	0.23	0.23
agriculture	-0.15	-0.13	-0.18	-0.18	-0.22	-0.18	-0.22	-0.23
composite	0.27	0.27	0.26	0.23	0.20	0.21	0.21	0.21

Note: All coefficients are significant at the 0.01% level, hence no * indication is given. Composite refers to a dummy that aggregates the three income categories. Dummies for which results are shown were defined using a 50 % threshold of the respective income(s) in total income of the tax unit. In order to have balanced panel observations for the seven-year average of total income, only the years 2004-2011 are included.

for all income categories[10] are included in the equation.Compliance ratios derived from the estimated coefficients generally increase over time, yet the level seems lower indicating higher underreporting than in the OLS specification. For example for business incomes, the ncompliance ratio is estimated at 54% in 2001 and 75% in 2014. This time trend is broadly in line with rising tax morale, as pointed out in the introduction.

A problem for the estimations is posed by the the wealth dummy which is supposed to fulfill the exclusion restriction: The main difficulty seems to be that only a very small fraction of tax units receives relevant capital gains, less than two percent of observations. We have tried to increase this share by including capital gains from different sources (not only those that are categorized as "other income" in German income tax law, but also some from business and self-employment), unfortunately to no avail. Logically, such a small fraction of observations is unlikely to explain the bulk of donating or not decisions. Hence, the robustness of the exclusion restriction is rather questionable: The wealth dummy is only signficant in half of the years in the 1st stage, it is correlated with the error terms in the 2nd stage and when included in the 2nd stage, it is often significant.[11]

As a consequence, we cannot assume that the exclusion restriction is fulfilled, so a possible selection into donating may not be fully explained by our approach. The resulting underreporting estimates should therefore be interpreted with caution. They only reliably compare taxpayers with donations on their tax return, not necessarily all tax payers. The former may not be representative for the latter.

[10]It should be recalled at this point, that due to the introduction of the withholding tax on capital incomes, the income tax data on capital revenues is seriously flawed from 2009 onwards.

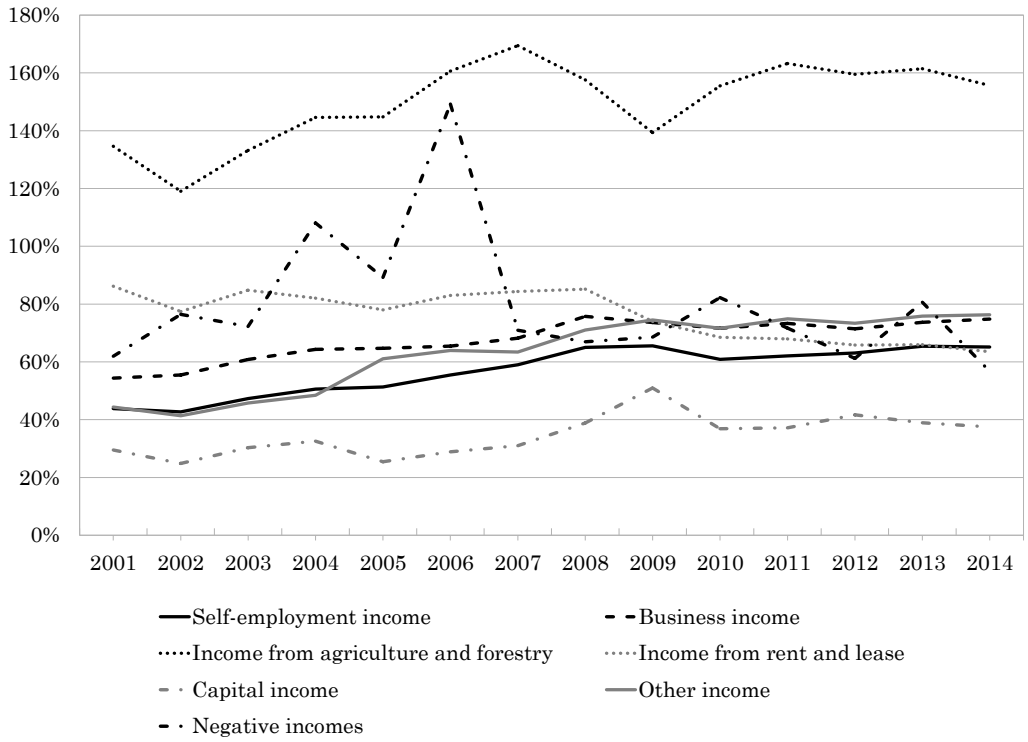
[11]So far, we have not found a better "instrument", because the range of possible variables in the TPP is limited. Any advice on this point is greatly appreciated.

Table 3.4: Coefficients and compliance ratios for some income categories, NLS baseline

Coef. k_j	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
self-employment	2.28	2.34	2.11	1.98	1.95	1.80	1.69	1.54	1.53	1.64	1.61	1.59	1.53	1.53
compliance ratio	44%	43%	47%	51%	51%	55%	59%	65%	66%	61%	62%	63%	65%	65%
business	1.84	1.80	1.64	1.55	1.55	1.53	1.47	1.32	1.36	1.40	1.36	1.40	1.36	1.34
compliance ratio	54%	55%	61%	64%	65%	65%	68%	76%	74%	72%	73%	71%	74%	75%
agriculture & forestry	0.74	0.84	0.75	0.69	0.69	0.62	0.59	0.63	0.72	0.64	0.61	0.63	0.62	0.64
compliance ratio	135%	119%	133%	145%	145%	161%	169%	158%	139%	156%	163%	160%	161%	156%
rent and lease	1.16	1.29	1.18	1.22	1.28	1.21	1.19	1.17	1.35	1.46	1.47	1.52	1.52	1.57
compliance ratio	86%	77%	85%	82%	78%	83%	84%	85%	74%	69%	68%	66%	66%	64%
capital	3.39	4.02	3.30	3.07	3.93	3.46	3.23	2.57	1.96	2.71	2.69	2.40	2.57	2.66
compliance ratio	29%	25%	30%	33%	25%	29%	31%	39%	51%	37%	37%	42%	39%	38%
other	2.25	2.42	2.18	2.06	1.64	1.56	1.58	1.41	1.34	1.40	1.33	1.36	1.32	1.31
compliance ratio	44%	41%	46%	48%	61%	64%	63%	71%	75%	72%	75%	73%	76%	76%
negative	1.61	1.31	1.38	0.93	1.12	0.67	1.41	1.49	1.46	1.22	1.39	1.63	1.24	1.76
compliance ratio	62%	76%	72%	108%	89%	149%	71%	67%	69%	82%	72%	61%	81%	57%

Note: All coefficients are significant at the 0.01% level, hence no * indication is given. Compliance ratios are given by $1/k_j$.

Figure 3.4: Compliance ratios derived from NLS baseline



Note: The percentage values indicate the compliance with respect to income from dependent employment, which is assumed to be correctly reported. Source: Own calculations based on the TPP.

Moreover, given the limited availability of demographic variables in the TPP dataset, it is likely that our estimation suffers from omitted variables bias. For instance, unobserved heterogeneity with respect to earners of self-employment and business income could bias our estimate of underreporting. As Bittschi et al. (2016) argue, these individuals could be more likely to be asked to donate (solicitation effect). Also, donating to charity may be a behaviour expected from them through social norms or out of business considerations (marketing). Hence, systematic differences between the dependly employed and self-employed are likely not fully captured by the controls available in the dataset.

Therefore, when interpreted as underreporting of income, these results should be viewed as an upper bound for tax evasion of earners of income from self-employment or business.

For the distributional regression, we do find a higher noncompliance for the top decile in almost all income categories (see table 3.16 in the appendix). For revenues from business and rent and lease, the compliance ratio of the Top10 is substantially lower. For self-employment, the additional effect for the Top10 becomes insignificant from 2008 onwards.

3.4.3.2 Panel estimations

When exploiting the panel dimension of the data we cannot expect that the errors are uncorrelated with the explanatory variables, which is why using fixed effects is appropriate. However, this entails the disadvantage that a lot of the variation between individuals that is interpreted as underreporting in the cross-section, is lost. This may explain why Bittschi et al. (2016) report a rather small effect of relevant income categories on donations. They show that a 10% increase in business income is associated with a 0.76% increase in donations, which may be interpreted as tax evasion[12].

As a simple baseline specification, we run several fixed-effects OLS specifications:

$$\ln(don_{it}) = \alpha + \beta \ln Y_{it} + \gamma X_{it} + FE_i + e_i \tag{3.6}$$

where Y_{it} is positive income from different categories, X_{it} are controls including the tax price of giving and FE_i are individual fixed effects. Alternatively,

$$\ln(don_{it}) = \alpha + \beta \ln Y_{it} + \gamma SE_{it} + \delta X_{it} + FE_i + e_i \tag{3.7}$$

where Y_{it} is total income again and SE_{it} is a self-employment dummy. As before, i denotes the individual, while now t additionally indicates the year.

Following Bittschi et al. (2016), we also estimate a fixed-effects Poisson model:

$$E(don_{it}|Z_{it}, FE_i) = exp(\alpha + \beta Y_{it} + \gamma X_{it} + T_t + FE_i + e_i) \tag{3.8}$$

where Z_{it} are all covariates, Y_{it} positive income from different categories and T_t time fixed effects.

Results

Coefficients from the OLS fixed-effects panels show that the estimated effect of self-employment on donations is slightly higher when self-employment is defined more broadly (25% share in total revenues rather than 50%). In line with our expectations, using the 7-year-average instead of current income decreases the self-employment coefficient size substantially (see table 3.5). Inversely, the importance of income rises. Full regression tables are provided in table 3.12 in the appendix.

[12]Bittschi et al. (2016) note that one may alternatively interpret the effect as that of the respective income types on donations, as the fixed effects arguably account for time-invariant tax evading behaviour.

Table 3.5: Coefficients from OLS panel

Total income	0,093
SE dummy, 50% share	0.164
Total income	0,093
SE dummy, 25% share	0.180
Total income, 7-year average	0.181
SE dummy, 50% share	0.086
Total income, 7-year average	0.181
SE dummy, 25% share	0.108

Note: Full regression tables are provided in table 3.12 in the appendix.

When applying the fixed-effects poisson model over the whole sample period, the effect of self-employment and business incomes on donations is much smaller. Descriptives statistics and results are given in tables 3.13 and 3.14 in the appendix. For the specification, we largely follow Bittschi et al. (2016). We have a longer sample period at our disposal, extending theirs by eight years. Moreover, when comparing the descriptive statistics, we probably employ the more representative sample: Our average income (as well as the mean donation) is much lower and closer to the population average in tax statistics. Consequently, we arrive at even lower coefficients: For a 10% increase in business income, we estimate a 0.39% increase in donations. For income from self-employment, said effect is roughly half in size. This difference to Bittschi et al. (2016) strengthens the interpretation that richer households drive the effect. Expectedly, the effect of changes in income over time is much smaller than the level effect in any yearly cross-section.

3.4.3.3 Macro implications

Based on the cross-sectional estimates from section 3.4.3.1, we gauge the losses incurred by the public coffers. As we have identified several issues that could bias our estimation of tax evasion upwards, we select the more conservate figures. For example, 7-year average incomes are preferred over 5-year, 3-year or current year incomes and the 50% share of self-employment income is selected over the 25% share.

We first perform some simple back-of-envelope calculations by applying the coefficients estimated in the cross-section regressions to assessed tax due [13]. This requires some simplifying assumptions: Firstly, we have to assume that the share and income category in total revenues is equivalent to its share in assessed income tax. Implic-

[13]For an application to published income tax statistics, see table 3.15 in the appendix.

itly, this means that the average tax rate is applied to the additional income which goes unreported. Secondly, the coefficient for composite and separate self-employment dummies must be assumed to reflect underreporting with respect to the respective income categories (self-employment, business and agriculture and forestry). Thirdly, it is assumed that income from dependent employment is correctly reported, as the compliance of other income categories is measured against it.

Additionally, we also exploit the micro dimension of the TPP, by applying the estimated coefficients from the different cross-section regressions directly at the individual tax units. We recalculate taxable income after deductions by adding the estimated underreported amount, and apply the tax schedule. We assume that only 75% of the underreported amount could be taxed, because taxpayers may be eligible for deductions or decrease their earnings by working less when facing a higher tax burden. By comparing the resulting tax due with originally assessed tax due, we get an alternative result for the tax loss. This method has the advantage of better reflecting the progressivity of the tax schedule, as well as differential estimates for taxpayers below and above the richest 10%.

Importantly, all macro estimates are static and do not ar at best partly and sweepingly consider behavioural responses of taxpayers.

The resulting tax losses are depicted in table 3.6, where columns 1, 3 and 5 show the simpler and columns 2, 4 and 6 the more nuanced estimates. Unsurprisingly, estimates based on current-year nonlinear least squares for all income categories are somewhat higher than those based on 7-year average income and a self-employment dummy. This holds also when only revenues from self-employment, business and agriculture are considered for the NLS estimates (table 3.15 in the appendix). Moreover, when comparing the two methodologies, taking into account the progressivity of the tax scheduale matters, as it increases the tax losses incurred.

On the time axis, the declining magnitude of income underreporting is confirmed. Tax losses in the NLS baseline simple estimate (column 5) decrease from EUR 21.3 bn in 2001 to EUR 15.8 bn in 2014. Relative to assessed income tax, this amounts to 12.0% in 2001 and 6.1% in 2014. If the avoided amount is included in the denominator, the implied tax gap is 10.7% in 2001 and 5.7% in 2014. In the estimate accounting for progressivity of the income tax (column 6), the amount is much higher but drops from EUR 70.2 bn in 2001 to EUR 32.4 bn in 2014. This implies a share of assessed income taxes of 39.6% in 2001 and 12.5% in 2014, and a tax gap relative to "true" tax due of 28.4% in 2001 and 11.1% in 2014. We can only speculate about factors that may explain this time trend: Possible explanations include rising tax morale, policy measures and measurement problems, which of course are neither mutually exclusive

nor a finite list.

It should be noted that the tax gaps are calculated relative to total assessed income tax, i.e. including wage tax levied on income from dependent employment, which is assumed to be 100% correctly reported in the FS-methodology. Hence, the estimated tax loss for the earlier years of our sample period, say up to the financial crisis of 2008, can be considered relatively large.

Table 3.6: Estimated tax losses from underreporting

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS composite SE dummy		OLS separate SE dummies		NLS all income categories	
	simple average	tax schedule	simple average	tax schedule	simple average	tax schedule
2001					21.3	70.2
2002					18.8	63.7
2003					16.6	50.6
2004	8.4	11.4	9.4	12.1	16.2	41.9
2005	9.5	12.5	10.4	13.0	18.1	49.4
2006	9.8	12.7	11.0	13.4	18.7	44.5
2007	9.9	12.7	11.0	13.2	20.2	50.3
2008	8.7	11.6	9.8	12.2	18.2	41.5
2009	8.1	10.9	9.3	11.7	13.4	28.3
2010	8.7	11.9	10.0	12.8	15.6	34.9
2011	9.5	12.7	10.9	13.6	16.1	35.0
2012					15.2	32.5
2013					14.5	29.1
2014					15.8	32.4

Note: SE = self-employment. Estimates based on OLS and NLS cross-section regressions using the TPP, as described in section 4.3.1. Coefficients from these estimations are applied to the assessed income tax for the "simple average" columns. For the "tax schedule" results, the coefficients are applied to taxable income and assessed tax is recalculated taking into account the tax schedule. For more details on the methodology, see text.

3.5 Conclusion

In this study, we combine different approaches to analyse the extent of income underreporting by German taxpayers. By comparing adjusted samples from the Taxpayer Panel and the Socioeconomic Panel, we find that incomes from self-employment and rent and lease reported to tax authorities are on average much lower than those reported in the anonymous survey. For wage incomes, in contrast, the discrepancy is negative and smaller. We furthermore find that the discrepancy for self-employment incomes increases along the income distribution. However, as income underreporting

to tax authorities might be only one of several possible explanations for the observed discrepancies, we also employ econometric approaches to estimate the degree of underreporting by non-wage earners.

Based on SOEP data, we estimate a food equation, relating the households' food expenditures to their income and other control variables. If the predicted food expenditures of self-employed differ significantly from the predicted food expenditures of wage earners, this might - everything else being equal - be interpreted as income underreporting by the self employed. For our data we do not find any signifcant differences in food expenditures between wage earners and self-employed and thus no indication of income underreporting. In contrast, we do find that self-employment is associated with higher average expenditures on electricity, heating, and warm water and with higher total housing cost. Provided that only a negligible share of self-employed works from home, this might indicate underreporting of self-employment income even in the SOEP and increase the gap between self-employent incomes between SOEP and TPP even more. However, the estimated coefficients are relatively small and the food regressions do not support such an interpretation.

As a third approach, we regress individuals' donations on their income and other control variables using the Taxpayer Panel. Results suggest that in particular receivers of income from self-employment and business donate more on average and that their propensity to donate out of the respective income is higher than the propensity to donate out of wage income. This might be interpreted as indication of income underreporting under the assumption that - ceteris paribus - only the level of income but not the source of income should determine taxpayers' preferences for making charitable donations. Unfortunately though, we are not fully able to control for heterogeneity with respect to receivers of different income types, because the tax micro data only contain a limited set of sociodemographics. Nonetheless, these findings call into question the equality of tax collection by income source and hence the progressivity of the tax schedule, because self-employment and business incomes are more concentrated at the top of the income distribution. This is in line with the literature, which tends to find underreporting of self-employed incomes in the range of 15-40% and increasing tax noncompliance with rising income. We estimate tax losses from income underreporting at EUR 15.8 to 32.4 bn in 2014, which implies a tax gap relative to true income tax due of 5.7 to 11.1%.

3.A Appendix

Table 3.7: Food-income ratio and key control variables

Definition of self-employed	Self-employment status	Number of households	Food-income ratio	Household income	Age of household head	Number of children
A	0	4025	0.16	3077	45	0.54
	1	600	0.16	4133	48	0.63
B	0	4025	0.16	3077	45	0.54
	1	431	0.16	4096	48	0.59
C	0	4142	0.16	3088	45	0.54
	1	487	0.16	4124	47	0.63

Note: Definitions of self-employed households: A: At least one person in the household defines herself as self-employed; B: The household head defines herself as self-employed; C: more than 25% self-employment in total household income. Source: SOEP_v35, own calculations

Table 3.8: Housing cost-income ratios and key control variables, 2013

Def. of self-employment	Self-employed	N	EHW-income ratio	total housing cost - income ratio	household income (3-year avg.)	age	number of children
A	0	9,586	0.074	0.32	2971	44	0.97
	1	801	0.071	0.3	4242	49	0.95
B	0	9,586	0.074	0.32	2971	44	0.97
	1	542	0.075	0.29	4138	49	0.91
C	0	9,587	0.074	0.32	2970	44	0.97
	1	633	0.072	0.31	4219	49	0.89

Note: Definitions of self-employed households: A: At least one person in the household defines herself as self-employed; B: The household head defines herself as self-employed; C: more than 25% self-employment in total household income. EHW are expenditures for electricity, heating and hot water. Age refers to the oldest working-age member of the household. Source: SOEP_v35, own calculations

Table 3.9: Food regressions

Dep var: ln (food expenditure)	(1)	(2)	(3)
	self-employment def. A b/se	self-employment def. B b/se	self-employment def. C b/se
self-employed	-0.013	-0.018	-0.017
	(0.020)	(0.023)	(0.022)
ln (HH income)	0.450***	0.455***	0.446***
	(0.017)	(0.018)	(0.017)
age	0.010***	0.010***	0.010***
	(0.001)	(0.001)	(0.001)
female	0.012	0.013	0.015
	(0.014)	(0.014)	(0.014)
n_children	0.128***	0.127***	0.127***
	(0.009)	(0.009)	(0.009)
n_adults	0.134***	0.131***	0.135***
	(0.010)	(0.010)	(0.010)
region2	0.000	-0.002	0.001
	(0.015)	(0.015)	(0.015)
region3	-0.134***	-0.135***	-0.134***
	(0.018)	(0.019)	(0.018)
education	-0.003	-0.003	-0.001
	(0.003)	(0.003)	(0.003)
married	0.045*	0.050*	0.048*
	(0.022)	(0.022)	(0.022)
widowed	-0.133*	-0.129*	-0.127*
	(0.052)	(0.052)	(0.052)
divorced	-0.034	-0.031	-0.036
	(0.024)	(0.024)	(0.024)
credit	-0.002	-0.003	-0.005
	(0.015)	(0.015)	(0.015)
rent	0.037*	0.039*	0.037*
	(0.015)	(0.015)	(0.015)
cons	1.623***	1.598***	1.636***
	(0.130)	(0.134)	(0.130)
r2	0.406	0.406	0.408
N	4573	4406	4577

Note: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. The variables age, female, education, married, widowed, and divorced refer to the oldest working-age household member. Education is proxied by the number of years of education. Rent is a dummy variable indicating whether the household is a renter i.e. not a home owner. The variable credit indicates whether the household is paying off a loan. Regressions includes only working age individuals and those households which can consistently be defined as either wage earners or self-employed between 2009 and 2011. Date source: SOEP_v35, own calculations.

Table 3.10: Housing-cost regressions - electricity, heating and hot water, 2013

Dep var: ln (EWH expenditure)	(1)	(2)	(3)
	self-employment def. A b/se	self-employment def. B b/se	self-employment def. C b/se
self-employed	0.099***	0.111***	0.098***
	(0.014)	(0.017)	(0.016)
ln (HH income)	0.151***	0.148***	0.149***
	(0.010)	(0.010)	(0.010)
age	0.006***	0.006***	0.006***
	(0.000)	(0.000)	(0.000)
female	0.042***	0.043***	0.043***
	(0.008)	(0.008)	(0.008)
n_children	0.096***	0.097***	0.097***
	(0.004)	(0.004)	(0.004)
n_adult	0.136***	0.139***	0.137***
	(0.006)	(0.006)	(0.006)
region2	0.016*	0.017*	0.016+
	(0.008)	(0.008)	(0.008)
region3	-0.005	-0.003	-0.006
	(0.011)	(0.011)	(0.011)
education	-0.010***	-0.010***	-0.010***
	(0.002)	(0.002)	(0.002)
married	0.045***	0.044***	0.045***
	(0.012)	(0.012)	(0.012)
widowed	0.141***	0.141***	0.141***
	(0.029)	(0.029)	(0.029)
divorced	0.055***	0.055***	0.055***
	(0.013)	(0.013)	(0.013)
credit	0.022**	0.022**	0.020*
	(0.008)	(0.008)	(0.008)
rent	-0.076***	-0.075***	-0.077***
	(0.009)	(0.009)	(0.009)
_cons	3.431***	3.447***	3.444***
	(0.074)	(0.075)	(0.075)
r2	0.288	0.286	0.288
N	10155	9898	9990

Note: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. EWH is expenditure for electricity, heating and hot water. The variables age, female, education, married, widowed, and divorced refer to the oldest working-age household member. Education is proxied by the number of years of education. Rent is a dummy variable indicating whether the household is a renter i.e. not a home owner. The variable credit indicates whether the household is paying off a loan. Regressions includes only working age individuals and those households which can consistently be defined as either wage earners or self-employed between 2012 and 2014. Date source: SOEP_v35, own calculations.

Table 3.11: Housing-cost regressions - total housing costs, 2013

Dep var: ln (total housing cost)	(1)	(2)	(3)
	self-employment def. A	self-employment def. B	self-employment def. C
	b/se	b/se	b/se
self-employed	0.053**	0.041*	0.032+
	(0.018)	(0.021)	(0.020)
ln (HH income)	0.353***	0.353***	0.350***
	(0.013)	(0.013)	(0.013)
age	-0.002***	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)
female	0.024*	0.026*	0.025*
	(0.010)	(0.010)	(0.010)
n_children	0.082***	0.082***	0.083***
	(0.005)	(0.005)	(0.005)
n_adult	0.045***	0.046***	0.047***
	(0.007)	(0.008)	(0.007)
region2	-0.017+	-0.017	-0.018+
	(0.010)	(0.010)	(0.010)
region3	-0.140***	-0.141***	-0.138***
	(0.014)	(0.014)	(0.014)
education	0.009***	0.008***	0.008***
	(0.002)	(0.002)	(0.002)
married	0.083***	0.076***	0.081***
	(0.015)	(0.015)	(0.015)
widowed	0.135***	0.127***	0.133***
	(0.036)	(0.036)	(0.036)
divorced	0.105***	0.099***	0.104***
	(0.016)	(0.016)	(0.016)
credit	0.084***	0.081***	0.083***
	(0.010)	(0.010)	(0.010)
rent	0.178***	0.179***	0.177***
	(0.011)	(0.011)	(0.011)
_cons	3.454***	3.452***	3.480***
	(0.093)	(0.093)	(0.094)
r2	0.223	0.222	0.219
N	10155	9898	9990

Note: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. total housing cost includes expenditure for electricity, heating and hot water, additional cost, rent, amortizations, maintenance. The variables age, female, education, married, widowed, and divorced refer to the oldest working-age household member. Education is proxied by the number of years of education. Rent is a dummy variable indicating whether the household is a renter i.e. not a home owner. The variable credit indicates whether the household is paying off a loan. Regressions includes only working age individuals and those households which can consistently be defined as either wage earners or self-employed between 2012 and 2014. Date source: SOEP_v35, own calculations.

Table 3.12: Panel OLS regressions, TPP 2001-2014

dependent variable: ln (donation)	(1)	(2)	(3)	(4)
	current income		7-year average income	
	50% SE share	25% SE share	50% SE share	25% SE share
ln (pos. total income)	0.093***	0.093***	0.181***	0.181***
	(0.001)	(0.001	(0.003	(0.003
self-employed	0.164***	0.180***	0.086***	0.108***
	(0.005)	(0.004)	(0.005)	(0.005)
ln (Taxprice giving)	-0.944***	-0.945***	-1.185***	-1.185***
	(0.009	(0.009)	(0.010)	(0.010)
Age	0.074***	0.074***	0.061***	0.061***
	(0.001)	(0.001)	(0.002)	(0.002)
Age^2	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
1 child	0.079***	0.079***	0.053***	0.053***
	(0.004)	(0.004)	(0.004)	(0.004)
2 children	0.176***	0.176***	0.112***	0.112***
	(0.005)	(0.005)	(0.006)	(0.006)
3 children	0.249***	0.249***	0.182***	0.182***
	(0.009)	(0.009)	(0.011)	(0.011)
4 or more children	0.315***	0.316***	0.229***	0.229***
	(0.016)	(0.016)	(0.020)	(0.020)
Gender	-0.277***	-0.275***	-0.280***	-0.279***
	(0.016)	(0.016)	(0.019)	(0.019)
East Germany	-0.178***	-0.178***	-0.173***	-0.174***
	(0.012)	(0.012)	(0.014)	(0.014)
No religion	0.028***	0.028***	0.013+	0.013+
	(0.007)	(0.007)	(0.008)	(0.008)
Catholic	0.138***	0.137***	0.128***	0.128***
	(0.009)	(0.009)	(0.009)	(0.009)
Protestant	0.084***	0.083***	0.071***	0.071***
	(0.007)	(0.007)	(0.007)	(0.007)
Other religion	0.036	0.035	0.009	0.008
	(0.101)	(0.101)	(0.132)	(0.132)
Constant	-1.948***	-1.941***	-2.560***	-2.556***
	(0.026)	(0.026)	(0.045)	(0.045)
Fixed effects	Yes	Yes	Yes	Yes
R2	0.238	0.239	0.267	0.269
N	16,062,058	16,062,058	9,851,091	9,851,091

Note: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. Robust standard errors are given in parentheses below the coefficients. A composite self-employment (SE) dummy is used, which aggregates information for revenues from self-employment, agriculture and forestry and business, depending on their share in total revenues.

Table 3.13: Descriptive statistics, TPP FEPM regressions

Variable	Mean	Std. dev.
donation	272.9	10860.6
Log positive income from		
Employment	8.782	3.938
Self-employment	0.825	2.711
Business	1.434	3.410
Agriculture	0.244	1.466
Rent and lease	1.261	3.001
Capital	1.092	2.749
Other sources	1.872	3.601
Log tax price of giving	-0.303	0.157
Dummy variables		
Age 15-24	0.043	0.202
Age 25-34	0.133	0.340
Age 35-44	0.236	0.425
Age 45-54	0.252	0.434
Age 55-64	0.186	0.389
Age 65 and above	0.159	0.366
Single female	0.004	0.060
Single male	0.010	0.102
Married	0.622	0.485
East Germany	0.100	0.300
Catholic	0.340	0.474
Protestant	0.263	0.440
Other religion	0.000	0.018
No religion	0.327	0.469
One child	0.175	0.380
Two children	0.177	0.382
Threechildren	0.047	0.212
Four or more children	0.013	0.111
Year 2002	0.070	0.255
Year 2003	0.073	0.260

Year 2004	0.075	0.264
Year 2005	0.078	0.268
Year 2006	0.078	0.268
Year 2007	0.078	0.268
Year 2008	0.077	0.267
Year 2009	0.077	0.267
Year 2010	0.077	0.266
Year 2011	0.075	0.264
Year 2012	0.059	0.236
Year 2013	0.058	0.234
Year 2014	0.057	0.232

Source: TPP 2001-2014, own calculations. Monetary values have been converted to 2014 Euros. The number of observations is 11,837,397 for all variables.

Table 3.14: Panel FEPM regressions, TPP 2001-2014

dependent variable: ln (donation)	(1)	(2)	(3)
Log income from employment			
Positive Income	0.000 (0.001)	0.001 (0.001)	0.000 (0.001)
Abs. neg. income			0.000 (0.008)
Log income from self-employment			
Positive Income	0.020*** (0.001)	0.022*** (0.001)	0.020*** (0.002)
Abs. neg. income			0.004 (0.003)
Log income from business			
Positive Income	0.039*** (0.002)	0.043*** (0.002)	0.047*** (0.002)
Abs. neg. income			0.016*** (0.002)
Log income from agriculture			
Positive Income	-0.002 (0.004)	0.001 (0.004)	-0.009 (0.006)

Abs. neg. income			-0.013 (0.011)
Log income from rent and lease			
Positive Income	0.005** (0.001)	0.014*** (0.001)	0.016*** (0.003)
Abs. neg. income			0.014** (0.004)
Log income from capital			
Positive Income	0.005*** (0.001)	0.005*** (0.001)	0.005*** (0.001)
Abs. neg. income			0.007 (0.006)
Log income from other income			
Positive Income	0.008*** (0.002)	0.008*** (0.002)	0.008*** (0.002)
Abs. neg. income			0.006 (0.004)
Total abs. neg. income		0.015*** (0.002)	
log tax price of giving	-2.190*** (0.018)	-2.198*** (0.018)	-2.192*** (0.018)
Dummy variables			
Age 15-24	-0.152*** (0.012)	-0.145*** (0.012)	-0.143*** (0.012)
Age 25-34	0.198*** (0.013)	0.192*** (0.013)	0.191*** (0.012)
Age 35-44	0.323*** (0.018)	0.304*** (0.019)	0.302*** (0.019)
Age 45-54	0.317*** (0.024)	0.291*** (0.026)	0.288*** (0.026)
Age 55-64	0.282*** (0.027)	0.253*** (0.028)	0.251*** (0.028)
Age 65 and above	0.326*** (0.029)	0.302*** (0.030)	0.300*** (0.030)
Single female	-0.434*** (0.030)	-0.421*** (0.030)	-0.421*** (0.030)
Single male	-0.411*** (0.020)	-0.397 *** (0.020)	-0.394*** (0.020)
Married	0.446*** (0.016)	0.434*** (0.016)	0.435*** (0.016)
East Germany	-0.197*** (0.056)	-0.197*** (0.056)	-0.199*** (0.056)
Catholic	-0.099*** (0.025)	-0.098*** (0.025)	-0.098*** (0.024)
Protestant	-0.043*** (0.012)	-0.043*** (0.012)	-0.044*** (0.012)
Other religion	0.000 (0.068)	0.001 (0.067)	0.002 (0.067)
No religion	0.024 (0.019)	0.023 (0.019)	0.023 (0.018)
One child	0.060*** (0.011)	0.058*** (0.011)	0.058*** (0.011)
Two children	0.108*** (0.013)	0.106*** (0.013)	0.105*** (0.013)
Three children	0.166*** (0.018)	0.165*** (0.018)	0.165*** (0.017)
Four or more children	0.183*** (0.020)	0.183*** (0.020)	0.183*** (0.020)

Table 3.15: Estimated tax losses (EUR bn), based on published tax statistics

	OLS composite SE	OLS separate SE	NLS baseline		
			only SE	total	tax gap (%)
2001			8.7	13.4	7.59
2004	6.5	7.1	7.7	11.0	6.06
2007	7.7	8.3	9.3	15.2	7.19
2010	6.8	7.5	12.4	17.1	8.29
2012			11.7	16.6	7.12
2013			11.0	16.0	6.52
2014			11.6	16.9	6.52

Note: SE = self-employment. Estimates based on OLS and NLS cross-section regressions using the TPP, as described in section 4.3.1. Coefficients from these estimations are applied to published income tax statistics. The estimated tax gap is based on total tax losses compared to total assessed income tax.

Year 2002	0.188*** (0.005)	0.189*** (0.005)	0.189*** (0.005)
Year 2003	0.120*** (0.006)	0.122*** (0.006)	0.122*** (0.006)
Year 2004	0.184*** (0.009)	0.186*** (0.009)	0.186*** (0.009)
Year 2005	0.295*** (0.006)	0.298*** (0.006)	0.298*** (0.006)
Year 2006	0.245*** (0.008)	0.248*** (0.008)	0.249*** (0.008)
Year 2007	0.373*** (0.012)	0.377*** (0.012)	0.379*** (0.012)
Year 2008	0.399*** (0.010)	0.404*** (0.010)	0.405*** (0.010)
Year 2009	0.412*** (0.010)	0.418*** (0.010)	0.420*** (0.010)
Year 2010	0.519*** (0.012)	0.526*** (0.012)	0.527*** (0.012)
Year 2011	0.529*** (0.013)	0.535*** (0.013)	0.536*** (0.013)
Year 2012	0.473*** (0.012)	0.479*** (0.012)	0.480*** (0.012)
Year 2013	0.561*** (0.012)	0.568*** (0.012)	0.569*** (0.012)
Year 2014	0.573*** (0.021)	0.579*** (0.021)	0.580*** (0.021)
Observations	11,837,397	11,837,397	11,837,397
Log pseudolikelihood	-2,05e+10	-2,05e+10	-2,05e+10

Note: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. Robust standard errors are given in parentheses behind the coefficients. Column 1 gives a baseline without negative incomes, columns 2 includes the latter as one aggregated variable while column 3 includes negative incomes for every income category.

Table 3.16: Coefficients and compliance ratios for some income categories, NLS with Top10-interaction

Coef. k_{tj}	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
self-employment	2.57	2.85	2.46	2.31	2.29	2.06	1.91	1.68	1.70	1.83	1.82	1.74	1.65	1.69
compliance ratio	39%	35%	41%	43%	44%	48%	52%	60%	59%	55%	55%	57%	60%	59%
business	1.73	1.73	1.50	1.47	1.45	1.41	1.37	1.24	1.29	1.32	1.30	1.35	1.30	1.28
compliance ratio	58%	58%	67%	68%	69%	71%	73%	81%	78%	76%	77%	74%	77%	78%
agriculture and forestry	0.68	0.77	0.67	0.61	0.62	0.56	0.56	0.60	0.68	0.61	0.58	0.63	0.60	0.63
compliance ratio	146%	130%	149%	163%	161%	177%	178%	165%	147%	163%	171%	158%	166%	159%
rent and lease	1.05	1.22	1.08	1.16	1.22	1.11	1.12	1.12	1.31	1.41	1.44	1.48	1.51	1.55
compliance ratio	95%	82%	93%	86%	82%	90%	89%	89%	76%	71%	70%	68%	66%	64%
capital	4.17	4.69	3.37	3.22	4.45	3.63	3.39	2.72	2.07	3.04	2.94	2.48	2.69	2.73
compliance ratio	24%	21%	30%	31%	22%	28%	29%	37%	48%	33%	34%	40%	37%	37%
other	2.17	2.41	2.10	2.06	1.63	1.53	1.55	1.37	1.31	1.37	1.30	1.33	1.27	1.26
compliance ratio	46%	42%	48%	48%	61%	65%	65%	73%	76%	73%	77%	75%	79%	79%
negative	2.02	1.59	1.58	1.09	1.32	0.72	1.59	1.61	1.59	1.42	1.57	1.71	1.30	1.84
compliance ratio	50%	63%	63%	92%	76%	139%	63%	62%	63%	70%	64%	58%	77%	54%
self-employment*Top10	0.61	0.75	0.76	0.55	0.17$^{\diamond}$	0.27	0.16	0.04$^{\diamond}$	0.03$^{\diamond}$	0.18	0.07$^{\diamond}$	0.08$^{\diamond}$	0.05$^{\diamond}$	0.02$^{\diamond}$
compliance ratio	31%	28%	31%	35%	41%	43%	48%	58%	58%	50%	53%	55%	59%	59%
business*Top10	2.25	2.63	2.62	1.83	1.53	1.51	1.07	0.60	0.70	0.85	0.65	0.47	0.40	0.39
compliance ratio	25%	23%	24%	30%	33%	34%	41%	54%	50%	46%	51%	55%	59%	60%
agriculture and forestry*Top10	-0.03$^{\diamond}$	0.13$^{\diamond}$	0.07$^{\diamond}$	0.10$^{\diamond}$	0.01$^{\diamond}$	-0.01$^{\diamond}$	-0.14	-0.04$^{\diamond}$	0.05$^{\diamond}$	-0.05$^{\diamond}$	-0.03$^{\diamond}$	-0.11$^{\diamond}$	-0.02$^{\diamond}$	-0.03$^{\diamond}$
compliance ratio	152%	112%	136%	140%	158%	179%	237%	179%	138%	177%	182%	190%	172%	167%
rent and lease*Top10	1.40	1.44	1.09	0.90	0.90	1.14	0.51	0.46	0.78	1.24	0.94	0.57	0.35	0.41
compliance ratio	41%	38%	46%	48%	47%	44%	61%	63%	48%	38%	42%	49%	54%	51%
capital*Top10	0.05$^{\diamond}$	8.93	10.85	9.46	7.37	5.64	5.13	1.28	7.47	6.07	7.55	5.95	3.81	4.91
compliance ratio	24%	7%	7%	8%	8%	11%	12%	25%	10%	11%	10%	12%	15%	13%
other*Top10	20.98	19.54	17.41	13.44	9.83	11.47	7.90	5.86	5.54	6.77	5.53	4.33	3.76	3.90
compliance ratio	4%	5%	5%	6%	9%	8%	11%	14%	15%	12%	15%	18%	20%	19%

Note: For readibility and as most coefficients are highly significant, in contrast to standard *-notation it is noted by $\diamond$ when a coefficient is insignificant (the unmarked coefficients are significant at the 0.1% level at least). Compliance ratios are given by $1/k_j$ for the non-interacted incomes. For the Top10-interacted ones, the compliance ratio is given by $1/(k_j + k_{tj})$.

Appendix A

On income tax avoidance: Technical Appendix

This technical appendix supplements chapter 2 “On income tax avoidance - a new micro data model for the German case". It aims to provide more detail concerning the data preparation and the model of the German income tax.

A.1 Data preparation

The Scientific-Use-File of the 2013 Income and Consumption Survey (EVS) that was applied for my analysis, is an 80% subsample of the full data. The only tangible disadvantage compared to the on-site-access version consists of reduced information for the regional variables: Due to confidentiality concerns, researchers have to pick between either a ranked population size of the community (EF4) or a ranked region density description (EF5). I chose the former.

A.1.1 Transformation to yearly values

As was mentioned in the paper, the EVS underwent some change over time: Before 1998, households were surveyed over the whole year. To reduce drop-out rates and improve representativeness, the surveying period was shortened to one quarter, starting in 1998. For my analysis, this introduces the challenge to transform the data from the quarterly to the yearly level.

For high-frequency variables like salaries, a simply multiplication is sufficient[1].

[1]The Federal Statistical Office does not apply any other method to get yearly values, according to personal correspondence on December 7, 2017.

Therefore, appropriate code is applied using R.[2]

A.1.1.1 The problem with low frequency variables

However, biases may arise if simply quadrupling all variables for two reasons: First, some items are observed less frequently than others in general, i.e. a purchase of a durable, say a dishwasher, is unlikely to take place four times a year. Second, there may be seasonal biases: As is documented in the paper, some items are more frequently observed for housholds surveyed in one quarter compared to those who participated in another. Adding to that, the average values may differ as well, as for instance Christmas bonuses do not only increase the occurrence of any lump-sum payment in Q4, but also the amount. Further problems arising from the quarterly periodicity of the EVS have been spelled out in more detail by Becker (2014, p. 14-16).

The following table lists all variables that I determine to be of low frequency, and which have been used in the subsequent analysis.

[2]R statistical software is an open source collaborative project with many contributors (R Core Team 2018). A major package I have used for coding is "Datatable", greatly enhancing speed of calculation and versatility of programming (Dowle & Srinivasan 2019).

Table A.1: Low frequency variables used in my analysis

Variable(s)	Description
Receipts	
EF110	one-time payments (christmas bonus, holiday pay, etc.)
EF112	indemnities, termination benefits, interim payments
EF113	profit-sharing (e.g. bonuses, performance fees)
EF118	income from secondary employment (gross value, including temporary jobs, minijobs, holiday jobs and expense allowances)
EF122-127	some non-cash benefits (solid fuels, gaseous fuels, district and central heating, electricity, liquid fuels, shares in the costs for heating and hot water)
EF180-182	interest payments, dividend distributions and other financial payouts
EF221-222	Other reimbursements (e.g. of energy or incidental costs, proceeds from carpools, or refunds of employer)
Expenses	
EF223-225	tax payments (church tax; income tax including prepayment, arrears, flat tax on capital payments; solidarity surcharge)
EF282-284	upkeep and repairs of real estate (differentiated by renter, owner occupier, third party owner)
EF317, 320, 322, 326	gas, liquid fuels and district heating and hot water
EF333-334	furniture and fixtures and their delivery and installment
EF335	laying of carpets and elastic floor covers
EF337	repair of furnitures, fixtures and floor covers
EF340-343, 345	Fridges, freezers, washing machines and dryers, dishwashers, ironing machines, other large white goods, installment of any of the former by a third party, repairs
EF368-369	therapeutic means and devices and the repair thereof
EF371	medical services (excluding dentist)
EF373-375	rent of therapeutic devices, other medical services outside of hospitals, hospital services
EF381	materials, components and spare parts of bicycles

EF386-387	third-party transportation services without air traffic (with and without overnight stay, respectively)
EF388-389	third-party air transportation services (with and without overnight stay, respectively)
EF401	data processing equipment, systems and application software
EF423	books and brochures, including downloads and apps
EF434	course fees (without acquisition of degree)
EF465	private liability insurance
EF467	disability insurance, including as a supplement
EF468	private accident insurance, including those with guaranteed return of premiums
EF471	term life insurance premium payments
EF472	membership fees for clubs, parties, etc.
EF473	monetary donations and other irregular transfers to non-profit organisations
EF474	voluntary alimony/support payments, money gifts
EF485	acquisition of houses, condos, real estate and garages; other expenses for construction of houses, garages and the like
EF487	maintenance that increases market value, for real estate not used by the owner
EF505	residential building and landowner liability insurance, for real estate not used by owner
EF507	large maintenance and construction works that do not increase market value, for real estate not used by owner

A.1.1.2 Quarterly adjustment with imputation

To adjust for the biases mentioned above, I apply an imputation approach[3]:

1. The dataset is split by quarters, getting datasets *DataQ1-DataQ4*.
2. The resulting four datasets are copied, and all variables that are of low frequency are replaced by missing values (NA). This gives *DataQ1_lowfreqNA - DataQ4_lowfreqNA*.
3. Four new quarterwise panels are created, where each has one quarter filled by *DataQ1-DataQ4* and the remaining quarters filled by low frequency missing val-

[3]Moreover, the yearly values are weighted by the quarter's mean divided by the mean of the whole sample, to correct for higher or lower levels of payments in some quarters. This seasonal adjustment is done for all items, regardless of frequency.

ues datasets. Then, they are combined into a long panel *Data_panel*, where the quarters are stacked below one another.

4. Within *Data_panel*, the missing value low frequency observations are imputed.

5. The values of variables are summed up over the quarters, creating a dataset with yearly values. I.e., the final value of variable EF110 of a household surveyed in Q1 consists of the value observed in Q1, plus the values imputed within the panel with observed values for Q2-Q4.

As a result, high-frequency variables are quadrupled, while low-frequency variables are imputed using the observations of other households.

The methodology of choice is predictive mean matching within a multiple imputation with chained equations[4]. Predictor choice is based on economic intuition[5]. For most of the variables to be imputed, these are: HH gross income EF60, quarter EF6, no. of persons in the HH EF7, social status of the main income earner EF37, number of children in the HH EF41-EF47, gender of HH head EF8u2, birth year of HH head EF8u3. The imputation is done using R's *mice*-package (van Buuren & Groothuis-Oudshoorn 2011).

As my problem is quite unique, suitable approaches are difficult to ascertain given that on average 3/4 of the observations have to be imputed for the affected variables. The latter entails a rather high computational load for many algorithms. Any ideas for improvement are highly welcome. From the implicates thus generated, one is drawn and the subsequent calculations are applied to this dataset. Generating more than a few, say 10, implicates while running 5 - 10 iterations of the algorithm, is computationally not feasible with standard CPUs and R's "mice"-package at the moment. For robustness, I have applied a couple of different implicates to the subsequent modelling exercise, and the results changed only slightly. Future research may achieve a little bit more robustness here, yet at the cost of a high computational or programming load.

On a theoretical level, of course single imputation is inferior to multiple imputation methods in terms of consistency, unbiasedness and precision (Rubin 1987, pp. 11-18;

[4]Previously, I inter alia tried a k-nearest neighbour single imputation, because it gave an acceptable fit for those variables that I deemed most important. It is simple and worked for the high number of variables that I originally imputed, including both those that were used in the later analysis and those that were not. However, it was discarded due to some disadvantages: Firstly, the algorithm that calculates the nearest neighbours is slow, it takes up a lot of computation time even with modern CPUs and for a modest number of observations. Moreover, the available R packages have not allowed me to introduce some random error to the imputed values. As a consequence, observations were partially lumpy, contributing to singularity issues later in the analysis. Most importantly, single imputation is theoretically inferior to multiple imputation and kNN in particular underperformed relative to PMM.

[5]A more structural approach could be a Lasso regression, but was not implemented by me because of cost-benefit considerations.

Van Buuren 2012): Generally, single imputation cannot account for uncertainty with respect to the model underlying the nonresponse. When filling a missing value with a single prediction, the estimate of the variance of the resulting variable will be too small. As a consequence, test statistics will be biased upwards, i.e. the true uncertainty of findings in later inference will be underestimated compared to multiple imputation.

A.1.2 Adjustments using items for previous year

As was mentioned in the paper already, some ad hoc adjustments for particular EVS items were proposed by Becker (2014): For income from self-employment, the previous year's income is used as a plausibility check to keep the current year's values within reasonable bounds. A similar procedure is applied to capital income: If the current year's self-employment income is stated as zero, it is replaced by the previous year's value if that was at least EUR 12,000. If the quadrupled quarterly value deviates by more than 30% from the previous year, 70% of the preceeding year's self-employed income is set for low incomes, 130% for high incomes exceeding EUR 9,000. Also if the current year's capital income is stated as zero, she replaces it by the previous year's value. If the previous year's proceeds from capital are zero, current year receipts are set to the quarterly value stated, assuming no other capital income was generated that year. If both t and t-1 show positive amounts, the quadrupled value is restricted to 125% of the previous year's proceeds.

This adjustment is possible, because income for the previous year is recorded for the following items. Note that they are given individually for the 1st and 2nd HH member, but only aggregated for the 3rd to 10th HH member.

- Gross bonus payments (christmas, 13th/14th month salary): EF625, 629, 633
- Gross holiday bonus: EF626, 630, 634
- Other bonuses/lump-sum payments: EF627, 631, 635
- Gross profit sharing payments: 628, 632, 636
- Gross self-employed income: EF638-640

Moreover, in the part of the EVS file that collects wealth measures (Geld- und Sachvermoegen), both stocks and flows of many items are recorded at the HH level dating to the 1st of January, 2013. Hence, the item EF601 (dividends and interest payments for monetary assets) can be used as previous year's capital income.

I partially follow Becker's approach by adjusting the relevant variables. Quadrupled self-employed incomes (summed-up items EF134-137) are restricted within bounds of

50% when deviating positively or negatively from previous year's incomes. Bonuses, lump-sum payments and indemnities are not quadrupled, but replaced by previous year's value when stated as zero. Also if they are more than 10% lower in t compared to t-1, the preceeding year's value is pasted. Considering a margin of 20%, the same is done with capital income. Moreover, it is restricted to 150% of the previous year's capital income, as proposed by Becker (2014).

A.2 The German income tax model

Evidently, the model I have constructed to calculate the taxes due requires the highest amount of explanation. I shall proceed by first summarising the general approach, then walking the reader through the calculation of the taxable income and tax due according to the income tax code (Einkommensteuergesetz - EStG).

A.2.1 General idea

The EVS contains a rich set of information on income, wealth and expenditures including tax payments. Many items are available on the personal level: some personal characteristics (EF8-15), most incomes (EF109-175) and some expenditures, especially taxes and social insurance contributions (EF223-242). Other variables exist only on the household level: most notably the detailed household expenditures (EF243-533), all wealth figures (EF570-623), household characteristics (EF1-7, 18-50, 534-569), income aggregates (EF51-72) and expenditure aggregates (EF72-106). Lastly, as has been mentioned above, for self-employed incomes and bonuses some variables for the previous year 2012 exist.

This structure of the data makes it necessary to allocate some of the HH-level variables to the individual HH members, because tax concepts usually refer to the individual. Therefore, the share of HH members in tax payments of the HH is calculated. Throughout the paper, this is applied as the standard allocation ratio of HH-level variables when necessary; exceptions are noted. My reasoning for this is that in the absence of better information, allocating tax-relevant (especially deductible) items according to reported individual tax payments seems to be rather conservative. Possible biases may arise for households with members that contribute substantial amounts to the household budget, but do pay little or not taxes for instance due to retirement. In these cases, tax-paying HH members might be attributed too high a share of HH-level expenditures for example. It should be noted that if anything, this would bias tax avoidance downwards, because allowed deductions were to be overestimated. Alternatives could be fixed allocation ratios by social status, especially employment.

To determine which of the aforementioned variables are relevant for calculating the income tax due, I predominantly used the standard works of Dittmann et al. (2014), which is employed by many tax counsellors. Moreover, the text of the income tax law in the version applicable for fiscal year 2013 was utilised of course (Deutsches wissenschaftliches Institut der Steuerberater e.V. 2014). Step by step, the necessary aggregates in the tax code were calculated (see table A.2): From the added-up profit

and surplus incomes (Gewinn- und Überschusseinkünfte), the relevant business and professionel expenses (Betriebsausgaben und Werbungskosten) are substracted to get the Sum of Revenues (Summe der Einkünfte). Further substracting the some reliefs (Altersentlastungsbetrag und Entlastungsbetrag Alleinerziehende), we arrive at the Total amount of revenues (Gesamtbetrag der Einkünfte). Substracting the special and extraordinary expenses (Sonderausgaben und außergewöhnliche Belastungen) yields the Income (Einkommen), of which the children's allowance is deducted to get the Taxable income (zu versteuerndes Einkommen). The final tax due is calculated by applying the tax schedule and further steps that take into account the progressivity proviso, the fifth rule and direct tax reliefs.

Table A.2: Calculation of tax due according to the German tax code, 2013

Profit income	**Surplus income**
Income from self-employment	Income from dep. employment
+ Income from agriculture and forestry	+ Capital income
+ Business income	+ Income from rent and lease
	+ Other income
− related business expenses	− related professional expenses
= **Sum of Revenues**	
− Proportional relief for elderly retired persons	
− Relief for single parents	
= **Total amount of revenues**	
− Special expenses	
− Extraordinary expenses	
= **Income**	
− Children's allowance	
= **Taxable income** (zvE)	
applying tax schedule under progressivity proviso yields tax due	
− any positive difference to scheduled tax if fifth rule is applied	
− direct tax reliefs (Steuermäßigungen)	
= **final tax due** (zu entrichtende Steuer)	

Source: Own table, based on Dittmann et al. (2014, p. 22).

The following section basically summarizes the R-code which contains additional remarks. It calculates taxes due, step by step according to table A.2, as well as taxes paid. All calculations are done on the individual tax payer level, taking into account single and joint filing. Aggregation on the household level is done later in the analysis.

Excursus: Estimating the number of children eligible for child benefits
For many tax purposes, it is necessary to know how many children make household members eligible to receive child benefits for said children. Several conditions apply:

1. Generally, child benefits are granted until the child is 18 years old.

2. This is extended up to the age of 21, if the child is unemployed and looking for a job.

3. Children are eligible until 25 years old, when undergoing education, lacking a place for vocational training or in voluntary service.

Hence, taking into account these conditions first all own children living in the HH are estimated, including step, adopted and care children. Since only one of the partners may receive the child benefits for their child, they are hence assigned to the 1st and 2nd person in the household. As default, all children are assigned to the HH head for tax purposes. If a married partner is present, the assignment is split up.

A shortcoming of the EVS is that child benefits are reported as income of the child, not of the parent that actually and legally receives it. This notwithstanding, cases of grown-up HH members without children in the HH, who however receive child benefits, can be identified. Exploiting this, I assume that they have children outside of the HH. Anyone who receives child benefits, is also eligible for the child tax allowance. Based on the amounts received, the number of own children living outside the HH is estimated.

Combining reported and unreported children yiels the estimated number of own (tax-relevant) children of HH members 1 and 2[6].

A.2.2 The Sum of Revenues

For determining the Sum of Revenues, both profit and surplus incomes have to be estimated, as well as business and professional expenses. As general rules, professional expenses of course can only be deducted for connected revenues and expenses have to be offset with any reimbursements.

A.2.2.1 Profit incomes and business expenses

The biggest chunk of business and self-employed incomes are reported in four different categories: withdrawals in kind (EF134), private withdrawals of agricultural self-employed (EF135) and self-employed (EF136), and other self-employed incomes

[6]There may be a few cases where grandparents are eligible for child benefits, but I cannot reliably distinguish those.

like professional fees (EF137). As they are given on the individual level, they do not require further adjustments beyond those mentioned in A.1.2.

A problem arises for business expenses: They are not collected in the EVS. Instead, the self-employed are asked to exclude them, i.e. they should basically directly report their Sum of Revenues in the survey [7]. Therefore, only allowances for self-employed that are committed to a honorary post are calculated from the relevant items. Some flat business expenses for certain professions (journalists, writers, but also some scientific and artistic side occupations) can be calculated, but are not used in the analysis to maintain consistency. As the tax payments are not observed and it is debatable whether they should be substracted from the amounts reported here, the direct tax relief for paid local business tax is not taken into account.

Adding to that, also income from the sale of solar power (EF176) is considered as profit income subject to tax. However, no information concerning the related expenses is given in the EVS, so they have to be estimated. For this purpose, several steps are taken:

1. As 5% of the acquisition costs can be deducted as depreciation, I need an estimate of said expenses. Therefore, from the average EEG remuneration in 2013[8], I calculate the kilowatt-hours that were sold.

2. Because some of the electricity is used for the household itself and not sold, fixed amounts depending on the HH size are added. These are based on Frondel et al. (2015, p. 44), reduced by roughly 1/3 because the sun does not shine every day.

3. From Wirth (2017), it is estimated that in Germany on average 900 kWh per kWp are generated. With this, the nominal power of the solar plants is calculated.

4. Schmole (2016) reports the following costs for installing solar panels in Germany: From Q4 2006 to Q4 2012, the average price for small plants with a nominal power up to 10 kWp fell from EUR 4,900 to 1,800 per kWp, the price for larger plants with 10-100 kWp fell from EUR 4,700 to 1,400 per kWp. For simplicity, I therefore assume rather high uniform acquisition costs of EUR 3,000 per kWp for the households owning solar power plants in 2013.

[7] "Self-employed, freelancers and farmers are kindly asked to only report expenses for private use. If you are unable to clearly seperate private and business expenses (e.g. for telephone or car usage), please fill in estimates for private expenses only." (Destatis 2016, p. 33, translated by the author).

[8] The EEG is the German law that was implemented to foster the use of renewable energies. It guarantees a fixed remuneration for electricity fed into the grid, depending on the year the power plant was built. The source of the average figure for 2013 is the federal Ministry of the Economy.

5. After calculating the ensuing acquistion costs by multiplying with the estimated nominal power, I uniformly multiply them with linear depreciation costs of 5%, plus 1% operating costs, to get the deductible expenses.

Substracting the deductible expenses from the solar electricity sale proceeds yields the amount that is subject to tax, which is subsequently attributed to the individual HH members using the default (share in tax payments).

Finally, adding up reported self-employed income and solar power sale income less related expenses, and substracting the honorary allowance (Ehrenamtspauschale) gives the Sum of Revenues for profit incomes.

A.2.2.2 Surplus incomes

Relevant incomes are from dependent work, capital, rent and lease, pensions and other revenue sources.

A.2.2.2.1 Income from dependent employment

The following income items are fully subject to tax and therefore directly collected into one variable first:

- EF109: basic salary (gross)
- EF112: indemnities/reparations, termination benefits, interim payments
- EF113: profit-sharing (e.g. bonuses, performance fees)
- EF114: other incomes (e.g. transport allowance, luncheon allowance, employer's contribution to maternity allowance)

One-time payments are subject to tax as other remunerations (sonstige Bezüge), so the corrected variable (see Section A.1) is added as well. However, it should be noted that these remunerations may be subject to the fifth rule[9], which is checked later in the code.

Capital-forming benefits (EF111) are tax exempt if they feed an occupational or private pension, but they are taxable otherwise (e.g. when invested in housing-saving or financial assets). Therefore, using EF242 it is checked whether the former applies so

[9]The fifth rule (Fünftelregelung) is profitable for tax payers with a comparatively large one-time payment (außerordentliche Bezüge). The additional income is fully taxed, but only 1/5 adds to the progressivity of the tax rate. This preferential treatment is only granted for a list of defined cases (§ 34 EStG).

the taxable observations of EF111 are determined. Unfortunately, some definition lack cannot be avoided because EF242 also contains capital-forming benefits contributions to other insurances for: life, education, burial, dowry.

Another case that requires some adjustment are the salaries of employees in partial retirement (Altersteilzeitentgelt - EF119). This is because only the base salary component of these employees is subject to tax, but not the top-up amount[10] which is paid for by the Federal Employment Agency (BA). Unfortunately, EF119 contains both parts, so some estimation about the split is needed. A rough division is done along the following reasoning: The partial retirement base salary consists of 50% of the last salary, plus at least 20% topped-up by the BA (if lower, the BA does not pay). So in total, the minimum remuneration of partially retired employees is most likely to be 60% of their last salary. Theoretically, the base salary could thus be anywhere between 50-83.3% and the top-up amount 50-16.6% of EF119. However, the base salary was often topped-up by the employer by up to 85% of the last salary due to labour contracts. Therefore, I assume that 75% of the variable consists of the base salary while the remaining 25% is the top-up amount, which seems to be rather conservative in the absence of better information.

Non-cash benefits (EF120-133) are subject to tax if they exceed EUR 44 per month, i.e. a yearly amount of EUR 528[11]. The relevant items are summed up and the condition is checked, so the taxable observations are determined.

Following are other incomes (EF118), which are already recorded at the personal level and therefore require no further adjustment. Income from carpools, which also falls into this category, is added later in the code once carpool users are determined.

As an intermediate result, all incomes from dependent employment (nichtselbstständige Beschäftigung) before deductions can be summed up from the items mentioned above.

A.2.2.2.2 Capital income

Next, capital incomes are considered. Payments of interest, dividends and other payouts from financial assets are taken over from section A.1.2.

[10] However, this amount is liable to progressivity proviso.

[11] To take into account all details, one has to differentiate once more between "Deputat" and "Endverbrauchsware" like the beverage for private use handed out to brewery workers for instance (Haustrunk), because the latter has a higher yearly allowance of up to EUR 1080. Only a small two-digit number of observations of the relevant categories EF128 (non-cash food benefits) and EF129 (non-cash beverage benefits) fulfills $528 > x \leq 1080$, but this is taken care of in the code nonetheless.

A problem arises for capital gains, because the available information in the EVS does not suffice to meet the relevant tax concepts. In principle, the sale of real estate (EF197) was liable to tax if it took place less than 10 years after acquisition (speculative period). For other economic goods like precious metals (EF199), the relevant time limit was one year. As one cannot determine from EVS information whether the requirement is fulfilled, no taxable income from such sales is included in the code.

The sale of financial assets (EF203) however, is taxable since 2009. Only capital gains from the sale of securities or forward contracts that were bought before 2009, remain tax exempt in 2013. It is of course uncertain how high the capital gain on the recorded sale is. For the sake of simplicity, I assume that it is 20% which add to taxable capital income. A more elaborate estimation using information from additional sources could be worthwhile, in particular to account for the share of sales that would be tax exempt because the assets were bought before 2009.

Another difficulty of yet even smaller importance for the overall target is posed by the sale of (unincorporated) company or coop shares (EF204). These are not included in my calculations, because there are a number of tax allowance provisions for people older than 55, permanently disabled and unable to work, and concerning the fifth rule. Some of these conditions cannot be checked with the information available, and presumably only a minor improvement could be achieved by attempting to include this item.

Coming into effect in 2009, Germany has a split income tax system where capital incomes are taxed with a flat rate of 25%. Therefore, the relevant capital income allowance (Sparerpauschbetrag) of EUR 802 is substracted directly at this point in the code, after apportioning the HH-level capital income to the HH members by default. This yields the Sum of Revenues for capital income.

Here, it makes sense to also calculate the resulting flat rate withholding tax on capital income[12] plus solidarity surcharge. For this calculation, it is taken into account that EVS participants should state income taxes after the capital income withholding tax (plus solidarity surcharge) is substracted.

A.2.2.2.3 Income from rent and lease

Next, incomes from rent and lease (EF177) are taken in. Note that also income from subletting (EF193) is subject to taxation if it generates a profit. There is a tax free

[12]Capital income of tax payers with a lower marginal tax rate than 25% according to the regular progressive schedule is taxed at this lower rate. This is checked automatically by the fiscal authorities (Günstigerprüfung), hence also later in the code. It is relevant for lower middle income households that typically enjoy rather small proceeds from capital.

amount of EUR 520 per year, so I make sure that only amounts beyond that are determined as taxable income. The resulting figures are again allocated to the HH members by default.

A.2.2.2.4 Other income

In this category, provision services, alimony receipts, income from other services rendered, from parliamentary pay and from speculative sales are considered as taxable. The latter are mostly not identifiable, exceptions are collected as capital income before; neither are incomes of deputies.

Nonetheless, the former three are contained in two items of the EVS: EF188 inter alia includes incomes from other services rendered, however also support/aid from sources like churches, unions, and other organisations which is not taxable. No further information on the distribution of any of these components was found, so it is simply assumed that only 1/3 of cases in EF188 contain some taxable component, which for these cases is estimated to make up 2/3 of the EF188 amount. As this is a HH-level variable, the relevant households are drawn randomly, the amount subsequently allocated to HH members according to their tax payment share. The tax free amount of EUR 256 per tax payer is taken into account as well.

Additionally, EF189 contains all supportive payments, money gifts, other support of other private HHs incl. life annuities. Of these items, alimony receipts are taxable under some conditions: The divorced couple has to agree to what is called "Realsplitting", where the paying party may deduct alimonies as special expenses, compensating the receiving party for having to pay tax on the received amount. This is mostly profitable for the ex-partners, but the receiving party has to consent. Another condition is that the procedure is only admissible for postmarital support payments, not for child support. No information could be found on how many divorced couples use it, so as a default I assume that 75% of them agree to do so. For those where EF189 plausibly contains some alimony receipt (i.e. divorced or living seperately), I furthermore assume that 90% of 189 are tax relevant here. As EF189 is a HH-level variable, it is divided up by the no. of divorced persons in the HH. Moreover, as child support payments are not subject to "Realsplitting", the amounts are adjusted for the number of own children living in the household[13]. The expenses allowance (Werbekostenpauschbetrag) of EUR 102 for such provisions is considered as well[14].

[13]For every own n children in the HH eligible for child benefits, the taxable component of EF189 is reduced by $1/(1+n)$. Of course, one cannot infer from the survey information to whom a child actually belongs. Facing this uncertainty, again the benevolent choice is made to rather allocate a lower taxable share of income to the tax payers.

[14]This amount is only granted once for all recurring payments received, including pensions, alimony

In theory, EF189 also contains taxable components of other support provisions that may be taxable, e.g. when some wealth was transferred from someohne who in return receives a life annuity. However, I do not find the share of this in EF189 to be plausibly identifiable, as no further information on the prevalence of such a type of transfer was found for the general population.

A.2.2.2.5 Pension income

The final component of surplus incomes that is included are pensions. The taxation of pensions is a complicated matter that was reformed in the first half of the decade after 2000. Since then, a fraction of pension income is liable to tax, depending on the year of first receipt, i.e. the age of entry. Since no information on the latter is directly available in the EVS, it is estimated from macro averages for all pensions of the respective type.

As a general rule, everyone below the average age of entry in 2013 is assumed to have entered her pension in 2013.

First, civil servant pensions (EF169-170) and occupational pensions (EF171-172) are collected (Versorgungsbezüge).

The allowance for civil servant/occupational pensions (Versorgungsfreibetrag) is 40% for those entering until 2005, afterwards it is reduced incrementally year by year. Moreover, an extra amount (Versorgungszuschlag) is added to the allowance to compensate for the abolition of the regular allowance (Arbeitnehmer-Pauschbetrag). For those who have entered until 2005, this amount is EUR 900, but equally reduced for later entry years. The following steps are taken to estimate the taxable share of these pensions:

1. The average age of entry of civil servants for the years 1993-2013 is taken from Destatis (2018*a*). As an example, someone entering in 2005 was 59.4 years old on average, i.e. born in 1945/1946.

2. Based on the year of birth of the individual (EF8u3, EF9u3, EF10u3, EF11u3, EF12u3), the entry age into pension for HH members 1-5[15] is determined[16].

receipts and others. Therefore, it is checked later whether pensioners may have already claimed the allowance.

[15] It is not necessary for HH member no. 6, because there is no one with a civil servant pension.

[16] As mentioned above, everyone below the average of entry, in this case everyone born after 1951, is assumed to have entered his pension in 2013. This inserts a downward bias to the allowance, i.e. it finally increases the estimated tax due. On the other hand, survivors are entitled to the allowance based on the year of death of the spouse (which of course cannot be determined from the survey). Since typically the older partner dies first, on average this might counteract the effect mentioned before, the magnitude of any of the two being unclear.

3. Following the tables provided in the tax code (§ 19 Abs. 2 Satz 3 EStG), the allowance including extra amount is calculated for every relevant individual.

Occupational pension receivers have to be at least 63 years old at entry to get the allowance, i.e. born in 1942 or before in 2005. Taking this into account, the allowances of the respective individuals are calculated following the same method as above.

Finally, the amounts that are liable to tax are calculated by substracting the allowance from the civil servant/occupational pension income.

Second, different types of public pensions (EF138-140) are collected.

Again, the taxable part is determined by the age of entry, starting at 50% if entered until 2005, increasing thereafter. The component that is liable to tax is determined as follows:

1. The average of entry for public pension receivers is taken from DRV (2017)[17]. In 2005, the average new pensioner was born in 1944.

2. Using the birthyear variables (EF8u3-EF13u3) once more, and applying the tables provided in the tax code (§ 22 Abs. 1 Satz 3 aa) EStG), the taxable share of the public pension of the relevant 1st-6th HH member is calculated.

Third, some private pensions (EF142, 143, 175) are considered.

The liability to tax of these incomes depends on several conditions. If the contract was signed before 2005, or for certain contracts that were concluded later and do not meet the requirements of the Rürup pension, the tax treatment during the qualifying period is decisive: if the contract was fiscally promoted, the payout is fully liable to tax, else only partially. The latter also applies if the contract is an annuity for life in consideration of property alienation (Veräußerungsleibrente). The taxable component of the relevant variables is then determined with the following steps:

1. For simplicity, in the absence of better information I assume that item EF175 (pensions from private life insurance) is fully composed of such type of pensions that are partially taxable only.[18] Items EF142 and 143 report the pension income from supplementary system of the public service (VBL, ZVK).

[17] It is slightly different for West and East, and males and females. Larger differences occur between old age pensions and pensions due to a reduction in earning capacity. Nonetheless, I take 61 years as the average for all entries in the years 2005-2013, as the additional small variation does not outweigh the cost of implementation.

[18] If the assumption were violated and some of the observations for this variable actually be fully liable to tax, this would introduce a downward bias to the final tax avoidance estimates.

2. The taxable component is determined by the profit share, which is based on the age at entry to the pension. Since no information on the latter is available, it is estimated from the DRV (2017) figures. Again, it is assumed that those younger than the average age at entry in 2013, entered the pension in 2013.[19]

3. Applying the birthyear variables and the relevant tables in the tax code (§ 22 Abs. 1 Satz 3 bb) EStG), the taxable component is calculated for HH members no. 1-6.

After collecting all tax-liable pension incomes, the expenses allowance (Werbekostenpauschbetrag) of EUR 102[20] is substracted to yield the Sum of Revenues for pensions.

A.2.2.3 Professional expenses

Next in the code, the deductible expenses for surplus incomes (Werbungskosten) are estimated.

A.2.2.3.1 Professional expenses related to income from rent and lease

The first type of proceeds considered here is income from rent and lease. The following items are part of the running costs of landlords and therefore completely deductible (luckily, the EVS differentiates here so all of them refer to not self-used real estate):

- EF312: residential home apportionment for condos
- EF480: interest on building loans and mortgages
- EF505: residential building and landowner liability insurance
- EF515: other operating expenses

Moreover, expenditures for a single building activity that increases value are fully deductible for the year of construction up to EUR 4,000. Taking this condition into account, the relevant deductions are calculated from EF487 (maintenance that increases market value).

Other expenses are only partially tax allowable. The first are acquisition costs (EF485), which are subject to deductions for expenses for wear and tear (AfA). In the

[19] This time, the assumption introduces a downward bias to the tax avoidance estimate, in opposite to the cases mentioned in footnote 16. This is because the younger someone is at the year of entry, the higher is the profit share and therefore the larger the taxable component.

[20] Higher expenses can be claimed, but rarely are. Moreover, they cannot be estimated properly with the EVS data.

past, higher rates of depreciation were allowed for tax purposes[21], however I assume that only the standard linear depreciation over 50 years (i.e. 2% per year) is applied. This introduces a downward bias to deductions of course, hence increasing the final tax avoidance estimate. Another problem that works in the same direction is that most houses generating income in 2013 were not built or bought in 2013. Therefore, a rather simple imputation is applied, using the ratio of the average estimated depreciation deductions $\bar{x_i}$ to the average rent proceeds $\bar{y_i}$ for the observations $i > 0$ (i.e. those landlords that report acquisition costs for 2013) to estimate the deductions for acquisition costs that cannot be observed:

$$\forall\ i\ where\ x_i = 0\ \&\ y_i > 0,\ x_i = y_i * \left(\frac{\bar{x_i}}{\bar{y_i}}|_{x_i>0}\right)$$

Additionally, the land value has to be substracted from the acquisition costs. In the absence of better information, a show case example of the BMF is used to assume that 1/3 of the costs reported in the EVS are for land, the remaining 2/3 for the building.

Adding to that, beyond the EUR 4,000 of building costs close to acquisition (anschaffungsnahe Herstellungskosten, up to three years after acquisition), the remainder of EF487 can be deducted using the standard 2% depreciation if it exceeds 15% of acquisition costs.

Logically then, those EF487 expenses that exceed EUR 4,000 but fall short of 15% of acquisition costs, can be deducted under another rule. They are part of maintenance expenses (Erhaltungsaufwand), as are large maintenance and construction works that do not increase market value (EF507). If they are large (zu verteilende Erhaltungsaufwendungen), they can be written off linearly over 2-5 years. No clear definition of "large" is given, so I compare it to gross rental income (EF177). As there seems to be substantial discretion that may decrease taxes when stretching the deductions over up to 5 years, my default assumption is to deduct the amount completely when it is below half of gross income, i.e. "small". It is stretched over 2 years when above that, over 3 years when $> 80\%$ of rental receipts, over 4 years when higher than rent and over 5 years when more than two times rental income. The write-off treatment of items EF487 and 507 is summarized in table A.3.

Finally, the deductible expenses of landlords (running costs, directly deductible building costs, acquisition cost depreciations, other depreciations) are summed up.

[21] Especially during the 1990s, investment in real estate in East Germany was incentivized through special tax allowances, for instance degressive depreciation. Moreover, in 2013 there were special rules for higher depreciation for areas affected by floods. Generally, people affected by fire and flood are allowed to write off more. However, this is impossible to ascertain from EVS information.

Table A.3: Tax treatment of maintenance and construction works for landlords

Condition	**Deductibility**
Value-increasing maintenance and construction (EF487)	
EF487 $\leq$ EUR 4,000	full deduction
EUR 4,000 $<$ EF487 $\leq 0.15 * 2/3 *$ EF485	deduction as Erhaltungsaufwand over up to 5 years
EUR 4,000 $<$ EF487 $> 0.15 * 2/3 *$ EF485	deduction with standard AfA (2% p.a.)
Value-preserving maintenance/construction (Erhaltungsaufwand = EF507 + some part of EF487)	
Erhaltungsaufwand $\leq 0.5 *$ EF177	full deduction
Erhaltungsaufwand $> 0.5 *$ EF177	deduction over 2 years
Erhaltungsaufwand $> 0.8 *$ EF177	deduction over 3 years
Erhaltungsaufwand $>$ EF177	deduction over 4 years
Erhaltungsaufwand $> 2 *$ EF177	deduction over 5 years

Variable names refer to the EVS 2013. EF177 are gross rental receipts, EF485 acquisition costs of real estate. Source: Own table.

For incomes from subletting, deductions may be claimed according to the m^2 used by the subtenant. I approximate this by comparing the sublease amount to the gross rent of the main flat (EF193 / EF263). Hence, I disregard a possible subletting of secondary domiciles due to lack of information. Please note that only the "Einkünfte", i.e. the profits generated from subletting, are taxable. So if a tenant for example adds up all the cost incurred by renting the flat, and splits them "fairly" over all the flatmates, not taxable income arises (for tax affairs, the split by m^2 matters, though). For simplicity though, I assume that 25% of the income from subletting is indeed taxable, i.e. 75% of the amount is assigned as deductions.

Substracting landlord and subletting deductions from rental receipts then gives the Sum of Revenues for income from rent and lease, which is allocated to the HH members by default.

A.2.2.3.2 Professional expenses for dependently employed

The next type of deductions that is considered are professional expenses of dependently employed individuals (Werbungskosten). There is a flat allowance of EUR 1,000 (Arbeitnehmer-Pauschbetrag) which is granted to everyone with income from not self-

employed work. Only if the claimed expenses exceed this amount, they are considered by the fiscal authorities. To identify whether this is case, I collect all items that may be claimed given the information available in the EVS.

For later calculations, first dummies for the (not-)self-employment of every individual are programmed, which in turn are used to calculate the share of employed members for each HH. These are necessary to attribute HH-level expenses to the HH members.

Next, the sample is restricted to HHs with at least one dependently employed member. Once the professional expenses are calculated, they are merged back with the remaining sample.

The first problem for the estimation of professional expenses are costs of home offices, which are tax allowable for some professions under certain conditions[22]. However, for the relevant items in the EVS (especially furniture) it is impossible to distinguish professional from private expenses and the conditions for eligibility cannot be checked properly. Moreover, for durable goods some imputation would have to be made because most purchases that are depreciated over many years are presumably not observed in 2013. For these reasons, these deductions are not included in my estimation.

Nonetheless, a plethora of other professional expenses remain to be included. The next one are costs of work equipment. As a general rule, relevant costs up to EUR 110 are accepted as allowance without further proof. Moreover, up to a price of EUR 410 for a single item the full costs may be deducted directly in the year of payment (geringwertige Wirtschaftsgüter). If they exceed that amount, the official depreciation tables (AfA) have to be used to depreciate it over several years.

The first item thus considered are computers, hardware and software (EF401) which must be depreciated over three years. I assume that half of the expenses in the HH-level item are tax-relevant, apply the conditions mentioned before and get the tax allowable amounts.

The next item are other consumer goods related to school and office, entertainment and leisure (EF426). Here, I suppose that all single purchases aggregated in the item have a price of less than EUR 410 and may therefore be fully deducted in 2013. Furthermore, it is assumed that 20% of the reported values are due to occupational purposes, allocated to the HH members using the HH employment share. As there

[22]The room has to be seperate from other rooms, must be used for professional occupation almost exclusively, be at home and be used for work. If no other workplace exists for the employee, expenses up to EUR 1,250 may be claimed. If the home office is the center of the entire occupation, full costs of the room are deductible.

may be expensive non-deductibles reported in EF426, a cap of EUR 500 is set for the tax allowable expenses of each HH member.

Similarly, expenses for writing material, paper and other expendable goods (EF427) are considered assuming that 10% are deductible. These are again attributed to individual HH members using the employment share, and a cap of EUR 250 is set for each individual.

Adding to that, also expenses for books, journals and brochures (EF423-425) are taken into account. Here, it is assumed that generally 10% are tax deductible, however for teachers and scientists this is increased to 20% using the economic sector items (EF8u19 - EF13u19).

Finally, all tax allowable expenses for work equipment are aggregated.

The biggest chunk of professional expenses are commuting costs. The treatment of these expenses was used to exemplify the estimation of deductions from EVS data in the main paper. Therefore, some aspects are doubled here, yet additional information concerning my approach is provided.

Households that receive income from dependent employment are entitled to deductions for the commuting distance they have travelled to work (Pendlerpauschale). In 2013, the allowance was EUR 0.30 for every completed kilometre one way, compensating the full round trip. In case public transport was used, the tax payer was allowed to claim the full cost as deductions, up to a limit of EUR 4,500 per year. Unfortunately, the EVS does neither contain the distance to work, nor expenses for commuting directly as separate items. As a consequence, one has to estimate them for all groups that may claim the deduction: car and motorcycle drivers, public transport users, bicyclists, pedestrians and carpool users. To get the deductible allowance, several steps are taken:

1. The relevant items are identified as expenses for car fuels and lubricants (EF383) and third-person transportation services (EF386).

2. A macro estimate for the shares of employed persons commuting with any of the means of transportation is taken from the sample census of 2012 (Destatis 2017*a*), then these shares are applied to the relevant group within the EVS (not self-employed persons) to get absolute numbers.

3. In the absence of more precise information concerning the distribution of commuters along other covariates (sex, age, income, etc), these absolute numbers of commuters of each type are drawn randomly in case of motorvehicle drivers (EF383), public transport users (EF386) and carpool users. For the latter, it is

assumed that half of them use their own (EF383) and half other people's (EF386) cars. Moreover, as the persons are drawn partially using the same variables, it is ensured that they are clearly seperated. I.e., I assume that someone either commutes alone in her car, or as a carpool member. To improve the random draw of carpool drivers, moreover EF 221 (mixed category of different refunds) is taken into account. The refunds are added as taxable income, assuming that for the drawn HH member 75% of the value (allocated to HH members by the tax share) are carpool-related.

4. For pedestrians and bicyclists, the kilometres driven are estimated ad-hoc using normal distributions. Without more precise information concerning their travel distance, this approach is justifiable on the grounds of the small ranges covered by typical commuters of this kind. As 25.4% of commuters in the sample census (Destatis 2017*a*) have a way to work below 5km and walking long distances is generally unlikely, I take 1km as the mean for pedestrians. Moreover 42.3% of employees commute no more than 10km (ibid) and bike users drive further, so for bicyclists 4km are assumed to be the mean.

5. For motorvehicle and public transport commuters, the share of the relevant items which are due to commuting have to be estimated. This is done using macro figures about the shares of purposes of kilometres travelled for different types of transport (Radke 2014). Some plausible assumption has to be made concerning the factor by which to increase the estimated share for actual motorized commuters compared to the macro average: My default value is a factor of 2[23]. Moreover, the federal state is taken into account, considering the average journey time to the nearest large city. Also, the size of the community is used to weight the share. See table A.4 for the details of these weights.

6. Motorvehicle users' kilometres driven are estimated from their gasoline expenses. Therefore, first the distribution of cars by fuel is determined, which was roughly 70.5% gasoline and 29.5% diesel in 2013 (Kraftfahrtbundesamt 2013). However, gas shows less kms travelled, 11km vs. 20.5km for diesel (Radke 2014). Hence, about 57.3% gas and 42.7% diesel cars were used for actual kilometres driven. The average fuel price in 2013 was 159ct/l for premium gasoline, and 143 ct/l for diesel (Radke 2014). Combining all these pieces of information yields a composite average of 152.1 ct/l for the average km driven. Dividing the estimated

[23]Lower and higher values ranging from 1.25 to 3 have been tried and final results were robust. Nevertheless, the precision of my estimate could profit from more granular information, especially from the tax statistics.

fuel expenses by this amount then gives the fuel consumption of the individual commuter. Average fuel consumption of cars in Germany was 7.3 l/km in 2013, i.e. 13.7 km/l (Radke 2014, p. 303). In turn, applying the latter to the estimated fuel consumption finally yields the kilometres driven, which are halved once more because only the one-way trip to work is compensated by the deduction.

7. Now, the commuter allowance can be computed for all relevant means of transport: The kilometres travelled by motorvehicles, pedestrians and bicycles are simply muliplied by EUR 0.30. The amounts spent by public transport users are claimed fully, under consideration of the maximum limit of EUR 4,500.

One might think that most of these adjustments are somewhat arbitrary, and that is correct. As is mentioned in the paper, substantial improvements to the precision of the commuter allowance estimation may be derived from directly matching work distance information from the IAB-Beschäftigtenhistorie, or the 2012 sample census questions that were used in the published aggregate here, or of course claimed amounts from the tax statistics.

The following type of deduction are professional expenses for further education (berufliche Fortbildungskosten), which are fully tax allowable if they are related to work. The first item to be considered here is the HH-level variable EF432 (examination fees, tuition fees, fees for training programs, etc). However if the values were allocated to employed HH members using the employment share only, there would be a substantial upward bias to the estimated deduction, because educational spending for children living in the HH is no eligible but likely making up a sizeable amount of EF432. For this reason, it is assumed that children between the ages of 16 and 25 reduce the share of professional education spending of other HH members. Excluding the household head, a reweighted share of the employed HH members + double-weighted children of that age is applied to allocate the educational spending collected in EF432.

Another variable that is taken into account here are fees for courses without degree (EF434). It is assumed that 25% of these costs are related to work, and subsequently allocated to the HH members using the reweighted educational employment share.

Occupational travel expenses are considered next. In contrast to the commuter allowance, the full round trip is deductible. Professional travellers may deduct a flat rate per km (Kilometerpauschale) or establish their actually incurred costs. For car drivers, I assume only the allowance is used because the costs are not fully ascertainable from EVS items[24]. In 2013, different allowances based on the mode of transport were legally

[24]This may introduce a small bias for users of expensive cars that could actually report higher

Table A.4: Adjustment of motorvehicle commuter's average share of commuting in total fuel consumption (EF383)

Condition	**Adjustment**
Base share from macro figures	
commuting on average: 21.7% of all passenger-kms driven by cars (Radke 2014)	multiplied by factor 2, i.e. to 43.4%
Community size	
population < 5,000	+ 10%pts
5,000 < population < 20,000	+ 5%pts
20,000 < population < 100,000	+ 2%pts
100,000 < population < 500,000	− 5%pts
population > 500,000	− 10%pts
population < 20,000	+ 5%pts
population > 20,000	+ 2%pts
population > 100,000	− 5%pts
(Non-City) Federal state	
Schleswig-Holstein	+ 3.6%pts
Niedersachsen	+ 3.9%pts
North Rhine-Westphalia	+ 3.0%pts
Hesse	+ 2.7%pts
Rhineland-Palatinate	+ 3.6%pts
Baden-Württemberg	+ 3.1%pts
Bavaria	+ 3.1%pts
Saarland	+ 2.8%pts
Brandenburg	+ 4.9%pts
Mecklenburg-West Pomerania	+ 3.7%pts
Saxony	+ 3.1%pts
Saxony-Anhalt	+ 4.3%pts
Thuringia	+ 3.9%pts

Variable names refer to the EVS 2013. Note that the community size ranges are mutually exclusive, and coded as such in the EVS (EF4). In case of the federal states, a 1%pt higher share is alloted for every ten minutes of the average driveway to the next city. The latter is taken from Canzler (2013, p. 313). Source: Own table.

valid: For cars 0.30 EUR/km, motorcycles 0.13 EUR/km, scooters 0.08 EUR/km, bikes

expenses.

0.05 EUR/km. In principle however, actual costs are deductible. Estimations similar to those for for the commuter allowance are thus performed to calculate tax allowable expenses:

1. From Radke (2014, p. 224), the macro shares of the modes of transport by purpose (in 2012) are taken. They are based on the passenger-kilometres travelled. Moreover, the share of employees with business trips in 2013 is calculated from ACTA (2015).

2. For travellers using their own motorvehicle, the resulting number is drawn randomly from the employed individuals with positive fuel expenses (EF383). For them, the share of passenger-kilometres that are due to business trips is multiplied with the fuel expenses thus calculated. Again, it is assumed that actual business travellers have a higher share than the macro average, by 50% adjusted upwards. Subsequently, from the fuel expenses the kilometres driven are calculated as in the commuter case. I assume that business travellers rarely use motorcycles, so for simplicity all observations are treated as car users. Hence, multiplying with EUR 0.30 per km gives the resulting full trip deductible business expenses.

3. A similar procedure is applied for pulic transport (including railway) users, but now the two relevant items EF386 (without overnight stay) and EF387 (with overnight stay) have to be considered. According to VDR (2014), on average 55% of business travel trips took only one day. This is taken into account when randomly drawing the eligible individuals using the relevant variables. Applying the upwards adjusted relevant share of business trips yields the resulting tax allowable expenses.

4. Alike, the case of business air passengers is treated. Again, there are variables with (EF388) and without (EF389) overnight stay. Additionally, it is assumed that actual business trip flyers are 1.5 times and those with overnight 2 times above the average macro share of flights due to business trips. Full costs are determined as deductible.

5. Bike users are drawn randomly from employed bike owners whose HH spent a positive amount on maintenance (both EF381 and EF539 > 0). The relevant parameters are taken over from the commuter case, and the EUR 0.08 allowance per kilometre is calculated for the full tround trip. Moreover, it is taken into consideration that there were 171 Mio. business trips for 10 Mio travellers in 2013, i.e. roughly 17 trips per capita on average (VDR 2014).

Travel and accomodation costs with respect to professional education may also be deductible.

Travel expenses are tax allowable under the same rules as for business trips. As they too are not captured directly, they have to be assessed similarly to business trips. Therefore, it is assumed that everyone for whom professional education expenses were established earler in the code, also has some travel expenses [25]. For motorvehicle users, the (relatively small) share of educational purposes in travelled passenger kilometres from Radke (2014) is adjusted upwards with the factor three, then multiplied with fuel expenses (EF383) to establish the deductible allowance as in the cases of commuting and business trips. In case of public transport users, the travelled passenger kilometre shares are multiplied by 1.5 for trips without overnight stay (EF386) and with 4 for trips with overnight stay (EF387), again deducting full estimated costs. Unfortunately, the travel costs of airplane users cannot be taken into account, because no macro statistics concerning the share of passenger kilometres travelled due to professional education are availabe.

Gauging the expenses for overnight accomodation with respect to business (-related educational) trips again requires using additional macro statistics. The relevant variable EF437 contains both private and professional expenditures, so the professional fraction has to be estimated. As there were 411.8 Mio. overnight stays of domestic and foreign citizens in Germany in 2013 (Destatis 2018*b*), and German companies' number of overnight stays was 43.9 Mio. inside Germany and 17.1 Mio. abroad (VDR 2014), one can estimate a 10.7% share of stays in Germany caused by German companies' business travels. However, this is not comparing the correct aggregates with respect to the EVS item EF437. One would need the number of private vs. business stays of German nationals. As this is not readily available, for now I simply assume that 20% of EF437 is a plausible share of business trips for persons with related business trip expenses.

Another deduction possibility are food expenses during business and professional educational trips (Verpflegungsmehraufwand). The allowance for this deduction is graded the duration of the trip: < 8h: 0, $8 - 14$h: EUR 6, $14 - 24$h: EUR 12, ≥ 24h: EUR 24 (per day) or EUR 40 (abroad, per day). Obviously, no information on the exact hours spent on business trips is available in the EVS. Hence, a simple approximation

[25] According to Destatis (2014*c*, p. 34), in 2012 55% of employees have participated in education related to their profession (betriebliche/berufsbezogene Weiterbildung). Combining this macro information with the additional assumption that 3/4 of professional education takes place outside of the workplace and applying the share to the relevant observations in the EVS yields a higher no. of people with professional educational expenses, than the no. of people estimated to incur educational expenses.

Table A.5: Deductions for educational and business trip expenses

	travel expenses
+	course fees
+	accomodation costs
+	food expenses
-	employer refunds
=	**deductions**

that accounts for the means of transport and amount of travel expenses is applied:

1. For air travellers, to account for trips longer than 24 hours it is assumed that there are EUR 24 of food allowance for every EUR 300 of airfares. Additionally, EUR 6 are granted for every EUR 50, the two not being mutually exclusive resulting in rather generous deduction possibilities.

2. To account for presumably lower costs for the average trip taken by train, the amount is halved: I.e., for every EUR 150 in train fares, EUR 24 are assigned as deductible food expenses. The amount for shorter period trips is the same as for air travel.

3. For car travellers, only the EUR 6 for every EUR 50 of travel expenses part is allowed for. No food expenses are deducted for bicycle users.

These deductible expenses are collected for both business and professional educational trips, and as usual for the 1st-6th person in the household.

From overall deductible expenses for professional travels, of course employer refunds have to be substracted. These are part of the mixed category EF222, which may also contain other refunds like expense allowances for donating blood or from statistical offices. As the latter are presumably small for people with professional travel expenses, I assume that 80% of EF222 are relevant refunds for business travel and professional development and substract them. Moreover, when allocating the HH-level item to the individuals, only dependently employed people are considered because the self-employed of course do net get employer refunds. The relevant calculations are summarised in table A.5:

The next item deductible as professional expenses are membership fees for unions and employer (and similar professional) organisations. These are contained in the mixed category EF472, which also involves membership fees for other organisations like parties, clubs, societies etc. Hence, again a plausible number of them are drawn randomly from the available observations. As the sample was restricted to households

with at least one dependently employed HH member, the share to be drawn is established by comparing the number of dependently employed persons subject to social security contributions of 29.6 bn in 2013 to an estimated number of union members and other eligible persons of 8.2 bn[26]. For the persons drawn as eligible, it is assumed that 75% of EF472 are tax-deductible membership fees.

Another deduction possibility presents itself for costs of motor vehicle repairs that are work-related. Of course, one cannot determine whether the repair expenses collected in variable EF384 are private or professional, hence the persons experiencing a repair related to work are drawn randomly using macro statistics again. Therefore, the number of accidents of 44.3 per 1000 motor vehicles in 2013 (Destatis 2014*b*) is combined with share of car usage for business and commuting reasons according to Radke (2014), which was 34.02%. As a rough estimate, it is assumed that half of the expenses reported for the affected households in EF384 are work-related and hence tax-deductible.

Adding to this, contributions for some insurances are deductible as professional expenses. For the professional share of the costs for accident insurance, 50% of the relevant variable EF468 are considered relevant here. As one cannot determine whether the payments are made for a private, professional or mixed insurance, I rather generously assume that all fall in the mixed category where regularly a fifty-fifty split is accepted by the fiscal authorities (Dittmann et al. 2014, p. 387).

Unfortunately, tax-deductible costs for a professional liability insurance cannot be considered, because they are not identifiable from the data[27]. The same holds for the costs of tax lawyer services, which are included in a mixed category of "other services" (EF456), along with a plethora of other items[28].

The next deductible item that can be derived are costs of moving due to professional

[26]Various sources were used to get this estimate. First and foremost, the IW Köln (2019) reports 6.1427 Mio members of DGB unions, plus 1.2764 Mio for the DBB and 0.27 Mio for CGB unions. The first is the main Federation of German Trade Unions, the next the German Federation of Civil Servants, the last the Christian Federation of Trade Unions. Adding to that, the union of dependently employed doctors (Marburger Bund) had some 0.115 Mio members, and there are a plethora of other possible federations. As the latter are too costly to study in detail, I simply add a generous further 0.5 Mio members eligible for deductions.

[27]In principle, these expenses are asked for in the EVS questionnaire (Destatis 2016, Item T/06 on p. 99), however they are not reported seperately in the microdata. Probably, the relevant insurance contributions are included in the mixed category item EF470, which is unfortunately too poorly documented to plausible estimate a share of liability insurance. In principle, liability insurance is widespread in Germany, however much less so for professional risks.

[28]These include costs for: any other lawyer services, job placement fees, funerals, decorative planting on graves, duplication expenses, newspaper advertisements, private detective services, typing office services, marriage counseling, other administration fees, energy pass, toilet fees, etc. (Destatis 2016, Item S/04 on p. 96). Therefore, deriving expenses for tax lawyer services from this seems futile.

reasons. They are partly included in variable EF385, which is another mixed category with "other services"[29]. To determine the tax-deductible moving costs, first the persons experiencing a move are randomly assigned. This is based on figures of Umzug AG (2014), who report a share of 15% of relocations due to professional reorientation, combined with an average share of people moving of 9.9% (Techem 2018). Additional conditions are applied to assure that only people reporting at least EUR 200 of expenses in EF385 and some inventory purchases or services related to them (EF333, 334, 340-343) are chosen. Next, it is assumed that for those assigned as moving, 75% of EF385 are actual moving costs. Additionally, the amount is restricted to EUR 2000 to account for the uncertainty induced by the variable's broadness. The EUR 2000 restriction is justified by case studies putting the highest price of transportation for moving inside Germany at EUR 1650 (Czycholl 2013).

Adding to the transportation costs, also expenses for new cooking stoves and heaters up to EUR 230 are tax-deductible and can be accounted for (EF343).

On top, people may also claim deductions for "other moving expenses" like aesthetic repairs, installation costs, etc. Here, the tax payer may simply deduct an allowance of EUR 687/695 for singles, or EUR 1374/1390 for married couples (before/after August 1, 2013). However, this is only allowed if the main home is moved. For simplicity, I assume that the latter is the case of there are transportation costs of more than EUR 300. The change of the allowance amount on August 1 is taken into account by assuming that households interviewed during Q3 and Q4 may claim the higher allowance, the others the lower one.

Nonetheless, some deductible moving costs are not identifiable in the EVS and therefore not included in my estimation: search costs like procuration fees, double payment of rent, additional schooling for children.

Moreover, also costs of a second home are tax-deductible if it is related to work. Then, costs up to the local comparative rent for 60m^2 may be deducted. Unfortunately, I have not identified these expenses in the EVS data.

Another item where deductions can be claimed are medical expenses related to occupational diseases and work accidents. Of course, the relevant EVS items (EF356, 361, 368, 369, 371, 373-375) again do not differentiate between private and occupationally related costs. Therefore, once more I draw some people randomly based on macro

[29] Except for costs of goods transport including furniture and rent of trucks and cars, it also includes expenses for: fees for bridges, roads, tunnels, ferrys etc., rent of motorcycles and bicycles, driving school, fees for driving licenses, motorvehicle registration fees, technical car inspection, parking fees, environmental badges, luggage carriers, luggage storage, courier services, rent of parking space related to work (Destatis 2016, Item J/10 on p. 84).

statistics. In 2013, 2.9% of employees were met with an accident related to work, the same number disaggregated by sex being 3.5% for men and 2.2% for women (Liersch 2014). Taking these differences into account, the persons are assigned randomly from those with positive medical expenses.

Moreover, also persons with an occupational disease are determined, because related expenses are fully deductible. Between 1992 and 2013, there were on average new cases of 8,040 women and 18,950 men with confirmed occupational diseases in Germany each year (Pfahl et al. 2016). The average length of employment in 2013 was 35.4 years for women and 40.2 years for men (Eurostat 2019). Multiplying the latter two figures gives a rough estimate of 0.285 Mio women and 0.762 Mio men with such a condition. Relating these figures to the 18.597 Mio employed women and 21.344 Mio employed men (in 2014, according to German statistical concept, Destatis 2015*b*), yields shares of 1.53% of employed females and 3.57% of employed males with an occupational disease. This is again a rather rough approximation[30], but a plausible one that Is use to draw shares from the relevant employed persons in the survey sample.

For the persons thus assigned an occupational disease, the tax-deductible part of the medical expenses is estimated. In the process, the household level expenses are divided by the number of people in the HH, assuming that everyone causes some sickness costs. As people with an occupational disease may cause higher expenses than the average HH member though, I adjust their share upwards by 50%.

A rather unimportant deduction that cannot be established from the data are costs of telecommunications, if plausibly related to occupation.

The next item that can be included are hospitality costs, of which 70% may be deducted if plausibly related to the occupation (by means of an expense receipt - Spesenbeleg). However, it is unlikely to be established precisely from the information in the EVS. Item EF435 includes all food and drinks in restaurants, cafés, ice cream parlours, food stalls and from delivery services that are consumed by everyone in the household. For simplicity, I divide by the number of HH members and take 20% of that as deductible professional hospitality expenses. As employer refunds have to be substracted, I reduce the amount by 20% of the item EF222, divided by the nunmber of dependently-employed in the HH.

The last item considered as professional expense are account fees, for which the fiscal authorities grant a EUR 16 allowance per year with no further proof needed. I

[30]Obviously, more granular information beyond sex could further disaggregate the picture along variables like occupation, income or age. Hence, by treating them all the same, probably deduction possibilities of lower income earners are underestimated, those of higher income earners overestimated.

Table A.6: Professional expenses of the dependently employed

	work equipment
+	commuting allowance
+	professional development
+	business travel
+	professional membership fees
+	job-related repair costs
+	professional accident insurance
+	job-related moving
+	job-related medical costs
+	professional hospitality costs
+	bank charges allowance
=	**Professional expenses**

Please note: Only includes deduction possibilites that could be considered for my model, given the data available.

include this for some occpuations[31].

Finally, all professional expenses of dependently employed persons are collected. Table A.6 gives an overview.

After all professional expenses are collected, also incomes from sidejobs in dependent employment (EF118) and refunds from carpools are added to taxable income.

Next, the flat-rate allowance of EUR 1,000 (Arbeitnehmerpauschbetrag) is considered. If the employee would claim less or no deductions, the amount is set as professional expense instead. Subsequently, the dataset is merged back together.

Until now, people with income from sidejobs or from honorary service remunerations (EF118), who are not otherwise dependently employed, have been excluded from the calculation of professional expenses. Their incomes are added, and the deductions calculated in a simplified way since their number is small and their incomes from this mostly small, too. Only the base allowance of EUR 1,000 is considered. Furthermore, for all dependently employed the honorary service allowance is included, under certain conditions[32].

[31] The EVS only reports branches. Those I have considered relevant are people in freelance, scientific and technical services (no. 13) and other economic services for firms and private persons (no. 14).

[32] Like for the self-employed, this allowance is intended only for income from honorary service that is charitable or serving the public good. Moreover, the occupation must be in a sidejob, i.e. less than 1/3 of a full-time job. I consider these conditions by only assigning the allowance to persons working in certain branches and under 14h per week or in a minijob.

A.2.2.4 Collection of the Sum of Revenues

At this point, the Sum of Revenues for all types of income can be calculated. This includes profit incomes (self-employed, business) and surplus incomes (dependent employment, rent and lease, capital, other) less the relevant deductions (business and professional expenses).

At this point in the code, it is useful to determine which married couples are opting for joint or single filing. This is because after the Sum of Revenues are determined joint filers are assessed together, i.e. most deduction possibilites are doubled. The default for married couples is joint filing, however they may opt for single filing instead. In that case, their remaining deductions are calculated sepererately, and the basic tax schedule for singles is applied.

Usually, joint filing is more useful for married couples: If partners have diverging incomes, especially if one of them is taxed at the top marginal tax rate for some part of his or her income, joint filing results in a lower tax rate on the common income, overcompensating the comparatively higher rate paid on the amount of the spouse with lower income. Single filing is profitable in fewer cases, e.g. when one partner has some transfer income (e.g. from unemployment insurance) that is subject to progressivity proviso and therefore increasing the tax rate on the income paid by the other partner. Or if one partner has negative income, hence eating up the allowances of the other if assessed together, one partner alone may file higher special expenses.

As no information on the filing status is given in the EVS, one has to either assume all couples use joint filing or make a profitability calculation and assign some filers to seperate filing. I do the latter and as a default, stick to a simple assumption: If one of the two partners differs by more than 20% from the common total income, the household uses joint filing.

A.2.3 The total amount of Revenues

Hereafter, some targeted allowances for certain groups are considered (single parents, elderly, farmers).

In 2013, farmers received a targeted allowance of EUR 670, if their income did not exceed EUR 30,700 (both amount doubled in case of joint filing). The relevant amounts are thus calculated and directly substracted from the Sum of Revenues.

Elderly persons in dependent employment are eligible for a proportional tax allowance, which is supposed to smooth the transition of upstream towards downstream

taxation of pension income in the German income tax system that was introduced starting in 2005 (§ 24a EStG). In my 2013 dataset, this was relevant for all tax filers born between 1940 and 1948, because the allowance depends on the year that follows after completing the 64th year of birth. The amount reaches 40% of their income for those becoming 65 in 2005 (max. EUR 1,900), decreased proportionally to 27.2% (max. EUR 1,292) for those turning 65 in 2013. The allowance is assessed seperately for joint filers. The determination base consists of gross income from dependent employment, plus the positive sum of other incomes[33] (all before deductions). Using the year of birth and the tables provided in the tax code, the relevant tax allowance for elderly employed persons (Altersentlastungsbetrag) is hence calculated.

The last special allowance to be included is the relief for single parents (Entlastungsbetrag für Alleinerziehende). The amount of EUR 1308 is granted if a person is single, widowed or divorced, living with at least one child in the same household and eligible for child benefits/allowance. Last but not least, the person must also not be living in a "household community" (Hausgemeinschaft), i.e. not together with other people whom they consume and pay together with[34]. Using the well-defined typecast item EF37, the single-parent households are identified and the allowance is granted if child benefits are received.

Substracting the tax allowance for the elderly and the relief for single parents from the Sum of Revenues, one now gets the total amount of Revenues (Gesamtbetrag der Einkünfte).

A.2.4 Special expenses

These are the quantitatively most important items among the tax deductions. Joint filers may derive some advantage here, because they may claim common special expenses and therefore higher amounts where applicable.

If no special expenses are claimed, a standard allowance of EUR 36 for single or EUR 72 for joint filers is considered.

Broadly, special expenses can be subdivided into precautionary expenses (for old age, social insurances and supplementary pensions) and other expenses (support payments, pensions rights adjustments, church tax, childcare, own or spouse's education, tuition fees, gratuities for political parties or tax-privileged organisations).

[33] I include the Sum of Revenues for income from self-employment, rent and lease and from other sources, but exclude capital income because the latter is only relevant here if the tax filer is subject to the personal income tax rate for her capital income rather than the flat rate of 25%.

[34] This condition had no practical relevance, because no such cases could be identified in the data.

A.2.4.1 Precautionary expenses

Retirement provision expenses are the first deductions in this category. Mostly, these are contributions to PAYGO public pension system including employer contribution, also to foreign public pension systems. Adding to that, contributions to insurance institutions of some occupations (farmers, pharmacists, doctors, solicitors, lawyers, etc) are included. Moreover, contributions to funded pension insurance systems (some private pensions, Rürup pension and occupational pension) may be deducted as special expenses. Contributions to public pension systems are captured in items EF229 and EF232, the latter included in variables EF234 (occupational pension) and EF526 (private pension).

The deductible amount is determined in several steps:

1. All potentially eligible contributions are collected[35].

   - For obligatory public pension contributions, this is simply done by taking employee contributions (EF229) and multiplying by 2 to also capture employer contributions.
   - The variable for private pension contributions (EF526) however is only available at the household level and has to be distributed among HH members. As default, I allocate it evenly among the dependently employed. Additionally, it is checked whether the household owns a Rürup (EF615) or some other private pension (EF613).
   - The contributions to public pensions of the self-employed, farmers and voluntarily employed are added. For farmers, public contributions subsidies (EF161) reduce the deductible amount, so they are substracted.
   - Voluntary public pension and occupational pension contributions (EF232 and EF234) are added.

2. The maximum is determined.

   - Regularly, it is EUR 20,000 for single and EUR 40,000 for joint filers.
   - For some occuptions however, the amount is cut: Civil servants, soldiers, judges, priests, Landtag/Bundestag/EP-deputies, proprietor-directors of GmbH and board members of joint stock companies[36]. The reduction amount is

[35] Theoretically, also contributions to additional insurance against occupational disability, reduced earning capacity or for survivor's provision are eligible under the condition that they are part of the pension insurance, and that the majority of the premium still pertains to a life annuity. However, the EVS lacks information to check this, of course.

[36] Of this group, only the latter three rather small groups cannot be determined from EVS items. The large rest is captured by variables EF8u8-EF13u8, category 4.

calculated as the virtual total contribution to the public insurance system (both employee and employer share), i.e. 18.9%*gross income, up to the assessment threshold of EUR 58,800. Hence, the maximum cut is EUR 11,113.

- Taking into account joint filing, the deductible retirement provision expenses are diminished by this reduction amount, yielding the applicable maximum (maßgebender Höchstbetrag).

3. Finally, the lower of the two amounts is taken. 76% of this less the tax-exempt employer share of public pension contributions, are the deductible retirement provisions.

The next category of deductions are other precautionary expenses, which again are subdivided into contributions to basic health and nursing insurance and the remainder. The former are fully deductible in principle, but some differentiations are made:

1. For obligatory contributions to public health insurance (EF226), the deductible amount is reduced by 4% if a claim on sick pay or similar benefits exists. No claim on sick pay exists for people in (partial) retirement, persons in obligatory military or community service, university students and interns. This personal characteristic is checked with EVS items EF8u8 - EF13u8. Additional conditions apply for some cases (all provided that there is a claim on sick pay):

   - Voluntarily publicly-insured self-employed have to substract 4% of their contributions (EF231) as well.
   - Voluntarily publicly-insured dependently employed have to substract 4% of their contributions (Ef231) less tax-free employer benefits for this purpose (EF116).
   - Voluntarily publicly insured retirees have to substract the employer benefits received from dependent work that assure claim on sick pay, or benefits received from the pension insurance[37].
   - Artists can substract benefits received from artist social insurance (EF141), from the base for the 4% cut of their voluntary public health insurance contributions (EF231).

2. The additional health insurance contribution that is set by each medical insurance (EF227), can be deducted fully.

[37] As there are less than ten observations for retired employees with such insurance contributions, I discard these cases.

3. In case of private health insurance contributions (EF235), only basic benefits are deductible, i.e. optional or comfort benefits are not.

4. Contributions to public nursing insurance (EF233) are fully deductible, as well as contributions to private obligatory nursing insurance (EF236).

The next subcategory are the remaining other precautionary expenses. All items that are relevant here are insurance contributions[38]:

1. the 4% cuts of public health insurance applied before, which are due to a claim on sick pay

2. the share of private health insurance due to optional or comfort benefits, and additional private health insurance including for travels (all included in EF462)

3. voluntary nursing insurance (EF463)

4. unemployment insurance (EF228)

5. accident insurance (EF468), if it has no premium refund (EF618) and if it is private not professional[39]

6. car liability insurance (EF464), but some additional steps are required in this case:

    - Only the part of the insurance that is private (i.e. not professional) is deductible, this is checked using the estimated professional expenses.
    - If the commuter allowance is claimed for using a private car, expenses for the car liability insurance may be claimed fully.
    - Unfortunately, EF464 is a mixed category that also contains premium payments for insurance on hull, which is not deductible. Some macro estimations are done using data on average premia for liability and hull insurance, to estimate the share of the variable that is deductible. According to GDV (2019), there were 60.2 Mio contracts for car liability insurance with an average premium of EUR 237, as well as 18.1 Mio contracts for partial coverage with an average premium of EUR 86 and 27 Mio contracts for full coverage with an average premium of EUR 292 for hull insurance. Considering

[38] All of these expenses are household level variables, which are allocated to the HH members evenly condiitioning on their employment status.

[39] In this case, it has to be deducted as professional expenses as outlined above. If such deductions are claimed, the amount is split in half for each deduction possibility.

this information, the average share of car liability insurance within variable EF464 should be 60.2%.

- As some households own several cars, I also divide by the number of cars.

7. private liability insurance (EF465)

8. disability insurance (EF467)

9. term life insurance (EF471)[40]

The sum of these items is only deductible, if the maximum amount for all "other precautionary expenses" is not exceeded yet. If this condition is met, the amount is filled up until the maximum amount is reached. This maximum is EUR 1,900 if one is either a receiver of benefits from the employer or from the public pension system for one's health insurance, or if one has a claim on state benefits for health costs (public officials, judges, priests, soldiers etc.). These conditions are checked using variables EF8u8 - Ef13u8, EF116u1 - EF116u6 and EF141u1 - EF141u6. Otherwise, the maximum is EUR 2,800 which mostly applies to people who pay for their health insurance on their own (i.e., the self-employed). The maximum amount of joint filers is the sum of their individual amounts. "Remaining other precautionary expenses" are calculated accordingly.

The last precautionary expenses that can be deductible as special expense are contributions to a some private pension insurance contracs, for the so called "Riester-Rente"[41]. The scheme and its tax treatment are rather complicated. It is lucrative mostly due to the state benefits that people receive on top of their contributions, the maximum base payment being EUR 154 per year. Moreover, for each child born before 2008 an additional amount of EUR 185 is granted, which was increased to EUR 300 for every child born in 2008 or afterwards. These amounts are conditioned on contributing at least 4% of previous year's income (minus the subsidy amount) and at least EUR 60, otherwise the maximum subsidy is cut proportionally if the applicable minimum is not reached. Moreover, the maximum amount for the state subsidy is EUR 2,100 per person.

[40]Conditions for some other capital/pension insurances are not testable with EVS information, especially whether a contracted was signed before 2005 or not.

[41]This type of contract was introduced in 2002 by then Minister for Employment and Social Affairs Walter Riester, with the intent to strenghten the third pillar of pensions in Germany. The "Riester-Rente" was supposed to add a widely-affordable capital-based retirement provision to the occupational pension and the public pension, the latter being cut to keep pension insurance contributions relatively constant in times of an ageing German demographic. The Riester-Rente is heavily subsidised, the distributional consequences being more regressive than hoped for (Corneo et al. 2018)

As I have no information on previous year's income, I use the current year's instead to calculate the minimum contribution to the Riester pension insurance that assures receiving the full subsidy. To correctly substract the subsidy amount for eligible children[42], some additional estimations are performed:

1. The number of own children living in the household born before/starting in 2008 are determined from their birth year (EF8u3 - EF15u3).

2. Being eligible for the child subsidy hinges on having received child benefits for that child for at least one month during the year. As no information on this is available I assign children to the first two persons in the household according to the following algorithm: 1 child → 1st HH member, 2 children → 1 for both, 3 children → 2 for 1st and 1 for 2nd, 4 children → 2 for both, etc. pp.

3. Moreover, people may be eligible for children that do not live with them, but for whom they receive child benefits. These were identified before in the code. As no information on their age is available, it is assumed they are all born in 2008 or afterwards and hence substantiate a claim on the higher EUR 300 state subsidy.

4. For joint filers, the basic allowance of EUR 154 may be transferred to the partner, who however also has to pay the minimun contribution then. Hence, the partners' minimum contributions are simply added up.

Unfortunately, variable EF526 is a mixed category for contributions to all private pension schemes, including Riester. As one cannot infer how the payments within it are distributed, I assume that people pay exactly the minimum amount of contributions for their Riester pension insurance. Only if the recorded contributions within EF526 are smaller than the minumum payment, they are taken.

Moreover, there is no item providing information on which individuals in the households actually own the Riester pension contract. Using variable EF614 that provides the wealth stock paid into all Riester contracts of the HH, it is therefore checked whether twice the possible minimum contributions payable by someone in the HH are smaller than this accumulated stock. If this is the case, I treat it as indicative that the Riester contract is not owned by that HH member, assigning her zero contribitions.

Another check concerns whether the stock is smaller than all possible Riester contributions of all HH members. The contributions collected in EF526 are probably often (partially) due to other private pensions (EF613, EF615). If double the amount of all calculated Riester contributions is smaller than the stock of Riester savings, and the

[42] I only correct the subsidies for the first two persons in the HH, because there is no way to tell whether other HH members have eligible children or not.

amount in stock for other private pensions savings is positive, Riester contributions are again set to zero for the individual.

Next, it is checked whether the Riester contributions thus estimated are below the previously established minimum contributions that make the individual eligible to the full state subsidy. Following the rules mentioned above, the basic state subsidy as well as the child bonuses are determined.

As a default, no deductions are claimed for the Riester contributions but the state subsidy is granted. However, tax payers may choose to forego the subsidy and instead claim an additional special expense for their Riester contributions, yielding a benefit particularly if their income is high and the state subsidy rather low (i.e. if they have no children). Whether using this is profitable for the tax payer, is checked automatically by the fiscal authorities. I have not included this in my estimation, but it could be possible with considerable programming cost.

A last item that could be included here are contributions to building societies (EF520), which can receive preferential tax treatment during the payout-phase. The latter however only applies to so-called "Wohnriester", i.e. if the arrangement is part of a Riester pension plan as described above. No information on these conditions is available, therefore I have not included this possible special expense in my estimation[43]

A.2.4.2 Other special expenses

A range of other costs can be deducted as special expenses. The first are payments based on pension rights adjustment after a divorce. Unfortunately, the calculation of possible deductions is complicated:

1. It has to be differentiated whether the pension is a public or private life annuity, or a payment from a pension fund.

2. The special expense is granted only when the split is based on a contractual agreement, not when it is due to a legally compulsory pension rights adjustment. However, variable EF237 could contain both variants so the share of people who may use the special expenses deduction has to be estimated. According to the German Public Pension System[44] , 80% of pension rights adjustments are legally compulsory while the remainder is split between contractual and internal agreements. Hence, one could assume that the contractual share is 20% and draw

[43]Please not that of roughly 30 Mio. building society contracts, only 1.15 Mio were "Wohnriester" in 2013 (BMAS 2019).

[44]These information were provided by E-Mail by Wolfgang Keck, responsible employee for the relevant statistics at Deutsche Rentenversicherung Bund.

those randomly.

3. But on top, EF237 also contains supportive payments to different groups (ex-partner, children, parents), not only pension rights adjustments. It is moreover unclear whether pension rights adjustment claims are contained in any EVS item at all (including EF237).

Due to the last point, I have abstained from including this deduction possibility in my estimation.

This notwithstanding, special expenses deductions for alimony payments are included in the estimation. These may be claimed up to EUR 13,805 when divorced or living permanently seperately. However, the receiver of the payment has to agree because she has pay tax on it. Usually, the arrangement (called "Realsplitting") is beneficial for both partners especially if their earnings differential is high, i.e. after a divorce in a classical sole-breadwinner marriage.

Unfortunately though, no information on the agreement of partners to this "Realsplitting" are available in the EVS, so the distribution of the types among the payments within variable EF237 has to be estimated. No macro statistics on the distribution of cases could be found, the first best source could probably be the tax statistics. So as a first best guess, I assume that half of the cases of divorced persons with positive payments in EF237 are relevant here. They are drawn randomly, further assuming that for them the full amount reported in the potentially mixed category is tax deductible, up to the maximum amount.

The next item deductible as special expense are costs for own/spouse'seducation. Up to EUR 6000 can be deducted if they pertain to the first professional education. I.e., expenses for a 2nd vocational training or a Master's programme on top of a Bachelor degree are not eligible here[45]. The latter is difficult to ascertain from EVS information, so some simplifying assumptions are made: It is checked whether individuals have no finished professional training or degree, whether they have not already claimed professional educational expenses, and whether the social status is in the appropriate category (student, worker or employee including persons undergoing vocational training). If all of these hold true, the following items are collected as eligible special expenses:

1. Tuition and exam fees (EF432) are fully deductible. The HH-level variable is allocated according to the number of employed persons in the HH, but adjusted upwards by 50% to reflect the presumably higher share of people undergoing their

[45]They may be deducted as anticipated professional expenses, though.

first professional education in comparison to average employees or self-employed persons.

2. Checking for ownership of the relevant hardware, the costs for PCs, laptops and printers (EF401) are collected as deduction. Again, the HH-level item is divided by the number of employed people, but adjusted upwards (by 25% in this case).

3. Costs of books, journals and alike (EF423-425) are allocated by dividing through the number of persons in the household, adjusted upwards by 50%.

4. Transportation costs are considered for car (EF383) and public transport (EF386) users, however only if the person has claimed educational expenses before in the code and has not claimed commuter or business travel expenses. The HH-level costs are divided by the number of employed persons in the HH. Additionally, the macro shares of driving and public transport due to education are taken from Radke (2014), adjusted upwards by 10 and 4 times, respectively, and applied to the costs calculated before.

5. Interest on student loans (not repayments) are also eligible. Variable EF607 contains the residual debt of educational loans, including interest payments. Since only the latter can be deducted its share has to be deducted. I assume a 4% annual interest rate for the contracts contained in EF607[46]. Since no further information on the type of loan or the duration is available, I simply multiply this interest rate with the residual debt to arrive at the deductible expenses.

Finally, all deductions in this category are summed up and the maximum amount condition is applied.

Coming to the next deduction possibility, 2/3 of childcare expenses (EF430 and 431) for children up to 14 years may be deducted, up to EUR 4,000 per child. The relevant expenses are thus calculated. By law, the apportionment between spouses should be according to actually incurred expenses. Since no such information is available, I allocate according to the tax payment share between the two instead.[47].

[46]This is based on the following thoughts: The most common educational loans in Germany are "KfW-Studienkredit" and "KfW-BaFöG-Darlehen", which were subject to different interest rates depending on the fixed interest period. The former's rate was 3.06-4.34% in 10/2012, the latter's 2.10-3.47% in 04/2012 (Studis Online 2019). For both, the span reports the variable rates and the 10 years fixed rates.

[47]It should be noted that there are few cases of households with childcare expenses but no children in the HH. These are probably connected to alimonies. However, there is no way to determine whether the payments pertain to children of an eligible HH member, so this issue is disregarded.

Also paid church tax (EF227) is eligible as special expense. No further processing is necessary since the variable is available on the personal level.

Another important special expense are membership fees and donations to associations that are charitable, serving the public good or churchly; and to political parties and registered eloctoral groups. For the latter, up to EUR 1,650 can be deducted. Totally, no more than 20% of the Sum of Revenues are deductible. This condition is checked after the following four components of this category are calculated.

First, I estimate political donations and memberships fees. For this purpose, a number of steps are necessary:

1. From Niedermayer (2017), the number of party members for the bigger parties is collected, additionally assuming some 100,000 members of other parties. Relating the sum to the population above 15 years gives a ratio of approximately 1.989% of people being members of a political party.

2. This share is drawn randomly from the persons in households reporting positive membership fees (EF472), thus establishing party members.

3. EF472 is a mixed HH-level category with membership fee payments for all sorts of clubs and associations. I allocate it by dividing through the number of HH members, adjusted upwards by 50% for party members' membership fees.

Second, the share of variable EF472 that goes to charitable associations is estimated as follows: Dues for large parts of all associations (Vereine) do not enjoy prefential tax treatment. This concerns for instance sports clubs (with some 24 Mio. members in 2013), automobile clubs (ADAC membership: 16 Mio.) or shooting clubs (1.5 Mio. members). Membership fees to other associations are tax-deductible though, for instance the German Red Cross (4.5 Mio. members), social association VdK (1.4 Mio. members) or charities like ASB (1.1 Mio. members). As a consequence, I assume that 1/3 of the employment-weighted membership dues reported in EF472 are tax-deductible as special expenses. For party members, the 1/3 pertains to the amount of EF472 after party membership dues have been substracted.

Third, donations to non-proft organizations (EF473) are apportioned to the HH members based on their share in tax payments. Almost all of this variable should be tax-deductible. According to Deutscher Spendenrat & GfK (2014), about 8.4% of donations in Germany in 2013 were given with other/no purpose given, and an additional 4.8% were donated for sports or culture/homeland (which nonetheless may be tax-deductible under some conditons). Therefore, it is assumed that 90% of EF473 is eligible as tax write-off.

Fourth, dues and donations[48] up to EUR 825 per spouse may be deducted as a direct tax relief also. Therefore, it is ensured that only the amount exceeding EUR 825 is eligible as a special expense.

A.2.4.3 Summing up special expenses

Finally, all special expenses are collected as in table A.7. If less special expenses are claimed, the standard allowance of EUR 36 or EUR 72 for joint filers is included.

Table A.7: Special expenses

	Old age provisions
+	provisional expenses for basic health insurance
+	provisional expenses for basic nursing insurance
+	other provisional expenses
+	alimony payments
+	expenses for own or spouse's education
+	childcare expenses
+	church tax payments
+	membership dues and donations
=	**Special expenses**

Please note: Only includes deduction possibilites that could be considered for my model, given the data available.

A.2.5 Extraordinary burden

The deductibles in this category are often case-specific, as the title already suggests. Some itemised cases are specified by the fiscal autorities, however these are mostly not identifiable with EVS information. The latter is the case for tax reliefs for disabled people, for persons who are nursing others and for the allowance for survivors.

Some of the general extraordinary burden cases are not identifiable in the EVS data, too: This holds for expenses for children's education who are grown up and staying outside of their parent HH, for home reconstruction costs in special cases (disability, sickness, danger, etc.) and for payments for the health insurance of other supported persons.

The first general extraordinary burden that can be identified in the data are expenses for nursing homes, which are fully tax-deductible if the inhabitant is not only

[48] Unfortunately, only dues can be determined. Estimating donations to political parties and registered voter groups may be added, slightly increasing the amount of (more valuable) direct tax reliefs at the cost of special expenses.

there for ageing reasons and paying for oneself or one's partner. Assets that were given to the tax payer in order to pay for the nursing cost have to be substracted here. However, less than ten cases with positive income could be detected, because people living in institutions are excluded from the survey. Therefore, and because the conditions requre further elaboration, this category is dismissed in my analysis.

What is included though, are expenses for ambulent nursing care (EF453). From them, payments from the public nursing insurance (EF162u1-u6) are substracted[49]. The resulting HH-level sum is then allocated to the HH members according to the tax payment share. This is arguably a rather crude apportionment, but in the absence of a better ad hoc solution it may be justified by the fact that the affected households have predominantly old age members[50].

The other big, and arguably most important, extraordinary burden item that is taken care of, are sickness costs. Most of the conditions that have to be met for eligibility cannot be checked with EVS information, for instance whether a simple or a qualified proof is necessary. Nonetheless, in the attempt to model a benevolent tax authority, tax write-offs for sickness costs are included in my estimation as follows:

1. All the medical expenses in the survey are summed up (EF356-375).

2. Because professional expenses have priority over extraordinary burden deductions, professional expenses due to a professional illness (as determined earlier in the code) are substracted.

3. Refunds received from private health insurance companies, which are recorded at the HH level, are substracted.

4. Finally, the amount is allocated to the HH members according to their tax payments, reduced by the individual level receipts from the public health insurance (EF145u1-u6 and EF146u1-u6). This yields the sickness costs deductible as extraordinary burden for each HH member.

One more category of expenses is deductible as an extraordinary burden: supportive payments, e.g. alimonies, that are not eligible as special expenses. These include in

[49]Unfortunately, payments from private nursing insurances are not recorded in a seperate item. They could theoretically be included either in EF185 (payments from private health insurances) or EF186 (payments from other private insurances, where only accident, car and household goods are named as examples). This possibly inflates tax-deductible expenses here, as it reduces the refunds that have to be substracted from the deduction.

[50]Unfortunately, health condition or disability is not observable in the data, hence the allocation cannot be easily more precise. Some, presumably minor, future improvement could be to only consider people above some age threshold and distribute the deduction according to the tax payment share among them.

particular payments for partners living abroad, children without income older than 25, relatives and mothers of extramarital children. Because this is a rather broad write-off opportunities, several conditions that are not testable with the EVS data, have to be met. It is however calculable how much a taxpayer may claim at the maximum, which is the so-called "sacrifice restriction" (Opfergrenze).

Two items come into consideration here. EF474 is a mixed category that contains voluntary support payments, but also money gifts. Because the share of the latter cannot be estimated easily, it is not included.Hence, only obligatory supportive payments (EF237u1-u6) that were not already subject to the special expenses deduction, are considered.

Next, the relevant net income for the "sacrifice restriction" is calculated according to table A.8. Based on this, for every EUR 500 of the net income, the "sacrifice restriction" is increased by one percentage point, up to 50% at the maximum. Moreover, the amount is reduced by 5%pts for a married partner and by another 5%pts for every relevant child.

Hence, taking into account the sacrifice restriction or alternatively EUR 8004 as the maximum, supportive payments deductions are calculated. Because no information on this is specified in the EVS, basic health and nursing insurance payments for the person supported cannot be included here [51].

Table A.8: Sacrifice restriction

	Gross employment remuneration less professional expenses (Sum of Revenues)
−	statutory deductions (income tax, solidarity surcharge, church tax, social security contributions)
+	tax-exempt wage-replacement benefits (sick pay, unemployment benefits, short-time benefits, maternity benefits, parental allowance)
+	child benefits
+	tax refunds
+	interest income less savings allowance and taxes
+	self-employed income
+	income from rent and lease
+	other income
+	non-taxable share of pensions
=	**relevant net income**

[51] Ideally, one would like to account for a number of further factors: Several people might be supported, their income and some public transfers would have to be substracted from the amount deductible, depending on who the supported person lives together with, whether they live abroad etc pp. Unfortunately, again there is too little information in the EVS to do that. Nonetheless, I at least crudely correct for HH heads with very high payments, assuming they support additional people. If the burden exceeds EUR 16,008, 60% of it are considered tax-deductible, 50% where it exceeds EUR 24,012.

Adding to this, also the reasonable burden (zumutbare Eigenbelastung) has to be calculated, i.e. the amount the fiscal authority deems acceptable for the tax payer to carry herself without further tax relief. It depends on the total amount of revenues of the taxpayer, i.e. taxable income less professional/business expenses and specific reliefs for elderly and single parents. The reasonable burden also depends on marriage status and the number of children, as laid out in table A.9.

Table A.9: Reasonable burden

Total amount of income	default case	Widowed[52]; joint filers	all w/ 1-2 children	all w/ > 2 children
$\leq 15,430$	5%	4%	2%	1%
$15,431 - 51,130$	6%	5%	3%	1%
≥ 51131	7%	6%	4%	1%

Source: Dittmann et al. (2014, p. 236)

To calculate their maximum burden, these conditions are tested for all individuals in the survey. All extraordinary burden items except alimony payments, which are not subject to the reasonable burden restriction but the sacrifice restriction derived earler, are then summed up. Extraordinary burden tax deductions are generated by determining the smaller of either the reasonable burden or the calculated expenses that qualify as a tax-write off in this category. Adding also the deductible alimony payments then yiels all extraordinary burden deductions.

A.2.6 Taxable income

At this point, the taxable income (zu versteuerndes Einkommen - zvE) may be calculated. Following the overview in table A.2, this is done by simply taking the Total amount of Revenues that was established in section A.2.3, and substracting the special expenses derived in section A.2.4 and the extraordinary burden deductions calculated in the previous section A.2.5.

Progressivity proviso

Some types of income are not liable to tax according to the law, but increase the tax rate on the income that is taxable. This so-called "progressivity proviso" (Progressionsvorbehalt) is applied as follows:

1. These incomes are added to the tax base, and tax due on this hypothetical income is generated.

2. The resulting tax rate is applied to the actual income without the tax-exempt components.

As a result of the progressive income schedule, a higher amount of tax has to be paid. The relevant income types that are subject to progressivity proviso are:

- unemployment benefits ALG I (EF147)
- sick pay (EF145)
- short-time compensation (EF148)
- payments from the European Social Fund (EF160)
- some employment promotion benefits (EF152)
- maternity pay (EF153)
- parent allowance (EF158)

Variable EF158 unfortunately is another mixed category that combines the federal level parent allowance (Elterngeld) with the sub-national level education allowance (Landeserziehungsgeld) that is paid in some federal states . The latter is not subject to progressivity proviso, which is taken into account as follows: In 2013, EUR 4.9 bn were paid out as parent allowance, while Landeserziehungsgeld was only handed out in three federal states (Saxonia, Bavaria, Baden-Württemberg). Hence for simplicity, I assume that in these states the latter makes up 20% of EF158.

Taking this into account, a variable that captures all taxable income including that which is subject to progressivity proviso, is calculated.

A.2.7 Calculating taxes due

The income tax schedule is summarised in table A.10. When calculating taxes due, progressivity proviso is taken into account by calculating tax due based on all taxable income plus the amount subject to progressivity proviso, then taking the resulting tax rate and applying it to the taxable income only. Further rules are applied subsequently.

This schedule is applied to every HH member with positive taxable income. Moreover, joint filing is taken into account: The common taxable income is halved, then these amounts are assessed according to the regular schedule and added up. In the code therefore, the common zvE is halved, the tax due is calculated and doubled.

Table A.10: German income tax schedule, 2013

Taxable income (zvE) in EUR	tax due ESt =
$\leq 8,130$	0
$8,131 - 13,469$	$(933.7 * y + 1,400) * y$
	$y = (\text{zvE} - 8,130)/10,000$
$13,470 - 52,881$	$(228.74 * z + 2,397) * z + 1,014$
	$z = (\text{zvE} - 13,469)/10,000$
$52,882 - 250,730$	$0,42 * \text{zvE} - 8.196$
$\geq 250,731$	$0,45 * \text{zvE} - 15.718$

Please note: zvE denotes the taxable income (zu versteuerndes Einkommen).

A.2.7.1 Child allowance

Taxpayers may choose between two options:

1. They may keep the child benefits that are paid out to parents and derive no further tax relief. This is usually beneficial to most households.

2. Alternatively, they may use the tax allowance for children of EUR 184 per child per month, and additionally the education allowance of EUR 110 per child per month. Every parent is eligible for said amounts. Then, the allowance is considered when determining the taxable income, a higher tax relief than the already paid-out child benefits is refunded.

The fiscal authorities automatically check which one of these options is more profitable for the taxpayer, and apply the rules accordingly. Hence, the same is done in my code.

For this purpose, several steps are taken:

1. For the 1st and 2nd person in the HH[53], the child benefits that were received are allocated[54].

2. The theoretical tax allowance per parent is calculated by assuming the person is eligible for the full year amount[55].

[53]It is not possible to determine whether there might be several married couples having children, and to whom those children pertain to, in the HH. Therefore, it is assumed the children are theirs.

[54]In the EVS, child benefits are not recorded as in the tax law or in practice, i.e. the relevant parent receives it. Instead, they are reported as belonging to the children living in the HH. Hence, all the individual level payments (EF167u1-u6) are summed up and allocated by fifty-fifty split if married couples do not use joint filing.

[55]Please note that this again works in the direction of granting more deductions than justified in

3. The resulting tax due and hence tax savings are calculated by reducing taxable income by the child allowance thus established, and applying the tax schedule.

4. Tax savings from the tax allowance are compared with paid-out child benefits, determining whether it is profitable to use the allowance. If this is the case, tax due is corrected accordingly.

A.2.7.2 The fifth rule

This rule is applied if a taxpayer receives a large one-time income in the respective tax year. Then, it may be profitable to only add one fifth of this payment to the taxable income, while the resulting additional amount of tax due is multiplied by five. However, only indemnities or payments for "long-time occupation", stretching at least twelve months over two different years, are eligible. The second case is not captured by the EVS, but to be conservative both indemnities (EF112) and one-time payments (EF110) are checked in my model. This is done similarly in both cases:

1. Taxes due with the relevant item excluded from taxable income are calculated using the tax schedule and taking into account joint filing.

2. Indemnities/one-time payments are added to taxable income according to the fifth rule; i.e. 20% of it. The tax schedule is applied accordingly.

3. Tax rates are derived and the tax due is calculated. The difference between the two is calculated, multiplied by five and added to the tax due without the relevant item.

4. The result is compared to the default where the additional payment is not taxed under the fifth rule, and the lower tax liability is assigned.

This is done only for the 1st and 2nd HH member, because there are very few relevant cases for the remaining HH members.

A.2.7.3 Tax reductions

These deductions are substracted directly from the tax due, i.e. they are the most valuable. Three different categories of expenses are eligible: Craftsmen services, employment and services close to home, as well as donations and membership dues for

reality. This is because actually, HHs might only be eligible for the allowance during part of the year. We cannot observe this, but the recorded child benefits may reflect it. So when comparing paid child benefits and the tax saving from the allowance, I might bloat up the latter, leading to additional unjustified tax savings.

political parties and registered voter groups. For the former two, typically 20% of expenses, up to a maximum amount, are deductible.

Craftsmen services

For an appartment or house that is used by the taxpayer herself, expenses for several items are eligible if the work is done in the home of the taxpayer. 20% of the costs are deductible, yet excluding material costs and transportation costs for material. The maximum amount of the tax relief is limited to EUR 1,200 per household, also for married couples. The following items in the EVS are (partially) eligible for the deduction:

1. Expenses for maintenance and repairs are collected in variables EF282-284. These include material costs, which are typically around 1/3 of such craftsmen bills[56]. Therefore, 2/3 of the expenses reported in these variables are considered here.

2. Expenses for the installation and transportation of furniture and lamps (EF333) are deductible only with the part that is due to installation, which I assume to make up half of the item. Moreover, the costs can be deducted only when they are not part of a new construction (erstmalige Fertigstellung des eigenen Haushalts). I assume that this is the case for 3/4 of the positive observations vor variable EF333, and draw them randomly.

3. Expenses for the laying out of carpets (Ef335) are subject to the same condition as those for furniture and lamps, so they are treated equally.

4. Expenses for the installation of large HH gadgets (EF342) are treated in the same way, too.

5. Expenses for the repair of furniture and fixtures and of floor covers (EF337), are notwithstanding assumed to take place at home of the taxpayer in 90% of cases. Hence, 10% are randomly drawn and excluded from tax deductibility.

6. Expenses for the repair of HH gadgets, including rented ones (EF345), are assumed to take place at home of the taxpayer in half of cases, and are drawn randomly accordingly.

Finally, all the craftsment service deductions are summed up and the maximum amount of EUR 1,200 per HH is checked.

[56]This assumption is based on searching online for said shares. For different works and trades, see for instance Immonet.de (2019)

Employment and services close to home

In this category, several items may be deducted: regular and "minor" employment at home, nursing home and care expenses, private moving costs, the latter three in case the service is performed by an alien party (firm). 20% of the combined expenses for regular employment and services are then deductible, up to EUR 4,000 in total.

For the variable capturing employment at home (EF355), a number of adjustments are necessary. This is first, because employing people at home is still mostly done illegally, estimates say by about 80-90% (Enste 2017). The same study reports that around 10-12% of all households use such services, which roughly squares with the 13.9% of households reporting some positive EF355 expenses in the EVS. Nonetheless, such employment can of course be registered lawfully. Then, 20% of the expenses up to EUR 510 are deductible if it is a "minor employment", i.e. part-time work with few hours. Alternatively, if it is subject to social security contributions i.e. a regular employment, it can be deducted more extensively. To determine the deductible employment expenses, the following steps are taken:

1. From Enste (2017), shares of correctly registered minijobbers (8.8%) and regularly employed (1.4%) among the HH employees are derived.

2. Assuming that they were employed at least half a month for EUR 225 in 2013, and up to the whole year with a remuneration of EUR 5,400 (EUR 450 * 12 months), correctly registered minijobbers are drawn from positive EF355 observations. 20% of these, up to EUR 510, are assigned as tax reduction.

3. Assuming minimum expenses, i.e. renumeration of employees, of at least EUR 1,000, lawfully registered regularly employed are drawn randomly from variable EF 355. Further, it is assumed that households either employ regularly, or a minijobber. 20% of expenses thus determined are recorded as tax reduction.

Next, expenses for services close to home are determined. The first item are costs of stationary and ambulant nursing, which are only eligible for a tax reduction if they have not been deducted as an extraordinary burden before. Here, only expenses due to ageing, not to sickness, are deductible. As before, this condition cannot be checked with EVS items, hence for ambulant nursing service expenses (EF453), the amount deducted as an exceptional burden is substracted to get the eligible amount. For expenses for nursing homes (EF452), no such correction is necessary. Of both amounts, 20% are calculated to be the relevant tax reduction.

The next item in this category are moving costs for private reasons. These are deductible only when an alien party, i.e. a firm, is used. The same macro statistics

as before, when determing professional costs due to moving, are considered: From Umzug AG (2014), it is known that 85% of relocations are due to private reasons, and only 25% of the movers use a private firm. Moreover, on average 9.9% of people in Germany moved in 2013 (Techem 2018). Some further steps are taken to determine the tax reduction from the relevant variable EF385:

1. As EF385 is a mixed category[57], I seek to target the HHs that actually move. Hence, it is checked whether HHs with EF385 expenses of at least EUR 500 also report inventory purchases or services related to them. Moreover, those who already claimed deductions due to professional moving are excluded. From the remainder, the relevant share of people moving for private reasons and using a private firm are drawn randomly.

2. Additionally, it is assumed that only 50% of EF385 are moving costs for the selected households (remember that EF 385 is a diverse collection). Moreover, the amount is upward-limited to EUR 2,500, which is justified by the fact that the highest transportation price for a scenario to move inside Germany was EUR 1,650 (Czycholl 2013).

3. Of the expenses thus established, 20% are deemed deductible.

Finally, all items deductible as tax reduction so far are summed up, and the maximum amount of EUR 4,000 per HH is applied.

Membership dues and donations to political parties

Here, 50% of expenses, up to EUR 825 per spouse, are eligible. The same holds once more for "independent voter groups", but this condition is of course untestable with EVS information. When the special expenses for membership dues were calculated, the amount that is eligible as a tax reduction was already determined. What is missing in my model though, are donations to political parties and independent voter groups.

A.2.7.4 Final tax due

As all direct tax reliefs are determined, substracting them from taxes due yields the corrected taxes due.

Moreover, now the solidarity surcharge of 5.5% on taxes due is added, taking into account the minumum taxes due of EUR 972 (EUR 1944 for joint filers).

Profitability calculation for capital income

Another profitability calculation that is performed automatically by the fiscal author-

[57] For more information, see footnote 29.

ities is the one that determines whether it is beneficial for the taxpayer to subject his capital income to his income tax rate rather than the flatrate withholding tax on capital income. Hence, this applies to rather small income earners whose marginal tax rate is lower than the 25% withholding tax. This test is implemented as follows:

1. The profitability consideration concerning the fifth rule and the child tax allowance are excluded from this particular profitability test[58]. The relevant taxable income including capital income is calculated including the progressivity proviso.

2. It adds the total amount of capital income, the total amount of all other taxable income and the income subject to progressivity proviso, and substracts special expenses and extraordinary burden deductions. Joint filing is of course aken into account, too.

3. Through application of the tax schedule, the tax rate is determined. Subsequently, the tax due on capital income is calculated and compared to the flatrate withholding tax. Hence, profitability calculation is completed.

After this is done, and to rule out that payers of flat rate withholding tax receive a tax reduction, the flat rate withholding tax payments are now added to overall taxes due.

This concludes the model of the German income tax code for 2013 using the EVS, on the side of what people should pay in taxes.

A.2.8 Estimating actual tax payments

The tax payments that are reported in the EVS contain the current year's payments only, and of course not yet the refunds for tax year 2013. Fortunately though, the refunds of the current year, i.e. for taxes paid in 2012, are recorded in the survey. From this variable EF183, current year's refunds are calculated as follows:

1. According to Destatis (2017*d*), average refunds for 11.5 Mio taxpayers in 2013 were EUR 935, while 1.5 Mio. taxpayers paid EUR 988 on average. In 2012, the relevant amounts were EUR 901 for 11.4 Mio. refunds and EUR 965 for 1.5 Mio. arrears. Hence, refunds rose by 3.8% and arrears by 3.4% between the two years.

[58]The additional precision derived from including them is presumably rather limited, but it could be implemented at rather high programming costs.

2. Unfortunately, arrears are not observed in the EVS[59]. Given their average growth rate, refunds can be taken care of. For this purpose, the HH-level variable EF183 is also allocated to the HH members according to tax payments.

Substracting these refunds from the reported tax payments in EF224 and EF225 then yields the actually paid tax in 2013.

In this part of the code, also the taxable income based on taxes paid is derived. Therefore, the tax schedule is inverted which approximately yields the following formula for the lower progression zone:

$$\text{zvE}^{paid} = 630 + 10000 * (T/933.7 + 0.56)^{1/2}$$

where T denotes the tax payment. For the upper progression zone, the formula is approximately

$$\text{zvE}^{paid} = -38927 + (T/228.74 + 23.02)^{1/2}$$

For the 1st proportional zone, the formula is simple:

$$\text{zvE}^{paid} = (50/21) * (T + 8196)$$

Also for the 2nd proportional zone, a simple formula may be derived:

$$\text{zvE}^{paid} = (20/9) * (T + 15718)$$

A.2.9 The estimation of tax underpayments

As described in the paper, the simple deduction of taxes paid from estimated taxes due yields the tax underpayment:

$$T_i^{loss} = T_i^{yall} - T_i$$

For further analysis, the most relevant variables are combined into a homogenous dataset containing all individuals born before 1996. These include income aggregates, income components, deductions, personal characteristics, household indicators and tax payments. The resulting dataset is structured such that the variables are the same for

[59]They are included in the current year's tax payments. This should not be a concern for my overall estimate, given that the amounts increased over time. If anything, it again contributes to the underestimation of tax avoidance.

all included individuals, i.e. the u1-u6 notation is eliminated.

Bibliography

ACTA (2015), Allensbacher Computer- und Technik-Analyse - ACTA 2015, Bericht, IfD Allensbach, Allensbach.

Albarea, A., Bernasconi, M., Marenzi, A. & Rizzi, D. (2020), 'Income underreporting and tax evasion in Italy: Estimates and distributional effects', *Review of Income and Wealth* **66**(4), 904–930.

Alstadsaeter, A., Johannesen, N. & Zucman, G. (2018*a*), Tax Evasion and Inequality, CEPR Discussion Papers 12781, C.E.P.R. Discussion Papers.

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

Alstadsæter, A., Johannesen, N. & Zucman, G. (2017), Who Owns the Wealth in Tax Havens? Macro Evidence and Implications for Global Inequality, NBER Working Papers 23805, National Bureau of Economic Research.

Alstadsæter, A., Johannesen, N. & Zucman, G. (2019), 'Tax Evasion and Inequality', *American Economic Review* **109**(6), 2073–2103.

Bach, S., Beznoska, M. & Steiner, V. (2017), *Wer traegt die Steuerlast? Verteilungswirkungen des deutschen Steuer- und Transfersystems*, Vol. 347, study edn, Hans-Boeckler-Stiftung.

Bach, S. & Buggeln, M. (2020), 'Geburtsstunde des modernen Steuerstaats in Deutschland 1919/1920', *Wirtschaftsdients* **100**, 42–48.

Bartels, C. & Jenderny, K. (2015), The Role of Capital Income for Top Incomes Shares in Germany, World Inequality Lab Working Papers 201501, HAL.

Bazzoli, M., Caro, P. D., Figari, F., Fiorio, C. V. & Manzo, M. (2020), Size, heterogeneity and distributional effects of self-employment income tax evasion in Italy,

FBK-IRVAPP Working Papers 2020-02, Research Institute for the Evaluation of Public Policies (IRVAPP), Bruno Kessler Foundation.

Becker, I. (2014), EVS und SOEP: methodische Aspekte bei Verteilungsanalysen, soeb-Working-Paper 2014-3, Forschungsverbund Soziooekonomische Berichterstattung.

Benedek, D. & Lelkes, O. (2011), 'The Distributional Implications of Income Under-Reporting in Hungary', *Fiscal Studies* **32**(4), 539–560.

Bittschi, B., Borgloh, S. & Moessinger, M.-D. (2016), On tax evasion, entrepreneurial generosity and fungible assets, ZEW Discussion Papers 16-024, ZEW - Leibniz Centre for European Economic Research.

BMAS (2019), 'Statistik zur privaten Altersvorsorge (Riester-Rente)'. Last accessed on January 2, 2020 at: `https://www.bmas.de/DE/Themen/Rente/Zusaetzliche-Altersvorsorge/statistik-zusaetzliche-altersvorsorge.html`.

BMF (2014), 'Übersicht über die Steuerrechtsänderungen seit 1964', Bundesministerium der Finanzen.

BMF (2015), 'Finanzbericht 2016', Bundesministerium der Finanzen.

BMF (2016), 'Finanzbericht 2017', Bundesministerium der Finanzen.

BMF (2017), 'Doppelbesteuerungsabkommen - DBA - sowie weitere staatenbezogene Veröffentlichungen', Bundesministerium der Finanzen. Last accessed on January 12, 2017 at: `https://www.bundesfinanzministerium.de/Web/DE/Themen/Steuern/Internationales_Steuerrecht/Staatenbezogene_Informationen/staatenbezogene_info.html`.

BMF (2020*a*), Datensammlungen zur Steuerpolitik 2019, Technical report, Bundesministerium der Finanzen.

BMF (2020*b*), Notwendigkeit, Potenzial und Ansatzpunkte einer Verbesserung der Dateninfrastruktur für die Steuerpolitik, Gutachten 5/2020, Wissenschaftlicher Beirat beim Bundesministerium der Finanzen, Berlin.

Bönke, T., Schröder, C. & Werdt, C. (2013), 'Compiling a harmonized database from Germany's 1978 to 2003 sample surveys of income and expenditure', *AStA Wirtschafts- und Sozialstatistisches Archiv* **7**(3), 135–168.

Canzler, W. (2013), *Räumliche Mobilität und regionale Unterschiede*, Datenreport, 2013 edn, Destatis, WZB, SOEP, bpb, chapter 10, pp. 305–322.

Casi, E., Spengel, C. & Stage, B. M. (2020), 'Cross-border tax evasion after the common reporting standard: Game over?', *Journal of Public Economics* **190**(C).

Cobham, A., Janský, P. & Meinzer, M. (2015), 'The Financial Secrecy Index: Shedding New Light on the Geography of Secrecy', *Economic Geography* **91**(3), 281–303.

Corneo, G. (2005), 'The Rise and Likely Fall of the German Income Tax, 1958-2005', *CESifo Economic Studies* **51**(1), 159–186.

Corneo, G., König, J. & Schröder, C. (2018), 'Distributional Effects of Subsidizing Retirement Savings Accounts: Evidence from Germany', *FinanzArchiv: Public Finance Analysis* **74**(4), 415–445.

Czycholl, H. (2013), Umziehen geht oft auch zum halben Preis, Newspaper article 21.04.2013, Die Welt. Last accessed on January 3, 2020 at: `https://www.welt.de/finanzen/verbraucher/article115477807/Umziehen-geht-auch-oft-zum-halben-Preis.html`.

de Willebois, E. v. d. D., Sharman, J., Harrison, R., Park, J. W. & Halter, E. (2011), *The puppet masters: How the corrupt use legal structures to hide stolen assets and what to do about it*, World Bank Publications.

Der Spiegel (1993), 'Service für Steuersünder'. Last accessed on 15 February 2018 at: `http://www.spiegel.de/spiegel/print/d-13689843.html`.

Destatis (2014*a*), *Statistisches Jahrbuch 2014*, Statistisches Bundesamt, Wiesbaden.

Destatis (2014*b*), Verkehr. Verkehrsunfälle 2013, Fachserie 8 Reihe 7, Statistisches Bundesamt, Wiesbaden.

Destatis (2014*c*), Weiterbildung 2014, Periodikum, Statistisches Bundesamt, Wiesbaden.

Destatis (2015*a*), Finanzen und Steuern, Jährliche Einkommensteuerstatistik, Sonderthema: Werbungskosten, Fachserie 14 Reihe 7.1.1, Statistisches Bundesamt, Wiesbaden.

Destatis (2015*b*), *Statistisches Jahrbuch 2015*, Statistisches Bundesamt, Wiesbaden.

Destatis (2016), Einkommens- und Verbrauchsstichprobe EVS 2013, Qualitätsbericht, Statistisches Bundesamt, Wiesbaden.

Destatis (2017*a*), Datenhandbuch zum Mikrozensus Scientific Use File 2012, Manual, Statistisches Bundesamt.

Destatis (2017*b*), Einkommens- und Verbrauchsstichprobe: Aufgabe, Methode und Durchführung, Fachserie 15 Heft 7, Statistisches Bundesamt, Wiesbaden.

Destatis (2017*c*), Finanzen und Steuern, Lohn- und Einkommensteuer, 2013, Fachserie 14 Reihe 7.1.1, Statistisches Bundesamt, Wiesbaden.

Destatis (2017*d*), 'Steuererklärung: Durchschnittliche Rückerstattung lag bei 935 Euro'. Last accessed on September 23, 2018 at: `https://www.destatis.de/DE/ZahlenFakten/ImFokus/OeffentlicheFinanzenSteuern/SteuererklaerungErstattung.html`.

Destatis (2018*a*), Finanzen und Steuern - Versorgungsempfänger des öffentlichen Dienstes 2017, Fachserie 14 Reihe 6.1, Statistisches Bundesamt, Wiesbaden.

Destatis (2018*b*), Touristik in Zahlen 2017, Periodikum, Statistisches Bundesamt, Wiesbaden.

Deutsche Bundesbank (1994), *Aufkommen und ökonomische Auswirkungen des steuerlichen Zinsabschlags*, Monatsbericht Januar, pp. 45–58.

Deutscher Spendenrat & GfK (2014), 'Bilanz des Helfens 2014'. Last accessed on January 3, 2020 at `https://www.spendenrat.de/wp-content/uploads/spendenrat/downloads/bdh/Bilanz_des_Helfens_2014.pdf`.

Deutsches wissenschaftliches Institut der Steuerberater e.V. (2014), *Handbuch der Steuerveranlagungen 2013: Einkommensteuer, Körperschaftsteuer, Gewerbesteuer, Umsatzsteuer*, C.H.Beck.

Dittmann, W., Haderer, D. & Happe, R. (2014), *Einkommen Steuererklärung 2013-2014*, Haufe, Munich.

Dörrenberg, P. & Peichl, A. (2018), Tax Morale and the Role of Social Norms and Reciprocity. Evidence from a Randomized Survey Experiment, CESifo Working Paper Series 7149, CESifo Group Munich.

Dörrenberg, P. & Peichl, A. (2021), 'Was gute Steuerpolitik braucht', Süddeutsche Zeitung. Last accessed on October 24, 2021 at: `https://www.sueddeutsche.de/wirtschaft/steuern-bundestagswahl-folgen-1.5408922`.

Dowle, M. & Srinivasan, A. (2019), *data.table: Extension of 'data.frame'*. R package version 1.12.0.
URL: *https://CRAN.R-project.org/package=data.table*

DRV (2017), Rentenversicherung in Zeitreihen. Kapitel Rentenzugang, Ausgabe 2017, Deutsche Rentenversicherung Bund, Berlin.

Engström, P. & Hagen, J. (2017), 'Income underreporting among the self-employed: A permanent income approach', *European Economic Review* **92**(C), 92–109.

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

Enste, D. H. (2017), Arbeitsplatz Privathaushalt. Minijobs und Schwarzarbeit von Haushaltshilfen, Kurzexpertise für die Minijob-Zentrale, Institut der deutschen Wirtschaft Köln Akademie GmbH.

Eurostat (2019), 'Duration of working life - annual data'. Last accessed on January 3, 2020 at `https://ec.europa.eu/eurostat/web/lfs/data/database`.

Fauser, H. (2019), On income tax avoidance – the case of Germany revisited, Working paper, mimeo.

Feldman, N. E. & Slemrod, J. (2007), 'Estimating tax noncompliance with evidence from unaudited tax returns', *Economic Journal* **117**(518), 327–352.

Fiorentino, E., Koch, C. & Rudek, W. (2010), Microdatabase: External position report of German banks, Technical report, Deutsche Bundesbank.

Fiorio, C. V. & D'Amuri, F. (2005), 'Workers' Tax Evasion in Italy', *Giornale degli Economisti e Annali di Economia* **64**(2-3), 247–270.

Frondel, M., Andor, M., Ritter, N., Sommer, S., Vance, C., Matuschek, P. & Müller, U. (2015), Erhebung des Energieverbrauchs der privaten Haushalte für die Jahre 2011-2013, Report, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI) und forsa Gesellschaft für Sozialforschung und statistische Analysen mbH.

Gallego-Granados, P. & Harnisch, M. (2017), Germany (DE) 2013-2017: EUROMOD Version H1.0, Country report, EUROMOD.

Garz, M. & Pagels, V. (2018), 'Cautionary tales: Celebrities, the news media, and participation in tax amnesties', *Journal of Economic Behavior & Organization* **155**, 288–300.

GDV (2019), 'Kfz-Versicherung. Überblick'. Last accessed on January 3, 2020 at: `https://www.gdv.de/de/zahlen-und-fakten/versicherungsbereiche/ueberblick-4660`.

Gravelle, J. G. (2015), Tax havens: International tax avoidance and evasion, Report, Congressional Research Service.

Gumpert, A., Hines, J. R. & Schnitzer, M. (2016), 'Multinational Firms and Tax Havens', *The Review of Economics and Statistics* **98**(4), 713–727.

Guyton, J., Langetieg, P., Reck, D., Risch, M. & Zucman, G. (2021), Tax Evasion at the Top of the Income Distribution: Theory and Evidence, NBER Working Papers 28542, National Bureau of Economic Research.

Hanlon, M., Maydew, E. L. & Thornock, J. R. (2015), 'Taking the Long Way Home: U.S. Tax Evasion and Offshore Investments in U.S. Equity and Debt Markets', *Journal of Finance* **70**(1), 257–287.

Hebous, S. & Johannesen, N. (2015), At Your Service! The Role of Tax Havens in International Trade with Services, CESifo Working Paper Series 5414, CESifo Group Munich.

Hemmelgarn, T. & Nicodeme, G. (2009), Tax Co-ordination in Europe: Assessing the First Years of the EU-Savings Taxation Directive, Taxation Papers 18, Directorate General Taxation and Customs Union, European Commission.

Immonet.de (2019), 'Modernisierung: Maßnahmen und Kosten'. Last accessed on January 3, 2020 at: `https://www.immonet.de/service/modernisieren-kosten.html`.

IW Köln (2019), 'Gewerkschaftsmitglieder'. Last accessed on January 2, 2020 at: `https://www.deutschlandinzahlen.de/tab/deutschland/arbeitsmarkt/tarifpolitik/gewerkschaftsmitglieder`.

Johannesen, N. (2014), 'Tax evasion and Swiss bank deposits', *Journal of Public Economics* **111**(C), 46–62.

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.

Keen, M. & Slemrod, J. (2017), 'Optimal tax administration', *Journal of Public Economics* **152**(C), 133–142.

Keynes, J. M. (1920), *The economic consequences of the peace*, Harcourt, Brace and Howe.

Kim, B., Gibson, J. & Chung, C. (2017), 'Using Panel Data to Estimate Income Under-Reporting by the Self-Employed', *Manchester School* **85**(1), 41–64.

Klautke, T. & Weichenrieder, A. J. (2010), 'Interest Income Tax Evasion, the EU Savings Directive and Capital Market Effects', *Fiscal Studies* **31**(1), 151–170.

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.

Korinek, A., Mistiaen, J. & Ravallion, M. (2006), 'Survey nonresponse and the distribution of income', *The Journal of Economic Inequality* **4**(1), 33–55.

Kraftfahrtbundesamt (2013), 'Bestand an Pkw am 1. Januar 2013 nach ausgewählten Kraftstoffarten'. Last accessed on January 3, 2020 at: `https://www.kba.de/DE/Statistik/Fahrzeuge/Bestand/Umwelt/2013/2013_b_umwelt_dusl_absolut.html?nn=1126794`.

Kudrle, R. T. (2008), 'The OECD's Harmful Tax Competition Initiative and the Tax Havens: From Bombshell to Damp Squib', *Global Economy Journal* **8**(1), 1–26.

Kudrle, R. T. (2009), 'Did blacklisting hurt the tax havens?', *Journal of Money Laundering Control* **12**(1), 33–49.

Kukk, M. & Staehr, K. (2017), 'Identification of Households Prone to Income Underreporting', *Public Finance Review* **45**(5), 599–627.

Lang, O., Nöhrbaß, K.-H. & Stahl, K. (1997), 'On income tax avoidance: the case of Germany', *Journal of Public Economics* **66**(2), 327–347.

Langenmayr, D. & Reiter, F. (2017), Trading Offshore: Evidence on Banks' Tax Avoidance, CESifo Working Paper Series 6664, CESifo Group Munich.

Larudee, M. (2015), 'Did capital go away? The contribution of capital flight to the decline in β during and after World War I', mimeo.

Lejour, A., Rabaté, S. & van 't Riet, M. (2020), Offshore Tax Evasion and Wealth Inequality: Evidence from a Tax Amnesty in the Netherlands, CPB Discussion Paper 417, CPB Netherlands Bureau for Economic Policy Analysis.

Leventi, C., Matsaganis, M. & Flevotomou, M. (2013), Distributional implications of tax evasion and the crisis in Greece, EUROMOD Working Papers EM17/13, EUROMOD at the Institute for Social and Economic Research.

Liersch, A. (2014), Arbeitsunfälle und arbeitsbedingte Gesundheitsprobleme. Ergebnisse einer Zusatzerhebung im Rahmen des Mikrozensus 2013, Wirtschaft und Statistik, Statistisches Bundesamt.

Lyssiotou, P., Pashardes, P. & Stengos, T. (2004), 'Estimates of the black economy based on consumer demand approaches', *Economic Journal* **114**(497), 622–640.

Matsaganis, M., Benedek, D., Flevotomou, M., Lelkes, O., Mantovani, D. & Nienadowska, S. (2010), Distributional implications of income tax evasion in Greece, Hungary and Italy, MPRA Paper 21465, University Library of Munich, Germany.

Meinzer, M. (2015), *Steueroase Deutschland: Warum bei uns viele Reiche keine Steuern zahlen*, C.H. Beck.

Menkhoff, L. & Miethe, J. (2017), Dirty money in new disguise: Capital flows into and out of tax havens. Unpublished Working Paper.

Menkhoff, L. & Miethe, J. (2019), 'Tax evasion in new disguise? Examining tax havens' international bank deposits', *Journal of Public Economics* **176**(C), 53–78.

Moore, J. C., Stinson, L. L. & Welniak, E. J. (2000), 'Income measurement error in surveys: A review', *Journal of Official Statistics* **16**(4), 331–362.

Niedermayer, O. (2017), Parteimitglieder in Deutschland: Version 2017, Arbeitshefte aus dem Otto-Stammer-Zentrum Nr. 27, Freie Universität Berlin.

Nöhrbaß, K.-H. & Raab, M. (1990), 'Quellensteuer und Kapitalmarkt: eine theoretische und empirische Untersuchung', *FinanzArchiv/Public Finance Analysis* (H. 2), 179–193.

OECD (1998), *Harmful tax competition: an emerging global issue*, OECD Publishing.

Pfahl, S., Hobler, D. & Weeber, S. (2016), Angezeigte und bestätigte Berufskrankheiten 1992 - 2014, WSI-GenderDatenPortal GH-08, Wirtschafts- und Sozialwissenschaftliches Institut.

Piketty, T. & Zucman, G. (2014), 'Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010', *The Quarterly Journal of Economics* **129**(3), 1255–1310.

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

R Core Team (2018), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Radke, S. (2014), *Verkehr in Zahlen 2014/2015*, Bundesministerium für Verkehr und digitale Infrastruktur, Berlin.

Reiter, F. (2017), Avoiding Taxes: Banks' Use of Internal Debt, Annual Conference 2017 (Vienna): Alternative Structures for Money and Banking 168115, Verein für Socialpolitik / German Economic Association.

Richter, N. (2016), 'Cheating, Divorce and Panama Papers', Süddeutsche Zeitung. Last accessed on February 3, 2017 at: `http://panamapapers.sueddeutsche.de/articles/5707ea36a1bb8d3c3495b999/`.

Rubin, D. B. (1987), *Multiple imputation for nonresponse in surveys*, John Wiley & Sons.

Schäfers, M. (2018), 'Finanzämter verschenken hunderte Millionen', Frankfurter Allgemeine Zeitung. Last accessed on October 24, 2021 at: `https://www.faz.net/aktuell/wirtschaft/rechnungshof-finanzaemter-schenken-steuerhinterziehern-geld-15888653.html`.

Schmole, H. (2016), Photovoltaik-Preismonitor Deutschland, German PV ModulePriceMonitor 2016, Ergebnisse 1. Quartal, Technical report, EuPD Research i.A. des Bundesverbands Solarwirtschaft e.V., Bonn.

Schneider, F. & Boockmann, B. (2018), Die Größe der Schattenwirtschaft - Methodik und Berechnungen für das Jahr 2018, Annual report, Johannes Kepler University Linz.

Sharman, J. C. (2010), 'Shopping for Anonymous Shell Companies: An Audit Study of Anonymity and Crime in the International Financial System', *Journal of Economic Perspectives* **24**(4), 127–140.

Silva, J. M. C. S. & Tenreyro, S. (2006), 'The Log of Gravity', *The Review of Economics and Statistics* **88**(4), 641–658.

Slemrod, J. (2007), 'Cheating Ourselves: The Economics of Tax Evasion', *Journal of Economic Perspectives* **21**(1), 25–48.

Slemrod, J. (2015), Tax administration and tax systems, Working Papers 2015/13, Institut d'Economia de Barcelona (IEB).

Spengel, C. (2017), 'Kollektivversagen: Cum/Cum, Cum/Ex und Hopp! [Collective Failure: The Problem of Dividend Stripping]', *Wirtschaftsdienst* **97**(7), 454–455.

Steiner, V., Wrohlich, K., Haan, P. & Geyer, J. (2012), Documentation of the Tax-Benefit Microsimulation Model STSM: Version 2012, Data Documentation 63, DIW Berlin, German Institute for Economic Research.

Studis Online (2019), 'Zinsentwicklung KfW-Bildungskredite. Zinsen sinken deutlich'. Last accessed on January 2, 2020 at: `https://www.bafoeg-rechner.de/Hintergrund/art-1395-kfw-zinsen-studis-april-2012.php`.

Techem (2018), 'Wer hat, der bleibt', Pressemitteilung Nr. 5. Last accessed on January 3, 2020 at: `https://newsroom.techem.de/fileadmin/de/newsroom/Pressemitteilungen/2018/PM_Techem_Umzugsquote_2017_final.pdf`.

Torregrosa-Hetland, S. (2020), 'Inequality in tax evasion: the case of the Spanish income tax', *Applied Economic Analysis* **28**(83), 89–109.

Torregrosa, S. (2015), Bypassing progressive taxation: fraud and base erosion in the Spanish income tax (1970-2001), Working Papers 2015/31, Institut d'Economia de Barcelona (IEB).

Umzug AG (2014), 'Über neun Millionen Bundesbürger ziehen jährlich um. Beim Umzug sind die Deutschen "Do-it-yourself"-Meister', Pressemitteilung. Last accessed on January 3, 2020 at: `https://www.presseportal.de/pm/51498/2793804`.

Van Buuren, S. (2012), *Flexible imputation of missing data*, 1st edn, Chapman and Hall/CRC.

van Buuren, S. & Groothuis-Oudshoorn, K. (2011), 'mice: Multivariate Imputation by Chained Equations in R', *Journal of Statistical Software* **45**(3), 1–67.
URL: *https://www.jstatsoft.org/v45/i03/*

VDR (2014), VDR-Geschäftsreiseanalyse 2014, Bericht, Verband Deutsches Reisemanagement e.V., Frankfurt am Main.

Vorgrimler, D., Gräb, C. & Kriete-Dodds, S. (2006), Zur Konzeption eines Taxpayer-Panels für Deutschland, FDZ-Arbeitspapier Nr. 14, Statistisches Bundesamt.

WDR (2016), 'Walter-Borjans will mehr gegen Steuerflucht tun'. Last accessed on February 5, 2017 at `http://www1.wdr.de/nachrichten/landespolitik/walter-borjans-steuerflucht-100.html`.

Wirth, H. (2017), Aktuelle Fakten zur Photovoltaik in Deutschland, Report, Fraunhofer-Institut für Solare Energiesysteme ISE, Freiburg.

Wooldridge, J. M. (2010), *Econometric analysis of cross section and panel data*, 2nd edn, MIT Press.

Zucman, G. (2013), 'The Missing Wealth of Nations: Are Europe and the U.S. net Debtors or net Creditors?', *The Quarterly Journal of Economics* **128**(3), 1321–1364.

Zucman, G. (2014), 'Taxing across Borders: Tracking Personal Wealth and Corporate Profits', *Journal of Economic Perspectives* **28**(4), 121–148.

Zucman, G. (2015), *The Hidden Wealth of Nations*, University of Chicago Press.

List of Figures

List of Tables

Deutsche Zusammenfassung

Die Dissertation analysiert Steuerhinterziehung und Steuervermeidung mit einem Fokus auf Deutschland. Dabei stehen die empirische Schätzung des Umfangs dieser Phänomene sowie ihre Verteilungswirkung im Vordergrund. Weiterhin werden Aufkommensverluste für die öffentlichen Kassen geschätzt, und einige der in diesem Feld in den ersten 15 Jahren des 21. Jahrhunderts umgesetzten Politikmaßnahmen untersucht. Aufgrund der allgemein schwierigen Datenlage für die Analyse von Steuerhinterziehung, sowie der im Fall deutscher Steuerdaten vergleichsweise limitierten Verfügbarkeit, restriktiven Zugangsmöglichkeiten und rechtlich-administrativ bedingter Analyseeinschränkungen, wurde dabei ausschließlich auf indirekte statistisch-ökonometrische Methoden zurückgegriffen.

Übergreifender Sinn der Arbeit ist es, über die Abschätzung des Umfangs und der Verteilungswirkung von Steuervergehen zu mehreren Zielen beizutragen: Erstens sollen Verzerrungen anderer Statistiken, etwa zur Verteilung von Einkommen anhand von Steuerstatistiken, korrigiert werden. Zweitens ist eine empirische Abschätzung von Steuerbetrug und Steuervermeidung im Interesse der Öffentlichkeit, deren demokratischer Wille sich repräsentiv vermittelt in der Steuergesetzgebung artikuliert. Zu dieser Willensbildung sollten wissenschaftlich fundierte Informationen beitragen. Zuletzt ist empirische Evidenz eine wichtige Voraussetzung, um zielgenaue, geeignete und angemessene Maßnahmen zur Bekämpfung illegitimer Praktiken zu entwickeln.

Kapitel 1 evaluiert Regulierungsmaßnahmen, die zur Eindämmung von internationaler Steuerhinterziehung getroffen wurden. Dabei werden vor allem Steuerinformationsaustauschabkommen in den Blick genommen, die bei hinreichendem Verdacht einen bilateralen Informationsfluss zwischen den Finanzbehörden der Unterzeichnerstaaten begründen. Insbesondere nach der Finanzkrise von 2007/2008 wurde eine Vielzahl davon geschlossen. Als Reaktion auf den Abschluss solcher Abkommen finden wir einen Rückgang der grenzüberschreitenden Einlagen aus Steueroasen in Deutschland. Wir bestätigen so mit einer neuen Identifikationsstrategie und monatlichen Daten der Bundesbank etablierte Ergebnisse aus der Literatur, in der bis dato mit Quartalsdaten der Bank für Internationalen Zahlungsausgleich ein Rückgang von 10 bis 40% geschätzt wurde. Dabei setzen wir eine neu zusammengestellte narrative Datenbank des deutschen regulativen Umfelds ein. Weiterhin zeigen wir disaggregierte Reaktionen einzelner Steueroasen und testen eine Reihe von Steuerrechtsänderungen und Datenlecks. SteuerhinterzieherInnen reagieren dabei nicht auf Steuersatzänderungen. Leaks zeigen zwar die erwarteten Effekte mit konsistenten Vorzeichen, sind aber kaum signifikant. Das bekräftigt den Ansatz, den Informationsaustausch in den Mittelpunkt der Analyse

zu stellen. Zu dessen Wirksamkeit ist abschließend allerdings anzumerken, dass wir eine Verschiebung der Einlagen in nicht-kooperierende Jurisdiktionen nicht ausschließen können. Empirische Befunde aus anderen Studien sowie anekdotische Evidenz deuten auf solche Verschiebungen als weit verbreitete Praxis hin.

Kapitel 2 analysiert Steuervermeidung entlang der Einkommensverteilung, anhand von Mikrodaten der Einkommens- und Verbrauchsstichprobe für das Jahr 2013. Die inhaltliche Breite der Befragung wird genutzt, um mit Angaben zu Einkommen, Ausgaben und Steuerzahlungen ein möglichst detailliertes Modell der deutschen Einkommensteuer zu programmieren. Somit werden die Bestandteile der zu versteuernden Einkünfte und die steuerlichen Abzüge so präzise wie möglich in einem Mikrodatenmodell geschätzt. Aus der Differenz der tatsächlich geleisteten und der sich aus dem Modell ergebenden eigentlich zu entrichtenden Steuerzahlungen wird der Steuerverlust berechnet. Die Ergebnisse bestätigen, dass Steuervermeidung mit zunehmendem Einkommen zunimmt: Der geschätzte Steuerverlust ist im obersten Dezil der Steuerzahlenden am größten, und beträgt dort rund 2,8% der steuerbaren Einkünfte (vor Abzügen) bzw. 15,8% der gezahlten Steuern. Wird der Steuerverlust statt auf der individuellen auf der Ebene der Haushalte betrachtet, sinkt der vom obersten Dezil vermiedene Betrag erwartungsgemäß etwas, auf 1,8% der steuerbaren Einkünfte bzw. 9,8% der gezahlten Steuer. Die aggregierten Steuerverluste für den deutschen Fiskus betragen mindestens 10,7 Mrd. Euro, was 4,4% der festgesetzten Einkommensteuer entspricht. Wichtig zur Interpretation dieser im internationalen Vergleich eher kleinen Steuerlücke ist, dass meine Schätzung mit großer Wahrscheinlichkeit eine Untergrenze darstellt: So ist die Datengrundlage verzerrt, weil Personen mit monatlichen Nettoeinkommen von mehr als 18.000 Euro nicht erfasst werden. Zudem ist das Modell konservativ geschätzt, sodass steuerliche Abzüge tendenziell großzügig gewährt werden und im Aggregat um rund 0,8 Prozentpunkte höher liegen als in der Steuerstatistik.

Kapitel 3 erweitert die Analyse von Steuerhinterziehung für Deutschland um weitere empirische Ansätze und einen intertemporalen Vergleich. Dabei wird der Unterschied zwischen verschiedenen Einkunftsarten in den Blick genommen, sowie erneut Verteilungsimplikationen analysiert. Wir vergleichen zunächst mithilfe der Diskrepanzmethode gewichtete Stichproben von Befragungs- und Steuerdaten, wobei wir das Sozioökonomische Panel (SOEP) sowie das Taxpayer-Panel (TPP) verwenden. Wir finden Muster, die zu anderen Untersuchungen passen: Durchschnittliche Einkommen aus Selbstständigkeit sowie Vermietung und Verpachtung sind im SOEP höher als im TPP, zunehmend in oberen Quintilen. Da Steuerhinterziehung nur eine mögliche Interpretation dieser deskriptiven Ergebnisse ist, nutzen wir zudem den Ansatz von Pissarides

& Weber (1989), der in der Literatur für eine Vielzahl von Ländern und Datensätzen angewandt wurde. Wir schätzen Regressionen für Nahrungsmittel und Wohnungskosten mit dem SOEP, sowie Gleichungen für Spenden mit dem TPP. Die Ergebnisse zeigen, dass Selbstständigkeit im SOEP zwar mit höheren Wohnungskosten, aber nicht mit höheren Nahrungsausgaben einhergeht. Mit dem TPP schätzen wir dagegen Gleichungen zur Erklärung der Höhe steuerlich abzugsfähiger Spenden, die konsistente und signifikante Effekte für Einkünfte aus Selbstständigkeit, Gewerbe und Vermietung und Verpachtung finden. Insbesondere für diese Einkommen sind die Spenden höher als bei vergleichbaren Lohneinkommen, was auf Steuerbetrug hindeutet. Für das oberste Dezil ist dieser Effekt stärker. Gleichzeitig ist zu beobachten, dass der Anteil der hinterzogenen Einkommen im Zeitverlauf abnimmt. So verhält es sich auch mit geschätzten Aufkommensverlusten des Fiskus, die in der höchsten unserer Schätzungen von rund 70,2 Mrd. Euro im Jahr 2001 auf 32,4 Mrd. Euro im Jahr 2014 zurückgehen. Gemessen an der geschätzten "wahren" festzusetzenden Einkommensteuer entspricht dies einer Steuerlücke von 28,4% im Jahr 2001 und von 11,1% im Jahr 2014 (bzw. 39,6% und 12,5% an der festzusetzenden ESt). Angesichts einer Vielzahl von ökonometrischen Einschränkungen sind diese Schätzungen daher im Gegensatz zu jenen des zweiten Kapitels eher als obere Grenze anzusehen.

English Summary (Abstracts)

Chapter 1: Germany's efforts to curb international tax evasion

We evaluate the impact of regulatory attempts by German authorities to combat international tax evasion and report a 32-34% reduction of tax haven deposits in German banks as a reaction to bilateral information exchange. We test for reactions in monthly cross-border liabilities of German banks against non-residents employing a new narrative database of the German regulatory environment. Our findings are comparable in magnitude to a number of reference studies which confirms the choice of Germany as a valid case study for international tax evasion. We show disaggregated reactions for a list of tax havens and find large reactions to information exchange for Guernsey, the Bahamas, and Jersey. We also test a number of tax changes and data leaks. Tax evaders do not react to changes in tax rates. Leaks show consistent signs but are hardly significant which confirms information exchange as the main focus of analysis.

Chapter 2: On income tax avoidance - a new micro data model for the German case

I study tax avoidance along the income distribution based on micro data from the Income and Consumption Survey (EVS) for the year 2013. The richness of the survey concerning income, taxes and expenditures is exploited by modelling the German tax code in terms of the items available in the EVS. I.e, components of taxable income and deductions are estimated as precisely as possible in a micro data model. Results confirm findings in the literature claiming that tax avoidance increases with rising income. The estimated amount of avoided tax is largest for the richest decile of individual tax payers, at around 2.8% of taxable income before deductions or 15.8% of taxes paid. Expectedly at the household level, the amount avoided by the highest income decile is reduced somewhat, to 1.8% of taxable income before deductions or 9.8% of taxes paid. Aggregate losses at the national level amount to at least EUR 10.7 bn, or 4.4% of assessed income tax revenues.

Chapter 3: Income tax noncompliance in Germany, 2001-2014

This chapter estimates income tax underreporting for the case of Germany, by income category and along the income distribution. Comparing weighted samples of survey and tax data, we find patterns that are in line with the literature: Average income from self-employment and from rent and lease in the survey is higher than in the tax data, increasing in upper quintiles. However, income underreporting to the tax authorities may be one of several possible explanations for these descriptive findings. We therefore expand our analysis with the Pissarides & Weber (1989) approach that has been applied

to a range of countries and data sources before. We use the German Socioeconomic Panel and the Taxpayer Panel, estimating food, housing cost and donation regressions. Results indicate that self-employment is associated with higher housing cost but not with higher food expenditure in the SOEP. In the TPP we find more robust indication of underreporting as self-employment and business incomes are significantly associated with higher donations and even more so for the top-income decile. We use our results to derive tentative estimates of aggregate tax revenue losses due to underreporting of self-employment and other non-wage incomes.

Erklärung an Eides statt

Wie bereits zu Beginn der Dissertationsschrift vermerkt, sind zwei der Kapitel (Papiere) zu jeweils gleichen Anteilen mit KoautorInnen verfasst wurden.

Gemäß §4 Abs. 2 der Promotionsordnung zum Dr. rer. pol. des Fachbereichs Wirtschaftswissenschaften der Freien Universität Berlin vom 13. Februar 2013 erkläre ich hiermit, dass ich mich noch keinem Promotionsverfahren unterzogen oder um Zulassung zu einem solchen beworben habe, und die Dissertation in der gleichen oder einer anderen Fassung bzw. Überarbeitung einer anderen Fakultät, einem Prüfungsausschuss oder einem Fachvertreter an einer anderen Hochschule nicht bereits zur Überprüfung vorgelegen hat.

(Ort, Datum, Unterschrift)

Gemäß §10 Abs. 3 der oben genannten Promotionsordnung erkläre ich zudem hiermit, dass ich für die Dissertation folgende Hilfsmittel und Hilfen verwendet habe:

- Statistiken, Regressionen, Simulationen: R, Stata
- Schriftsatz und Formatierungen: LaTeX, Texmaker, Overleaf
- Darstellungen: R, Stata, Excel, Inkscape

Auf dieser Grundlage habe ich die Arbeit selbstständig verfasst.

(Ort, Datum, Unterschrift)