Empirical Essays on Tax Compliance

Inauguraldissertation

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2014

vorgelegt

von

Diplom-Volkswirt Philipp Dörrenberg

aus

Ratingen

Referent: Prof. Dr. Clemens Fuest
Korreferent: Prof. Dr. Dirk Sliwka

Tag der Promotion: 04. November 2014

Acknowledgements

Doing a Ph.D. is not easy and can be a time full of uncertainty, frustration and pressure. I was lucky to have people around me who made this thesis happen and supported me in many different ways. **Andreas Peichl**, so many thanks for so many things. You are a great mentor, promoter, supervisor, motivator, co-author, boss, teacher and friend. I don't believe my time as a doctoral student would have been as productive and enjoyable without your steady support and guidance. **Denvil Duncan**, thank you very much for sharing the risk of getting into lab experimental public finance. I very much enjoy developing ideas and writing papers with you. My research visit to Bloomington was great, both from an academic and a private perspective. I am also grateful for guidance with many career issues and improving my English. **Clemens Fuest**, thanks for supervising my thesis (despite quite some distance between Cologne and Oxford), steady support and bringing me to Mannheim. **Christoph Feldhaus**, thanks for sharing all the pain, frustration and fun one has as a (doctoral) student. We have to write a paper together at some point; maybe about the (un-?)importance of economic research. **Andreas Lichter**, good times at Bruesseler. Thanks for listening to private and academic issues during all these years. **Anna Rauch**, you are a great friend and co-author. Good luck with finishing your Ph.D. to the three of you! I really enjoyed my time as part of the "Peichl crew" at IZA. Thanks to **Mathias Dolls**, **Dirk Neumann**, **Judith Niehues**, **Nico Pestel**, **Max Loeffler**, **Sebastian Siegloch** (soccer is certainly good for everybody.), and **Eric Sommer**. **Christian Weyand** and **Christopher Zeppenfeld**, fun at CGS. I am grateful for unhelpful comments and suggestions in the Mateclub. **Jan Schmitz**, there will soon be an important evasion contribution using data from a field experiment in Slovenia. **CGS cohort of 2010**, **Jan Mueller**, you are superb, Spain was a blast. I further thank **CGS**, **IZA** and **ZEW** for financial support, office space, and a nice working environment. Thanks to **Dagmar Weiler** for keeping CGS together and welcoming me even after my time on a scholarship. **Dirk Sliwka**, thanks for serving as my second advisor and academic head of CGS. I am also grateful to Cologne professors **Oleg Badunenko**, **Felix Bierbrauer**, **Helge Braun** and **Bettina Rockenback** for helpful academic guidance and motivation. The final months of this dissertation were written at ZEW and I thank the Redistribution and Taxation departments. Thanks are particularly due to **Kathi Finke** (great MaTax co-coordinator) and **Friedrich Heinemann** (thanks for hiring me and familiarizing me with policy consultancy). **My parents**, you literally made this happen. Nothing I have and do would have been possible without you. I am very happy and lucky to have you. Thanks! **Moritz**, the best and coolest brother I can imagine. Thanks for being there and sticking close to me. You are great. **Johanna**, my love, Scherz, partner, roomie, Helene and friend. You supported me during these (sometimes difficult) Ph.D. years. Thank you very much! I am so happy we found each other.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Tax non-compliance[1] is a significant problem in most countries across the world. For example, the Internal Revenue Service estimated the tax gap – i.e., the difference between taxes owed and taxes paid in a fiscal year – in the United States in 2006 to be around 450 billion US Dollars which is roughly 17% of all true tax liabilities (IRS 2012). For the UK, the government agency HM Revenue & Customs reports that the 2010/2011 tax gap is 32 billion Pounds (about 54 billion US Dollars) which corresponds to 6.7% of all tax liabilities (HMRC 2012). The Swedish National Tax Agency published tax gap estimates in the range of 10% of tax liabilities in 2008 (SNTA 2008) and Kleven et al. (2011) report that the overall share of taxpayers who underreport a positive amount of income was 10.7% in Denmark in 2007/08. Tax non-compliance is suggested to be an even more severe problem in developing and transitional countries. Schneider and Enste (2000), for example, estimate the average size of the shadow economy to be 23% of GDP in transition countries and 39% in developing countries. Anecdotal evidence additionally suggests that tax evasion has always existed and is as old as taxes themselves (Slemrod 2007).

While it is in the nature of tax non-compliance that its magnitude is difficult to measure, all attempts of quantification point in the direction that governments in almost all countries are forced to renounce significant amounts of tax revenue. This restricts public spending and possibly also leads to altered market outcomes and economic distortions (Andreoni et al. 1998; Pomeranz 2013). Given the prevalence of evasion, it is no surprise that a large academic literature has explored various aspects of evasion (see below for a broad literature review). This literature can be broadly separated into three strands (Andreoni et al. 1998; Slemrod and Weber 2012; Alm 2012): The first strand of literature aims to quantify the extent of non-compliance on an aggregate level. The

[1] Following Slemrod and Weber (2012), the term *tax non-compliance* refers to tax evasion *and* the shadow economy. While the shadow economy and tax evasion are two distinct concepts, they are clearly related and the size of the shadow economy is sometimes even used as a proxy for the amount of tax evasion (Alm 2012). Tax non-compliance entails illegal activities and does not cover legal tax avoidance.

second strand looks at the determinants of (non-)compliance while the third explores the consequences of non-compliance and its effects on market outcomes.

1.1 Research questions and agenda

Despite the large and growing literature on tax evasion, there remain important unanswered questions in each strand. This dissertation consists of four self-contained papers that aim to fill some of this gap, particularly in the second and third strands of the literature.[2] The research questions and methods in each chapter are *briefly* described in the following.[3]

This **Chapter 1** provides an overview of the research papers included in the dissertation and broadly places the papers in the large literature on tax compliance. Given the inherent difficulty of measuring non-compliance, Chapter 1 particularly aims to give an overview of empirical methods and discusses the empirical approaches that are used in the subsequent chapters.

Chapter 2 is placed in the large literature on the determinants of tax compliance and particularly contributes to a better understanding of the non-pecuniary motives for tax compliance, which are the least well understood of all determinants (Dwenger et al. 2014). The Chapter poses the question of whether the usage of tax revenues affects compliance behavior. The empirical tool is a laboratory compliance experiment in which subjects are randomly assigned to four different treatment groups that differ only in how the generated tax revenue is spent.

The remaining chapters study how tax non-compliance affects market outcomes and tax policy. **Chapter 3** examines the question of whether the canonical textbook laws of tax incidence hold in the presence of evasion opportunities. To explore the effect of evasion opportunities on tax incidence, a laboratory experiment is conducted in which subjects trade fictitious goods in competitive double auction markets. The impact of evasion on tax incidence is identified by comparing the equilibrium prices in groups with and without evasion opportunities.

Heterogeneity in evasion behavior may also affect other behavioral responses to taxation. In this context, **Chapter 4** examines the extent to which labor supply elasticities with respect to tax rates depend on evasion opportunities. A laboratory experiment is employed in which the labor supply response to taxes is compared between groups with and without access to evasion.

Whereas Chapters 3 and 4 show that tax evasion may impact responses to tax

[2]The question of how tax evasion can be measured is not at the center of this dissertation, and is discussed only briefly below. See Andreoni et al. (1998), Slemrod and Weber (2012) and Alm (2012) for more detailed overviews of different methods.

[3]See Section 1.3 for a more detailed description as well as the findings.

reforms, **Chapter 5** asks whether tax policy is affected by heterogeneity in access to evasion and intentions to be compliant. More specifically, this chapter examines the relationship between 'tax morale' – defined as the intrinsic motivation to comply – and tax policy. The chapter uses a unique observational data set and employs an instrumental variable strategy to address endogeneity concerns.

1.2 Literature review and discussion of methods

In this section, I present an overview of the literature on tax non-compliance and show how the research questions posed in my dissertation fit into the literature. Since it is in the very nature of tax non-compliance that it is illusive and difficult to observe empirically, I particularly survey the literature with respect to empirical data and methods, and critically discuss the empirical tools that are used in my dissertation. I also provide a brief overview of the main findings that have been presented in this extensive evasion literature.[4]

1.2.1 Theoretical literature

The large literature mainly emerges from theoretical work by Allingham and Sandmo (1972) which is in the spirit of Becker's (1968) economics-of-crime approach. Allingham and Sandmo's (A-S) seminal contribution models compliance decisions as risky gambles with fully rational taxpayers who maximize after-tax incomes in the presence of exogenous and random audit probabilities and penalties. The model predicts that higher penalties and audit probabilities improve compliance, whereas the effect of higher tax rates is ambiguous. Several theoretical studies extend this stylized model to clarify the effects of various parameters on evasion and improve its realism. For example, Yitzhaki (1974) shows that the structure of the penalty affects how the tax rate affects compliance. Pencavel (1979) and Sandmo (1981) incorporate endogenous labor supply in the A-S framework and show that the effect of all enforcement variables becomes ambiguous. An important extension of the A-S approach is to account for interactions between tax authorities and taxpayers, which eventually result in endogenous audit probabilities (Andreoni et al. 1998).

Kleven et al. (2011) also allow for endogenous audit probabilities. Their model accounts for the fact that the risk of being detected critically depends on the type of income source: whereas workers with self-reported income (usually the self-employed) have low propensities to be audited, workers in industries with third-party reporting face

[4]This section shows how the research questions posed in this dissertation generally fit into the development of the evasion literature. The single chapters contain literature reviews that are more specific to the respective chapter.

an audit probability near one and are therefore predicted to evade very little. Chapters 3 to 5 are partly motivated by this type of theory and show empirically how heterogeneity in evasion access and compliance behavior affects market outcomes. While the literature usually acknowledges that non-pecuniary motives – which are relevant in Chapters 2 and 5 of this dissertation – also matter for compliance decisions, theoretical literature incorporating these motives is rare. An exception is Traxler (2010) who adds social norms to the A-S framework. Chapter 5 of this dissertation contributes to the theoretical literature in this regard as it is among the first papers to incorporate intrinsic motivations to pay taxes – tax morale – into a 'standard' optimal taxation framework.[5]

1.2.2 Empirical literature

The theoretical literature discussed above is the basis for a large empirical literature examining where and why theory fails to explain real-world compliance behavior and studying various possible determinants of non-compliance. Given the inherent difficulty of measuring non-compliance, a variety of data sources and methods are used for empirical analyses. Because this dissertation contributes to the empirical literature on tax compliance, I briefly describe and critically discuss the most commons sources of evidence and methods in the following. The methods used in the literature are usually based on either *i)* observational or *ii)* experimental data, and both are employed in my dissertation. The discussion of methods and data motivates the use of and gives special emphasis to survey data and laboratory experiments as these are the main data/methods used in the dissertation. After the review of empirical data and methods, I briefly summarize the findings of the empirical literature with respect to the two strands of literature that I mainly relate to in my dissertation: determinants of tax compliance and the effects of non-compliance on market outcomes.[6]

Methods based on observational data

Macro-based indicators A widely used approach to explore the aggregate magnitude of non-compliance and its correlates uses macro-based indicators. For example, several studies predict how indicators of true economic activity such as electricity demand (e.g., Kaufmann and Kaliberda 1996, Johnson et al. 1997 and Duncan and Sabirianova-Peter 2014) or money circulation (e.g., Tanzi 1983 and Feige 1990) translate to true economic outcome or income. These predictions are then compared to official output/income statistics and the gap is interpreted as an estimate of the shadow economy. The underlying as-

[5] See Andreoni et al. (1998), Slemrod and Yitzhaki (2002) and Sandmo (2005) for more thorough overviews of the theoretical compliance literature.

[6] The following section is partly based on Schneider and Enste (2000), Slemrod and Weber (2012) and Alm (2012).

sumptions on the translation of electricity, respectively currency, to true economic output may, however, be critical. Another concern is regarding the measurement of true income. Official GDP statistics may not be suitable because many countries adjust their measure of GDP for underground activities. The so-called Model or MIMIC (multiple-indicators, multiple causes) approach accounts for multiple causes and effects of underground activities to approximate the magnitude of shadow activities. Pioneered by Frey and Weck-Hanneman (1984) in the context of the shadow economy, this method has its roots in the statistical theory of unobserved variables. The basic idea is to first define observed causes (e.g., the burden of taxation or attitudes towards the government) and effects (e.g., worker participation in the official economy or additional monetary transactions) of the unobserved shadow economy and then connect these causes and effects through the unobserved shadow economy variable using a set of structural equations.[7] The MIMIC approach, however, critically depends on the selection of the observed causes and effects and the estimations are not transparent. Estimates for many shadow economies around the world based on MIMIC methods are presented and discussed in Schneider and Enste (2000) and a critique is provided by Breusch (2005a, 2005b).

Audited tax returns The most precise source of information on individual tax compliance is based on tax returns audited by the tax authorities. In order to derive a representative picture of evasion, this approach requires (exogenous) random audits rather than endogenous audits that are targeted at suspicious taxpayers. While most countries mostly perform endogenous audits, the US Internal Revenue Service (IRS) used to conduct audits of a stratified random sample of about 50,000 taxpayers every three years between 1965 and 1988. Clotfelter (1983) and Feinstein (1991) use these data to estimate the effects of taxes and audit probabilities on evasion. However, deriving causal estimates using these data is very difficult because the data are only collected every three years, variation in tax rates and audits is unlikely to be exogenous and the data lack information on personal characteristics.

Surveys Another possibility to learn about individual evasion behavior is to rely on anonymous, self-reported household surveys in which respondents are either directly asked for their compliance behavior or, more indirectly, for their attitudes towards evasion. As opposed to audited tax returns, surveys have the main advantage that they usually include many socioeconomic and demographic variables, that they cover representative samples over time, and that they allow cross-country comparisons. Surveying evasion

[7] Note that the estimates for the shadow economy that are derived using the MIMIC approach should not be used as dependent variables in order to identify determinants of the shadow economy because all possible determinants of the shadow economy are supposed to enter the MIMIC method itself and are therefore mechanically correlated with the MIMIC based estimate of the shadow economy.

behavior directly, however, is suspected to yield very imprecise information. Elffers et al. (1987), for example, link survey responses to tax audit data and show that the correlation between actual compliance and self-reported compliance is very low. One reason for this finding might be that evaders wish to excuse their behavior or that respondents fear (irrationally or not) that their answers may be forwarded to tax authorities.

Due to these concerns regarding direct survey evidence on non-compliance, a wide literature, including Chapter 5 of this dissertation, use questionnaires which do not survey compliance behavior directly, but instead intend to measure attitudes towards non-compliance. This procedure is expected to be less blunt and therefore more reliable. The most common survey in this context is the World Values Survey (Minkov 2012) which is a world-wide survey that includes the following question on attitudes towards tax compliance: *Please tell me for the following statement whether you think it can always be justified, never be justified, or something in between: 'Cheating on taxes if you have the chance'.* The question is measured on a ten-scale index with one (1) meaning 'never justifiable' and ten (10) meaning 'always justifiable'. A wide literature uses the answer to this question as a proxy for 'tax morale' and to estimate the determinants of compliance or tax morale (see e.g., Slemrod 2003, Slemrod 2003, Richardson 2006, Torgler 2006, Halla 2012, and Doerrenberg and Peichl 2013).

Research by Torgler and Schneider (2009) and Halla (2012) validates the World Values Survey measure of tax morale by providing evidence that there is indeed a causal link between tax morale as surveyed in the World Values Survey and actual compliance levels. While survey measured tax morale is unlikely to be appropriate to quantify levels of evasion, these studies suggest that it can be a valid indicator for heterogeneity in compliance behavior across countries or individuals. Chapter 5 of this dissertation exploits the fact that the World Values Survey measure of tax morale is associated with actual inclinations to evade and uses it as an explanatory variable to explain variance in tax rates across groups with different levels of tax morale in a cross-country approach. The survey data structure allows to control for time and country specific effects, demographic and socio-economic variables and, in the spirit of Lubian and Zarri (2011), the use of an instrumental variable that is correlated with tax morale but unrelated to tax policy decisions. In general, however, estimating causal effects based on survey data is challenging because they lack random variation and often do not allow the inclusion of person specific fixed effects.

Traces of true income A different method to derive estimates of individual non-compliance uses traces of true income. The basic idea is to compare groups with and without evasion opportunities, usually wage earners and self-employed workers, with respect to variables that are assumed to represent a constant share of true income. For example, Pissarides and Weber (1989) compare the ratio of reported food expenditures to reported income using self-reported survey data. An estimate of tax non-compliance among the self-employed is backed out by assuming that employees report their true income correctly, the self-employed underreport income, food expenditures are correctly reported by both type of workers, and the self-employed and employees spend the same share of true income for food.[8] Feldman and Slemrod (2007) use a similar approach but have more accurate (tax return) data and use the ratio of reported charitable income to true income. The underlying assumptions are equivalent; it is especially assumed that the type of income source – wage earner vs. self-employment – is unrelated with preferences for charitable giving. While this method clearly adds to the literature on estimating the extend of evasion, its underlying assumption is very strong. Workers self-select into industries and employment types and it is therefore likely that wage earners and the self-employed are different along many dimensions and not only with respect to their evasion opportunities. As a result, it is difficult to assume that employees and the self-employed have identical preferences for food expenditures or charitable contributions.[9]

Methods based on experimental data

Field experiments The previously discussed sources of evidence face two problems when studying the determinants of non-compliance: non-compliance is not directly observable and identifying causal effects is difficult due to the lack of control groups, panel data and randomized variation in the potential determinants. Randomized field experiments overcome both these problems. Randomization of the explanatory variable of interest allows to identify the causal effect. The problem of not observing the dependent variable is also circumvented by randomization: a simple difference-in-differences strategy, in which the change in reported income of a control group before and after the treatment intervention is compared to the change in reported income of a treatment group, can be employed to estimate the change in evasion due to to the treatment intervention. Slemrod et al. (2001) and Blumenthal et al. (2001) are the first to run a randomized field experiment in the context of tax compliance. Taxpayers in randomly assigned treatment groups received letters containing normative appeals or information that tax returns are audited.[10]

Field experiments have the advantage of combining external with internal validity; taxpayers are treated in their natural environment while randomization ensures credible

[8] In a similar approach, Gorodnichenko et al. (2009) also use consumption expenditure data to estimate the evasion response to tax reforms.

[9] Research on entrepreneurship for example shows that personality traits are different between wage earners and entrepreneurs (Cramer et al. 2002).

[10] Other examples of field experiments on tax compliance behavior include Iyer et al. (2010), Kleven et al. (2011), Carrillo et al. (2013), Pomeranz (2013), Dwenger et al. (2014) and Hallsworth et al. (2014).

identification. This makes them a very powerful tool to study the determinants of tax compliance. On the downside, however, they are time-consuming, expensive and require the cooperation with tax authorities. Their largest disadvantage though is that they only allow the study of certain research questions. One reason for this is that many countries have strict obligations that require equal treatment of all tax payers. A second reason is that some potential determinants such as trust in government or voting rights over tax revenue spending can not be induced in a field experiment. Field experiments further do not allow the study of questions on the effects and implications of tax non-compliance because it is not feasible to randomly assign taxpayers to groups with and without access to evasion opportunities.

Laboratory experiments Another widely used tool to explore questions related to tax compliance is the laboratory experiment. The laboratory offers opportunities to exploit randomized variation and to observe levels of compliance directly. As opposed to field experiments, the laboratory further allows the exploration of a wider range of questions and is cheaper, less time-consuming and less elaborate. The design of lab experimental compliance decisions differs with respect to many features across studies in the literature, but the basic idea is similar: Subjects in an artificial laboratory are either endowed with an amount of gross income or gross income is endogenously earned in a real-effort task. A tax is due on gross income and subjects must decide how much gross income to report for tax purposes. There is a probability that the reporting decision is audited. In the case of an audit a penalty applies that is a function of the underreported amount whereas in the case of no audit subjects keep the amount not reported and only reported income is taxed. After completion of the experiment, subjects are paid based on their decisions and the realization of the audit outcome.

External validity of laboratory experiments While the internal validity of lab experiments is high due to randomization and the controlled circumstances, factors such as the artificial environment of the lab or the (usual) use of (undergraduate) college students are often suspected to challenge the external validity of the results. Since three chapters in this dissertation employ lab experiments, I discuss the most common point of criticism regarding compliance lab experiments in some detail here and intend to provide an assessment of the type of research questions that may or may not be addressed using lab experimental methods.[11]

First, economic compliance lab experiments are money-incentivized and include au-

[11] I do not discuss the external validity of lab experiments in general, but instead focus on the most important concerns in the case of compliance experiments. For more general discussions see Levitt and List (2007), Falk and Heckman (2009) and Kessler and Vesterlund (2014). The lab experimental chapters of this dissertation discuss the external validity of the specific experimental designs.

dit risks and penalties. They therefore mimic the crucial features of real-world compliance decisions. When making compliance decisions, taxpayers in reality as well as in the lab face a trade off between more money through cheating and less money in the case of detection. Since the majority of real-world tax evasion offenses is likely to be in a range that implies money penalties rather than jail in the case of detection, the decisions in the lab are comparable to real-world compliance decisions. In this regard, the use of college students may even increase the results' external validity because their choices are more likely to be sensitive to money stakes in the range of USD 20 per hour of lab work (Falk and Heckman 2009 put forward a similar argument in the context of lab experiments in general). In addition, most compliance experiments include a money earning stage in which subjects have to work for their gross-income and therefore develop a sense of ownership.

Second, compliance experiments do not intend to approximate the magnitude of real-world tax non-compliance, the levels of lab experimental tax evasion are not directly interpretable and the magnitude of treatment effects should be treated with caution. Instead, lab experiments are appropriate to study the underlying mechanisms behind compliance behavior and can help to understand the direction of real-world effects. For a very simple example, consider an experiment in which the effect of audits is tested: while such an experiment may show that higher audits are likely to improve compliance, one has to be careful with deriving conclusions about the size of the effect (in the sense of "doubling the audit probability increases compliance by X %"). Interpreting the direction rather than size of laboratory treatment effects is not specific to compliance experiments. The recent overview article by Vesterlund (2014) argues in a similar vain for lab experiments in general.

Third, the use of student subjects is often subject to criticism. However, there is not much evidence that students behave differently in the lab than other groups (Alm 2012; Falk et al. 2013). In addition, student subjects are usually very homogeneous with respect to factors such as age, education, political views, etc. Any differences between groups in lab experiments are therefore entirely driven by the treatment effect and not by unobserved variables or unlucky randomization. As noted before, students are also more likely to be responsive to experimental money stakes.

Finally, another concern is based upon the well-established result in the literature that relatively high compliance levels are due to the fact that taxpayers value public goods and the use of tax money use. Well-executed experiments therefore usually spend lab tax revenues for well-perceived purposes and inform subjects about the use of their tax money (see Chapter 2 of this dissertation for more discussion).

Lab experimental literature Paying regard to these points and pioneered by Becker et al. (1987), a relatively large literature uses lab experiments to study compliance related research questions (Torgler 2002, Alm and Jacobson 2007 and Alm 2012 provide overviews). The seminal lab experimental compliance papers by Alm, McClelland, and Schulze (1992) and Alm, Jackson, and McKee (1992a) study questions evolving around audit probabilities, public good provision, efficiency, and uncertainty. More recent contributions explore tax compliance in the context of social interactions (Fortin et al. 2007), enforcement spillovers (Alm et al. 2009), emotions (Coricelli et al. 2010), the interaction between evasion and charitable giving (Hsu 2008), information of compliance behavior of others (Lefebvre et al. 2014), and credence markets (Balafoutas et al. 2014). Chapters 2, 3, and 4 of this dissertation relate to this strand of literature by using lab experiments in the context of tax compliance. These chapters address research questions which are – arguably – relevant, but difficult, if not impossible, to address with observational data or field experiments. This makes the lab the natural tool to rely on.

Determinants of tax compliance

Governments all across the world strive to reduce levels of non-compliance and demand advice regarding policies that are effective in combating evasion. Identifying the determinants of compliance is therefore important both from an academic and policy perspective. In the following, I briefly discuss previous findings in the literature which uses the above sources of evidence. The overview particularly serves to show how Chapter 2 of this dissertation, which contributes to the understanding of the determinants, fits into the literature. In general, the modern literature classifies motives for tax compliance into three categories: *i)* deterrence parameters and limited cheating opportunities, *ii)* false perceptions of enforcement parameters and lack of information about cheating possibilities, and *iii)* non-pecuniary and intrinsic motives (Andreoni et al. 1998; Alm 2012; Dwenger et al. 2014).

First, **deterrence parameters and limited cheating opportunities** deter taxpayers from non-compliance. The relevance of deterrence measures is shown by lab experimental (Alm 2012) and field experimental (Slemrod et al. 2001; Kleven et al. 2011) evidence finding that increased audit probabilities improve tax compliance.[12] The effect of higher audits is found to be particularly strong among the self-employed or workers who work in cash-based industries; that is, among those workers with good evasion opportunities. This finding motivated Slemrod (2007) and Kleven et al. (2011) to study whether missing opportunities to evade, for example due to third-party reported income

[12] Note that it is extremely difficult to study the effect of audit rates on compliance using non-experimentally observational data because audits are endogenous and targeted at risk groups of taxpayers.

of wage earnings, may explain why observed levels of evasion are higher than predicted by Allingham and Sandmo (1972) type of models. Their research suggests that most taxpayers are partly compliant simply because they do not have an opportunity to evade. Based on these findings, Dwenger et al. (2014, page 1) conclude that "the notion that deterrence is weak is to some extent an illusion". Despite showing that evasion is substantial among the self-employed, Kleven et al. (2011) additionally find that higher tax rates have a positive effect on the evasion of self-employed income. Along with corresponding lab experimental results showing that evasion rates are considerable and respond positively to higher tax rates (Alm 2012), it can be concluded that taxpayers who have the opportunity to cheat do so at least partly for pecuniary reasons.

Second, as opposed to theoretical assumptions, taxpayers may have **false perceptions of enforcement parameters** and lack **information about cheating possibilities**. The former argument builds upon the work of Tversky and Kahneman (1974) who show that most people heavily overestimate the probability of a risky event to occur. This result may indicate that potential evaders perceive the chance of detection to be higher than it actually is and therefore refrain from evasion. Erard and Feinstein (1994) and Scholz and Pinney (1995) provide empirical evidence in the context of tax compliance that confirms this notion. Another reason for unexpectedly high compliance rates may be that taxpayers are not informed about their evasion opportunities. Paetzold and Winner (2014), for example, present quasi-experimental evidence from Austria that tax evasion of workers who start a new job goes up if the knowledge of evasion opportunities is high among the existing co-workers in the new firm. This result suggests that information about evasion opportunities are likely to be relevant for compliance.[13]

Third, taxpayers may be compliant with tax rules for **non-pecuniary and intrinsic motives**. Important issues in this context are conditional cooperation and social norms. Taxpayers seem to be more compliant if they perceive evasion levels among their peers to be low and vice versa (Frey and Torgler 2007; Traxler 2010; Alm 2012; Hallsworth et al. 2014; Lefebvre et al. 2014). Intrinsic motivations or tax morale are further reasons for non-pecuniary motives to comply. Torgler and Schneider (2009) and Halla (2012) show that survey measured intrinsic motivations are positively related to actual levels of tax compliance. Dwenger et al. (2014) find that 20% of individuals pay their tax liability even in a setting with zero deterrence, suggesting that intrinsic motives are substantial. Other non-pecuniary factors that seem to matter for compliance decisions, and that are likely to interact with tax morale, are among others: trust in government (Slemrod 2003; Feld and Frey 2007), patriotism (Konrad and Qari 2012), religious beliefs (Torgler 2006), service orientation of tax authorities (Alm et al. 2010), social interactions (Fortin et al.

[13] Other studies such as Chetty and Saez (2013) and Chetty et al. (2013) confirm that some taxpayers lack knowledge about the tax system.

2007; Bo et al. 2014), efficiency of redistribution (Alm, Jackson, and McKee 1992a), and voting possibilities over tax revenue spending (Alm and Jackson 1993; Lamberton et al. 2014). Chapter 2 of my dissertation examines the role of tax revenue usage for compliance and therefore contributes to this strand of literature that explores non-pecuniary motives for tax compliance.

Market outcomes in the presence of non-compliance

Identifying the drivers of compliance is obviously important. However, as noted by Andreoni et al. (1998, page 818), from a public finance perspective tax compliance "spans the notions of equity, efficiency and incidence" and is of further academic interest because it is presumed to affect economic outcomes. For example, the deadweight loss of taxation is likely to be higher in the presence of non-compliance since more distortionary taxes are required to meet given revenue goals. Efficiency losses do also occur when taxpayers waste resources to shelter taxable income. This suggests that in addition to studying its determinants, it is important to study the impact of non-compliance on market outcomes. One issue that is particularly relevant in this context is heterogeneity in compliance behavior. Slemrod (2007) and Kleven et al. (2011) emphasize that some taxpayers, in particular those with self-reported income, have considerably better opportunities to underreport taxable income than other taxpayers. The literature on intrinsic motivations to comply with the tax law additionally shows that such intrinsic motives are also heterogeneously distributed across the population of taxpayers (Torgler 2007).

It is widely suspected that this heterogeneity implies that taxes affect different taxpayers differently or that markets where evasion is an option converge to different equilibria than others. Andreoni et al. (1998), for example, propose several intuitive examples where heterogeneity in compliance behavior might affect market outcomes. It is instanced that tax systems are effectively less redistributive than legislated if the wealthy are able to evade a larger share of their income. This is confirmed by several empirical and theoretical studies (Roine 2006; Johns and Slemrod 2010; Alm and Finlay 2013; Traxler 2012; Duncan 2014; Doerrenberg and Duncan 2014a). Andreoni and co-authors further note that the incidence of taxation may be affected because those who are able to shelter their incomes pay a smaller share of taxes. Kopczuk et al. (2013) present suggestive empirical evidence that non-compliance indeed affects incidence. Labor market outcomes are likely to be altered as well if certain taxpayers have better evasion opportunities than others. For example, workers with a high willingness to evade may self-select into industries with good evasion opportunities or taxes could have differential effects on labor supply of workers with and without access to evasion.

While these examples in Andreoni et al. (1998) have intuitive appeal and show that the literature speculates that non-compliance is likely to affect market outcomes, evidence on many questions regarding the consequences of tax non-compliance is rather scarce.[14] In Chapters 3, 4, and 5 of this dissertation, I contribute to closing this gap and study the implications of tax non-compliance, and in particular heterogeneous compliance behavior, in different contexts. One reason for scarce empirical evidence on the effects of evasion opportunities and compliance behavior may be that identifying causal effects in this context is usually very difficult because workers self-select into industries and markets; that is, they select the industry/market that matches their preferences. As a result, taxpayers in different occupations/markets are very likely to differ not only with respect to their evasion opportunities but also along other dimensions such as risk aversion and other personality traits. Simply comparing occupations/markets with and without evasion options therefore does not identify a causal effect. Chapters 3 to 5 of this dissertation intend to circumvent this fundamental problem by conducting laboratory experiments with random assignment to evasion opportunities (Chapters 3 and 4) and an instrumental variable strategy with observational data (Chapter 5).

1.3 Overview of chapters and main findings

The preceding section elaborates based on the literature that the study of both the determinants and consequences of tax (non-)compliance is important. This section provides an overview of each chapter of this dissertation. It particularly summarizes the main findings and shows how the chapters contribute to both strands of literature; determinants and consequences of tax non-compliance. Incorporating the insights of the literature review and discussion of methods, I additionally briefly motivate the empirical tool that is chosen in the respective chapter.

1.3.1 Chapter 2: Does the use of tax revenue matter for compliance behavior?

Chapter 2 of this dissertation contributes to a better understanding of the non-pecuniary motives for tax compliance, which are not very well understood. While it is often speculated that the usage of tax revenue is among the non-pecuniary determinants of tax compliance, the literature is relatively silent in this regard. Chapter 2 of this dissertation contributes to fill this gap in the literature and explores the role of tax revenue usage for compliance behavior. The empirical tool used is a laboratory experiment which is the appropriate method for the study of this question because it is difficult, if not impossible, to study the effect of revenue use with observational data: random assignment of different types of tax money usage to taxpayers is difficult to achieve. Another justification to use

[14] Chapter 3, 4 and 5 of this dissertation provide more detailed and specific literature reviews.

a lab experiment is that the experiment only requires to compare compliance behavior across treatment groups whereas the actual levels of evasion are not interpreted. In the experiment, subjects make tax reporting decisions and are randomly assigned to different treatment groups that differ only in how the generated tax revenue is spent. 1) Tax revenue is equally *redistributed* among all subjects, 2) it goes to the experimenters' *scientific research fund*, 3) it is donated to the *Red Cross*, 4) it is transferred to the *German federal budget*. The motivation for the choice of these four treatment groups is twofold: First, they reflect set-ups that are either used in the "real-world" (group 4 with transfer to the federal government) or have been used in the lab experimental compliance literature and hence allow an evaluation of whether the use of tax revenue matters for previously conducted experiments (groups 1 to 3).[15] Second, they reflect four situations that differ with respect to the degree that subjects benefit from their tax payment; something which is speculated to matter for compliance behavior.

The findings of this Chapter indicate that average compliance levels are higher in the groups in which tax revenue is spent for research and charity purposes, relative to the two other groups with redistribution and tax payment transfer to the government. These differences are large and economically sizable. For example, relative to the group with redistribution among subjects, the level of compliance is about 40% and 35% higher in the Research and Red Cross groups, respectively. Although these differences between groups are economically meaningful, several non-parametric and parametric (regression) analyses reveal that they are not different from zero in a statistical sense. This paper therefore provides some suggestive evidence that the recipient of tax payments matters for compliance behavior, but the experimental data do not allow to reject the null hypothesis of no treatment effect. While it cannot be ruled out that other types of tax revenue use would make a larger and more precisely estimated difference, this chapter shows that the most common types in the literature and a realistic transfer to the government are likely to affect compliance behavior. The results may further suggest that taxpayers prefer spending tax revenues for charitable, rather than redistributive, purposes. The chapter contributes to the literature in several ways. First, it provides new evidence on the determinants of tax compliance. Second, it benefits the methodological side of compliance lab experiments in that it shows that the recipient of tax revenue is likely to make a difference in compliance lab experiments. Third, it adds to the behavioral literature on pro-social behavior and giving by showing that a different type of pro-social behavior – tax compliance – may depend on the type of return (i.e., the type of recipient) that is granted in exchange for pro-social behavior.

[15] For example, Alm, Jackson, and McKee (1992a) use a public good structure where all tax payments are redistributed among subjects, Fortin et al. (2007) transfer paybacks to scientific research funds, and Doerrenberg and Duncan (2014b) donate tax revenues to the Red Cross.

1.3.2 Chapter 3: Tax incidence in the presence of tax evasion

After adding to the literature on the determinants of tax compliance in Chapter 2, the next chapters study the consequences of evasion behavior on different outcomes. Chapter 3 examines the question of whether the canonical textbook laws of tax incidence hold in the presence of evasion opportunities. Intuitively, access to evasion opportunities allows taxpayers to lower their effective tax rates, which should result in different responses to tax rates for taxpayers with and without evasion opportunities. Despite its intuitive appeal, evidence on the effect of tax non-compliance on tax incidence is scarce. One reason for scarce evidence is likely to be rooted in the difficulty of identifying this effect with observational data. Causal inference with observational data would require random variation in access to evasion across otherwise similar markets. This is obviously difficult to achieve since access to tax evasion is most likely one of the dimensions of a market that determines whether buyers and sellers select to participate in that market. The chapter is therefore based on results of a laboratory experiment, in which subjects trade fictitious goods in competitive double auction markets. Subjects are randomly assigned roles of buyers and sellers in treatment and control groups. A per-unit tax is imposed on sellers, and sellers in the treatment group are provided the opportunity to evade the tax whereas sellers in the control group are not. Therefore, the impact of evasion on tax incidence is identified by comparing the equilibrium prices in the treatment and control group.

The results of the experiments show that the market equilibrium price in the treatment group with evasion opportunities is economically and statistically lower than in the control group where sellers are not provided an option to underreport. Accordingly, the number of traded goods is higher in the treatment group. These results are consistent with a theoretical model that is set up to improve the understanding of the mechanisms behind the results. In the model, access to evasion opportunities reduces the effective tax rate and therefore dampens real behavioral responses. The findings of the lab experiments further suggest that the benefits of tax evasion are not limited to the side of the market with access to evasion but are partly shifted to the non-evading side of the market.

1.3.3 Chapter 4: Evasion opportunities and labor supply elasticities

Heterogeneity in evasion opportunities may also affect other behavioral responses to taxation. Agents with evasion opportunities have more margins to respond to tax changes and it is therefore likely that there is an interaction between the sensitivity of different adjustment margins, such as labor supply or deduction behavior, with tax evasion. Chapter 4 of the dissertation examines the extent to which labor supply elasticities with respect to tax rates depend on evasion opportunities. Self-selection of individuals into markets with evasion opportunities, along with non-trivial identification of labor supply elasticities with respect to taxation, make it difficult to estimate a causal effect using observational data. This chapter therefore also employs a laboratory experiment to examine the research question. In the experiment, all participants undertake a well-established real-effort task over several rounds. Subjects face a tax rate that varies across rounds and are required to pay taxes on earned income. The treatment group is given the opportunity to underreport income, while the control group is not.

The lab experimental results provide evidence that subjects with an evasion opportunity indeed respond differently to changes in the net-of-tax rate than participants who do not have access to evasion. This confirms the expectation that an extra channel to adjust to changing tax rates affects the tax sensitivity of other adjustment channels. The results further show that this differential effect is more prevalent when tax rates go down. Additionally, the direction of the treatment effect is dependent on the order and evolution of tax rate changes. As expected, the elasticity of taxable income is considerable larger in the treatment group with evasion opportunity than in the control group.

In general, chapters 3 and 4 provide evidence that economic responses to tax rates depend on access to evasion opportunities; individuals who are able to evade have an extra margin to respond to tax rate changes and therefore adjust their behavior differently than individuals without this extra margin.

1.3.4 Chapter 5: Effect of compliance heterogeneity on tax policy

Chapter 5 of this dissertation asks whether tax policy is affected by the fact that different taxpayers have different inclinations to be non-compliant. Since it is difficult to mimic tax policy making in the laboratory using student subject pools, this chapter employs observational rather than experimental data. A single observational data set is, however, not readily available for an exploration of the question at hand, which is why a unique cross-country data set based on the World Values Survey and the World Tax Indicators is constructed. The resulting rich data set offers the opportunity of measuring various relevant variables – in particular tax burden, attitudes towards evasion, and income – and endogeneity concerns are addressed through an instrumental variables approach which exploits variation that is unrelated to the outcome variable of interest. The Chapter additionally makes a theoretical contribution as it is one of the first that incorporates tax morale into a 'standard' model of optimal taxation.

The findings of the empirical analysis show that income groups with high tax morale face higher average and marginal tax rates, conditional on other tax relevant variables.

This result is somewhat surprising and three possible mechanisms are proposed which may help to explain the result: *i)* an inverse elasticity argument where governments seek to minimize distortions, *ii)* a political economy argument where governments take voting behavior into account, and *iii)* an administrative costs argument where taxing high morale groups is more cost efficient.

1.4 Structure and co-authors

Chapter 2 studies how tax revenue usage affects compliance behavior. This chapter is solo-authored and not publicly available yet. **Chapter 3** examines whether the incidence of taxes is affected by evasion opportunities. It is co-authored with Denvil Duncan (Indiana University) and circulates as Doerrenberg and Duncan (2014c). **Chapter 4** deals with the question of whether tax evasion opportunities affect labor supply responses to taxation. This chapter is also co-authored with Denvil Duncan and was published as Doerrenberg and Duncan (2014b). The last **Chapter 5** studies whether heterogeneous levels of intrinsic motivation to pay taxes affect tax policy. It is joint work with Denvil Duncan, Clemens Fuest (ZEW Mannheim) and Andreas Peichl (ZEW Mannheim), and was published as Doerrenberg et al. (2014).

Chapter 2

Does the use of tax revenue matter for compliance behavior? Experimental evidence

2.1 Introduction

Identifying the drivers of tax non-compliance is one of the key aims of governments across the world and is also at the center of the economic literature on tax evasion. This paper examines whether the use of tax revenue is among these drivers and of relevance for compliance behavior. As opposed to the degree of control over private spending, taxpayers only have very limited information over the use of their tax payments. This lack of information over the use of tax revenue is suspected to be *among* the reasons why we observe considerable tax non-compliance (Alm 2012).[1] Another related reason for non-compliance may be that taxpayers have different preferences regarding the use of tax payments than policy-makers who administer tax expenditures. In this paper, I explore the relationship between tax usage and tax evasion and study whether the nature of tax revenue spending affects (lab experimental) compliance behavior.

Since it is difficult – if not impossible – to study the effect of tax money use on tax compliance with observational field data,[2] I address this question in the framework of a standard tax evasion lab experiment. Subjects first receive an endowment which is subject to a tax, and they are then given a tax reporting decision. The novelty of my experiment is that each subject is randomly assigned to one of four treatment groups (between-subjects design) that differ only in how the generated tax revenue is spent: 1)

[1]For example, Lamberton et al. (2014) report that the tax gap amounts to 385 billion US Dollars in the US and 42 billion pounds (about 70.8 billion US Dollars) in the UK.

[2]This has obvious reasons: First, evasion is, by its very nature, not observed. Second, one would need a field experiment where different taxpayers are randomly informed about the exact use of their tax payments.

Tax revenue generated in the experiment is equally *redistributed* among all subjects, 2) All tax revenue goes to the experimenter's *scientific research fund*, 3) Tax revenue is donated to the *Red Cross*, 4) Tax revenue is transferred to the *German federal budget* (administered by the Ministry of Finance). Since assignment to treatment groups is random and everything except the use of tax revenue is held constant, any differences in compliance behavior can be attributed to the way tax revenue is spent.[3]

The motivation for the choice of these four treatment groups is twofold: First, they reflect set-ups that are either used in the "real-world" (federal budget) or have been used in the lab experimental compliance literature and hence allow an evaluation of whether the use of tax revenue matters for previously conducted experiments. For example, Alm, Jackson, and McKee (1992a) use a public good structure where all tax payments are redistributed among subjects, Fortin et al. (2007) transfer paybacks to scientific research funds, and Doerrenberg and Duncan (2014b) donate tax revenues to the Red Cross. Second, they reflect four situations that differ with respect to the degree that subjects benefit from their tax payment. This is likely to matter for compliance behavior. Torgler et al. (2010), for example, show that there exists a negative correlation between the level of local tax autonomy and shadow economy measures, suggesting that compliance may be higher if taxpayers are under the impression that their tax payments are spent locally and hence to their direct benefit. In the first group, redistribution among all participants of a session mimics a situation in which tax money is spent locally and to the direct benefit of a well-defined group. In the second group, participants benefit less directly: more money for research marginally increases the reputation of the participant's university and may lead to more lab experiments in which subjects can participate and earn money. In the group with donations to the Red Cross, subjects benefit only if they have a warm glow for giving to charity (e.g., Andreoni 1990). The fourth group with transfers to the budget of the federal government mimics the most realistic tax payment set up, and it is likely that tax money is perceived to benefit an individual subject only very marginally.[4]

The experimental results indicate that average compliance levels are higher in the groups in which tax revenue is spent for research and charity purposes, relative to the two other groups with redistribution and tax payment transfer to the government. These differences are large and economically sizable. For example, relative to the group with redistribution among subjects, the level of compliance is about 40% and 35% higher in the Research and Red Cross groups, respectively. Although these differences between groups are economically meaningful, several non-parametric and parametric (regression)

[3]That is, the *recipient* of tax revenues is varied in my experimental design. Since most subjects are likely to have a broad perception of how the different recipients *use* tax revenues, I argue that this link between tax revenue recipient and tax revenue use allows me to infer conclusions not only about the effect of tax revenue recipient but also of tax revenue use on compliance.

[4]For example, Mettler (2011) shows that many people under-recognize the benefits of tax-funded policies.

analyses reveal that they are not different from zero in a statistical sense. This paper therefore provides some suggestive evidence that the recipient of tax payments matters for compliance behavior, but the experimental data do not allow to reject the null hypothesis of no treatment effect. While it cannot be ruled out that other types of tax revenue use would make a larger and more precisely estimated difference, my paper shows that the most common types in the literature and a realistic transfer to the government are likely to affect compliance behavior. The results may further suggest that taxpayers prefer spending tax revenues for charitable, rather than redistributive, purposes.

This paper contributes to the literature in several ways. First, it adds to the literature on tax evasion in general by presenting more evidence on the potential determinants of tax evasion.[5] For example, Frey and Torgler (2007) find a strong correlation between survey measured tax morale and institutional quality and Torgler and Schneider (2009) show that measures of the shadow economy are correlated with the level of institutional quality as well. The large-scale field-experiment by Hallsworth et al. (2014) provides further evidence that compliance depends on tax revenue use. Taxpayers who receive a letter that reminds them that tax payments are used to finance public goods like roads and schools pay their tax liabilities quicker than taxpayers who do not receive such a letter. These studies using observational data from the field suggest that compliance depends on the institutions that administer tax payments and is affected by the salience of tax payment use, but they do not inform whether the type of tax money use matters as well.

There is also lab experimental literature regarding the use of tax payments that I relate to and which I extend. Alm and Jackson (1993) find that compliance increases if taxpayers are able to vote over the spending of tax payments. A study by Lamberton et al. (2014) explores a related question and shows that compliance increases even if subjects are only provided a hypothetical voice over tax payments that is not binding, i.e., is not directly implemented in the actual tax payment use. Other studies such as Alm, McClelland, and Schulze (1992) and Alm, Jackson, and McKee (1992b) show that compliance increases if complying yields efficiency gains. In their experiments, the sum of all tax payments is increased by a multiplier to increase efficiency and then equally redistributed among the subjects.[6] These papers provide evidence that use of

[5]Note that I do not claim that the use of revenue is the *only* reason for non-compliance but I test if it is *among* the drivers. See Alm (2012) or Slemrod and Weber (2012) for overviews on the determinants of compliance.

[6]The procedure is therefore similar to standard public good experiments (following Marwell and Ames 1981). Alm, McClelland, and Schulze (1992) additionally show that compliance increases with the public good multiplier, suggesting that the level of tax payment efficiency matters as well. Note that the tax money recipients in my design are not necessarily perceived to differ in efficiency. I intentionally do not introduce a public good multiplier in the treatment group with redistribution of tax money among subjects because redistribution is usually not perceived to be this efficient, some subjects might rather have the "leaky bucket" (Okun 1975) picture in mind where tax money is lost in the process of

tax payments generally matters for compliance behavior[7] and I add to this literature by examining if the type of recipient affects evasion. In particular, subjects in the present experiment are neither provided a voice over alternative spending alternatives, nor are they faced with a situation in which complying is associated with an efficiency gain. In line with "real-world" applications, taxpayers in my experiment are confronted with different tax payment recipients and there are no obvious efficiency gains.

A second set of results in Alm and Jackson (1993) is more closely related to my paper. They conduct a survey among students before their lab experiments to determine the popularity of various campus organizations. Subjects in all treatments are first informed about the survey and which organizations were voted to be most- and least-popular. In one treatment, participants were then told that tax payments would be donated to the most-favored organization, whereas in another treatment it was told that tax money would go to the organization that was surveyed to be the least-favored. The results show that compliance is higher when tax payments go to the more popular organization. This suggests that individual compliance behavior is affected by the majority opinion, and that social norms regarding the recipient matter. While I relate to this study in that I compare compliance between treatments that differ with respect to the tax money recipient, I do not test if the public opinion on the recipient affects compliance. In my design, subjects decide independently of each other and it is neither known if the recipient is considered to be popular nor is it made salient to the participants that tax money could have been spent in a more popular way.[8]

This paper further benefits the methodological side of the lab experimental literature on tax evasion in that it tests whether researchers should be concerned about how they spend the "tax revenue" generated in their experiments. Given the nature of tax evasion, it is very difficult to study evasion with observational data that suffer from false reporting and randomized or quasi experiments. Field experiments (e.g., Slemrod et al. 2001; Kleven et al. 2011; Pomeranz 2013) are promising but only allow to study certain questions – besides being expensive and time consuming.[9] The broad literature on lab experimental compliance experiments (e.g., Alm, Jackson, and McKee 1992a, Fortin et al. 2007, Alm et al. 2009, Doerrenberg and Duncan 2014b, Balafoutas et al. 2014) shows

redistribution.

[7]This is further supported by Cowell and Gordon (1988) who show theoretically that introducing public goods to the standard Allingham and Sandmo (1972) model changes the effect of tax rate changes on evasion. Becker et al. (1987) find in a lab experiment that evasion increases if a taxpayer supposes that she receives less tax financed public transfers than other tax payers.

[8]It can well be presumed that subjects in my design do not know that the experiment is intended to examine the effect of varying tax money recipients. Not making the purpose of the experiment salient is also one of the reasons why I exploit between-subject variation rather than within-subject variation.

[9]For example, many countries have strict obligations that require the equal treatment of all tax payers. This often makes the conduction of field experiments with randomized treatments legally difficult, and certainly limits the range of questions that can be examined.

that the lab is a reliable environment to study compliance behavior. While it is critical to derive conclusions about actual levels of compliance from lab experiments, it is possible to study various mechanisms that help to understand compliance behavior (see Slemrod and Weber 2012 or Alm 2012 for similar assessments).[10] This paper evaluates whether previously used ways of spending lab tax revenues (see above for references) affect compliance behavior. It provides suggestive evidence that the use of tax revenue may matter for lab experimental evasion decisions. Researchers conducting lab experiments in the context of tax compliance should be aware that the choice of tax revenue recipient may impact the average level of compliance.

Third, the paper speaks to the behavioral economics literature on pro-social behavior (i.e., paying taxes), donating and altruism (note that fully evading is the dominant rational strategy in my experiment). For example, Carpenter and Myers (2010) and Chetty and Saez (2014) present evidence from the field that pro-social behavior depends on the – monetary or non-monetary – benefit that people receive in return. These results correspond with the broad literature that uses lab experiments to study pro-social behavior (see Vesterlund 2014 for a recent survey). Several lab experimental studies show that the type of recipient matters for giving behavior in dictator games and voluntary donations to government or charity agencies (e.g., Eckel and Grossman 1996; Fong 2007; Fong and Luttmer 2011; Li et al. 2011).[11] Despite imprecise treatment effects, my paper somewhat confirms the results of this literature in that it shows that another type of pro-social behavior – tax compliance – may also depend on the type of return (i.e., the type of recipient) that is granted in exchange for pro-social behavior. In a broader perspective, the paper further suggests that some of the results in behavioral economics are also likely to apply to the economics of tax evasion.

The paper proceeds as follows. Section 2.2 introduces the experimental design and procedures. Parametric and non-parametric results are presented in Section 2.3, and Section 2.4 discusses the results and concludes the paper.

2.2 The laboratory experiment

This section describes the experimental design and procedures. I first provide a brief overview of the design and then lay out the flow of each round of the experiment. In

[10]This argument is in line with for example Kessler and Vesterlund (2014) who explore the issue of external validity of laboratory experiments. They argue that lab experiments inform about the direction of "real-world" effects whereas the magnitudes of most lab experiments should be treated with caution. Fortin et al. (2007), Slemrod and Weber (2012) and Doerrenberg and Duncan (2014b) present more detailed discussions of the external validity of tax compliance lab experiments.

[11]Donating for real charity purposes is not standard in the lab experimental literature on giving. Fong and Luttmer (2009) and Grossman and Eckel (2012) are among further examples in which lab donations are donated to real charity organizations.

a next step, the payment structure and calculation of total tax revenue are described and details on the treatment groups are presented. The organizational procedure of the experiment and summary statistics are described at the end of this section.

2.2.1 Overview of the experiment

The laboratory experiment I design consists of one practice round and eight payoff-relevant rounds. Each round has two stages: In the first stage, subjects are endowed with an amount of money that represents their pre-tax gross income. This endowment is the same for all subjects in all groups but differs across rounds. Gross income is subject to a tax rate of 30%. In the second stage, all subjects make a tax reporting decision. There is a 10% audit probability of being audited and a penalty is due if a subject is audited and income was underreported. This basic set-up of the experiment is identical for all subjects in all experimental sessions. In order to study the research question, I randomly assign subjects to one of four treatment groups – between subjects design – that differ only in how the generated tax revenue in a session is used. The four treatment groups are as follows (see section 2.2.4 for more detailed information): 1) Tax revenue is equally *redistributed* among all subjects, 2) all tax revenue goes to the experimenter's *scientific research fund*, 3) tax revenue is donated to the *Red Cross*, 4) tax revenue is transferred to the *German federal budget* (administered by the Ministry of Finance). All subjects within a given session are always in the same treatment group, and the relevant tax revenue consists of the tax payments of all subjects in one session.

2.2.2 Overview of a round

Endowment

Subjects receive an exogeneous endowment, which constitutes their pre-tax gross income, in the first stage of each round. As opposed to several recent tax evasion experiments such as e.g., Alm et al. (2009) and Doerrenberg and Duncan (2014b), I do not implement a real-effort task in the first stage of the experiment. In the literature, real-effort tasks are used to induce a sense of ownership and to simulate a real-world situation where money is earned. However, contributions using a real-effort task study research questions where there is no concern that the treatment variation may have an effect on the effort supply, and hence gross incomes, of subjects. This is different in the context of the present paper where the recipient of tax money is varied across treatment states. If one tax recipient is perceived to be more reliable or in need, subjects may work harder in these treatment groups and generate higher gross incomes. This would bias the results because subjects in different groups would not only differ with respect to the tax money recipient but also with respect to their effort levels and gross incomes. Since evasion decisions are likely to depend on effort and wealth, differences in compliance levels across groups could not be solely attributed to different tax money recipients.

Endowing subjects with an exogenous gross income and not allowing for endogenous income earning follows the early lab experimental evasion literature (e.g., Alm, Jackson, and McKee 1992a) and is also in line with more recent contributions such as Fortin et al. (2007) or Balafoutas et al. (2014). In order to examine if the treatment effect of interest depends on the level of gross income, I vary the level of endowments across rounds. Before the experiment I chose four different levels of endowment and randomized the assignment of these four levels to the eight rounds of the experiment. The chosen endowment levels are 65 ECU (13 EUR), 58 ECU (11.60 EUR), 51 ECU (10.20 EUR) and 44 ECU (8.80 EUR).[12] Note that the levels and order of endowments is the same for all subjects in all sessions and treatment groups. 58 ECU was randomly drawn to be the endowment level in the practice round. Subjects were informed in the instructions before the start of the experiment that the endowment may vary from round to round and that endowments are in the range of 44 and 65 ECU. The endowment levels in each round are displayed in Table 2.1.

Table 2.1: Endowment levels in each round

Round:	1	2	3	4	5	6	7	8
Endowment (ECU):	44	65	58	58	51	65	51	44

Notes: Reported are the endowment levels (pre-tax gross incomes) in each payoff-relevant round. Numbers are in Experimental Currency Units (ECU) where 5 ECU correspond to 1 EUR. The four endowment levels 65, 58, 51, 44 were randomly assigned to the eight rounds of the experiment. 58 ECU was randomly drawn to be the endowment in the practice round.

The tax reporting decision

Subjects' gross incomes (endowments) are subject to a proportional tax rate of 30%.[13] However, subjects in all treatment groups are faced with a tax reporting decision in the second stage of each round. This reporting stage follows good practice in the lab experimental evasion literature (see Torgler 2002 and Alm 2012 for overviews). Subjects are first informed about their gross income in this round and are then asked to indicate an amount between zero and their true gross income for tax purposes. There is an exogenous probability of 10% that their reporting decision is audited, and in the case of an audit,

[12] These endowment levels are relatively high, especially considering that mostly undergraduate students participated in the experiment and that the sessions lasted about 45 minutes. For an impression of the purchasing power of these endowment levels, note, for example, that a minimum hourly wage of 8.50 EUR is discussed to be introduced in Germany. A full lunch menu at the University of Cologne's cafeteria costs about 2,50 EUR.

[13] 30% is chosen as the tax rate because it is likely to be perceived as a realistic tax rate in Germany. Lamberton et al. (2014), for example, also use a 30% tax rate.

all underreported income is detected and a penalty is due. The chosen audit probability of 10% is commonly used in the literature (e.g., Alm et al. 2009; Fortin et al. 2007; Doerrenberg and Duncan 2014b). As in almost all previous compliance lab experiments, the penalty in case of detected underreporting is proportional to the evaded amount. However, while many recent studies use a fine rate of 2 – meaning that detected subjects pay taxes on their reported income plus *twice* the evaded amount – , I employ a fine rate of 3 in order to generate a sufficiently high level of compliance that allows for variation in compliance levels in all directions (e.g., Alm, Jackson, and McKee 1992b also use a fine rate of 3). After each reporting decision, the computer determines randomly if a reporting decision is audited. Subjects are informed about the parameters of the tax system (tax rate, audit probability and penalty structure) in the instructions and the computer screen also reminds them before each round. All tax reporting decisions are completely anonymous and neither other subjects nor the experimenters are able to know during a session if and how much a subject underreported.

Net income

The final income subjects receive in each round, i.e., a round's "net income", depends on whether the reporting decision is audited or not. If it is not audited, subjects earn their gross income minus taxes paid on reported income. In case of an audit, a subject has to pay the true tax liability plus a penalty that equals twice the evaded amount.[14] Audited subjects who have not underreported income simply pay taxes on their true gross income. That is, a subject i receives a net income Y_i^{net} in a given round that can be summarized as follows:

$$Y_i^{net} = \begin{cases} = G(1-\tau) - 2\tau(G-R_i) \text{ with probability p} \\ = G - \tau R_i \text{ with probability (1-p)} \end{cases} \tag{2.1}$$

where G is the exogenous gross income (endowment) that does not vary between subjects, R_i is the amount reported for tax purposes in this round, τ is the proportional tax rate of 30% and p is the exogenous and constant audit probability of 10%.

Tax payment and total tax revenue

The tax payment of a subject depends on her gross income, the tax rate, penalty structure and audit outcome, as well as the compliance decision. Tax payment of each subject i in

[14] Note that this is equivalent to paying taxes on *reported* income plus three times the evaded amount, i.e., a fine rate of 3.

a round is denoted with t_i and can be written as:

$$t_i = \begin{cases} = \tau G + 2\tau(G - R_i) \text{ with probability p} \\ = \tau R_i \text{ with probability (1-p).} \end{cases} \tag{2.2}$$

The total tax revenue generated in a round is the sum of tax payments from all subjects in that round. The number of subjects in a session shall be N so that total tax revenue T is written as

$$T = \sum_{i=1}^{N} t_i. \tag{2.3}$$

T is the relevant tax revenue from one round that is used depending on the respective treatment group (the treatment groups are described further below).

2.2.3 Final payment

Subjects do not receive the sum of all eight payoff-relevant rounds' net incomes, but their final pay-off is instead based on the net income of one randomly chosen round. This payment structure avoids wealth effects, satiation and unreliable decisions once subjects have achieved a certain desired pay-off. It further allows to have higher stakes involved in each of the reporting decisions (see e.g. Blumkin et al. 2012 and Doerrenberg and Duncan 2014b who have a comparable payment structure). After the completion of all rounds in a session, the computer randomly selects one of the eight rounds and subjects are informed of the draw. The randomly chosen round is the same for all subjects within one session; that is, the computer randomly selects a round at the end of each session that determines the payment-relevant round for all participants of this session.[15] Experimental Currency Units (ECU) are used as the currency during the experiment. The net income (in ECU) for the selected round is then converted to EUR using the publicly known exchange rate of 5 ECU for 1 EUR. In the treatment group with redistribution among subjects, tax revenue of the payoff-relevant round in the session is split equally among all subjects and each subject's share is also converted to EUR and added to the final payment. In addition to this, subjects in all groups receive a show-up fee of 2.50 EUR which is added to the selected net income. See the summary statistics in section 2.2.6 for information on final payments.

The amount of tax revenue that is spent depending on the respective treatment group is also based on the randomly selected round of a session; hence, the tax revenue generated in the randomly selected round of one session is used in the way that was indicated to the subjects of a treatment group.

[15] An alternative would have been to select rounds individually for each subject. However, this would not have allowed the implementation of the redistribution treatment where tax revenue is split equally among all subjects.

2.2.4 Treatment groups

The flow of each round, including payment structure and generation of tax revenue, is identical for all subjects in all sessions. In order to identify how the usage of tax revenue matters for the evasion decisions in the reporting stage, I employ a between-subjects design in which each subject is randomly assigned to one of four treatment groups that only differ in how tax revenue generated in a session is used. This design ensures that all differences across the four groups are solely due to the use of tax revenue and not endogenously confounded by any other variables.[16] The usage of tax revenue is public information and made salient to the subjects by indicating it in the instructions as well as on the computer screen when the reporting decision is made. That is, the instructions and computer screens are identical for all subjects except for single sentences that are regarding the usage of tax revenue. The instructions (see Appendix 2.6) describe in an extra section labeled ”Use of tax revenues” how tax payments are spent. In addition, while subjects make their reporting decision, they are reminded on the computer screen how their tax payments will be used. I describe the four treatment groups in more detail in the following:

1. **Redistribution**: Tax money generated in this group is equally redistributed among all subjects in one session. Subjects are informed that after the determination of the relevant payment round the computer calculates total tax revenue in the selected round and splits it equally among all subjects of the session; that is, the share of tax revenue that a subject receives is only indirectly linked to her own performance or compliance choices. This treatment simulates a set-up where tax revenue is spent within a small and well-defined group of individuals who, although they most likely had not met before, have seen each other before and in the lab.[17] In addition, tax revenue in this group is used such that it directly benefits the monetary payoff of the subjects. The procedure in this group is similar to previous compliance experiments (e.g., Alm, Jackson, and McKee 1992a) and experiments in the large literature on public good provision following Marwell and Ames (1981), except that I do not add a public good multiplier which increases the efficiency of redistribution.

2. **Research**: Subjects in this group were informed that all tax revenue generated in the selected payment round of this experimental session is spent for research purposes. In particular, it was stated in the instructions and on the computer screen

[16] Charness et al. (2012), for example, argue that a between-design is usually superior to a within-design where each subject is faced with different treatments. In a between-design such as mine, order effects in the treatment conditions do not affect the results and it is not made salient to the subjects that other treatments exist.

[17] Note that while this treatment induces a personal component to the experiment, compliance decisions and labor effort performances are of course anonymous so that it is never known to any other participant how much each subject contributed to the total tax revenue.

that tax revenue goes to the research fund of the experimenters and will be spent on future research projects (without specifying a particular project). This treatment group intends to stimulate a situation in which tax money is spent locally at the University where the experiment was conducted and where most of the participants are enrolled as students. As a result, tax payment used here does not directly benefit the subject but may only have a small indirect impact through a small improvement of the university's research reputation or an increased probability that more experiments are conducted in which subjects can participate and earn money. Subjects may have an idea of how research is conducted at their university and may know university researchers in person from classes and seminars. That is, subjects may have a sense what is done with the money but this sense is clearly less developed than in the *redistribution* group. This way of spending laboratory tax payments is in line with Fortin et al. (2007).

3. **Red-Cross**: In this treatment group, subjects are informed that total tax revenues from the selected payment round are donated to the German Red Cross. The Red Cross is a non-ideological charity organization that is usually perceived as reliable and transparent. It is used as the tax recipient in lab experimental evasion studies by, for example, Doerrenberg and Duncan (2014b) to make the compliance decision as realistic as possible. A paper by Eckel and Grossman (1996) shows that dictators share more in experimental dictator games if the recipient is the American Red Cross, suggesting that subjects do act differently in lab experiments where money goes to a charity organization such as the Red Cross. The donations are not targeted at a specific project, country or the such, but go to a general Red Cross fund and Red Cross Germany decides independently how to use the donation from this experiment. Participants in the present experiment may have a vague idea which projects are supported by the Red Cross but they do not know the exact usage of the money. That is, they only have a diffuse idea how their tax money is spent. It is in any case save to assume that subjects will not directly benefit from the donated money unless they have a warm glow for giving.

4. **Federal Government**: All tax revenue collected in the payoff relevant round of this treatment group is transferred to the budget of the federal German government. Tax payments therefore go to the same general fund as any federal taxes (such as the German income tax) and spending is decided upon by the federal government after tax payments have been received. This payment situation therefore mimics the real-world set-up where federal taxes go to the general fund and are not targeted at specific purposes. Transferring to the federal government without a specific purpose

is easily possible for any person willing to give money.[18]

2.2.5 Organization

The lab experiment was run in the Cologne Laboratory for Economic Research (CLER) at the University of Cologne, Germany in June 2014. A random sample of the laboratory's subject pool of approximately 4000 persons was invited via email – using the recruitment software *ORSEE* (Greiner 2004) – to participate in the experiment. Potential participants signed up for participating in a session on a first-come-first-serve basis. A total of 126 subjects participated in the experiment (see section 2.2.6 for summary statistics). Neither the content of the experiment nor the expected payoff were stated in the invitation email. The computerized experiment was programmed utilizing *z-tree* software (Fischbacher 2007).

All subjects in one session were in the same treatment condition and I conducted one session of 32 or 31 subjects for each of the four treatment groups, giving a total of four sessions.[19] Each session included a practice round, in which subjects could familiarize themselves with the experiment and in particular the reporting decision, eight payoff-relevant rounds, and lasted about 45 minutes (including reading of the instructions and payment of all participants). Upon entering the lab, random assignment to computer booths was implemented by asking each subject to draw an ID number out of a box. Subjects received a hard copy of the instructions when they entered the lab (see appendix 2.6) and were allowed as much time as they needed to familiarize themselves with the procedure of the experiment. They were then given the opportunity to ask any clarifying questions in private.

2.2.6 Data and summary statistics

Table 2.2 presents summary statistics separated by treatment groups. Several demographic and attitudinal variables were surveyed through a questionnaire at the end of the experiment. The demographic variables include age, gender and native language. I further asked one question on risk aversion[20] and one on tax morale.[21]

The summary statistics show that males and native German speakers make up

---

[18]A simple bank transfer is required to give money to the federal government's budget. Such transfers are administered by the so-called *Bundeskasse*. See `http://www.kkr.bund.de` for more information.

[19]It was intended to have 32 subjects in all sessions but there were two no-show ups.

[20]The measure of risk aversion is obtained by asking subjects to choose between a certain pay-off of EUR 50 and a gamble that pays EUR 100 with probability of 0.5 and EUR 0 with probability of 0.5.

[21]The tax morale question is adopted from World Values Survey (Minkov 2012). "Please tell me for the following statement whether you think it can always be justified, never be justified, or something in between: 'Cheating on taxes if you have the chance'." This is the most frequently used question to measure tax morale in observational studies (e.g., Slemrod 2003, Alm and Torgler 2006, and Halla 2012).

39.7% and 86.5% of the sample, respectively, and that the average age is 24.52 years. Randomization into treatment groups worked mostly well as the variables are fairly balanced across treatment groups. Non-parametric pair-wise Wilcoxon rank-sum tests for differences in distributions reveal small statistical significances in age between treatment group 4 (government) and the other groups, as well as in the share of native German speakers between group 1 and the other groups. All other group wise comparisons yield non-significant differences. In order to ensure that the results are not driven by these differences, I provide regression results that control for all demographic variables.

The table also depicts how much tax revenue was generated in each of the groups (average per round: 309.23 ECU, 61.85 EUR) and how much tax revenue was eventually used in the respective treatment group (Variable *Tax amount used*. Recall that this was based on one randomly selected round). That is, in the group with redistribution 283.80 ECU (56.76 EUR) were equally redistributed among all 31 subjects. In the other groups, 351.60 ECU (70.32 EUR) were kept for research purposes, 255.01 ECU (51.00) EUR) were donated to the Red Cross and 215.52 ECU (43.10 EUR) were transferred to the budget of the German federal government.

Table 2.2: Summary Statistics by Treatment Status

	Treatment group				
	Redistribution	Research	Red Cross	Government	Total
Age	23.26	24.22	23.59	27.06	24.52
	(4.993)	(3.599)	(3.387)	(8.771)	(5.723)
Male	0.387	0.531	0.375	0.290	0.397
	(0.495)	(0.507)	(0.492)	(0.461)	(0.491)
German native	0.968	0.813	0.844	0.839	0.865
	(0.180)	(0.397)	(0.369)	(0.374)	(0.343)
Tax morale	7.935	7.250	7.594	7.645	7.603
	(2.190)	(2.300)	(2.662)	(2.715)	(2.460)
Risk	1.323	1.281	1.219	1.484	1.325
	(0.653)	(0.634)	(0.491)	(0.811)	(0.655)
Tax revenue avg.	308.2	337.7	290.0	300.7	309.2
	(0)	(0)	(0)	(0)	(17.88)
Tax amount used	283.8	351.6	255.01	215.52	276.48
	(0)	(0)	(0)	(0)	(41.22)
Observations	31	32	32	31	126

Notes: Reported are the means of demographic and attitudinal variables by treatment group. Standard deviations in parentheses. Variable *Tax revenue* is the average amount of tax revenue generated over all rounds. Variable *Tax amount used* is the tax revenue in the randomly selected payoff-relevant round, which was spent as indicated in the respective treatment group. All information, except *Tax revenue avg.* and *Tax amount used*, were surveyed after the end of the experiment.

2.3 Results

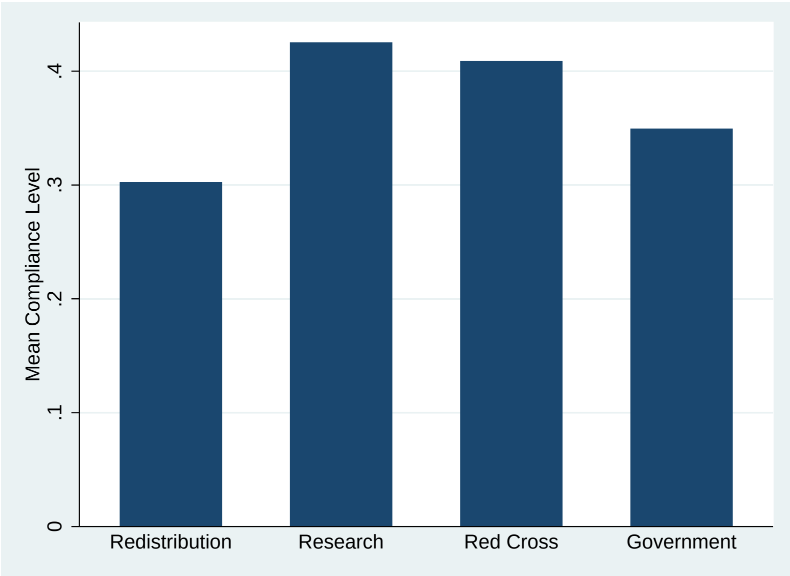
2.3.1 Non-parametric comparison of treatment groups

In this section, I provide an overview of compliance levels and compare the treatment groups with respect to evasion behavior. Figure 2.1 depicts the mean share of reported income, i.e., ratio of income reported for tax purposes to endowment, in each treatment group. The figure suggests that average compliance levels are different across the groups. I observe that the share of gross income that is reported for tax purposes is 30.22% in the ”Redistribution” group, 42.52% in the ”Research” group, 40.87% in the ”Red Cross” group, and 34.94% in the ”Government” group.[22] While the compliance levels are

[22]The fact that the ranks of groups in terms of tax revenue (reported in the previous section) do not mimic the ranks in compliance levels is due to the fact that tax revenues include penalty payments. For example, while tax revenue in the Red Cross group is relatively low despite the relative high compliance rate, subjects in this group were subject to less (random) audits and therefore had to pay less penalty fees.

similar in groups 1 and 4, they are larger in groups 2 and 3. That is, subjects in the "Research" and "Red Cross" treatments are on average more compliant than subjects in the "Redistribution" and "Government' groups. These effects are fairly large and economically relevant. For example, compliance in the "Research" and "Red Cross" groups is about 41% and 35%, respectively, higher than in the "Redistribution" group.

Figure 2.1: Compliance Rate by Treatment Group



Notes: Reported are the means of compliance rates by treatment group. Compliance rate is defined as income reported for tax purposes divided by gross income. $N = 126$. Treatment groups differ in the use of tax revenue.

Figures 2.2 to 2.5 in the Appendix report histograms of the compliance levels. The Figures show that most participants in all four treatment groups chose the extreme values of compliance behavior; either full compliance or full evasion.[23] In all four groups, the number of full evaders is higher than the number of full compliers. The figures further reveal that the differences in means are partly driven by the relatively low share of full evaders in group "Research" and the relatively high share of full compliers in groups "Research" and "Red Cross". The number of full evaders in the group with transfers to research is particularly low relative to the other groups. This corresponds with the results for the average compliance levels, which is highest in this group.

In a next step, I first average the data by subject,[24] and then provide an overview of means in compliance levels for each group along with the respective standard deviations in Table 2.3. The table depicts that deviations from the mean are high in all groups. For example, the standard deviation is 0.35 in the "Redistribution" group which has an

[23]Tobit regression analyses below account for this distribution of compliance levels.

[24]This procedure yields one observation per subject. Averaging is necessary because a subject's decisions are very likely to be correlated across the eight periods; this implies that one should not treat the eight decisions of a subject as independent observations.

average compliance level of 0.302. These high variances put into question whether the descriptive differences in means reflect significant differences in a statistical sense. In order to test statistical significance, I first conduct a Kruskal-Wallis test which evaluates the hypothesis that samples from all four groups are drawn from the same distribution. The test yields a p-value of 0.44, suggesting that there are no *statistical* differences between all four groups. This result is confirmed by pairwise Wilcoxon rank-sum tests in which each group is non-parametrically compared to each other group (not reported in the Table): none of the pairwise comparisons show a statistically significant effect. The difference between groups 1 and 2, which have the lowest and highest compliance levels respectively, is close to conventionally defined significance; however, the p-value of 0.134 does not allow to reject the null hypothesis of no effect even for these two groups.[25] While there are sizable differences in compliance levels between groups, the results therefore suggest that these differences do not reflect statistically significant differences.

Table 2.3: Compliance and Standard Deviations by Treatment Status

	Treatment group					
	Redistribution	Research	Red Cross	Government	Total	K-W
Compliance	0.302	0.425	0.409	0.349	0.372	0.44
	(0.350)	(0.369)	(0.411)	(0.392)	(0.380)	
Observations	31	32	32	31	126	

Notes: Reported are the means of compliance rates by treatment group. Standard Deviations in parentheses. Compliance rate is defined as income reported for tax purposes divided by gross income. Treatment groups differ in the use of tax revenues. *K-W* is the p-value of a Kruskal-Wallis test for differences in distributions between the treatment groups (H0: no differences).

2.3.2 Parametric regressions

This section explores whether the non-parametric results translate into estimates from parametric regressions. In the following, I present results based on the following regression model:

$$\phi_{igt} = \alpha + \beta \cdot Treat_g + \phi \cdot \mathbf{X}_{ig} + \eta_t + \epsilon_{ig}, \tag{2.4}$$

where subscripts indicate a subject i in treatment group $g = 1, 2, 3, 4$ (with $g = 1$: redistribution, $g = 2$: research, $g = 3$: Red Cross, $g = 4$: government) and round t. The dependent variable ϕ_{igt} is the share of gross income that is reported for tax purposes, i.e. the compliance rate. $Treat_g$ is a categorical indicator variable for a subject's treatment

[25]The p-values of all pairwise Wilcoxon rank-sum tests are: Group 1 vs 2: 0.134, 1 vs 3: 0.441, 1 vs 4: 0.948, 2 vs 3: 0.695, 2 vs 4: 0.215, 3 vs 4: 0.902.

group[26] and β is the coefficient of interest. The coefficients for each treatment group will be relative to the omitted group with $g = 1$, in which tax revenues are redistributed among all subjects. ϵ_i is an error term. Some specifications also include a set of round fixed effects η_t, and controls for demographic and attitudinal variables that are captured in $\mathbf{X}_{ig}$. I use standard pooled OLS panel regressions as well as Tobit regressions that account for the censored (between zero and one) nature of the dependent variable. Standard errors in all regressions are clustered on the subject level to adjust statistical inference for the interdependency (serial correlation) of decisions across the eight rounds of the experiment.[27]

Specification (I) of Table 2.4 shows OLS panel estimates of the effect of treatment status on the share of reported income (without including any further variables). The estimated coefficients of the treatment variables, all relative to the omitted reference group "Redistribution", confirm the non-parametric analysis: while the coefficients mirror the fairly large differences in means, relatively high standard errors do not allow to reject the hypothesis of no effect. As in the non-parametric comparisons, the difference between groups "Redistribution" and "Research" is fairly close to conventional significance, but does not cross the 10% threshold (p-value: 0.171). These results are virtually unchanged in the presence of period fixed effects in specification (II). Specifications (III) and (IV) add control variables for demographics (age, gender language) and attitudinal variables (tax morale and risk surveyed after the experiment). The main results still remain stable: compliance is higher in groups "Research" and "Red Cross" but these differences are not precisely estimated. Table 2.5 in the Appendix displays the coefficients for the control variables. In line with many findings, males in this experiment were less compliant than women. All other variables are close to zero and not statistically significant. The histograms discussed above indicate that many subjects chose extreme compliance behavior at either zero or one. This motivates the use of Tobit regressions that account for censoring at zero and one as a robustness check. Results from such Tobit regressions, presented in Table 2.6 in the Appendix, are in line with all previous findings: compliance levels are different, the difference between groups 1 and 2 is close to conventional significance, and all other coefficients are not not statistically significant.

[26] Recall that I employ a between-subjects design where each subject is exclusively in one of the four treatment groups. This also permits including individual fixed effects in the regressions.

[27] Note that in contrast to the non-parametric analysis, I use the panel data set with eight observations per subject in the regressions.

Table 2.4: Panel Regressions: Effects of treatment on compliance

	(I)	(II)	(III)	(IV)
Reference group: Treatment Redistribution				
Research	0.123	0.123	0.140	0.140
	(0.090)	(0.090)	(0.088)	(0.088)
Red Cross	0.106	0.106	0.102	0.102
	(0.095)	(0.095)	(0.093)	(0.093)
Government	0.047	0.047	0.018	0.018
	(0.093)	(0.094)	(0.097)	(0.097)
constant	0.302***	0.327***	0.217	0.241
	(0.062)	(0.065)	(0.207)	(0.210)
Period FE		X		X
Controls			X	X
N	1008	1008	1008	1008
R2	0.014	0.019	0.046	0.051

Notes: Pooled OLS panel regressions based on equation 2.4. Standard errors in parentheses clustered by individuals. Estimates are based on a sample of 126 individuals and eight periods. Significant levels are $* < 0.10$, $** < 0.05$, $*** < 0.01$. Dependent variable: Compliance rate defined as income reported for tax purposes divided by gross income. Independent variables of interest are the dummy variables for the treatment groups "Research", "Red Cross" and "Government". Coefficients of interest are relative to the omitted treatment group "Redistribution" (in "constant"). Treatment groups differ with respect to the usage of tax revenue. Specifications (II) and (IV) include period fixed effects. Specifications (III) and (IV) include control variables for age, gender, language, tax morale and risk attitude. The results for these coefficients are depicted in Table 2.5 in the Appendix.

Heterogeneity in gross income levels In a next step, I test whether treatment effects depend on the level of endowment; something which may of relevance for future compliance experiments. In a first step towards the exploration of the role of gross income, I regress the compliance rate on gross income. The estimate from this regression is slightly negative, but marginally significant (p-value: 0.095). In order to explore whether the level of endowment is relevant for the treatment effects, I then interact the level of gross income with each of the treatment dummies. The results, presented in Table 2.7 in the Appendix, reveal that the coefficients on the interaction variables are always very close to zero and never statistically significant. This suggests that the effect treatment effect of tax revenue is unlikely to depend on the level of gross income.

2.4 Discussion and concluding remarks

While the literature sometimes speculates that the type of tax money recipient affects compliance decisions, evidence on this question is surprisingly scarce. This paper aims to contribute to filling this gap and examines if lab experimental compliance decision depend on the nature of tax expenditures. I run a "typical" tax compliance lab experiment and solely vary how generated tax revenues are spent: 1) Redistribution among all subjects in a session, 2) transfer to the experimenters' research fund, 3) donation to the German Red Cross and 4) transfer to the general budget of the federal government. These treatment groups reflect set-ups that have previously been used in the experimental compliance literature as well as a realistic scenario where tax money is transferred to the federal government.

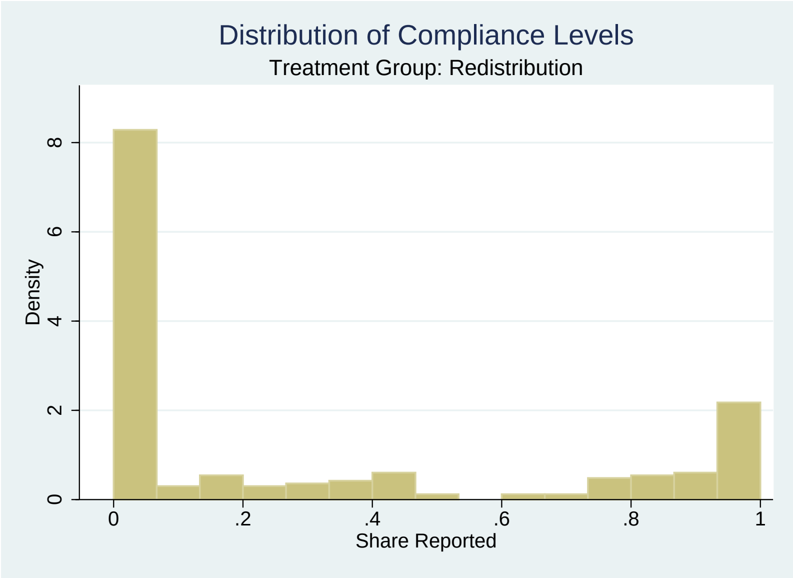
The experimental results indicate that average compliance levels are higher in the groups in which tax revenue is spent for research and charity purposes (Red Cross). These effects are fairly sizable. For example, compliance in the "Research" and "Red Cross" groups is about 40% and 35% higher than in the group with redistribution among subjects. Non-parametric and parametric (regression) analyses, however, depict that these differences are not precisely estimated and do not differ from zero in a statistical sense. These results suggest that the type of tax recipient is likely to matter for lab experimental compliance decisions, but the present experimental data do not allow the statistical rejection of the null hypothesis of no treatment effect. Although the effects are imprecisely estimated, the large differences in means along with p-values fairly close to conventionally defined significance yet show that future lab experiments should be aware that the type of recipient may make a difference. While it cannot be ruled out that other types of tax revenue use would make a larger and more precisely estimated difference, my paper depicts that the most common types in the literature and a realistic transfer to the Government *may* affect compliance behavior. The imprecision in the estimated differences may to some extend be due to the relatively low number of observations. Follow-up studies that study the present research question with a larger sample may be promising and shed more light on my results.

The paper's results are in line with the large literature on pro-social donation behavior in that they provide further evidence that the type of recipient is likely to matter for pro-social behavior. The low compliance level in the group with redistribution among anonymous subjects along with the high compliance level in the "Red Cross" group is particularly complementary to Eckel and Grossman (1996) who find that transfers in dictator games are substantially higher when the recipient is the Red Cross, relative to a situation with anonymous recipients. A general conclusion from this may be that the economics of tax compliance might be able to benefit from the studies on charitable giving and vice versa.

The reasons for higher compliance levels in the "Research" and "Red Cross" groups are speculative. The extent to which subjects directly benefit from their tax payments does not seem to make a difference: tax compliance is higher in groups without a direct link between tax payment and a subject's own benefit. However, given that the government redistributes the vast share of tax revenues, the results may indicate that taxpayers prefer using tax money for charitable donations and giving – to research and the Red Cross – over the redistributive use of tax money – as in groups "Redistribution" and "Government". The average compliance rate in groups 1 and 4 where tax payments are used for redistributive purposes is 0.326, whereas it is 0.417 in the other two groups where tax payments are donated. Another explanatory factor may be concrete knowledge about how the respective recipient spends the allocated tax money. Subjects are familiar with the type of research that is conducted at their university, and they are likely to have a clear perception about donation use of the Red Cross. In contrast, since government money is transferred to the large general fund, subjects in this group may lack information about the concrete usage of their money. As suggested by the literature on charitable giving, anonymity may be among the reasons as well. Whereas subjects are familiar with researchers and their work through lectures and seminars, it is unlikely that subjects knew other lab experiment participants. There likely is not a personal relationship to the Red Cross either, but it is a large charity organization and their work and representatives are fairly well-known. A further reason may be rooted in differing opinions on the reliability and efficiency of the recipients. In particular, public government finances are often perceived to be inefficient and wrongly targeted in large Continental European welfare states (Algan et al. 2014)

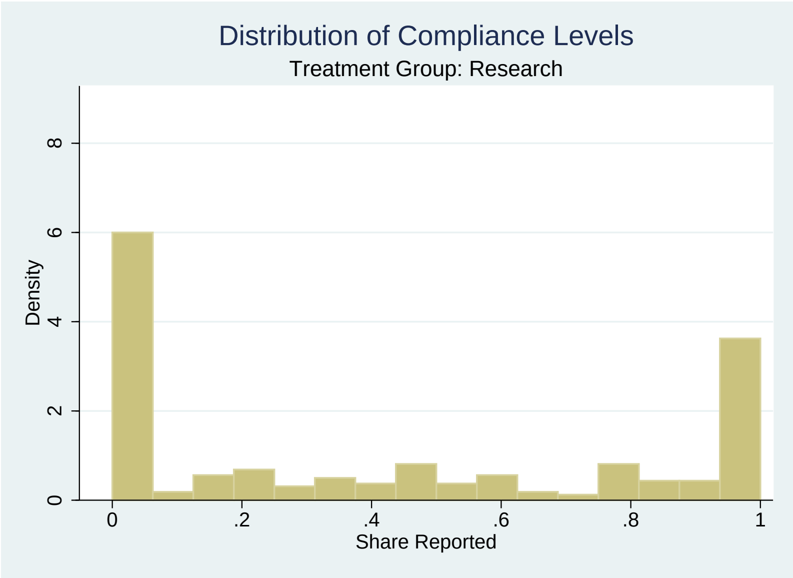
2.5 Appendix A: Tables and Figures

Figure 2.2: Distribution of Compliance Rates: Group Redistribution



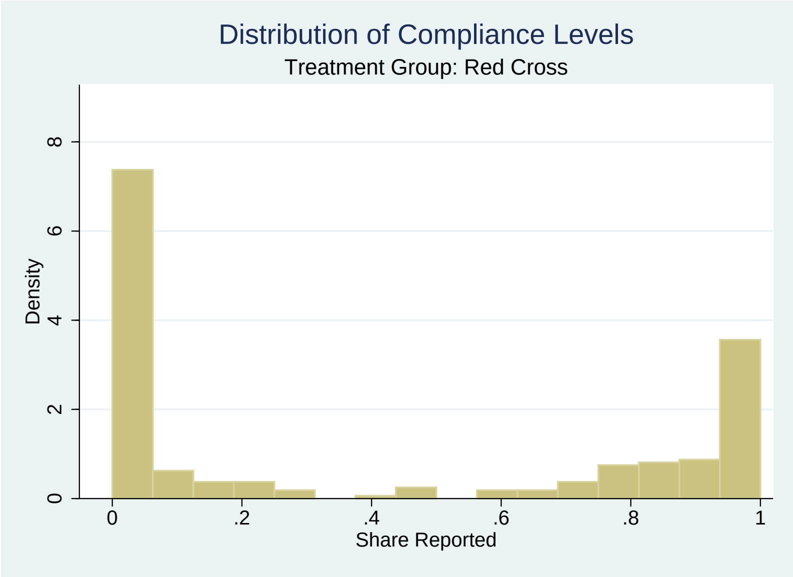
Notes: Histogram of Compliance Rates. Compliance rate is defined as reported income divided by gross income.

Figure 2.3: Distribution of Compliance Rates: Group Research



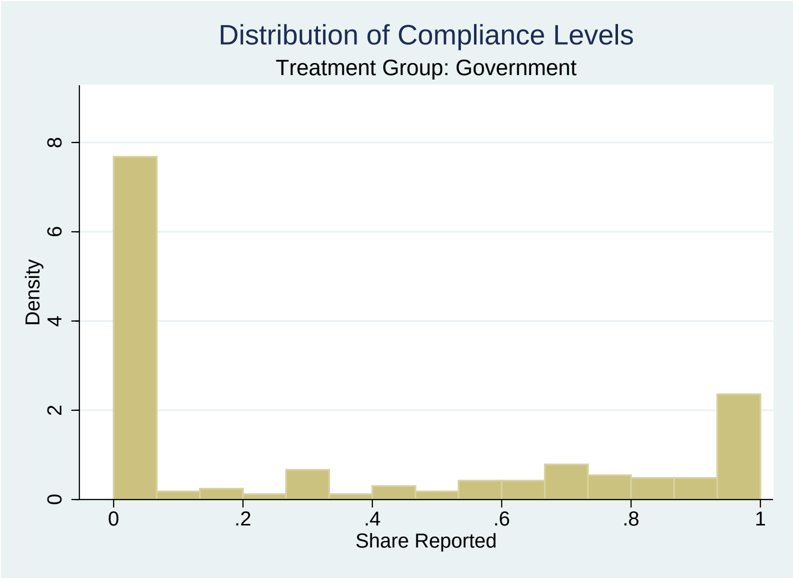
Notes: Histogram of Compliance Rates. Compliance rate is defined as reported income divided by gross income.

Figure 2.4: Distribution of Compliance Rates: Group Red Cross



Notes: Histogram of Compliance Rates. Compliance rate is defined as reported income divided by gross income.

Figure 2.5: Distribution of Compliance Rates: Group Government



Notes: Histogram of Compliance Rates. Compliance rate is defined as reported income divided by gross income.

Table 2.5: Panel Regressions: Effect of treatment on compliance (detailed results)

	(I)	(II)	(III)	(IV)
Reference group: Treatment Redistribution				
Research	0.123	0.123	0.140	0.140
	(0.090)	(0.090)	(0.088)	(0.088)
Red Cross	0.106	0.106	0.102	0.102
	(0.095)	(0.095)	(0.093)	(0.093)
Government	0.047	0.047	0.018	0.018
	(0.093)	(0.094)	(0.097)	(0.097)
Age			0.004	0.004
			(0.006)	(0.006)
Male			-0.135*	-0.135*
			(0.070)	(0.070)
German			-0.032	-0.032
			(0.101)	(0.102)
Tax Morale			0.010	0.010
			(0.013)	(0.013)
Risk			-0.008	-0.008
			(0.055)	(0.055)
constant	0.302***	0.327***	0.217	0.241
	(0.062)	(0.065)	(0.207)	(0.210)
Period FE		X		X
N	1008	1008	1008	1008
R2	0.014	0.019	0.046	0.051

Notes: Pooled OLS panel regressions based on equation 2.4. Standard errors in parentheses clustered by individuals. Estimates are based on a sample of 126 individuals and eight periods. Significant levels are $* < 0.10$, $** < 0.05$, $*** < 0.01$. Dependent variable: Compliance rate defined as income reported for tax purposes divided by gross income. Independent variables of interest are the dummy variables for the treatment groups "Research", "Red Cross" and "Government". Coefficients of interest are relative to the omitted treatment group "Redistribution" (in "constant"). Treatment groups differ with respect to the usage of tax revenue. Specifications (II) and (IV) include period fixed effects. Specifications (III) and (IV) include control variables for age, gender, language, tax morale and risk attitude.

Table 2.6: Tobit Regressions: Effects of treatment on compliance

	(I)	(II)	(III)	(IV)
Reference group: Treatment Redistribution				
Research	0.299	0.300	0.332	0.332
	(0.217)	(0.217)	(0.211)	(0.211)
Red Cross	0.213	0.213	0.199	0.200
	(0.233)	(0.233)	(0.228)	(0.228)
Government	0.086	0.086	-0.007	-0.006
	(0.232)	(0.232)	(0.239)	(0.239)
Age			0.015	0.015
			(0.014)	(0.014)
Male			-0.351**	-0.351**
			(0.174)	(0.174)
German			-0.177	-0.178
			(0.226)	(0.226)
Tax Morale			0.026	0.026
			(0.032)	(0.032)
Risk			0.021	0.021
			(0.130)	(0.129)
constant	0.025	0.074	-0.261	-0.216
	(0.161)	(0.167)	(0.515)	(0.517)
Period FE		X		X
N	1008	1008	1008	1008
Pseudo R2	0.006	0.008	0.026	0.028

Notes: Tobit panel regressions based on equation 2.4. Standard errors in parentheses clustered by individuals. The tobit regressions account for censoring in the compliance level at zero and one. Estimates based on a sample of 126 individuals and eight periods. Significant levels are $* < 0.10$, $** < 0.05$, $*** < 0.01$. Dependent variable: Compliance rate defined as income reported for tax purposes divided by gross income. Independent variables of interest are the dummy variables for the treatment groups "Research", "Red Cross" and "Government". Coefficients of interest are relative to the omitted treatment group "Redistribution". Treatment groups differ with respect to the usage of tax revenue. Specifications (II) and (IV) include period fixed effects. Specifications (III) and (IV) include control variables for age, gender, language, tax morale and risk attitude.

Table 2.7: Panel Regressions: Heterogeneity w.r.t. gross income

	(I)	(II)	(III)	(IV)
Reference group: Treatment Redistribution				
Research	0.237	0.096	0.254*	0.114
	(0.147)	(0.191)	(0.144)	(0.189)
Red Cross	0.187	0.047	0.183	0.042
	(0.117)	(0.170)	(0.113)	(0.166)
Government	0.023	-0.117	-0.006	-0.146
	(0.117)	(0.170)	(0.123)	(0.174)
Research × Inc	-0.002	0.000	-0.002	0.000
	(0.002)	(0.003)	(0.002)	(0.003)
Red Cross × Inc	-0.001	0.001	-0.001	0.001
	(0.001)	(0.002)	(0.001)	(0.002)
Govn't × Inc	0.000	0.003	0.000	0.003
	(0.001)	(0.003)	(0.001)	(0.003)
Age			0.004	0.004
			(0.006)	(0.006)
Male			-0.135*	-0.135*
			(0.070)	(0.070)
German			-0.032	-0.032
			(0.101)	(0.102)
Tax Morale			0.010	0.010
			(0.013)	(0.013)
Risk			-0.008	-0.008
			(0.055)	(0.055)
constant	0.302***	0.339***	0.217	0.253
	(0.062)	(0.070)	(0.207)	(0.211)
Period FE		X		X
N	1008	1008	1008	1008
R2	0.014	0.019	0.047	0.051

Notes: Pooled OLS panel regressions. Standard errors in parentheses clustered by individuals. Estimates are based on a sample of 126 individuals and eight periods. Significant levels are $* < 0.10$, $** < 0.05$, $*** < 0.01$. Dependent variable: Compliance rate defined as income reported for tax purposes divided by gross income. Independent variables of interest are the dummy variables for the treatment groups "Research", "Red Cross" and "Government", as well as the interaction variables between these treatment dummies with the level of gross income ("Inc"). Coefficients of interest are relative to the omitted treatment group "Redistribution" (in "constant"). Treatment groups differ with respect to the usage of tax revenue. Specifications (II) and (IV) include period fixed effects. Specifications (III) and (IV) include control variables for age, gender, language, tax morale and risk attitude.

2.6 Appendix B: Instructions

The following pages display the English translation of the instructions (the original German version is available upon request). The instructions for the four treatment groups are almost identical and only differ with respect to the information about tax revenue use. These difference are indicated in the following instructions.

Instructions

Welcome and thank you for participating in our experiment. From now on until the end of the experiment, please refrain from communicating with other participants. If you do not abide by this rule, we will have to exclude you from the experiment.

We kindly ask you to read the instructions thoroughly. If you have any questions after reading the instructions or during the experiment, please raise your hand and one of the instructors will come to you and answer your question in person. Your payment and your decisions throughout the experiment will be treated confidentially. None of the other participants is informed, neither during nor after the experiment, about your decisions in the experiment or your payment.

You can earn money in this experiment. During the experiment, your payments will be calculated in a virtual currency: Experimental Currency Units (ECU). 5 **ECU corresponds to 1 EUR.** After the experiment, your pay-off will be converted to Euro and given to you in cash. Additionally, you will receive a show-up fee of 2.50 EUR.

The Experiment

Overview

The experiment consists of one practice round and **eight** payoff rounds. You cannot earn money in the practice round. Each round has two stages. In the first round, you receive an income. This income is subject to a tax of 30% and in the second stage of the experiment you are asked to make an income reporting decision for tax purposes. There is a random chance of 10% that your reporting decision is checked. If your reporting decision is checked and you have not reported the true amount of income, you will have to make an extra payment.

[Treat 1] The tax revenue generated in the experiment is equally redistributed among all participants of the experiment.

[Treat 2] The tax revenue generated in the experiment is used for future research projects of the experimenters.

[Treat 3] The tax revenue generated in the experiment is donated to the German Red Cross.

[Treat 4] The tax revenue generated in the experiment is transferred to the German federal government.

The details of the experiment are described in the following.

Proceeding of a round

Stage 1: Gross income

In the beginning of each round, you receive an income. This income is subject to a tax and is therefore called gross income. The level of gross income may vary from round to round, but is always between 44 and 65 ECU. The level of gross income is the same for all subjects in a given round.

Stage 2: Income reporting decision

The gross income you receive in stage 1 is subject to a tax rate of **30%**. This tax rate is the same throughout the entire experiment.

You are now asked to make an income reporting decision for tax purposes and report the level of gross income that you received. You report an amount that shall be taxed at the tax rate of 30%. This reported amount can be between zero and your full gross income.

After you have completed the tax reporting decision and confirmed through pressing "OK", the computer determines whether it is checked if the reported amount equals the true amount of gross income. To do so, the computer randomly selects a natural number between 1 and 10. Your decision is **only** checked if the number **1** is selected. That means there is a random chance of 10% that the reporting decision is checked. The experimental investigators of course cannot see whether or not you reported your full gross income.

Calculation of net income

Your payment at the end of a round is called net income.

After your tax reporting decision you are shown which number was randomly selected by the computer. Based on this random selection of a number you are faced with one of the following two scenarios for your net income:

Scenario a): The computer selects the **number 1**:

Your reporting decision *will be checked* to determine whether you reported your full gross income for tax purposes. Depending on your previous decision, there are two possibilities for your net income:

- **Too little gross income reported:**

  If your reported income is *lower* than your full gross income, then you will have to pay the tax liability based on your full gross income and an extra payment. This extra payment is the higher the lower the reported amount is. The extra payment equals twice of the underreported tax liability. Hence:

  → Net income = gross income – (gross income x 0.30) – [2 x 0.30 x (gross income – reported income)]

- **True gross income reported:**

If your reported income *equals* your full gross income, then your net income consists of your gross income less your tax liability. Hence:

→ Net income = gross income – (gross income x 0.30)

Scenario b): The computer selects a **number between 2 and 10**

Your reporting decision *will not be checked* to determine whether you reported your full gross income for tax purposes. Your net income, in this case, consists of your gross income less the tax payment. The tax payment depends on your income reporting decision. The tax payment is the reported income multiplied with the tax rate of 0.30. Hence:

→ Net income = gross income – (reported income x 0.30)

Use of tax revenue

Your tax payment in a given round is calculated as gross income minus your net income. Hence:

→ Tax payment = gross income – net income

That is, the amount of tax payment depends on the level of gross income, the amount you reported for tax purposes and whether your reporting decision was checked.

The tax payments of all participants sum up to tax revenue that is generated in this experiment.

[Treat 1] The overall tax revenue, plus potentially paid extra payments, in a randomly chosen round are equally redistributed among all participants in the laboratory. In addition to your net income you hence receive a share of total tax revenue. This share is calculated by summing up the tax payments of all participants and then dividing by the number of participants.

[Treat 2] The overall tax revenue, plus potentially paid extra payments, in a randomly chosen round are used for research purposes. The tax payments are used by the experimenters for future research projects.

[Treat 3] The overall tax revenue, plus potentially paid extra payments, in a randomly chosen round are donated to the German Red Cross. The donation to the Red Cross is not targeted at a specific aid project.

[Treat 4] The overall tax revenue, plus potentially paid extra payments, in a randomly chosen round are transferred to the federal budget of the Federal Republic of Germany. The tax payments are transferred to the *Bundeskasse*, where all regular federal tax payments in Germany go. The transfer to the federal budget is not targeted at a specific purpose.

Payoff

Your final payoff, which will be paid to you at the end of the experiment, does **not** consist of the sum of all net incomes in the eight payoff rounds. Instead, after the end of the experiment a round will randomly be selected based on which you are paid. This selected payoff relevant round is identical for all participants.

After the end of the eighth payoff round, the computer randomly determines the payoff relevant round. Each of the eight rounds can be selected with the same probability. Your final payoff is based on your net income in the selected round. That is, if, for example, number "2" is selected, your final payoff consists of your net income in round two. The selected net income is converted to EUR and paid to you in cash. In addition, you receive a show-up fee of EUR 2,50. The use of tax revenue also depends on the tax payments made in the same randomly selected round.

Final Remarks

At the end of the experiment, you will be asked to complete a short questionnaire while we prepare the payments. All information collected through this questionnaire, just like all data gathered during the experiment, are anonymous and exclusively used for scientific purposes.

Chapter 3

Tax incidence in the presence of tax evasion

3.1 Introduction

The standard textbook theory of taxation makes the prediction that the economic incidence of a tax solely depends on the relative elasticity of demand and supply. The tax burden will fall disproportionately on consumers if demand is more inelastic than supply and more on producers if supply is more inelastic than demand. However, there is evidence that the incidence of a tax also depends on other factors such as tax salience (Chetty 2009) and remittance policy (Slemrod 2008).[1]

Another factor likely to drive the incidence of taxes is the prevalence of tax evasion opportunities. Intuitively, access to tax evasion allows taxpayers to lower their effective tax rate by reducing their tax base. As a result, the “real" behavioral responses that determine tax incidence are likely to differ between taxpayers who can evade and those who cannot.[2] Understanding this possible source of deviation between observed and standard theoretical economic incidence is important given the prevalence of tax evasion in both developed and developing countries (Slemrod 2007; Schneider et al. 2010; Kleven et al. 2011). Although the impact of tax evasion on incidence has intuitive appeal and is policy relevant, theoretical evidence is mixed (Marrelli 1984; Yaniv 1995; Bayer and Cowell 2009), and there is very little empirical evidence.

The objective of this paper, therefore, is to contribute empirical evidence on the effect of tax evasion on tax incidence. The specific research question is, do sellers with tax evasion opportunities trade their goods and services at different prices than they otherwise would if no tax evasion opportunities existed? In other words, are equilibrium

[1] The literature is reviewed in detail further below.

[2] By real response we are referring to behavioral changes that directly affect decisions such as leisure, consumption or production (Slemrod 1995).

prices different in markets where evasion is an option relative to markets without evasion opportunities? We design an economic laboratory experiment with between-subject variation in which subjects trade fictitious goods in a double auction. Subjects are randomly assigned roles as sellers or buyers in treatment and control groups. A per-unit tax is imposed on all sellers. In the treatment group, sellers are able to under-report the number of units sold, whereas in the control group the tax liability is paid automatically (as with exact withholding). The only difference between the treatment and control group is that sellers in the treatment group can evade the sales tax. Therefore, we identify the impact of evasion on tax incidence by comparing the equilibrium prices in the treatment group with that in the control group.

The experimental laboratory has been used extensively to study the economic incidence of taxes. In fact, various studies have found that the theoretical results of tax incidence – without evasion – hold in competitive experimental markets such as a double auction (Kachelmeier et al. 1994; Borck et al. 2002; Ruffle 2005). We therefore introduce tax evasion to an environment that has been shown to provide credible results in the context of tax incidence.[3] Relying on the controlled environment of the laboratory also means that we are able to avoid much of the econometric problems of observational data analyses and thus produce clean identification of the treatment effect. Achieving causal identification using observational data requires random variation in access to evasion across otherwise similar markets. This is obviously difficult to achieve since access to tax evasion is most likely one of the dimensions of a market that determines whether buyers and sellers select to participate in that market.

We place our empirical question in a theoretical framework based on the standard textbook partial equilibrium approach to tax incidence. The model shows that access to tax evasion changes the incidence of the tax. More specifically, an increase in the statutory tax rate leads to a smaller increase in the market price if sellers are non-compliant relative to a market with full compliance; this impact on incidence is increasing in the non-compliance rate. The simple reason for this result is that sellers with an evasion option are able to reduce their effective tax rate relative to those without evasion. As a result, the tax causes the supply curve to shift up by less in the case with evasion. In our specific context, a per-unit tax on sellers who can evade taxes reduces the share of the statutory tax burden that is passed on to buyers.

Our empirical results confirm the theoretical prediction. In particular, the equilibrium price in the treatment group with tax evasion is statistically and economically lower than in the control group. Accordingly, the number of units traded is higher in the case with evasion. By under-reporting sales, sellers reduce their effective tax burden, which

[3]We employ an experimental double auction similar to Grosser and Reuben (2013). Riedl (2010) provides an overview of experimental tax incidence research.

allows them to sell at lower prices. Although the statutory burden on buyers is lower in the presence of tax evasion, we find evidence that sellers shift the full expected effective tax rate onto buyers. As expected, the smaller impact on the market equilibrium results in a much lower partial equilibrium excess burden in the presence of tax evasion.[4]

Addressing the research question posed in this paper makes several important contributions to the literatures on tax incidence, tax salience and tax evasion. First, several studies have attempted to identify the incidence of taxes using observational data. For example, Alm et al. (2009) and Marion and Muehlegger (2011) find that the incidence of the fuel tax in the US is fully shifted to final consumers and related to supply and demand conditions, Saez et al. (2012) find that tax equivalence does not hold in the context of the Greek payroll tax, and Fuest et al. (2013) find that the burden of local business taxes in Germany partly falls on employees via lower wages.[5] To overcome the challenges of identifying causal effects using observational data, several studies explore the question of economic incidence in a laboratory setting. For example, Kachelmeier et al. (1994), Quirmbach et al. (1996), Borck et al. (2002), and Ruffle (2005) find that the theoretical predictions of tax incidence hold true in a competitive laboratory market with full information.[6] We add to this strand of the literature by introducing tax evasion to a standard competitive experimental double-auction market and show that this changes the incidence of the tax. This finding is important because it suggests that tax equivalence, which is the focus of the existing laboratory tax incidence literature, is unlikely to hold in the real world where buyers and sellers have different access to evasion (Slemrod 2008).

Two studies more closely related to ours in that they estimate economic incidence in the presence of tax evasion are Alm and Sennoga (2010) and Kopczuk et al. (2013). The latter provides empirical evidence that the stage of production at which the tax on diesel is collected in the US affects the economic incidence of the tax. Although they suggest that this difference is driven by variation in access to evasion across production stages, reliance on observational data makes it difficult to cleanly identify whether this effect is fully due to variation in compliance behavior. Alm and Sennoga (2010) use a computable general equilibrium (CGE) model to simulate the economic incidence of tax evasion for a “typical" developing country. They find that the benefits of evasion generally do not stay with the evader if there is free entry, which suggests that evasion changes the incidence

[4]Tax evasion opportunities also have a negative effect on generated tax revenue; revenue is higher in the control than in the treatment group. However, the foregone tax revenue represents a transfer to private agents and does not affect welfare as long as one does not impose a welfare function that gives higher weight to tax financed public goods relative to private consumption.

[5]Other examples of observational tax incidence studies include Evans et al. (1999), Gruber and Koszegi (2004), and Rothstein (2010).

[6]Kerschbamer and Kirchsteiger (2000) and Riedl and Tyran (2005) find that the laws of tax incidence do not translate to non-competitive experimental market.

of taxes. Since we rely on the controlled environment of the lab, our empirical approach provides precise control over the market institutions, which allows us to randomize access to evasion and measure non-compliance accurately. As a result, we are able to offer cleaner identification of the impact of tax evasion on the economic incidence of the tax than these two studies. Nonetheless, we view our work as complementary to these papers. The illusive nature of tax evasion implies that consistent results across multiple techniques is required if we are to draw firm conclusions about causes and consequences of tax evasion. We argue that our results, combined with Kopczuk et al. (2013) and Alm and Sennoga (2010), provide evidence that the standard textbook model of tax incidence does not hold up in many real world applications.

Second, our paper is related to the tax salience literature, which shows that the standard theoretical incidence result may not hold if taxes are not salient. For example, Chetty (2009) show in the field that people under-react to taxes that are added at the register relative to taxes that are shown in the displayed price of goods. Similarly, Sausgruber and Tyran (2005) find in a lab experimental context that indirect taxes exhibit a different incidence than direct taxes because indirect taxes are less visible.[7] Similar to the salience effects described in Chetty (2009), we provide evidence that the willingness and opportunity to evade taxes dampens "real" behavioral responses to taxation. Taxpayers who evade taxes do not adjust their economic behavior in response to changes in the tax rate by as much as if evasion was not an option. The dampening effect on "real" behavioral responses changes the incidence of the tax just as if the tax was not salient to the taxpayer.

Finally, our paper adds to the general tax evasion literature. Naturally, obtaining credible causal evidence in the context of tax evasion is very difficult using observational studies (Slemrod and Weber 2012). A broad strand of literature has therefore employed lab experiments to study evasion (e.g., Fortin et al. 2007; Alm et al. 2009; Balafoutas et al. 2014).[8] However, unlike most of the tax evasion literature, we focus on the implications of tax evasion (e.g., Doerrenberg and Duncan 2014b) rather than on explaining tax evasion (e.g., Alm 2012). In particular, like Doerrenberg and Duncan (2014b), we show that real responses to taxes are small in part because of income shifting responses such as tax evasion.

Our findings also have important policy implications related to the distribution of tax burden and the effectiveness of tax policy as a tool for influencing behavior. First, understanding the impact of tax evasion on the economic incidence of taxes is important for the correct evaluation of the distributional effect of tax policies aimed at reducing tax

[7] Further examples of salience effects in the field are Finkelstein (2009) and Gallagher and Muehlegger (2011). Abeler and Jaeger (2014) and Blumkin et al. (2012) study questions related to salience effects in a laboratory setting.

[8] Andreoni et al. (1998) and Torgler (2002) provide surveys on tax compliance in experiments.

evasion. The results we present here further suggest that accounting for tax evasion in incidence studies may lead to a re-evaluation of the progressivity/regressivity of various taxes. For example, a sales tax where the benefits of evasion stay with sellers is likely to be more regressive than one where the benefits are shifted to buyers, especially if the evading sellers sell mostly to lower income individuals. This finding is important for empirical analysis of the distributional implications of tax reform proposals.

Second, taxes aimed at influencing real behavior are likely to be less effective if the market participants responsible for remitting the tax have access to tax evasion opportunities. Because the effective tax rate is lower among evaders, “real" behavioral responses to the tax are dampened, which limits the ability of the tax to achieve desired behavioral outcomes. For example, a number of states in the U.S. are now considering the adoption of road mileage user-fees as a replacement for fuel taxes (Duncan and Graham 2013). The salience of road mileage user-fees suggests that they are likely to reduce vehicle miles traveled. According to our results, this desirable outcome is unlikely to occur if mileage user-fees are administered in ways that facilitate tax evasion via odometer tampering, say. More generally, to the extent that tax evasion cannot be fully eliminated, our findings suggest that it might be optimal to apply higher tax rates to goods sold in markets with evasion opportunities (e.g., Cremer and Gahvari 1993). Not only could this be more efficient, but it might also achieve the desired adjustments in behavior.

The relevance and importance of these policy implications is especially obvious when one considers the prevalence of tax evasion across the world (Slemrod 2007; Schneider et al. 2010; Kleven et al. 2011). Transaction taxes, which we focus on in our study, are of particular interest in this context. For example, sales tax gap estimates range from 2 percent to 41 percent for the value added tax in the European Union and 1 percent to 19.5 percent for the retail sales tax in the United States (see Mikesell 2014 for a review of sales tax evasion estimates). Additionally, it is generally accepted that ‘use’ tax evasion by both businesses and individuals is much higher than retail sales tax evasion; e.g., GAO (2000) assume non-compliance rates of 20 to 50 percent among businesses and 95 to 100 percent among individuals in a study of the potential revenue losses of e-commerce.[9]

Therefore, our results are relevant in countries such as the United States where, for example, a number of states have adopted, or are in the process of adopting, legislation aimed at (a) restricting the sale of “zappers", which are used to evade sales taxes, and (b) requiring online traders to register as sales tax collectors. Our findings suggest that such measures are likely to result in higher prices as affected sellers fully adjust to the retail sales tax. While we focus on sales taxes here, the findings also suggest that other anti-tax evasion initiatives, such as the Foreign Account Tax Compliance Act (FATCA), are likely

[9]Consumers in the United States are required to pay ‘use’ tax in lieu of the retail sales tax if the seller is not required – by law – to register as a tax collector in the consumers’ state.

to affect the level of economic activity as affected parties respond to the reduced evasion opportunities.

The remainder of the paper is as follows. We describe the experimental design in section 3.2, the theoretical framework in section 3.3 and the results in section 3.4. Our findings are discussed in section 3.5 and section 3.6 concludes.

3.2 Experimental Design

3.2.1 Overview

The experimental design reflects a standard competitive experimental double auction market as pioneered by Smith (1962).[10] The auction and the parameters in our experiment are based on Grosser and Reuben (2013). In each round of the double auction market, 5 buyers and 5 sellers trade two units of a homogeneous and fictious good. Sellers are assigned costs for each unit and buyers are assigned values. The roles of sellers and buyers as well as the costs and values are exogenous and randomly assigned to the lab participants. We impose a per-unit tax on sellers to this set-up and give sellers in the treatment group the opportunity to evade the tax whereas sellers in the control group pay the per-unit tax automatically (as with exact withholding). We employ a between-subjects design where each participant is either in the control or treatment group. Further details on the experimental design are provided in the next subsections.

3.2.2 Organization

The experiment was conducted in the Cologne Laboratory for Economic Research (CLER), University of Cologne, Germany. A large random sample of all subjects in the laboratory's subject pool of approximately 4000 persons was invited via email – using the recruitment software *ORSEE* (Greiner 2004) – to participate in the experiment. Participants signed up on a first-come-first-serve basis. Neither the content of the experiment nor the expected payoff was stated in the invitation email. The experiment was programmed utilizing *z-tree* software (Fischbacher 2007). We ran eight sessions over two regular school days in November and December 2013.[11] Each session consisted of either a control or treatment group market and lasted about 100 minutes (including review of instructions and payment of participants).

[10] Double auction markets mimic a perfectly competitive market. Dufwenberg et al. (2005), for example, rely on an experimental double auction to study financial markets. Holt (1995) provides an overview.

[11] There are two regular semesters at the tertiary level in Germany; winter semester lasting from October to March and summer semester between April and July. Therefore, the experiment was implemented during the regular semester.

We conduct four control and four treatment sessions for a total of 80 subjects.[12] Experimental Currency Units (ECU) are used as the currency during the experiment. After the experiment, ECU are converted to Euro with an exchange of 30 ECU = 1 EUR and subjects are paid the sum of all net incomes (see below) in Euro. It was public information that all tax revenue generated in the experiment would be donated to the German Red Cross.

At the beginning of each session, subjects are randomly assigned to computer boothes by drawing an ID number out of a bingo bag upon entering the lab. The computer then randomly assigns each subject to role as buyer or seller, as well as her costs or values which stay constant during the experiment. Subjects are given a hard copy of the instructions when they enter the lab and are allowed as much time as needed to familiarize themselves with the procedure of the experiment. They are also allowed to ask any clarifying questions.

3.2.3 Description of a session

Each session includes 1 market that is either a control or treatment group market. Each market has five buyers and five sellers who each have 2 units of a fictitious good to trade. Sellers and buyers are randomly assigned costs and values for both of their units; the roles as buyer or seller and the assigned values and costs are exogenously determined and stay constant for the entire experiment. All ten subjects in one session/market first trade in 3 practice rounds and then 27 payoff relevant rounds.

Trade in the Double Auction As is common in experimental markets, subjects are given demand and supply schedules for a fictitious good at the beginning of the session (Ruffle 2005; Cox et al. 2012; Grosser and Reuben 2013). The demand schedule for buyers assigns a value to each of two items and the supply schedule for sellers assigns a cost to each of two items. The cost/value of the units vary across items and subjects as illustrated in Table 3.1. This allows us to induce demand and supply curves for each market, which are depicted in Figure 3.1. The schedules are chosen such that demand and supply elasticities are equal in equilibrium. The demand and supply schedules remain fixed across periods in a given session, and they do not differ between control and treatment markets.

Subjects trade the good in a double auction market that is opened for two minutes in each period. During this time, each seller can post an "ask" that is lower than the current ask on the market, but higher than the cost of the item to the seller. In other

[12] While this appears to be low, note that Grosser and Reuben (2013), for example, only implemented four markets and yet have sufficient statistical power to identify a treatment effect. See section 3.4.2 for summary statistics on demographic characteristics of the participants.

words, sellers cannot trade an item below its cost. Additionally, sellers must sell their cheaper unit before they sell their more expensive unit. Similarly, each buyer can post a "bid" that is higher than the current bid on the market, but lower than the value of the item to the buyer. Therefore, buyers cannot buy an item at a price that exceeds its value. Buyers must also buy their most valued item before their least valued item. The lowest standing ask and the highest standing bid are displayed on the computer screen of all ten market participants.[13]

An item is traded if a seller accepts the standing buyer bid or a buyer accepts the standing seller ask. Subjects are not required to trade a minimum amount of items, items that are not traded yield neither costs nor profits. Traders are not allowed to communicate with each other. This trading procedure is identical for the treatment and control groups.

Income: Control Group Gross-income in each period consists of the sum of the profit on each unit traded. Sellers' gross profit on each unit is equal to the difference between the selling price and cost, while buyers' profit on each unit is the difference between value and price paid. All subjects (buyers and sellers) are told that sellers have to pay a per-unit tax for each unit sold, that the tax rate is fixed across all periods at $\tau = 10$ ECU per-unit and that the tax is collected at the end of every third trading period. In other words, subjects complete three rounds of trading then tax is collected from sellers, then three more rounds of trading then another tax collection and so on. This yields 27 trading periods and 9 tax collections; we discuss this design feature below. We define total gross profit in each trading period i ($i = 1, 2, 3, ..., 25, 26, 27$) as

$$\Pi_i^s = P_{i1}d_1 + P_{i2}d_2 - C_1d_1 - C_2d_2, \tag{3.1}$$

for sellers and

$$\Pi_i^b = V_1d_1 + V_2d_2 - P_{i1}d_1 - P_{i2}d_2, \tag{3.2}$$

for buyers. Superscripts s and b indicate seller and buyer, respectively, $d_j = 1$ if good j is traded and 0 otherwise, P_{ij} is the price of good j in period i, C_j is the cost of good j and V_j is the value of good j.

Because taxes are collected at the end of every third trading period, a seller's net income for each tax collection period k ($k = 3, 6, 9, 12, 15, 18, 21, 24, 27$) is equal to:

$$\pi_k^s = \Pi_k^s + \Pi_{k-1}^s + \Pi_{k-2}^s - \tau U, \tag{3.3}$$

where U is the total number of units sold in the last three rounds and $\tau = 10$ is the

[13] Figure 3.8 in the appendix depicts a screenshot of the experimental market place for a seller in the treatment group with evasion opportunity.

statutory per-unit tax rate. Because buyers do not pay a tax, their net income for each tax collection period may be written as:

$$\pi_k^b = \Pi_k^b + \Pi_{k-1}^b + \Pi_{k-2}^b \tag{3.4}$$

Both buyers and sellers are shown their gross income after every trading period and their net income after every tax collection period. Subjects' final payoff is the sum of their net incomes from the nine tax collection periods.

Income: Treatment Group Since buyers do not pay the tax, the calculation of gross and net income for buyers in the treatment group is identical to that of the control group: see equations (3.2) and (3.4). Sellers, on the other hand, make a tax reporting decision at the end of every third round. In other words, subjects complete three rounds of trading then sellers make a reporting decision; then three more rounds of trading then another reporting decision and so on.

One advantage of allowing subjects to report after every third trading period is that it increases the probability that every subject has a positive amount to report and must therefore explicitly decide if they wish to under-report sales for tax purposes. Another advantage of using every third round is that it yields 9 reporting decisions. This is advantageous because it means that subjects can learn the implications of tax evasion for their profits and update their beliefs about the probability of being caught. As a result, we can be assured that the market equilibrium in the evasion treatment reflects the impact of tax evasion on the behaviour of market participants. Although reporting every period would maximize the number reporting decisions, we opted against this option because excess supply in the market implies that some subjects will sell zero units in a given trading period, which trivializes the reporting decision. Another option is to have subjects make a single reporting decision at the end of the experiment. While this approach maximizes the chance that everyone has a positive amount to report, having a single reporting decision would not allow subjects to learn or update their beliefs. We opted for every third round as a reasonable compromise between these two extremes.[14]

Sellers can report any number between 0 and the true amount sold in the previous three trading periods, and the reported amount is taxed at $\tau = 10$ ECU per-unit. Sellers face an exogenous audit probability of $\gamma = 0.1$ (10 percent) and pay a fine, which is equal to twice the evaded taxes if they underreport sales and are audited. The tax rate, audit probability, and fine rate are fixed across periods and sessions, and all subjects – buyers and sellers – in the treatment group receive this information at the beginning of the experiment.

[14]Although subjects in the control group do not make a reporting decision, we collect taxes and report their net profits at the end of every third period to ensure comparability with the treatment group.

Therefore, unlike sellers in the control group who must pay taxes on each unit sold, sellers in the treatment group are able to evade the sales tax by underreporting sales. Sellers' gross income in any trading period i is the same as in equation (3.1), but their net income in each tax collection period is rewritten as:

$$\pi_k^s = \begin{cases} \Pi_k^s + \Pi_{k-1}^s + \Pi_{k-2}^s - \tau R & \text{if not audited,} \\ \Pi_k^s + \Pi_{k-1}^s + \Pi_{k-2}^s - \tau U - \tau(U - R) & \text{if audited,} \end{cases} \tag{3.5}$$

where R is the reported number of units sold, U is the number of units actually sold over the last three rounds, and $\tau = 10$ is the statutory per-unit tax rate. Subjects' final payoff is the sum of their net incomes from the nine tax collection periods.

3.2.4 Market Equilibrium

The demand and supply schedules described in Table 3.1 and displayed in Figure 3.1 can be used to determine the competitive equilibrium price and quantity with and without the per-unit tax. Theoretically, we expect the market to clear with 7 units traded at any price in the range 48 ECU to 52 ECU in the case without taxes. We obtain a range of prices in equilibrium because the demand schedule is stepwise linear (Ruffle 2005; Cox et al. 2012; Grosser and Reuben 2013).[15]

A per-unit tax on sellers increases the cost of each unit by 10 ECU and thus shifts the supply curve to the left as shown in Figure 3.1. In the absence of tax evasion opportunities, this theoretically produces a new equilibrium quantity of 6 units, which is supported by an equilibrium price in the range of 53 ECU to 57 ECU. Because the demand and supply schedules have equal elasticity in equilibrium, the incidence of the tax should theoretically be shared equally between buyers and sellers; buyers pay an extra 5 ECU and sellers receive 5 ECU less (after paying the tax).

The question we seek to answer is whether this equilibrium outcome is affected by the presence of tax evasion opportunities among sellers. The next section provides a theoretical discussion for why tax evasion may or may not affect the incidence of the tax.

3.3 Theoretical Context

This section places our experimental design in the context of a simple theoretical set-up based on the textbook partial equilibrium analysis of tax incidence. It is straight forward

[15]Grosser and Reuben (2013) conducted an experiment using the same demand and supply schedule as we do and find that the "no-tax" equilibrium is equal to that predicted by the theory. Therefore, although we do not implement the "no-tax" treatment here, we expect that our "no-tax" equilibrium is in line with theoretical expectations.

to show that the standard text book results hold for the control group because they have no evasion opportunity. In particular, the incidence of the tax is determined solely by the relative elasticity of demand and supply. We want to know if and how these predictions change when the person legally responsible for remitting the tax has access to evasion opportunities. Therefore, the analysis that follows focuses on the treatment group where the opportunity to evade the tax exists. However, as shown below, the model describing the control group is nested within the model described here.

3.3.1 Buyers' and Sellers' Decision

Consider a market where buyers and sellers trade q units of a good at price p in a given trading period. Sellers have to pay a per-unit (excise) tax for each unit they sell, but are provided a tax reporting decision. As a result, the per-unit tax τ is paid on r $(0 \leq r \leq q)$, which is the number of units reported by a seller. The tax reporting decision is audited with probability γ, and because all audits lead to the full discovery of actual sales, a fine equal to twice the evaded taxes, $2(q-r)\tau$, must be paid if audited.

Buyers Buyers are never allowed an evasion opportunity so their only choice variable is the number of traded units; they simply maximize

$$\pi^B = (v - p)q \tag{3.6}$$

with respect to q, where v is the assigned value the buyer receives from each unit bought. This is maximized where the price paid for the last unit is equal to the value of that unit to the buyer $v = p$. Aggregated, this suggests that market demand is a function of price, $D(p)$.

Sellers On the other hand, sellers maximize expected profits π^S, which are given by:

$$\pi^S = (p - c)q - r\tau - 2\gamma(q - r)\tau, \tag{3.7}$$

where c is the assigned cost of each unit sold by a seller. Equation (3.7) can be rewritten as:

$$\pi^S = (p - c)q - q\underbrace{\tau(\phi + 2(1 - \phi)\gamma)}_{effective\ tax\ t_e}, \tag{3.8}$$

where $\phi = r/q$ $(0 \leq \phi \leq 1)$ is the share of actual sales that is reported for tax purposes. Sellers face an expected effective tax rate $t_e = \tau(\phi + 2(1-\phi)\gamma)$ for each unit they actually sell.[16] It is clear that the effective tax rate differs from the statutory tax rate as long

[16] For reasons of brevity, we henceforth refer to t_e as the effective tax rate, rather than *expected* effective tax rate.

as $\phi \neq 1$. In fact, the effective tax rate is lower than the statutory rate as long as sales are underreported ($\phi < 1$) and the chance of being audited is less than one half $\gamma < 1/2$. Additionally, the effective tax rate is positively related to the statutory tax rate τ ($\frac{\delta t_e}{\delta \tau} \geq 0$), the share of reported income ϕ ($\frac{\delta t_e}{\delta \phi} > 0$ for $\gamma < 0.5$) and the audit probability γ ($\frac{\delta t_e}{\delta \gamma} \geq 0$). Note that because sellers in the control group do not have an opportunity to evade the tax, their reported sales is equal to their actual sales; i.e., $r = q$, which implies that $\phi = 1$. As a result, the effective tax rate t_e is equal to the statutory tax rate τ in the control group.

Sellers chose the number of traded units q and the share of reported sales ϕ to maximize their expected profits π^S. Maximizing equation (3.8) yields the following first order conditions (F.O.C.) with respect to q and ϕ:

$$\begin{aligned} p &= c + t_e \\ \gamma &= 1/2 \end{aligned} \tag{3.9}$$

These F.O.C. define the optimum and have the usual interpretations: sellers trade until the revenue from the last unit sold is equal to its cost, and some amount of sales is underreported if the chance of being caught is no more than one half.

Note that the effective tax rate t_e equals the statutory tax rate τ when the audit probability is $\gamma = 1/2$. This implies that the optimal output q^* is defined by $p = c + \tau$, which is identical to the q^* in the control group.[17] The optimal output in the treatment group is also identical to that in the control group if sellers report honestly, which is predicted to occur if $\gamma > 1/2$. Intuitively, the effective tax rate is greater than the statutory tax rate if the audit probability is greater than one half, except when $\phi = 1$. Therefore, sellers can maximize their expected profit by reporting honestly when the chance of being audited is greater than one half. By reporting honestly in this case, sellers ensure that $t_e = \tau$, which yields the same q^* as in the control group. Therefore, the opportunity to evade the tax has no effect on sellers output decisions if the audit probability is equal or larger than 50%. This further implies that the opportunity to evade has no effect on the incidence of the tax.

However, the optimal strategy for any audit probability γ smaller than 1/2 is to fully evade. Since γ is 0.1 in our experiment, the model predicts that sellers do not report any of their sold units. With this optimal strategy of full evasion ($\phi = 0$), the effective tax rate reduces to $t_e = 2\gamma\tau$, which is less than the statutory tax τ since $\gamma = 0.1$. In this case, the F.O.C. with respect to q reduces to $p = c + 2\gamma\tau$. Because unit costs and the statutory tax rate are equal in both treatment and control groups, it follows that the

[17] To see this, recall that $\phi = 1$ in the control group because there is no opportunity to evade the tax. Substituting this value into equation (3.8) and optimizing over q yields $p = c + \tau$.

price in the treatment group (evasion possible) is lower than in the control group (evasion not possible).

This optimal solution assumes that sellers are fully rational and therefore evade all units due to the low audit probability. However, the existing tax evasion literature shows that most empirically observed evasion rates are lower than suggested by standard theory (Alm 2012). Our experimental results also support this observation; despite the low audit probability of 0.1 we observe an average compliance rate of 7%. That is, ϕ is larger than zero in our experiment. However, only 33 out of 40 sellers report non-zero sales in any given reporting period, which suggests that most sellers behave rationally as predicted by the theory and that $\phi < 1$. Therefore, the non-zero compliance rate reflects the effect of a combination of full and partial evaders.[18]

So how does the incidence of the tax in the treatment group compare to that in the control group? We address this question in the next section for a general case where market compliance is greater than zero and smaller than one, i.e. $0 \leq \phi \leq 1$.

3.3.2 Market outcome

Consider the empirical case where $0 \leq \phi \leq 1$. The F.O.C. with respect to q implies that market supply is implicitly defined not only by the price p but also by the effective tax rate t_e, which in turn depends on τ, ϕ and γ. This allows us to express market supply as a function of p and t_e: $S(p, t_e)$. We already showed that market demand is defined by $D(p)$ because each buyer's q^* depends on the price only. As has been shown in the literature, an experimental double auction market clears the market so that demand equals supply:

$$D(p) = S(p, t_e) \tag{3.10}$$

The equilibrium condition is differentiated implicitly with respect to p and τ to determine the incidence of the statutory excise tax τ:

$$\frac{dp}{d\tau} = \frac{\partial t_e}{\partial \tau} \frac{\partial S}{\partial t_e} \frac{1}{\frac{\partial D}{\partial p} - \frac{\partial S}{\partial p}}. \tag{3.11}$$

The incidence result is equivalent to the textbook case except for parameter $\frac{\partial t_e}{\partial \tau} = \phi(1 - 2\gamma) + 2\gamma$. $\frac{\partial t_e}{\partial \tau}$ equals 1 as long as the effective tax rate is equal to the statutory tax rate. This is the case in the control group where there is no evasion opportunity.[19] Therefore, the incidence of the tax in the treatment group is lower than in the control group if $\frac{\partial t_e}{\partial \tau}$

[18] One could define the difference between the "irrational" $\phi > 0$ and the rational $\phi = 0$ as a parameter for an exogenously given level of irrationality.

[19] Note that this is also the case if everyone in the treatment group reports honestly. However, we do not consider this case here since our empirical observation shows less than full compliance.

is smaller than 1. It is easy to show that, conditional on $\gamma < 1/2$, this is the case as long as $\phi < 1$. In other words, we expect the market equilibrium price in the presence of tax evasion to be lower than in the case where tax evasion is not possible as long as some units are evaded ($\phi < 1$). This follows from the fact that the statutory tax rate is only due on reported units and therefore has less of an impact if a positive amount of units is evaded, i.e., the effective tax rate will be lower than the statutory rate as long as we do not see full compliance.

To see this more clearly, rewrite equation (3.11) in terms of elasticities:

$$\frac{dp}{d\tau} = \frac{\partial t_e}{\partial \tau} \frac{\epsilon_{S,\tau}}{\epsilon_{S,p} - \epsilon_{D,p}} \tag{3.12}$$

where $\epsilon_{S,\tau}$ is the supply elasticity with respect to the tax rate τ, and $\epsilon_{S,p}$ and $\epsilon_{D,p}$ are the supply and demand elasticities with respect to the price. In our experimental design, supply elasticities equal demand elasticities in equilibrium. Assuming that $\epsilon_{S,\tau} = \epsilon_{S,p}$, the incidence result can hence be rewritten:[20]

$$\frac{dp}{d\tau} = \frac{\partial t_e}{\partial \tau} \frac{1}{2} \tag{3.13}$$

In the control group where evasion is not possible $\frac{\partial t_e}{\partial \tau} = 1$ and a 1 unit increase in the tax rate increases the price by 1/2 units. On the other hand, a 1 unit increase in the tax rate in the treatment group - where evasion is possible - increases the price by $(\phi(1-2\gamma)+2\gamma) \times 1/2$ units, which is less than 1/2 as long as $\phi < 1$ (recall that $\gamma < 0.5$ in our experiment).

The intuition for this result is fairly straight forward. Since $\phi < 1$ implies that the effective tax rate is lower than the statutory tax rate, sellers who evade are less responsive to changes in the statutory tax rate than sellers who report honestly. Therefore, as long as some firms evade the tax, the industry supply curve shifts up by a smaller margin than would be observed in the absence of tax evasion. This is illustrated in Figure 3.2. First, consider panel A in Figure 3.2, which represents the control group where evasion is not possible. The supply curve generally shifts up by the effective tax rate. Because the effective tax rate equals the statutory tax rate in the case without evasion, the supply curve shifts up by the full amount of the statutory rate. This results in a new market equilibrium (p_c^*, q_c^*); subscript c indicates control group.

On the other hand, the supply curve in the treatment group – shown in panel B of Figure 3.2 – shifts up by the effective tax rate, which is less than the statutory rate as long as $\phi < 1$. This results in a new market equilibrium (p_t^*, q_t^*) where $(p_t^* < p_c^*)$ and $(q_t^* > q_c^*)$; subscript t indicates treatment group. Also note that the difference between

[20]This is a fair assumption because the tax is highly salient in our experiment.

the two equilibria increases with evasion and is maximized if all sellers in the treatment group fully evade the tax. This difference in price suggests that some of the benefits of evasion are shifted to buyers in the form of lower prices.

3.4 Empirical Strategy and Results

Recall that we are interested in identifying the impact of tax evasion opportunities on the economic incidence of a sales tax. We describe the empirical strategy used to identify this treatment effect in section 3.4.1 and our findings in section 3.4.2.

3.4.1 Empirical Strategy

Given the discussion in section 3.3, we are particularly interested in knowing whether the market clearing price in the treatment group is different from the price in the control group. Therefore, the first step in our empirical strategy is to define the market price. The experiment produced one price for each unit sold in a given market-period, which allows us to create three measures of market price. The first measure is simply the price at which each item is sold, which we denote P. We also calculate the mean and median price in a given market-period and denoted them $\overline{P}$ and P_{50}, respectively. Therefore, our data set has one observation per market-period when price is measured by $\overline{P}$ or P_{50} and n observations per market-period when market price is measured by P, where n is the number of units sold in that market-period.[21]

Second, due to random assignment to groups and markets, any (non-parametric) difference in these prices between the treatment and control groups is taken as evidence of the presence of treatment effects. We also test for treatment effects parametrically by regressing each measure of price, separately, on a treatment dummy. The baseline model for $\overline{P}$ is specified as follows:

$$\overline{P}_{i,m} = \beta_0 + \delta T_m + \epsilon_{i,m}, \tag{3.14}$$

where $\overline{P}_{i,m}$ is the mean price of the good in period i (with $i = 1, ..., 27$) of market m (with $m = 1, ..., 8$). T_m is a dummy for the treatment state, which is equal to one if treatment group and zero if control group. $\epsilon_{i,m}$ is a standard error term. Our coefficient of interest is δ, which represents the difference in market price between the two groups. More precisely, δ indicates the causal effect of evasion opportunity on the equilibrium market price. This causal interpretation follows from the fact that the groups are identical except for access to evasion and random assignment of participants to the two groups.

[21] The minimum number of units sold in a market period is 5 and maximum is 7. Therefore, n ranges from 5 to 7.

We set up our data as a panel with 27 periods per market and run pooled ordinary least squares (OLS) regressions with standard errors clustered on market. OLS regressions yield unbiased results because random selection into treatments cause the error term to be uncorrelated with our explanatory variable of interest in each period.[22] Because the treatment status of each market and hence the participants in that market is always the same, the treatment effect (parametric and nonparametric) is identified using a between-market design.[23] We include period fixed effects in some specifications.

3.4.2 Results

Demographics of Participants

After the experiment, subjects reported their age, gender, native language, level of tax morale and field of study. Tax morale is determined using a question very similar to one used in the World Values Survey (Minkov 2012).[24] Each of these variables is summarized in Table 3.2. Casual observation of the data shows that randomization into the treatment states worked well. This is confirmed by non-parametric Wilcoxon rank-sum tests for differences in distributions between groups (Wilcoxon 1945; Mann and Whitney 1947); we do not observe any statistically significant differences in gender, age, share of participants whose native language is German, tax morale or field of study across the two groups. While we do not explicitly measure other attitudinal variables such as social norms or preferences, randomization implies that these omitted variables are also balanced across groups and therefore do not have any effect on our results.

Among all participants, approximately 51% were male, 77% indicated German to be their native language, and the average age was 26 years. Approximately 24% of subjects stated that cheating on taxes can never be justified and 48% indicated that economics is their major field of study.

---

[22]Note that estimators that allow for censoring, such as Tobit models, are unnecessary since the market price is not censored. Although the market price could be no lower than 18 and no higher the 82, the distribution of market prices suggest that these prices were never binding; the lowest market price is 30 and the highest is 63.

[23]Notice that this also implies that it is not possible to estimate the treatment effect in the presence of market fixed effects. Each individual is randomly assigned to a market and everyone in the market has the same treatment status. Therefore, the treatment status of a market is the same as the treatment status of the individuals trading in that market.

[24]"Please tell me for the following statement whether you think it can always be justified, never be justified, or something in between: 'Cheating on taxes if you have the chance'." This is the most frequently used question to measure tax morale in observational studies (e.g., Alm and Torgler 2006, Halla 2012 and Doerrenberg and Peichl 2013).

Compliance and effective tax rate

Our hypothesis that markets with evasion opportunities clear at lower prices requires that sellers actually underreport sales. The empirical results show that this condition is satisfied. We find that every subject evaded some positive amount of sales at least once and 33 of the 40 subjects in the treatment group fully pursued the profit maximizing rational strategy of full evasion in every reporting period. As a result the mean compliance rate is approximately 7% among all sellers in treatment group and 61% among those who report non-zero sales.[25]

Recall that the effective tax rate t_e can be written as $\tau(\phi + 2(1-\phi)\gamma)$. Substituting the exogenously determined statutory tax rate ($\tau = 10$ ECU) and audit probability ($\gamma = 0.1$), and rearranging the expression, yields $t_e = 2 + 8\phi$. Therefore, the average effective tax rate depends on the average compliance rate and, in our case, is approximately $t_e = 2.56$ ECU $(= 2 + 8 \times 0.07)$.[26]

We combine this effective tax rate with the demand and supply schedules shown in Table 3.1 and Figure 3.1 in order to predict the equilibrium price and quantity in the treatment group. As shown in Figure 3.1, a per unit tax shifts the supply curve up. However, unlike the control group, the supply curve only shifts up by 2.56 ECU in the treatment group, thus producing a new equilibrium with price in the range of 50.65 to 52 ECU and 7 units. Consistent with the theoretical model, this price is lower than the predicted equilibrium price range for the control group: 53 to 57 ECU (see section 3.2.4).

Price

Non-parametric results The non-parametric results presented in Figures 3.3 and 3.4 and Table 3.3 show clearly that the price in the treatment group is lower than in the control group. Figure 3.3 reports the mean market price by period for the treatment and control groups. The data show that the mean market price varied a lot in both groups in the first 10 to 14 trading periods. This is consistent with the existing literature, which generally finds that double auction markets take approximately 8 to 10 rounds to converge (Ruffle 2005).

Although price in both groups converged in roughly same number of periods, the

[25] This level of evasion is at the high end of evasion estimates in the experimental tax evasion literature (e.g., Fortin et al. 2007; Alm et al. 2009; Alm et al. 2010; Coricelli et al. 2010). However, these studies focus on income taxes and are therefore not directly comparable to our results. We do not know of any sales tax experiments in the tax evasion literature. Evidence from the real world suggest that our compliance rates are not unreasonable. For example, the compliance rate in our experiment is comparable to the compliance rate for the 'use' tax in the United States; 0 to 5 percent among individuals (GAO 2000).

[26] The average effective tax rate is obviously lower among those who fully evade ($t_e = 2$ ECU) and higher among those who partially evade ($t_e = 6.88$ ECU). In either case, the effective tax rate is sufficiently different from the statutory tax rate to generate a treatment effect as predicted by the theoretical model.

evolution of prices is different. Price increased steadily to equilibrium in the treatment group, and behave erratically in the control group. For this reason, and as is common in the literature, our primary results are based on data from trading periods 14 to 27; we provide results for the full sample for illustrative purposes. The mean market price in both groups stabilized after round 14: at approximately **54.35 ECU** in the control group and **51.63 ECU** in the treatment group (see panel B of Table 3.3). These observed market prices are well within the equilibrium price ranges predicted by the theoretical model; see sections 3.2.4 and 3.4.2. This confirms that the experimental results are consistent with the theoretical predictions, which increases our confidence in the results.

More importantly, the mean market price in the treatment group is 2.72 ECU lower than in the control group. This represents the estimated treatment effect and it is statistically different from zero at the 1% level according to the Wilcoxon rank-sum test. In other words, we find that markets with access to tax evasion trade at lower prices than markets without access to tax evasion.[27] Figure 3.4 and the second column of Table 3.3 show that the treatment effect is qualitatively the same when we use median market price instead of mean market price; in this case the treatment effect is ECU 3.00.[28]

Parametric results We extend the analysis above by estimating equation (3.14) for the mean market price as the dependent variable. The estimated treatment effect reported in Panel B of Table 3.4 ranges from -2.65 ECU to -2.70 ECU and is statistically different from zero at the 1 percent level.[29] Additionally, the estimates are robust to the inclusion of period fixed effects (model 2), demographic covariates (model 3) and both period fixed effects and demographic covariates (model 4). They are also robust to the definition of price as demonstrated by the results in Table 3.5. Estimating equation (3.14) with the median market price P_{50} as our dependent variable yields treatment effects of -1.60 ECU to -2.10 ECU that are statistically different from zero at the 1% level (see Panel A of Table 3.5). Although these estimates are approximately 0.70 to 1.00 ECU smaller than that reported in Panel B of Table 3.4, they remain economically meaningful.[30] These results

[27]Note that the estimated treatment effect is larger for the full sample (panel A). Because this sample includes data before the market price converges, we prefer the estimate in panel B.

[28]Further evidence that tax evasion affects the incidence of a tax is provided in Figures 3.6 and 3.7, which report the cumulative distribution of mean and median market prices, respectively, for the treatment and control groups. Both figures show clearly that the price in the control group is not drawn from the same distribution as that in the treatment group. This conclusion is supported by the Kolmogorov-Smirnov test for equality of distribution functions; in both cases we reject the null that the distributions are equal. This result also holds when we use the individual ask prices (P) instead of mean or median prices; results available upon request.

[29]Panel A of Table 3.4 reports the results for the full sample. These results are reported for illustrative purposes only since the market does not clear until around period 14.

[30]We also estimate the model with the ask price for each unit sold as the dependent variable and report the results in Panel B of Table 3.5. The estimated treatment effect in this case is -2.66 ECU to -2.72 ECU, which is almost identical to that for the mean market price as reported in Panel B of Table

confirm our earlier non-parametric findings that the market price in the treatment group is lower than in the control group.

Units sold

Non-parametric results The theoretical predictions in section 3.3 suggests that treatment markets will clear at a lower price and higher quantity than the control group. We have already demonstrated that the market clearing price is lower in the treatment group. This section shows that the treatment group also sold more units than the control group. The results in Table 3.3 show that the mean number of units sold per period in the control group is 5.96, which is close to the 6 units predicted by the theoretical model. On the other hand, the treatment group sold an average of 6.5 units per period. Although this is slightly below that predicted by the theoretical model, it is statistically different from the number of units sold in the control group at the 1% level. In other words, the estimated treatment effect of 0.5 units is statistically different from zero. The difference in sales between the two groups is more obvious when we look at the total number of units sold by each group. Again, restricting attention to trading periods 15 to 27 (after the market clears), we find that the treatment group sold a total of 336 units while the control group only sold 308 units. Corresponding numbers for periods 1 to 27 are 704 and 647 in the treatment and control group, respectively.

Parametric results These results are supported by results from a regression analysis that are reported in Table 3.6. Focussing on Panel B, which reports results for periods 15 to 27, we find a treatment effect of 0.6 units; relative to the control group, the treatment group sold approximately 0.6 more units per period.

3.5 Discussion

The results presented in section 3.4.2 show that markets with sellers who have the opportunity to evade taxes trade more units and do so at lower prices than markets where tax evasion is not possible. Section 3.5.1 explains the incidence results in the context of the theoretical model and section 3.5.2 describes the welfare implications. The external validity of our findings is discussed in section 3.5.3.

3.5.1 Incidence

The treatment effect identified above is very much consistent with the theoretical model in section 3.3. According to the theoretical framework, tax evasion lowers the effective tax

3.4.

rate facing sellers, thus allowing them to trade at lower prices in a competitive market. As a result, the final tax burden shifted to buyers is lower than it would otherwise be in the absence of tax evasion. This is exactly what we find; the mean compliance rate ranges from 0% among full evaders to 61% among partial evaders with an average of 7% among all sellers. This implies an average effective tax rate of 2 ECU among full evaders and 6.88 ECU among partial evaders for an average of approximately 2.56 ECU among all sellers. Because the market is competitive, sellers facing these lower effective tax rates trade at lower prices in an effort to maximize their profits.

So how does this response among sellers affect the incidence of the tax? In order to answer this question, we fist have to determine the incidence of the tax in the control group, which requires knowing the market equilibrium in the absence of the tax. Although we did not run a “no-tax" treatment, we are able to estimate this “no-tax" equilibrium by relying on evidence from Grosser and Reuben (2013) who run a “no-tax" treatment using a comparable double auction market with the same demand and supply schedule as we do.[31] In line with the theoretical expectation, they find a mean market price of 49 ECU and 7 units in the “no-tax" equilibrium. On the other hand, the market in our control group (with tax but no evasion opportunity) cleared with a mean price of 54.35 ECU and 5.96 units, which is well within the equilibrium predicted by the theory: 53 ECU to 57 ECU with 6 units traded.

More importantly, this equilibrium price is approximately 5 ECU above the “no-tax" equilibrium of 49 ECU. This suggests that the incidence of the tax burden in the control group is shared equally between buyers and sellers since the tax rate is 10 ECU per unit. Again, this is consistent with the theoretical framework; since the demand and supply schedules have equal elasticity in equilibrium, the burden is expected to be shared equally between buyers and sellers.

The next step is to determine the extent to which access to evasion affected the incidence of the tax. The mean market clearing price in the treatment group (with tax and evasion opportunity) is 51.63 ECU, which is within the 50.65 to 52 ECU price range we predicted given the empirical compliance rate (7%) and mean effective tax rate 2.56 ECU (see section 3.4.2). Considering the statutory tax rate of 10 ECU per unit, this implies that buyers in the treatment group pay 26.4% $(51.64 - 49)/10$ of the *statutory* tax burden, compared to the 50% in the case without evasion. In other words, access to evasion reduced the statutory tax burden on buyers by about 23 percentage points.

This result would seem to suggest that sellers bear the greater share of the tax

[31] The experimental design in Grosser and Reuben (2013) differs from ours in that they use a within subject design where each subject trades in a market with and without the tax. We are aware that within subject and between subject designs may yield different results (Charness et al. 2012). However, we argue that their “no-tax" estimate is a reasonable baseline to use in our incidence analysis, especially since they randomized the order of tax and “no-tax" treatments. Additionally, their result is in line with the theoretical prediction which is further support for using their result as a baseline result.

burden despite equal supply and demand elasticities. To see that this is incorrect, consider the following exercise. For simplicity, consider the mean effective tax rate of 2.56 ECU. If sellers with evasion opportunity continued to share the effective tax burden 50-50, we would expect the price in the treatment group to increase by approximately 1.28 ECU (= 2.56/2) relative to the “no-tax" equilibrium; that is to 50.28. However, this is not what we observe. The price in the treatment group is 51.63 ECU, which suggest that sellers shift the full effective tax rate onto buyers; buyers bear 2.63 ECU ($= 51.63 - 49$) even though the effective tax rate is 2.56 ECU. As a result, about 102.7% ($= (51.63 - 49)/2.56$) of a seller’s effective tax rate is shifted onto buyers.

3.5.2 Welfare Implications

Because we find that markets with access to evasion trade a lower prices and higher quantity, we expect the efficiency cost of the tax to be lower in market where tax evasion is possible. Our induced demand and supply curves along with our estimated equilibrium price and quantity allows us to do a back-of-the-envelope calculation to approximate the impact of evasion on the deadweight loss of the per unit tax. For simplicity, we assume linear demand and supply schedules in the following calculations and estimate the partial equilibrium deadweight loss as the area of the *Harberger triangle*.[32] The estimated equilibria for the average trading period are (49 ECU, 7 units), (54.35 ECU, 5.96 units) and (51.65 ECU, 6.5 units) for “no-tax", control, and treatment, respectively. This implies that the deadweight loss of the 10 ECU per unit tax in the control group where there is no access to evasion is approximately 5.50 ECU ($= 1/2 * 10 * (7 - 5.9)$) per trading period. This scales up to approximately 71.5 ECU across all market clearing trading periods in the control group.

For the treatment group we consider the case where the effective tax rate is 2.56 ECU, which is over-shifted to buyers; recall that buyers pay 2.63 ECU more than in the “no-tax" baseline. Using the mechanics of the simple partial equilibrium model, the supply curve has to shift up by 5.26 ECU for the buyers’ price to increase by 2.63 ECU and quantity to fall to 6.5 units. Using these numbers, we get a deadweight loss of 1.32 ECU ($= 1/2 * 5.26 * (7 - 6.5)$) per trading period. This scales up to approximately 17.1 ECU across all market clearing trading periods in the treatment group. Therefore, the deadweight loss of the tax is much lower in the treatment group than in the control group. Table 3.7 summarizes our results.

There are three important things to note about the calculations shown above. First, the calculations assume that tax evasion is costless except for the fine. In particular, tax

[32] Note that these are uncompensated estimates of deadweight loss, which overestimates the traditionally favored compensated estimates. For this reason, we focus on the difference in deadweight loss between the two groups rather than the level of the deadweight loss.

evasion does not require any real resource costs in our experiment. Therefore, the excess burden is limited to the changes in the market outcome: quantity sold and price. To the extent that tax evasion requires real resource costs, the reduction in excess burden caused by evasion is likely to be partially offset by the cost of real resources used to facilitate evasion.[33] Second, we exclude the impact on tax revenue since the foregone tax revenue represents a transfer to private agents and does not affect welfare as long as one does not impose a welfare function that gives higher weight to tax financed public goods relative to private consumption.

Third, our calculations hold the tax rate constant rather than tax revenue. It can be shown that the difference in excess burden between the treatment and control group would be smaller if a revenue requirement is imposed. This follows from the fact that a lower tax rate would be required in the control group in order to generate the same amount of revenues as the treatment group. Since the excess burden is increasing in the square of the tax rate, a lower tax rate would imply a substantially smaller excess burden in the control group than the amount shown above. In fact, it is possible for markets without tax evasion opportunities to generate smaller excess burdens than markets with evasion opportunities, if the compliance rate is very low.

3.5.3 External Validity

As with all economic laboratory experiments, there remains doubt about the external validity of our results.[34] One major concern is that the setting in the lab is abstract and artificial. However, the literature shows that laboratory double auctions generate very plausible equilibria (e.g., Smith 1962; Holt 1995; Dufwenberg et al. 2005; Grosser and Reuben 2013.). Although subjects trade in fictitious goods, they receive actual money pay-offs and thus face incentives similar to buyers and sellers in actual markets. Furthermore, the question of tax incidence has been widely studied in the laboratory setting (e.g., Riedl and Tyran 2005; Ruffle 2005; Cox et al. 2012; Grosser and Reuben 2013) and shown to lead to results that reflect theoretical predictions very well.

In order to make the tax evasion decision as realistic as possible we used actual tax terminology and announced to the participants that all tax revenue would be donated to the German Red Cross, a non-ideological charity organization that is usually perceived as reliable and transparent.[35] Although evasion may occur among buyers as well, the real-

[33] The absence of real resource costs in our experiment is comparable to 'use' tax evasion by individuals who purchase goods online. Purchasing goods online in order to evade the sales tax, arguably, involves smaller resource costs than visiting a store front in person.

[34] See Levitt and List (2007) for a critical discussion of the generalizability of lab experiments. Falk and Heckman (2009) offer a defense of most concerns, some of which are also discussed here.

[35] Tax morale research (Torgler 2007) finds that taxpayers are more likely to comply with tax laws if they believe that the tax revenue is spent transparently. Eckel and Grossman (1996) show that dictators

world problem seems to be more relevant among sellers; sellers are usually responsible for remitting sales taxes to the government. In this sense, our laboratory setting mimics the operation of most transaction taxes in the real world and thus reflects the more relevant "real-world" sales tax evasion concerns.

It is also often argued that the stakes in lab experiments are too small to interpret the outcomes as realistic. This is unlikely to be true in our case because our average pay-off of EUR 19.63 is relatively high and, to give an idea of its purchasing power in Cologne, roughly corresponds to eight full lunch meals in the student cafeteria at Cologne University.[36] Furthermore, many experiments conducted in locations where the stakes were equivalent to more than a month's earnings find very similar results to conventional "small-stake" experiments (Slonim and Roth 1998). Additionally, it is questionable whether high stakes render a more realistic setting, since most "real-life" decisions do not involve massive amounts of money either (Falk and Heckman 2009). Another concern is the reliance on university students as participants. Many experiments with non-student populations find results comparable to experiments with students (Charness and Kuhn 2011; Falk et al. 2013). Furthermore, Alm et al. (2011) compare students to non-students in tax compliance experiments and find that the reporting responses of students to policy innovations are largely the same as those of non-students in identical experiments and "real" people in non-experiment environments.

3.6 Conclusion

We use data generated in an economic laboratory experiment to identify the effect of tax evasion among sellers on the economic incidence of a per-unit tax. We find strong evidence that access to evasion opportunities affect the incidence of a per-unit tax. In particular, sellers who are able to evade a per-unit tax trade at lower prices and sell more units. In fact, relative to the baseline case where buyers face 50% of the statutory tax burden, buyers in the treatment group only face approximately 26.4% of the statutory tax burden. Although buyers pay lower prices than they otherwise would, we find that sellers fully shift the expected effective tax onto buyers. Partial equilibrium welfare analysis reveal that evasion options reduce the excess burden of taxation and increase welfare. However, the latter result may change if a revenue requirement is imposed.

Our findings suggest that access to evasion reduces the effectiveness of taxes that are implemented with the specific intent of changing the activity level of market participants. Furthermore, because evasion reduces the amount of the tax that is shifted onto

share more in dictator games if the recipient is the American Red Cross. Overall, we donated EUR 288 to the Red Cross.

[36]The show-up fee is equivalent to one meal. The cafeteria at University of Cologne is the most popular spot for students to buy their daily lunch.

buyers, our findings also suggest that sales taxes may be more or less regressive than we think depending on which part of the income distribution benefits most from the evading activities of sellers.

The results also imply that policy makers do not necessarily have an easy choice when deciding whether to pursue evasion reducing strategies or to exploit the potential efficiency gains of evasion. For example, Cremer and Gahvari (1993) show that the optimal Ramsey rule in the presence of tax evasion calls for higher tax rates on the good with the tax evasion opportunity. The argument is that evasion lowers the real behavioral response and thus lowers excess burden; this is confirmed by our results. However, given that governments often face revenue requirements along with the fact that tax evasion may require real resource costs, we prefer a strategy that seeks to minimize tax evasion opportunities. This is especially important in cases where the policy objective is to influence real behavior. Evasion reducing strategies may also make sense on revenue grounds. Although revenues represent a simple transfer from an economic welfare perspective, revenues are used to produce public goods/service that are likely to be underproduced or not produced at all as tax revenues decline.

Finally, while we show that tax evasion opportunities affect tax incidence, it is not clear that the magnitude and effect is the same across all types of taxes. Conditional on the ease with which taxes can be evaded, it is also possible that the mechanism of evasion matters. For example, Tran and Nguyen (2014) show that Vietnamese firms evade VAT by artificially increasing their sales and material costs, which is facilitated by colluding with other producers in the supply chain. The presence of collusion as a means of evasion suggests lower competitive pressure, which may lead to different incidence outcomes under a VAT compared to retail sales taxes where collusion among firms is not necessary for evasion. Given recent calls for the adoption of VAT in the USA, we argue that this potential difference is worth investigating in future research. More generally, it would be interesting to know if and how evasion mechanisms in different tax systems affect the incidence of taxes.

Tables and Figures

Tables

Table 3.1: Demand and Supply Schedules

Buyer			Seller		
Subject	Value 1	Value 2	Subject	Cost 1	Cost 2
1	82	52	1	18	48
2	77	72	2	23	28
3	67	37	3	33	63
4	62	42	4	38	58
5	57	47	5	43	53

Notes: Reported are demand and supply schedules.

Table 3.2: Summary statistics of Demographic Variables

	Gender	Age	German	Tax Morale	Econ
	Control Group (Non-Evaders)				
Mean	0.43	24.90	0.72	0.25	0.43
St. Dev.	0.50	6.87	0.46	0.44	0.50
N. of Subjects	40	40	39	40	40
	Treatment Group (Evaders)				
Mean	0.60	26.93	0.83	0.23	0.53
St. Dev.	0.50	12.25	0.38	0.42	0.51
N. of Subjects	40	40	40	40	40
P-value	0.12	0.23	0.26	0.79	0.37

Notes: Reported are the mean characteristics of treatment and control groups. Gender is a dummy that is equal to 1 if male, German is a dummy that is equal to 1 if native language is German, tax morale is a dummy that is equal to 1 for subjects who believe cheating on taxes can never be justified and Econ is a dummy that is equal to 1 if field of study is economics. One subject in the control group did not report his/her language. P-value is for the Wilcoxon rank-sum test; null hypothesis is that there is no difference in the characteristics between the two groups.

Table 3.3: Summary Statistics of Choice Variables

	Price			Units sold		N.
Group	Mean	Median	Std. Dev.	Mean	Std. Dev.	Obs.
			Panel A: Full Sample			
Non-evader	54.98	54.00	4.99	6.04	0.53	647
Evader	51.24	51.00	2.50	6.56	0.51	704
			Panel B: Period>14			
Non-evader	54.35	54.00	3.92	5.96	0.47	308
Evader	51.63	51.00	1.69	6.50	0.50	336
P-value	0.00	0.00	–	0.00	–	–

Notes: Reported is the mean and median market price and mean number of units sold in each group. Price is the price at which each unit in a given market period is sold. Units sold is the number of units sold in a market period. Panel A uses all completed contracts from periods 1 to 27 and panel B uses all completed contracts in periods 15 to 27. P-value is for the Wilcoxon rank-sum test; null hypothesis is that there is no difference between the two groups.

Table 3.4: Impact of treatment on mean market price

	Model 1	Model 2	Model 3	Model 4
	Panel A: Full Sample			
Treat	-3.750***	-3.750***	-4.300***	-4.300***
	(1.009)	(1.077)	(0.347)	(0.371)
Constant	55.008***	54.181***	48.868***	48.040***
	(0.727)	(1.247)	(2.632)	(3.407)
R2	0.499	0.517	0.737	0.754
N	216	216	216	216
	Panel B: Period>14			
Treat	-2.701***	-2.701***	-2.651***	-2.651***
	(0.795)	(0.847)	(0.075)	(0.081)
Constant	54.362***	54.297***	49.508***	49.443***
	(0.539)	(0.516)	(0.572)	(0.750)
R2	0.553	0.563	0.884	0.894
N	104	104	104	104
Control variables	No	No	Yes	Yes
Period FE	No	Yes	No	Yes

Notes: Robust standard errors adjusted for clustering at the session level are in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Estimates are based on equation (3.14) with the dependent variable defined as mean market price in a given market period. Panel A uses all completed contracts from periods 1 to 27, panel B uses all completed contracts in periods 15 to 27. Period FE is period fixed effects.

Table 3.5: Impact of treatment on median and ask market price

	Model 1	Model 2	Model 3	Model 4
	Panel A: Median Ask Price (P_{50})			
Treat	-2.087***	-2.087***	-1.589***	-1.589***
	(0.625)	(0.665)	(0.218)	(0.233)
Constant	53.779***	53.918***	60.175***	60.314***
	(0.089)	(0.222)	(1.655)	(1.809)
R2	0.538	0.563	0.853	0.878
N	104	104	104	104
	Panel B: Ask Price (P)			
Treat	-2.720***	-2.721***	-2.662***	-2.660***
	(0.798)	(0.808)	(0.065)	(0.069)
Constant	54.354***	54.255***	49.500***	49.481***
	(0.543)	(0.486)	(0.491)	(0.593)
R2	0.173	0.176	0.276	0.279
N	644	644	644	644
Control variables	No	No	Yes	Yes
Period FE	No	Yes	No	Yes

Notes: Robust standard errors adjusted for clustering at the session level are in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Estimates are based on equation (3.14) with the dependent variable defined as median market price in a given market period in panel A; and the market price for each good in each market period in Panel B. All panels use completed contracts from periods 15 to 27. Period FE is period fixed effects.

Table 3.6: Impact of treatment on units sold

	Model 1	Model 2	Model 3	Model 4
	Panel A: Full Sample			
Treat	0.336***	0.334***	0.320***	0.324***
	(0.064)	(0.068)	(0.027)	(0.035)
Constant	6.088***	6.525***	6.701***	7.186***
	(0.059)	(0.144)	(0.406)	(0.277)
R2	0.090	0.292	0.100	0.301
N	1,006	1,006	1,006	1,006
	Panel B: Period>14			
Treat	0.402***	0.403***	0.598***	0.594***
	(0.125)	(0.125)	(0.051)	(0.056)
Constant	5.939***	6.177***	7.891***	8.102***
	(0.118)	(0.323)	(0.756)	(0.878)
R2	0.148	0.262	0.191	0.303
N	476	476	476	476
Control variables	No	No	Yes	Yes
Period FE	No	Yes	No	Yes

Notes: Robust standard errors adjusted for clustering at the session level are in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Estimates are based on equation (3.14) with the dependent variable defined as the number of units sold in a given market period. Panel A uses all completed contracts from periods 1 to 27, panel B uses all completed contracts in periods 15 to 27. "Period FE" is period fixed effects.

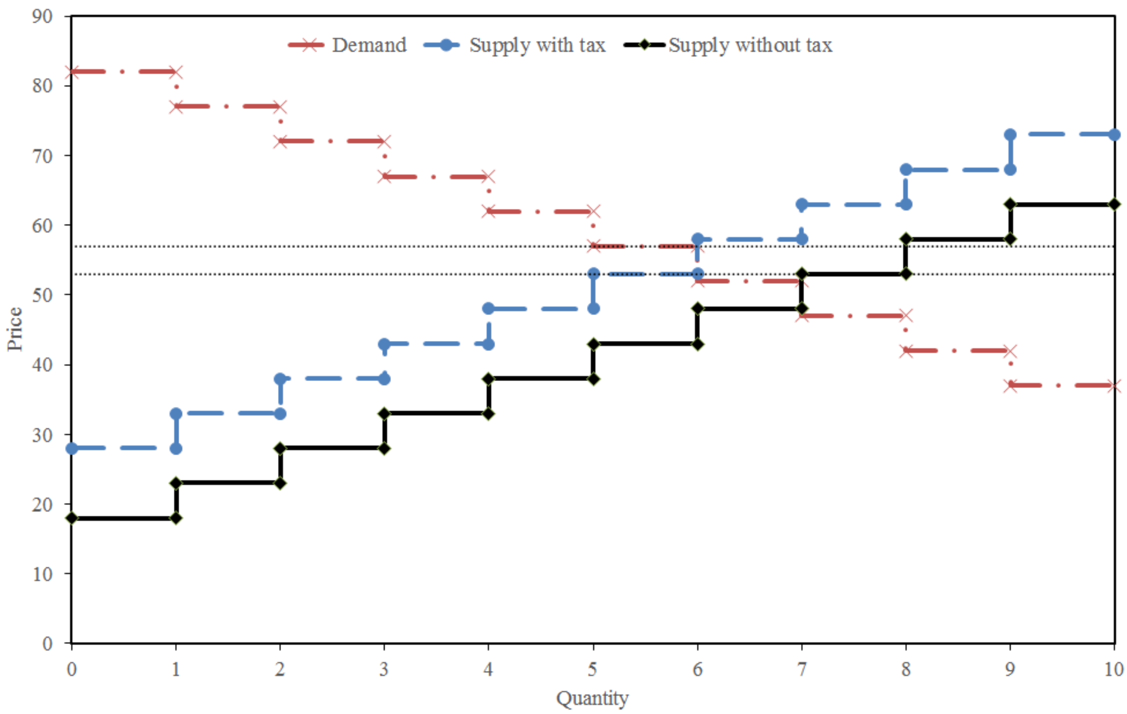
Table 3.7: Overview of Results

Condition	Price	Units	EB
No Tax	49	7	//
Control	54.35	5.96	5.50
Treatment	51.63	6.50	1.32
Treat Effect	-2.72	0.54	-4.18

Notes: The results in "No Tax" row are from Grosser and Reuben (2013) who use identical supply and demand schedules in an experimental double auction without taxes. "Control" and "Treatment" refer to the groups without and with evasion opportunity, respectively. Reported are the mean prices and number of units traded. The excess burden "EB" of the per-unit tax is calculated based on the simple *Harberger triangle*. "Treat Effect" indicates the non-parametric treatment effect defined as the difference between treatment and control group. All numbers expressed in Experimental Currency Units.

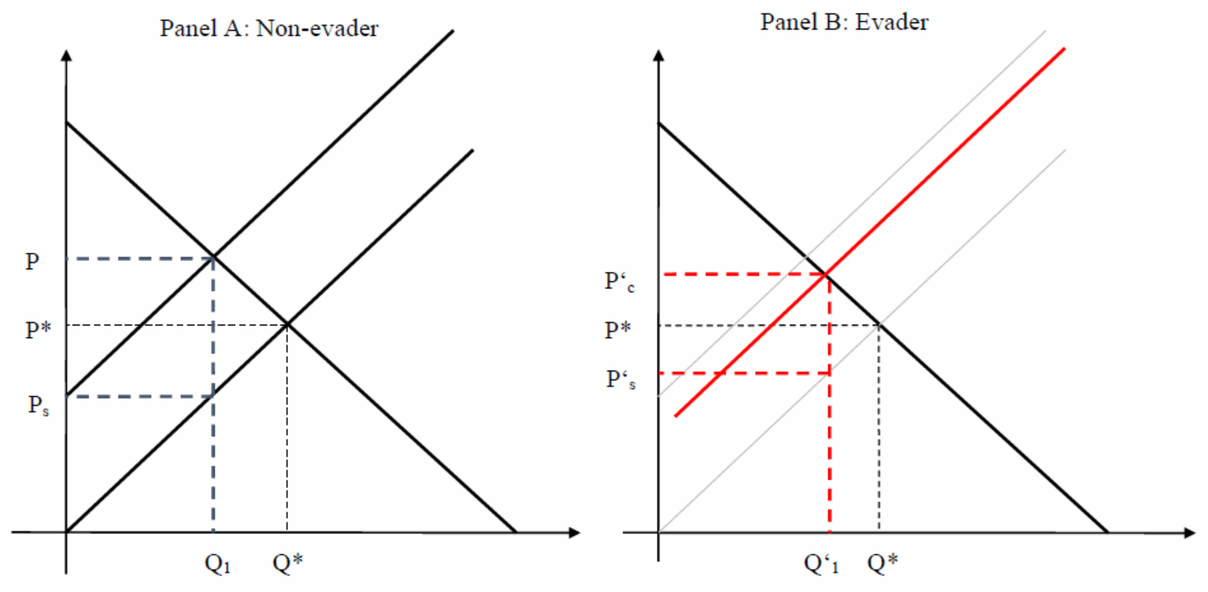
Figures

Figure 3.1: Supply and Demand Schedule



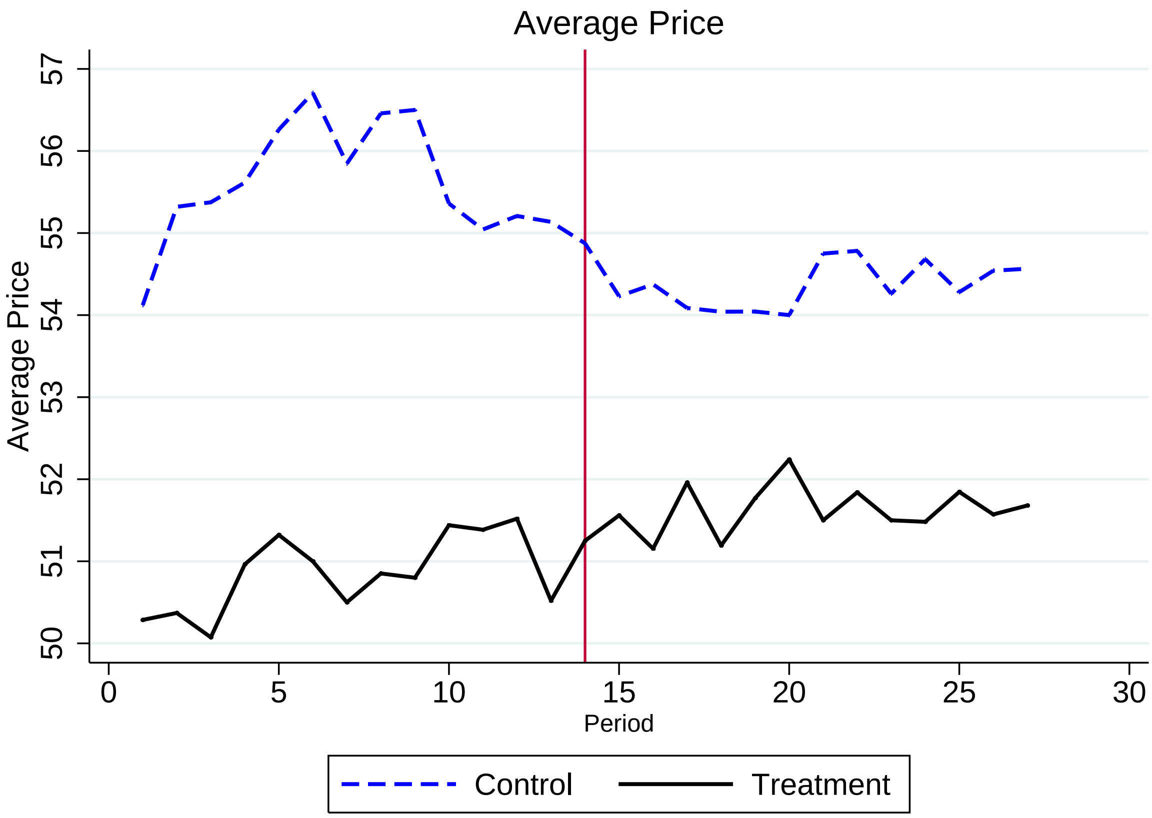
Note: The figure is adapted from Grosser and Reuben (2013, page 42, Figure 1). It shows the demand schedule for buyers and the supply schedule for sellers with and without the per unit tax. The predicted equilibrium occurs where the curves intersect: quantity $q = 7$ and price p between 48 and 52 without tax and quantity $q = 6$ and price p between 53 and 57 with the ECU 10 per unit tax.

Figure 3.2: Economic incidence of tax on seller



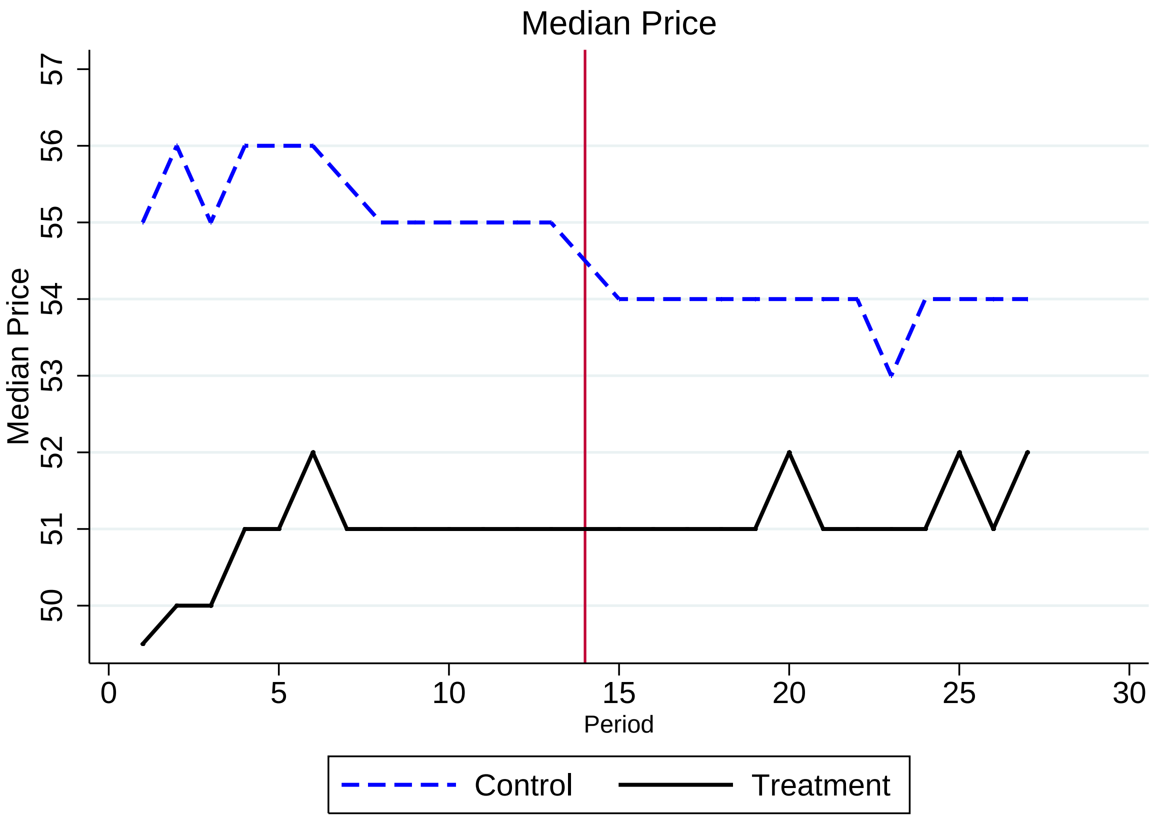
Notes: The imposition of a per-unit tax would ordinarily cause the supply curve to shift to the left and the market equilibrium to move from point (P^*, Q^*) to (P_c, Q_1) as illustrated in panel A. Because sellers are able to evade the tax, the supply curve shifts by a smaller amount causing the equilibrium to move from (P^*, Q^*) to $(P_c^{'}, Q_1^{'})$ as illustrated in panel B, where $P_c^{'} < P_c$.

Figure 3.3: Average market price by period and treatment



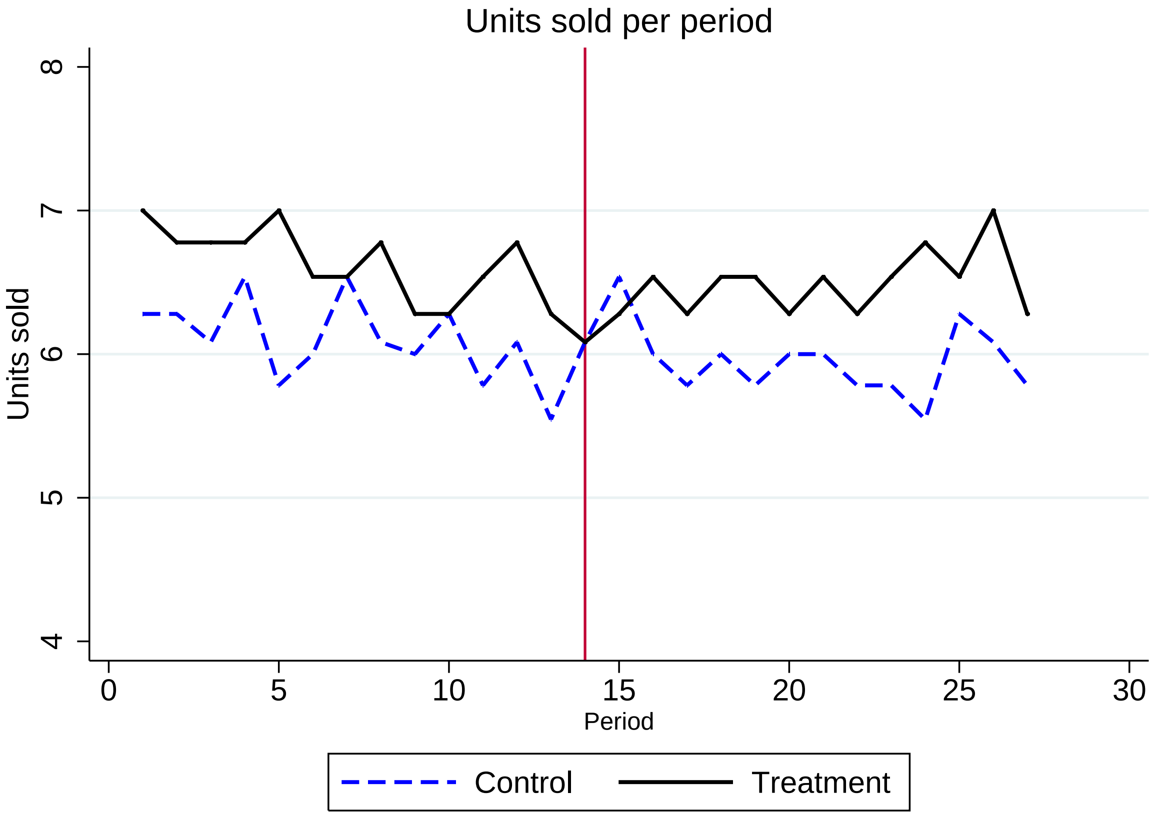
Notes: Reported is the average market price $\overline{P}$ in each period for the treatment and control groups. The vertical line indicates period 14; empirical results are based on market periods 15 to 27.

Figure 3.4: Median market price by period and treatment



Notes: Reported is the median market price P_{50} in each period for the treatment and control groups.The vertical line indicates period 14; empirical results are based on market periods 15 to 27.

Figure 3.5: Units sold by period and treatment



Notes: Reported is the number of units sold in each period for the treatment and control groups.The vertical line indicates period 14; empirical results are based on market periods 15 to 27.

3.7 Appendix A: Tables

Table 3.8: Impact of treatment on market price

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Ask Price		Mean Ask Price		Median Ask Price	
Treat	-2.662***	-2.660***	-2.651***	-2.651***	-1.589***	-1.589***
	(0.065)	(0.069)	(0.075)	(0.081)	(0.218)	(0.233)
Age	-0.367***	-0.370***	-0.368***	-0.368***	-0.641***	-0.641***
	(0.017)	(0.017)	(0.021)	(0.022)	(0.059)	(0.064)
Gender	-21.352***	-21.389***	-21.435***	-21.435***	-17.990***	-17.990***
	(0.177)	(0.174)	(0.219)	(0.234)	(0.633)	(0.676)
German	29.607***	29.642***	29.663***	29.663***	22.833***	22.833***
	(0.347)	(0.329)	(0.410)	(0.438)	(1.186)	(1.267)
Tax Morale	-1.274***	-1.258***	-1.245***	-1.245***	-0.921	-0.921
	(0.219)	(0.222)	(0.254)	(0.271)	(0.735)	(0.786)
Economics	5.126***	5.141***	5.156***	5.156***	2.562***	2.562***
	(0.153)	(0.162)	(0.183)	(0.195)	(0.529)	(0.565)
Constant	49.500***	49.481***	49.508***	49.443***	60.175***	60.314***
	(0.491)	(0.593)	(0.572)	(0.750)	(1.655)	(1.809)
R2	0.276	0.279	0.884	0.894	0.853	0.878
N	644	644	104	104	104	104
Period FE	No	Yes	No	Yes	No	Yes

Notes: Robust standard errors adjusted for clustering at the session level are in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Estimates are based on equation (3.14) with the dependent variable defined as the market price for each good in each market period in Models 1 and 2; mean market price in a given market period in Models 3 and 4; and median market price in a given market period in Models 5 and 6. All panels use completed contracts from periods 15 to 27. Period FE is period fixed effects. Gender is a dummy that is equal to 1 if male, German is a dummy that is equal to 1 if native language is German, tax morale is a dummy that is equal to 1 for subjects who believe cheating on taxes can never be justified and Field of study is a dummy that is equal to 1 if field of study is economics.

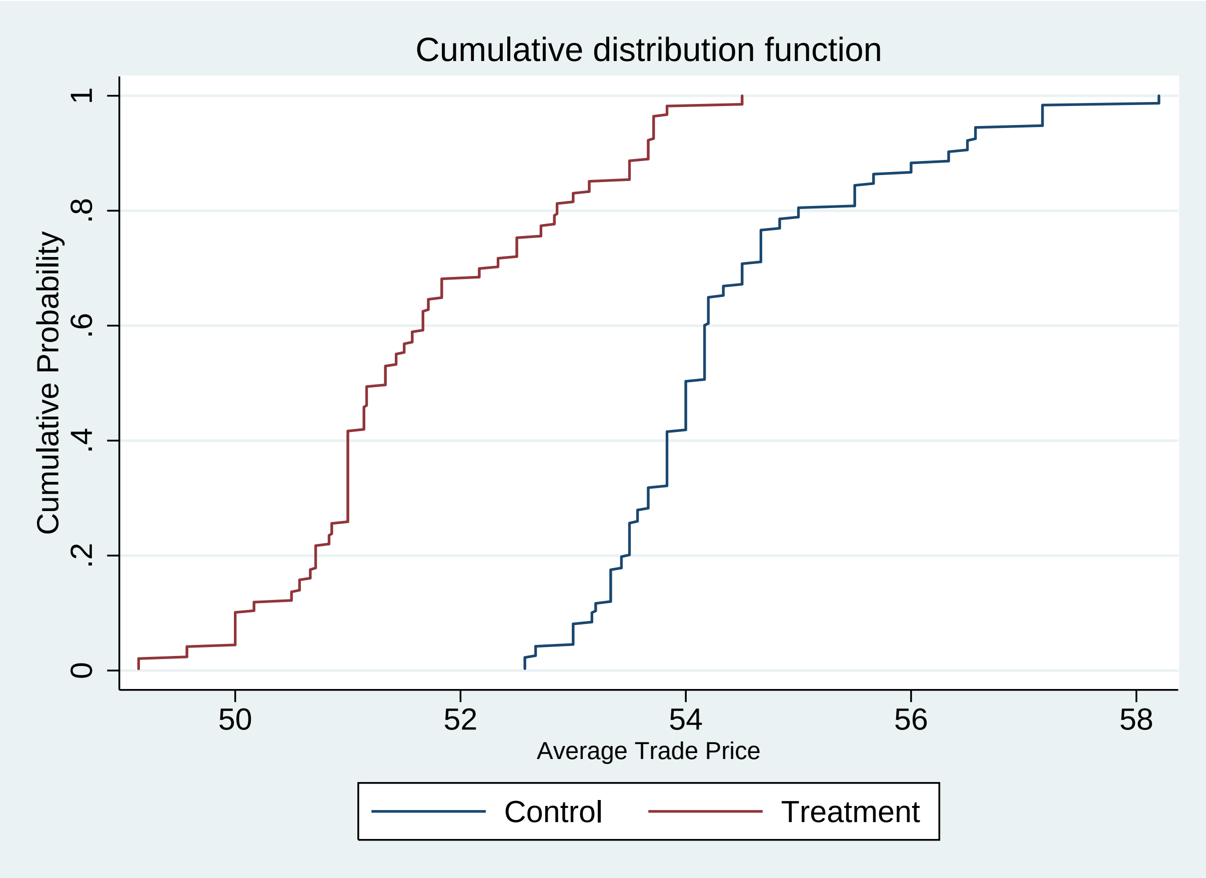
Table 3.9: Impact of treatment on units sold

	Model 1	Model 2	Model 3	Model 4
Treat	0.539***	0.540***	0.385***	0.383***
	(0.171)	(0.173)	(0.131)	(0.131)
Age			-0.017	-0.017
			(0.035)	(0.035)
Gender			2.349***	2.343***
			(0.353)	(0.363)
German			-2.000***	-1.973***
			(0.691)	(0.691)
Tax Morale			0.495	0.479
			(0.436)	(0.448)
Economics			-0.351	-0.349
			(0.305)	(0.305)
Constant	5.961***	6.147***	6.832***	7.005***
	(0.088)	(0.231)	(0.978)	(1.064)
R2	0.235	0.315	0.352	0.433
N	644	644	644	644
Control variables	No	No	Yes	Yes
Period FE	No	Yes	No	Yes

Notes: Robust standard errors adjusted for clustering at the session level are in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Estimates are based on equation (3.14) with the dependent variable defined as the number of units sold in a given market period. Estimation is based on all completed contracts in periods 15 to 27. Period FE is period fixed effects. Gender is a dummy that is equal to 1 if male, German is a dummy that is equal to 1 if native language is German, tax morale is a dummy that is equal to 1 for subjects who believe cheating on taxes can never be justified and Field of study is a dummy that is equal to 1 if field of study is economics.

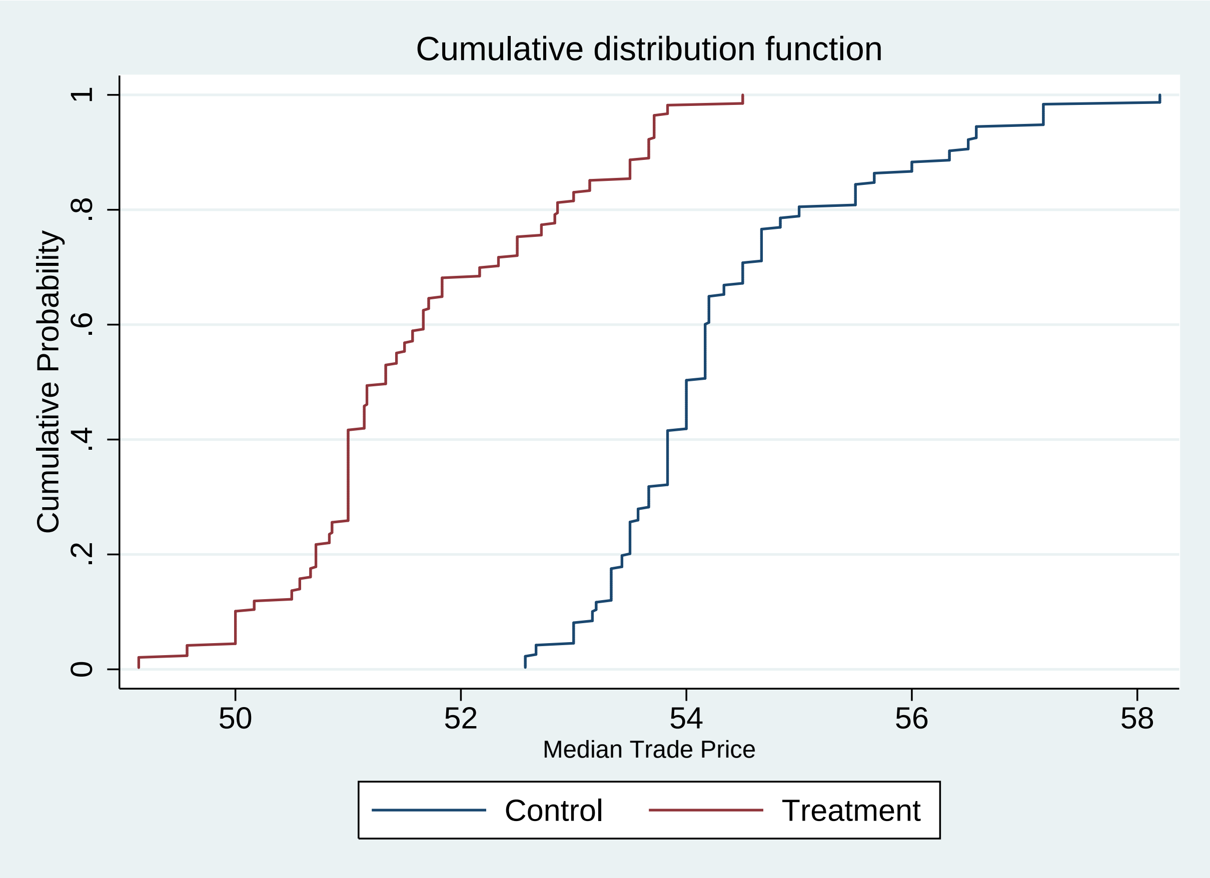
3.8 Appendix B: Figures

Figure 3.6: Cumulative distribution of market price by treatment



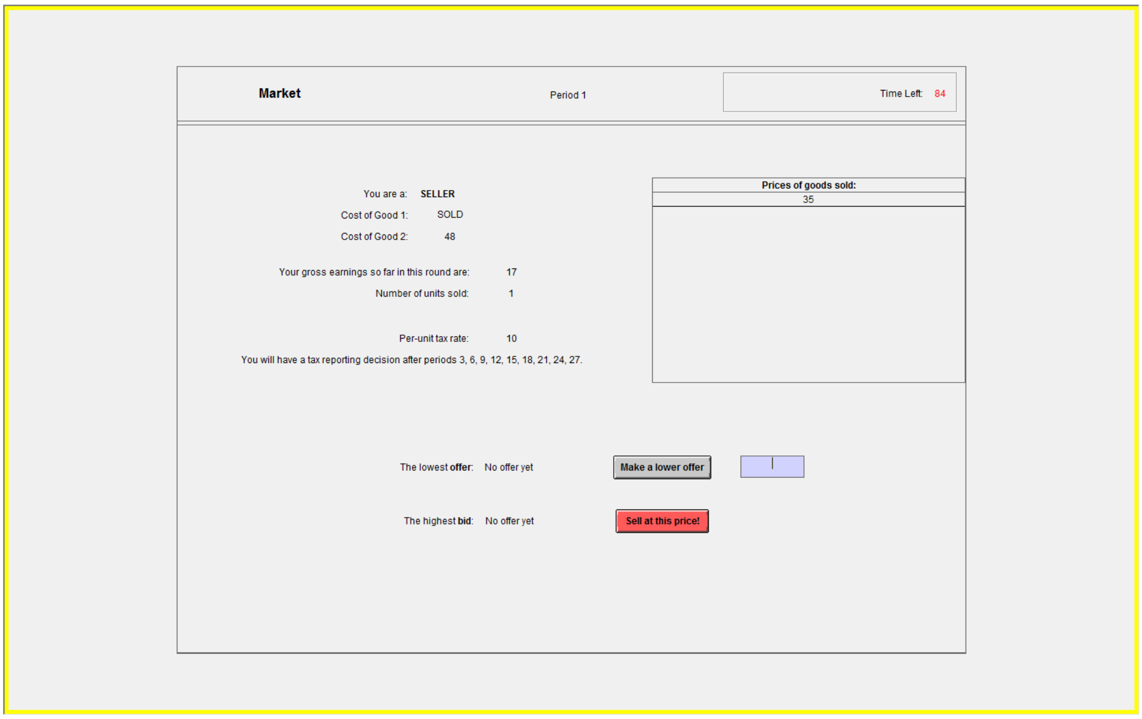
Notes: Reported is the cumulative distribution of average market price $\overline{P}$ for the treatment and control groups. Distributions are based on data from market periods 15 to 27. Two-sample Kolmogorov-Smirnov test for equality of distribution functions reports a maximum difference in distributions of 0.770 with pvalue of 0.000. This implies that the null hypothesis that the distributions are equal is rejected.

Figure 3.7: Cumulative distribution of market price by treatment



Notes: Reported is the cumulative distribution of median market price P_{50} for the treatment and control groups. Distributions are based on data from market periods 15 to 27. Two-sample Kolmogorov-Smirnov test for equality of distribution functions reports a maximum difference in distributions of 0.751 with pvalue of 0.000. This implies that the null hypothesis that the distributions are equal is rejected.

Figure 3.8: Screenshot of the Market Place



Note: Screenshot of the lab experimental double-auction market place. The screen displays the market place for a seller in the treatment group with evasion opportunity. The seller has sold her first unit at a price of 35. The cost for the first unit was 18, yielding a current gross-income of 17. Her second unit with cost 48 is not traded at this point. The screen shown is translated to English, the original experiment was conducted in German. The market place is based on Grosser and Reuben (2013).

3.9 Appendix C: Instructions

The following pages contain the translated instructions for both treatment groups. The original German versions are available from the authors upon request.

3.9.1 No-Evasion opportunity control group

Welcome and thank you for participating in our experiment. From now on until the end of the experiment, please refrain from communicating with other participants. If you do not abide by this rule, we will have to exclude you from the experiment.

We kindly ask you to read the instructions thoroughly. If you have any questions after reading the instructions or during the experiment, please raise your hand and one of the instructors will come to you and answer your question in person. Your payment and your decisions throughout the experiment will be treated confidentially.

You can earn money in this experiment. How much you earn depends on your decisions and the decisions of other participants. During the experiment, your payments will be calculated in a virtual currency: Experimental Currency Units (ECU). **30 ECU correspond to 1 Euro**. After the experiment, your pay-off will be converted to Euro and given to you in cash. Additionally, you will receive a show-up fee of 2.50 Euro.

The Experiment

Roles

At the beginning of the experiment, the computer will randomly assign five participants to the role of **"sellers"** and five other participants to the role of **"buyers"**. Therefore, you will either be a buyer or a seller. Your role as seller or buyer will remain the same throughout the experiment. You will only know your own role and not the roles of other participants.

Overview

The experiment consists of 3 practice rounds and 27 paying rounds. At the beginning of each round, all buyers and sellers trade a fictitious good in a **market place**. As a buyer, you can buy units of the fictitious good and as a seller you can sell units. You can earn ECU in the market place and your earnings depend on your decisions and the decisions of the other participants. Each unit sold will be subject to a per **unit tax of 10 ECU** for sellers. The tax rate is the same for all sellers and is due at the end of every third round. Details on the market place will be explained further below. All tax revenues paid by you and all other participants will be donated to the German Red Cross.

The Market Place

Basics

The market place is opened for two minutes at the beginning of each round. All buyers and sellers trade a fictitious good. In each market period, each **seller can sell two units** of the fictitious good and each **buyer can buy two units** of the good.

Units, costs, and values

If you are a seller, you will be given the **costs** for two units of a fictitious good at the beginning of the experiment. These units shall be denoted ”Unit 1” and ”Unit 2”, where Unit 1 costs less than Unit 2. The cost of these units to you is the same in all rounds. However, the cost of each seller's units will differ from the cost of other sellers' units. Each seller only knows her own costs.

If you are a buyer, you will be given the **values** for two units of a fictitious good at the beginning of the experiment. These units shall be denoted ”Unit 1” and ”Unit 2” where Unit 1 values more than Unit 2. The value of these units to you is the same in all rounds. However, the value of each buyer's units will differ from the value of other buyers' units. Each buyer only knows her own values.

Asks, Bids, and Transactions

Sellers can make ”asks” and Buyers can make ”bids” during the trading period. All asks and bids are visible to everyone through the screen that appears during the two minutes of trading. This screen will also state your type (Seller or Buyer), the time left in the trading period and the costs or values that you were assigned for each Unit. Each Seller can first sell Unit 1 and afterward Unit 2. Accordingly, Buyers can first buy Unit 1 and then Unit 2.

Sellers cannot sell goods at prices lower than the assigned cost for the respective Unit.
Buyers cannot buy at prices that exceed their assigned value for the respective Unit.

Sellers can make asks at any time during the trading period but each ask has to be lower than the current lowest ask on the market. Similarly, Buyers can always propose bids as long as they are larger than the current largest bid on the market.

To realize a **transaction**, Sellers can either accept a bid or buyers can accept an ask. The transaction price for the unit will then be equal to the accepted ask or bid.

(Gross) Earnings in the Market Place

Units that are not traded do not yield any earnings. Gross earnings for each Unit are as follows:

For Sellers:
Gross Earnings from selling Unit 1 = transaction price of Unit 1 - cost of Unit 1
Gross Earnings from selling Unit 2 = transaction price of Unit 2 - cost of Unit 2

For Buyers:
Gross Earnings from buying Unit 1 = value of Unit 1 - transaction price of Unit 1
Gross Earnings from buying Unit 2 = value of Unit 2 - transaction price of Unit 2

Screenshots from trading market

Sellers:

Here Screenshot Sellers

Sellers can accept a current bid by pressing "Sell at this Price". To make a new ask, Sellers have to enter their ask price into the field to the right of the "Make a smaller ask" button and press the button to submit the ask.

Buyers:

Here Screenshot Buyers

Buyers can accept the current ask by pressing "Buy at this Price". To make a new bid, Buyers have to enter their bid into the field to the right of the "Make a smaller bid" and press the button to submit the bid.

Calculation of Net Income for Sellers
After three consecutive trading periods, the screen shows how many units of the fictitious unit you have traded over the previous three rounds and the resulting gross income from the previous three periods. For each unit traded in the three previous periods, a per-unit tax of **10 ECU is due for sellers**
Therefore, a seller's payment – the net income – , consists of her sum of all gross earnings from the three previous rounds, (henceforth denoted "sum gross income") minus the tax payment. The tax payment is the number of units sold over the previous three periods multiplied by the tax rate of 10 ECU. Hence:

Net Income = sum gross income - (number of units sold in previous 3 rounds * per-unit tax rate)

After every third round, sellers are informed about the net income that they earned over the previous three periods.

Payment

The first 3 rounds serve as practice rounds, in which you cannot earn money. The subsequent 27 rounds are paying rounds.

Buyers do not pay taxes so that gross earnings equal net earnings. A buyer's payoff hence equals the sum of gross earnings from all 27 trading periods.

Sellers receive a payoff that consists of the sum of all net incomes, each of which is earned after every third paying round (i.e., after paying rounds 3, 6, 9, 12, 15, 18, 21, 24, 27.)

You will be paid the payoff in cash at the end of the experiment. Additionally, each participant receives a show-up fee of 2.50 Euro. If the sum of all gross or net incomes is negative or zero, you will be paid the show-up fee; that is, you cannot make losses and will earn a minimum amount of 2.50 Euro.

Final Remarks

After the completion of all 30 rounds – 3 practice round plus 27 paying rounds – the experiment is finished. You will be asked to complete a short questionnaire at the end of the experiment while we prepare the payments. All information collected through this questionnaire, just like all data gathered during the experiment, are anonymous and exclusively used for scientific purposes. After you have completed the questionnaire, please remain seated at your booth until we call you to come up front to pick up your payment.

3.9.2 Evasion opportunity treatment group

Welcome and thank you for participating in our experiment. From now on until the end of the experiment, please refrain from communicating with other participants. If you do not abide by this rule, we will have to exclude you from the experiment.

We kindly ask you to read the instructions thoroughly. If you have any questions after reading the instructions or during the experiment, please raise your hand and one of the instructors will come to you and answer your question in person. Your payment and your decisions throughout the experiment will be treated confidentially.

You can earn money in this experiment. How much you earn depends on your decisions and the decisions of other participants. During the experiment, your payments will be calculated in a virtual currency: Experimental Currency Units (ECU). **30 ECU correspond to 1 Euro**. After the experiment, your pay-off will be converted to Euro and given to you in cash. Additionally, you will receive a show-up fee of 2.50 Euro.

The Experiment

Roles

At the beginning of the experiment, the computer will randomly assign five participants to the role of **"sellers"** and five other participants to the role of **"buyers"**. Therefore, you will either be a buyer or a seller. Your role as seller or buyer will remain the same throughout the experiment. You will only know your own role and not the roles of other participants.

Overview

The experiment consists of 3 practice rounds and 27 paying rounds. At the beginning of each round, all buyers and sellers trade a fictitious good in a **market place**. As a buyer, you can buy units of the fictitious good and as a seller you can sell units. You can earn ECU in the market place and your earnings depend on your decisions and the decisions of the other participants. Each unit sold will be subject to a per **unit tax of 10 ECU** for sellers. The tax rate is the same for all sellers and is due at the end of every third round. At the end of every third round, sellers are asked to report the number of units that they sold in the previous three market rounds. There is a 10% chance that the reported decision will be checked for accuracy. Details on the market place will be explained further below. All tax revenues paid by you and all other participants will be donated to the German Red Cross.

The Market Place

Basics

The market place is opened for two minutes at the beginning of each round. All buyers and sellers trade a fictitious good. In each market period, each **seller can sell two units** of the fictitious good and each **buyer can buy two units** of the good.

Units, costs, and values

If you are a seller, you will be given the **costs** for two units of a fictitious good at the beginning of the experiment. These units shall be denoted "Unit 1" and "Unit 2", where

Unit 1 costs less than Unit 2. The cost of these units to you is the same in all rounds. However, the cost of each seller's units will differ from the cost of other sellers' units. Each seller only knows her own costs.

If you are a buyer, you will be given the **values** for two units of a fictitious good at the beginning of the experiment. These units shall be denoted ”Unit 1” and ”Unit 2” where Unit 1 values more than Unit 2. The value of these units to you is the same in all rounds. However, the value of each buyer's units will differ from the value of other buyers' units. Each buyer only knows her own values.

Asks, Bids, and Transactions

Sellers can make ”asks” and Buyers can make ”bids” during the trading period. All asks and bids are visible to everyone through the screen that appears during the two minutes of trading. This screen will also state your type (Seller or Buyer), the time left in the trading period and the costs or values that you were assigned for each Unit. Each Seller can first sell Unit 1 and afterward Unit 2. Accordingly, Buyers can first buy Unit 1 and then Unit 2.

Sellers cannot sell goods at prices lower than the assigned cost for the respective Unit.
Buyers cannot buy at prices that exceed their assigned value for the respective Unit.

Sellers can make asks at any time during the trading period but each ask has to be lower than the current lowest ask on the market. Similarly, Buyers can always propose bids as long as they are larger than the current largest bid on the market.

To realize a **transaction**, Sellers can either accept a bid or buyers can accept an ask. The transaction price for the unit will then be equal to the accepted ask or bid.

(Gross) Earnings in the Market Place

Units that are not traded do not yield any earnings. Gross earnings for each Unit are as follows:

For Sellers:

Gross Earnings from selling Unit 1 = transaction price of Unit 1 - cost of Unit 1
Gross Earnings from selling Unit 2 = transaction price of Unit 2 - cost of Unit 2

For Buyers:

Gross Earnings from buying Unit 1 = value of Unit 1 - transaction price of Unit 1
Gross Earnings from buying Unit 2 = value of Unit 2 - transaction price of Unit 2

Screenshots from trading market

Sellers:

Here Screenshot Sellers

Sellers can accept a current bid by pressing "Sell at this Price". To make a new ask, Sellers have to enter their ask price into the field to the right of the "Make a smaller ask" button and press the button to submit the ask.

Buyers:

Here Screenshot Buyers

Buyers can accept the current ask by pressing "Buy at this Price". To make a new bid, Buyers have to enter their bid into the field to the right of the "Make a smaller bid" and press the button to submit the bid.

The Reporting Decision for Sellers

After three consecutive trading periods, you will be shown the number of units traded over the three previous trading rounds and the respective gross earnings on those units. For each unit traded in the three previous periods, a per-unit tax of **10 ECU is due for sellers**.

Sellers will then be asked to report the number of units sold in the previous three rounds for tax purposes. The reported amount may be between zero and the total number of units that were actually sold over the previous three rounds. After the reporting decision is submitted by pressing the "OK" button, the computer will determine if it is checked whether the reported number equals the actual number of units sold over the last three periods. The computer makes this call by randomly selecting an integer number between 1 and 10. The reporting decision will **only** be checked if the computer selects the number 1. Therefore, there is a random chance of 10% that the reporting decision will be checked.

Calculation of Net Income for Sellers

Sellers will be informed of the outcome of the random draw, and will be faced with one of the following two scenarios:

1. Computer selects a number between 2 and 10 (2, 3, 4, 5, 6, 7, 8, 9 or 10):

The reporting decision will *not* be checked. A seller's earnings after taxes – the net income –, in this case, consists of the sum of all her gross earnings from the three previous periods (henceforth denoted ”sum gross income”) minus the tax payment. The tax payment is the **reported** number of units sold multiplied by the tax rate of 10 ECU. Hence:

<u>Net income</u> = sum gross income - (reported number of units sold * per unit tax rate)

2. Computer selects number 1:
The reporting decision *will* be checked. A seller's earnings after taxes – the net income –, in this case, consist of sum of all her gross earnings from the three previous periods (henceforth denoted ”sum gross income”) minus the tax payment. The tax payment is based on the number of units **actually** sold over the last three periods. If the number of units was **not** reported correctly, a seller will additionally have to pay a penalty that is equal to the amount of **tax liability that was not paid**. Hence:

<u>Net income</u> = sum gross income - (actual number of units sold * per unit tax rate) - (number of units *not reported* * per unit tax rate)

<u>Payment</u>
The first 3 rounds serve as practice rounds, in which you cannot earn money. The subsequent 27 rounds are paying rounds.
Buyers do not pay taxes so that gross earnings equal net earnings. A buyer's payoff hence equals the sum of gross earnings from all 27 trading periods.
Sellers receive a payoff that consists of the sum of all net incomes, each of which is earned after every third paying round (i.e., after paying rounds 3, 6, 9, 12, 15, 18, 21, 24, 27.)

You will be paid the payoff in cash at the end of the experiment. Additionally, each participant receives a show-up fee of 2.50 Euro. If the sum of all gross or net incomes is negative or zero, you will be paid the show-up fee; that is, you cannot make losses and will earn a minimum amount of 2.50 Euro.

<u>Final Remarks</u>
After the completion of all 30 rounds – 3 practice round plus 27 paying rounds – the experiment is finished. You will be asked to complete a short questionnaire at the end of the experiment while we prepare the payments. All information collected through this questionnaire, just like all data gathered during the experiment, are anonymous and exclusively used for scientific purposes. After you have completed the questionnaire, please remain seated at your booth until we call you to come up front to pick up your payment.

Chapter 4

Experimental evidence on the relationship between tax evasion opportunities and labor supply

4.1 Introduction

Labor supply elasticities observed in empirical analyses are usually smaller than responses along other margins (Slemrod 1994) and are often heterogeneously distributed across different types of workers.[1] While there are several reasons why this is the case, this paper explores one possible explanation: access to tax evasion opportunities. It is well known that access to evasion opportunities varies across workers, thus making it easier for some workers to hide income relative to other workers. For example, whereas many wage earners are subject to third-party reporting, rendering tax evasion almost impossible, the self-employed and workers in industries that rely on cash payments have considerable access to evasion. The objective of the present paper is to test whether these differences in evasion opportunities affect the responsiveness of labor supply to changes in tax rates.

We use the theoretical framework of Pencavel (1979) to show that the responsiveness of labor supply to taxes is likely to vary with the opportunity to evade because workers with evasion opportunities are able to adjust their taxable income by exploiting two inter-related margins: labor supply and evasion. In particular, the theoretical framework suggests that while both types of workers respond to tax rate changes via standard income and substitution effects, workers with access to evasion are additionally affected by evasion-induced effects. As a result, evaders' labor supply response to tax changes is expected to differ from that of non-evaders. However, opposing income and substitution

[1] For example, there is evidence that labor supply elasticities vary by gender, with relatively larger estimates for females (Keane 2011), tend to decrease in income (Roed and Strom 2002; Aaberge and Colombino 2013) and vary by marital status (Bargain et al. 2014).

effects prevent us from obtaining any clear predictions about the relative magnitude of the effect of access to evasion on the labor supply response.

We rely on an empirical approach to answer the research question because of this ambiguous theoretical effect. This is achieved by combining well-established laboratory experimental designs from the tax evasion and labor literatures. In each of ten rounds, 205 subjects first complete a real-effort labor task, as in Gill and Prowse (2012), and then make a tax payment.[2] We vary the tax rate across rounds in three specific ways that include both tax increases and tax decreases; each variation constitutes a tax evolution. Subjects are randomly assigned to tax evolutions, as well as treatment and control groups that are identical in every respect except access to evasion; while subjects in the treatment group are able to evade taxes by underreporting their earned income, subjects in the control group cannot. Following Fortin et al. (2007) and Alm et al. (2009), for example, subjects in the treatment group face an exogenous audit probability and penalty if they are detected. This experimental design allows us to determine if the responsiveness of effort supply with respect to tax rates varies between the two groups. Since access to evasion is the only difference between the two groups, differences in the responsiveness to tax rate changes can be attributed to the difference in evasion opportunity.

Reliance on experimental methods to answer our research question is motivated by the near impossibility of answering this question with observational data. Tax evasion opportunities are hardly observable[3], and the standard labor supply elasticities are usually difficult to estimate. Additionally, even if we had good information on evasion opportunities and labor supply responses, clean identification would require us to solve self-selection into jobs with greater access to evasion. The advantage of using an experimental approach is that we are able to randomly assign subjects to treatment states and control the variables of interest, which allows us to cleanly identify the effect of evasion opportunities on labor responses to taxation. Using economic laboratory experimental techniques to analyze supply of effort and tax evasion is not new; see Charness and Kuhn (2011) for a comprehensive survey of the labor effort literature and Graetz et al. (1986), Alm, Jackson, and McKee (1992a), Fortin et al. (2007) and Alm et al. (2009) for tax evasion examples.[4]

[2]Modeling labor effort instead of labor supply is common in the laboratory experimental literature on labor supply and is usually done because of the difficulty in modeling labor supply as measured by hours of work; see Dickinson (1999), Falk and Fehr (2003) and Charness and Kuhn (2011) for references and discussion. This issue is discussed further in sections 4.2.4 and 4.5.3.

[3]Slemrod and Weber (2012) survey the empirical tax evasion literature and conclude that it is difficult to obtain credible causal evidence in observational studies. Experiments are one possibility for approaching tax evasion issues in a causal manner.

[4]However, in contrast to our work, most experimental contributions in this field look at the amount of evasion as the outcome of interest. Andreoni et al. (1998) and Torgler (2002) provide surveys on tax compliance in experiments. Additionally, our work is different from studies that look at the effect of changes in gross wages, net wages and tax rates; see section 4.5.2 for a more detailed discussion.

We find evidence that access to evasion affects the extent to which individuals' labor efforts respond to changes in the net-of-tax-rate, and that this effect depends on the the institutional setting regarding how tax rates evolve across rounds. The estimated treatment effect, i.e. the difference between the two groups' effort response to a change in the net-of-tax rate, is negative when subjects experience tax increases followed by a tax decrease. On the other hand, the estimated treatment effect is positive when subjects experience tax decreases followed by a tax increase. We also find that the treatment effect is more obvious for tax decreases than for tax increases. Finally, we find that the elasticity of taxable income is considerably higher in the treatment group relative to the control group. The internal validity of the experiment, randomization of subjects into groups, and lack of evidence that subjects did not understand the incentives of the experiment all point to causal treatment effects.

We are among the first to empirically examine the labor supply implications of the observed evidence that tax evasion opportunities are heterogeneously distributed across workers. Whereas most studies based on the seminal theory of Allingham and Sandmo (1972) assume that all taxpayers operate in an environment in which underreporting is available, more recent studies contradict this view. For example, Slemrod (2007) and Kleven et al. (2011) find evidence of almost no evasion among individuals subject to third-party reporting but substantial evasion among the self-employed.[5]

Our paper is related to the literature at the intersection of labor supply and tax evasion. As opposed to theoretical (Sandmo 1981; Cowell 1985) and empirical (Lemieux et al. 1994; Frederiksen et al. 2005) contributions that compare formal and informal labor markets, we compare formal labor markets that have two distinct levels of access to evasion. In this respect, our paper is more in line with the theoretical contributions of Pencavel (1979) and Slemrod (2001), who extend the standard labor supply model with taxes to account for tax evasion and avoidance, respectively. Both papers provide theoretical evidence that the ability to reduce one's tax liability through legal or illegal means affects labor supply decisions.

However, the paper most closely related to ours is Collins et al. (1992), which uses a laboratory experimental approach to measure the effect of access to evasion opportunities on the change in labor effort when individuals move from a system with no taxation to a system with either a proportional, "mildly progressive" or "steeply progressive" tax system. In other words, their experiment focuses on the progressivity of the tax structure. Their results indicate that the opportunity to evade has a positive effect on labor effort

[5]Other studies show that bunching around kinks in the tax schedule is mostly prevalent among the self-employed, allowing for the interpretation that other types of workers simply do not have the opportunity to adjust their taxable income due to lack of evasion opportunities (Saez 2010; Bastani and Selin 2014; Chetty et al. 2013). Additionally, the tax morale literature shows that the intrinsic willingness to pay taxes is considerably lower among the self-employed relative to wage earners (e.g., Alm and Torgler 2006 and Konrad and Qari 2012).

when taxes are proportional or steeply progressive and a negative effect when taxes are mildly progressive. However, it is difficult to tell what the subjects are responding to when the tax regimes in Collins et al. (1992) are implemented; are they responding to the fact that they must now pay taxes, the progressivity of the schedule, the multiple brackets and rates, or the top rate? As a result, the differential effect in effort observed across compliance groups cannot be interpreted as the causal effect of a change in tax rates. Unlike Collins et al. (1992), we use a flat tax rate that varies across rounds in all sessions. We are therefore the first to directly and cleanly measure how the labor effort response to tax rates varies with access to evasion.

The remainder of the paper is organized as follows. Section 4.2 describes the experimental design and set-up. We discuss theoretical predictions in section 4.3 and present the empirical strategy and results in section 4.4. We discuss the empirical findings and issues of external validity in section 4.5. Section 4.6 concludes the paper.

4.2 Experimental Design

In the experiment we designed, subjects earn income by completing a labor task and then pay taxes on their income. In order to answer our research question, one group of subjects is given the opportunity to underreport their income, while the other is not. We refer to the group that is given the opportunity to evade as the treatment group; the group that has no evasion opportunity is called the control group. A detailed description of the experimental design is provided below.

4.2.1 Organization

The experiment was conducted in the Cologne Laboratory for Economic Research (CLER), University of Cologne, Germany. All subjects in the laboratory's subject pool of approximately 4000 persons were invited via email—using the recruitment software *ORSEE* (Greiner 2004)—to participate in the experiment. Potential participants could sign up on a first-come-first-serve basis. A total of 205 subjects, mostly undergraduate students from the University of Cologne, participated in our experiment (see section 4.4.2 for summary statistics). Neither the content of the experiment nor the expected payoff was stated in the invitation email. The computerized experiment was programmed utilizing *z-tree* software (Fischbacher 2007).

We conducted 14 sessions over five school days in June 2012 and July 2013. Each session included one practice round, 10 paying rounds and 15 subjects; each lasted approximately 70 minutes on average (including the review of instructions and payment of participants). Random assignment to computer boothes was implemented by asking each subject to draw an ID number out of a box upon entering the lab. The subjects' decisions and payments were linked to their ID; the experimenter had no way of matching this information to their names. Subjects also received a hard copy of the instructions when they entered the lab (See appendix 4.10) and were allowed as much time as they needed to familiarize themselves with the procedure of the experiment. They were then given the opportunity to ask any clarifying questions.[6]

4.2.2 Treatment Effect and Variation in Tax Rates

The 14 experimental sessions consisted of seven treatment sessions in which underreporting was available and seven control sessions with no option to underreport income. In order to identify the differential response to varying taxes between the control and treatment group, we require variation in tax rates. In six sessions—three treatment and three control—the tax rate was set at 15% in the first three paying rounds, 35% in rounds 4 to 6, 50% in rounds 7 to 9 and 15% in the last round. We use two additional tax evolutions in order to check whether the treatment effect depends on the way tax rate changes across rounds.[7] The second evolution of tax rates, which we use in two treatment and two control sessions, was 50% in rounds 1 to 3, 35% in rounds 4 to 6, 15% in rounds 7 to 9 and 50% in the last round. In the final four sessions—two treatment and two control—the tax rate evolved as follows: 35% in rounds 1 to 3, 50% in rounds 4 to 6, 15% in rounds 7 to 9 and 35% in the last round. The tax rate was 15% in the practice round of all sessions. Table 4.1 provides and overview of the variation of the treatment status and tax rates.

4.2.3 Overview of a single Round

At the beginning of each round, participants in both the treatment and the control groups were told the tax rate for that round. In the instructions, participants were told that the "tax rate may, but does not have to, vary from round to round". Therefore, they did not know the tax rate for a given round until the beginning of that round. This option was chosen because we did not want subjects' labor effort decision in round t to be influenced by the tax rate in round $t + 1$.

Each round had two stages in the *treatment group*: a labor task and a reporting decision. In the *labor task* stage, subjects undertook a real-effort task and earned money

[6] Doerrenberg and Duncan (2014a) use parts of the same data to study the distributional implications of tax evasion. In particular, they show that tax evasion drives a wedge between observed and actual income inequality, which implies that making tax policy decisions on the basis of observed income inequality can be misleading.

[7] This also allows us to separate the tax effect from possible learning effects and thus identify the of tax rate changes in the control group. However, it is possible to identify the treatment effect in the presence of a correlation between tax rates and learning as long as learning follows the same path in both groups.

Table 4.1: Overview of Treatments and Tax Rates

Tax	Tax Rates (%):				Number of Participants		
Evolution	1-3	4-6	7-9	10	Treat	Control	Overall
1	15	35	50	15	45	45	90
2	50	35	15	50	30	28	58
3	35	50	15	35	30	27	57
Total Number of Participants:					105	100	205

Tax rates in rounds 1-3, 4-6, 7-9 and 10, as well as number of participants per group for each of the three tax evolutions. 14 sessions were conducted at the Cologne Laboratory for Economic Research in June 2012 (Tax Evolution 1) and July 2013 (Tax Evolutions 2 and 3). Each participant was either in the treatment or control group. Five subjects did not show up for the control sessions in tax evolutions 2 and 3.

depending on their performance on the task. Their gross-income in each round consisted of their labor income plus a fixed amount of non-labor income.[8] In the *reporting decision* stage, subjects in the treatment group had to report an income amount between zero and their true earned gross income; the reported amount was taxed at the tax rate for that round. These individuals faced an exogenous audit probability of ten percent and paid a fine equal to twice evaded taxes if audited. The audit probability and fine rate were fixed across rounds, and participants were given this information at the beginning of the experiment.

Participants in the *control group* undertook the same labor task as the treatment group, but were not given a reporting decision. That is, they first performed the labor task and earned money based on their performance. Their gross income, earned money from the labor task plus the fixed amount of non-labor income, was then automatically taxed at the tax rate for that round. The tax payment was withdrawn and subjects were informed of their net income for the round. The two stages are explained in more detail in the following two sections.

4.2.4 The Labor Task Stage

The labor task, which was completed by both groups in the first stage of each round, was designed by Gill and Prowse (2012) and involves moving a set of sliders across a computer screen (see figure 4.4 in the appendix for a screenshot of the task).[9] The sliders

[8]Although every subject has two minutes to earn labor income by completing the labor task, we provide each subject with non-labor income to ensure that every subject has an income reporting decision in every round. This design feature mimics income generating processes in the real world since individuals generally have labor and non-labor income.

[9]Gill and Prowse (2013) provide details and show how to implement the slider task. It has been used widely since its introduction: Riener and Wiederhold (2011), Gill and Prowse (2012), Hammermann

were initially positioned at zero and could be repositioned to any integer between 0 and 100, inclusive. Subjects received feedback on the current position of the slider only, which was indicated at the rightmost end of each slider. The arrow keys on the key board were disabled to ensure the subjects only used the left mouse key to complete the task; use of the arrow keys makes the task trivial. For each round, subjects were given 2 minutes to align 48 sliders at position 50. As in Gill and Prowse (2012), “the 48 sliders are arranged on the screen so that no two sliders are aligned exactly one under the other." This prevented subjects from positioning one slider at 50 and then visually matching the other sliders at this position. The number of correctly aligned sliders was taken as a measure of labor effort.

The slider task has a number of advantages, which are described in Gill and Prowse (2012, 2013). It is easy to explain and implement, does not require prior knowledge, does not allow guessing, and is identical across rounds, treatments and subjects. Most importantly, the slider task generates a lot of variation, and it is virtually impossible to move all sliders correctly within the time frame of two minutes. As a result, performance on the task can be interpreted as labor effort. It is important to note that we do not measure hours worked nor do we give subjects the opportunity to substitute “off-the-job” leisure (e.g., taking a day off to go to the beach) for effort. This is similar to labor markets in the “real world” that require individuals to work a fixed number of hours each day. Individuals working under these types of contracts often respond to market incentives by changing jobs or by changing their effort on the job. For example, Dickinson (1999) notes that workers can substitute effort for on-the-job leisure such as relaxing in their chairs without working. Therefore, our experiment provides participants the opportunity to enjoy on-the-job leisure. We acknowledge that our measure of labor effort cannot be generalized to more traditional measures of labor supply, such as hours of work, without caution. However, the choices between labor effort and hours of work share many characteristics and are isomorphic in many ways, so that our measure of labor effort can serve as a good proxy for labor supply (Charness and Kuhn 2011).

4.2.5 The Reporting Decision Stage

Subjects in the treatment group were informed of their gross income and were asked to make an income reporting decision for tax purposes in the second stage of the experiment. The language used in the instructions and on the computer screen explicitly stated that reported income would be taxed and the instructions also disclosed that all tax revenues would be donated to the German Red Cross. Donating tax revenues to the Red Cross may

et al. (2012), Gill and Prowse (2014), Gill et al. (2013) and Cettolin and Riedl (2013), among others. Djawadi and Fahr (2012) also use the slider task in the context of tax compliance, but they examine a different research question than we do and employ compliance as the dependent variable.

be viewed as public good, which has been shown to have a positive effect on willingness to pay taxes.[10] To the extent that subjects view the Red Cross as benevolent, they may evade less than they would if the money was donated to a less acknowledged organization. This implies that we derive lower bound estimates of the treatment effect.

As is standard in the experimental tax evasion literature, reported income could be any amount between 0 and true gross income. After the reporting decision was complete, student helpers walked up to each computer booth with a 10-sided die, which each subject threw. The student helper then entered the number thrown on the die in the computer, which then determined the subject's audit status and calculated his or her tax liability, penalty, if any, and net income.[11] Allowing subjects to throw a 10-sided die like we do is one way of ensuring that the evasion opportunity is credible. In other words, subjects are more likely to believe that their audit status is random if their audit status is determined manually with a die as compared to automatic randomization with a computer. We acknowledge that the presence of student helpers in the lab may yield an "experimenter effect" that could potentially cause disturbances. However, the interaction between the student helper and the subjects was limited to throwing the die and wait at most 2 seconds for subjects to hit "enter". The student helper did communicate with the subjects and had no opportunity to view subjects' choices.[12]

Subjects who underreported income and threw a one were audited and had to pay a fine equal to twice their evaded taxes (i.e., the underreported amount multiplied by twice the tax rate). This implies an audit probability of 10 percent, which, together with the fine rate, is a commonly used penalty structure in the tax evasion literature Alm et al. (2009). All other subjects who either reported honestly or underreported but threw a die number between two and ten received a net income equal to true gross income less the tax rate multiplied by the reported gross income. Of course, the student helper could not see if a subject underreported his or her income because the student helper stepped away before the screen summarizing the round's payment appeared.

Unlike those in the treatment group, subjects in the control group did not have the opportunity to underreport their income and thus could not evade taxes. In other words, their tax liability was "withheld at source". A corollary in the "real" world would

---

[10]See footnote 25 and our conclusion for a discussion and references. We donated 541.69 EURO to the German Red Cross after all 14 sessions were completed.

[11]The exact procedure was as follows: The student helper entered the number showing on the face of the die; the subject confirmed this number by hitting "ENTER", which resulted in a pop-up screen with one of the following sentences: *You have been audited* or *You have not been audited*. The subject had to press "ENTER" again to see the screen summarizing the round's payment. By this time the student helper would have moved on to the next subject.

[12]It should be noted that the outcomes do not support the presence of an experimenter effect; the level of evasion is consistent with existing estimates of evasion in the literature (e.g., Fortin et al. 2007; Alm et al. 2009; Alm et al. 2010; Coricelli et al. 2010), and the supply of labor effort is similar to that found in other studies that use the slider task (e.g., Gill and Prowse 2012; Gill and Prowse 2014).

be individuals whose only source of income is labor earnings that are subject to third party reporting and withholding and who receive pre-populated tax forms. For example, taxpayers in Germany must enter a tax identification number into their tax forms when filing taxes at the end of the tax year. The tax form is then populated with income information that has been subjected to third-party withholding, such as labor income. Therefore if the only source of income is labor income subject to third-party reporting, there is no opportunity to underreport income.[13]

Since we are only interested in the extent to which the responsiveness of labor effort varies with evasion status, both groups faced an identical proportional tax rate that varied by round (see Table 4.1). Subjects were informed of the parameters of the tax system at the beginning of each round. Again, the only difference between treatment and control group was the ability to evade, and the only difference between rounds was the tax rate. The empirical analysis in section 4.4 exploits this between subject variation in treatment.

4.2.6 Payment

Subjects in both groups earned experimental currency units (ECU) throughout the experiment. Their *gross income* in each round had two parts: earned labor income and non-labor income. The non-labor income was set at five ECU per round and remained constant across subjects and sessions. The labor income was based on performance in the slider task, and subjects were paid six ECU for every correctly aligned slider. The *net income* in each round was a function of gross income, the tax rate, and, in the treatment group, the reporting decision and audit outcome.

After all rounds of a session were completed, each subject threw a 10-sided die to determine which of the 10 rounds she would receive payment for. Paying subjects their net income for a single randomly chosen round instead of the sum of net incomes across all rounds is advantageous because it allows us to avoid wealth effects, satiation and irrational decisions that generally occur once a certain expected payoff is achieved (see e.g. Blumkin et al. 2012 for a recent example that handles payment in a similar manner). Additionally, this payment method, along with the fact that subjects were not informed about the prevailing tax rate until the start of a round, avoids adding an intertemporal

[13] We acknowledge that taxpayers in Germany and most other countries who file taxes still have the opportunity to evade taxes by overstating their deductions. However, this is not an option in our experiment. Even if taxpayers have the opportunity to enter their income on tax forms as is customary in the United States, it would make little sense to underreport income that has been subject to third-party reporting since the tax authority has independent information on the correct labor earnings. In other words, the probability of detection is 100 percent for these individuals, which means that underreporting income for tax purposes will lead to lower net income than if gross income were reported accurately. Workers facing these conditions almost always report their true gross income. For example, Slemrod (2007) shows that income subject to third-party reporting in the United States has a compliance rate of over 90 percent. We designed the experiment to be similar to the institutional setting of Germany since this was where the experiment was implemented.

component to our experiment. Therefore, the experiment matches our static theoretical framework, and the differential effects that we derive ought to be interpreted as differences in Marshallian elasticities, as opposed to intertemporal labor substitution elasticities. The total ECU for the selected round was converted to EUROs using an exchange rate of 1 ECU to 0.1 EURO. A show-up fee of 2.5 EURO was added to the subject's net income to determine the final payment. Final payment ranged from 2.5 EURO to 25.2 EURO with an average of 10.11 EURO in the control group and 11.98 EURO in the treatment group.

4.3 Theoretical Framework

This section of the paper sets up a theoretical framework that matches the experimental design discussed in the previous section. We rely on this framework to highlight the main channels through which access to evasion affects the responsiveness of labor supply. The model is informed by Pencavel (1979), who merges the standard neo-classical labor supply model with the Allingham and Sandmo (1972) model of tax evasion. While social norms or preferences certainly matter for tax evasion decisions, we exclude such variables from our theoretical discussion because randomization in our experiment implies that they are balanced across groups and therefore do not have a differential effect on the treatment and control groups. We also make reference to results from the standard labor supply model with taxation but no evasion.

4.3.1 Model

Individuals make a labor supply decision L, which yields labor income wL and non-labor income M; w is the wage rate. They then report $R \leq (wL + M)$ to the tax authority to determine their tax liability. Reported income R is audited with probability p and, because we assume all audits lead to the full discovery of true income, a fine equal to twice the evaded taxes must be paid if audited. Assuming individuals consume all of their income implies:

$$Consumption = \begin{cases} C_a & = (wL + M) - \tau R - 2\tau(wL + M - R) \\ & = (wL + M)(1 - 2\tau) + \tau R \text{ with probability p} \\ C_n & = wL + M - \tau R \text{ with probability (1-p),} \end{cases} \tag{4.1}$$

where subscripts a and n indicate audited and not audited, respectively, and τ is the proportional tax rate.

As in Pencavel (1979), we assume individuals choose L and R to maximize an expected utility function that satisfies the standard assumptions of the neo-classical labor supply model; continuous, twice differentiable, and concave. We also assume the utility function is strongly separable in consumption and labor (Pencavel 1979). With these assumptions in mind, the individual maximization problem is specified as follows

$$\begin{aligned} maxEU = pU(C_a, L) + (1-p)U(C_n, L) \\ st. \\ C_a = (wL + M)(1 - 2\tau) + \tau R \\ C_n = wL + M - \tau R, \end{aligned} \tag{4.2}$$

This model reduces to the standard labor supply model with taxation if there is no opportunity to evade. Differentiating equation (4.2) with respect to R and L, respectively, yields first-order conditions, which can be used to determine the effect of evasion on the responsiveness of labor supply. Since the derivations are similar to those presented in Pencavel (1979), we discuss the implications here and leave a formal derivation of the main results for appendix 4.9.

4.3.2 Prediction

The derivations presented in appendix 4.9 identify two channels through which evasion affects the responsiveness of labor supply to tax rate changes. The first channel pertains to the standard income and substitution effects, which apply to both groups. An increase in the tax rate reduces the return on effort, which reduces effort via a substitution effect and increases effort via an income effect. Because these effects work in opposite directions, their net effect on the responsiveness of effort supply is ambiguous. These standard income and substitution effects are similar to those derived in the standard labor supply model, and they apply to both types of workers.

While the discussion above represents the total effect on workers who do not have tax evasion opportunities, there is a second channel for workers who are able to hide income from the tax authorities. For this group of workers, higher taxes are likely to lead to higher evasion (Alm 2012), which implies a higher net wage. This evasion-induced effect on net wages has income and substitution effects on the responsiveness of labor effort. Increased evasion implies higher net wage and thus higher income, which implies lower work effort via the income effect if leisure is a normal good. At the same time, the higher net wage implies that leisure is more costly, thus leading to a substitution towards greater effort. The signs of these evasion-induced income and substitution effects are not only opposite each other, but also opposite the signs of the standard income and substitution effects. Assuming leisure is a normal good, the evasion-induced income effect reinforces the standard substitution effect while the evasion-induced substitution effect reinforces the standard income effect. Therefore, the net effect on evaders depends on the size of evasion-induced effects relative to the standard effects.

Because the effect of taxes on both groups is ambiguous, it is not clear whether workers with evasion opportunities are more or less responsive than workers without such opportunities. It is also possible that the two opposing effects cancel each other so that there is no difference in responsiveness among workers with and without evasion opportunities. This ambiguity makes sense, since workers with evasion opportunities have two interrelated means of changing their taxable income in response to a change in the tax rate: work less and/or report less income. We are therefore forced to rely on an empirical analysis to answer our research question.

4.4 Empirical Strategy and Results

Recall that we are interested in identifying the differential effect of tax rates on participants with an evading opportunity relative to those without access to evasion. We begin this section by describing the empirical strategy that allows us to identify this treatment effect (Section: 4.4.1). The relevant summary statistics and empirical results (4.4.2) for the full sample of subjects are presented next, followed by results for each tax evolution.

4.4.1 Empirical Strategy

Participants in our experiment are randomly assigned to either the treatment or control group. Because the treatment status of a each participant is always the same, we employ a between-subjects design to identify our coefficient of interest. Furthermore, because we are only interested in the labor effort changes with respect to varying tax rates, we average (collapse) the decisions of each individual by rounds without tax rate changes.[14] Therefore, our data set has four observations per person: average behavior over rounds 1, 2, 3; over rounds 4, 5, 6; over rounds 7, 8, 9 and finally behavior in round 10. It follows that we compare average behavior across changes in the tax rate as is usually done in the ETI literature (Saez et al. 2012). Analyzing the data in this way also allows us to abstract from learning effects occurring within rounds for which there are no tax changes. It is further consistent with the notion that people adjust to the prevailing tax rate across periods (i.e., years) without tax changes, whereas policymakers are mostly interested in the effects of an actual change in the tax rate.

As our research question is ultimately about changes in labor effort in response to tax rates, we first generate a variable for each person that measures the changes in labor effort between periods t and $t-1$. Due to random assignment, any (non-parametric) difference in this generated variable between the treatment and control groups

[14]Recall that the tax rates in all sessions change after rounds three, six and nine.

will be a first test for the presence of treatment effects. We also test for treatment effects parametrically by regressing labor effort on the net-of-tax rate (NTR), which is defined as (1 - tax rate).[15] The estimated model is specified as follows:

$$L_{i,r,s} = \beta_0 + \beta_1(1-\tau_r) + \beta_2 T_s + \delta(1-\tau_r)T_s + \epsilon_{i,r,s}, \tag{4.3}$$

where $L_{i,r,s}$ indicates labor effort of individual i (with $i = 1, ..., 205$) in treatment state s (with $s = 0, 1$) and period r (with $r = 1, 2, 3, 4$). $(1 - \tau_r)$ stands for the net-of-tax rate NTR, and T_s is a dummy for the treatment state: either control (dummy is zero) or treatment (dummy is one) group. $\epsilon_{i,r,s}$ is a standard error term. Our coefficient of interest is δ, which represents the difference in the effect of taxes on supply of effort across the two groups. More precisely, δ indicates the effect of the NTR in the treatment group relative to the control group. Due to the random assignment of participants to the two groups, our experimental set-up allows us to *causally* identify the treatment effect given by the parameter δ.

We set up our data as a panel with four observations per individual and run random effects (GLS) regressions with standard errors clustered on individuals. Random effects regressions yield unbiased results because random selection into treatments and exogenous tax rates cause the error term to be uncorrelated with our explanatory variable of interest in each time period (Wooldridge 2010).[16] Obviously, because the treatment status of a single individual never changes, it is not possible to estimate the treatment effect in the presence of individual fixed effects.

4.4.2 Results

The following subsections provide summary statistics and then results for the full sample of 205 participants and each tax evolution separately.

Summary Statistics

Panel A of Table 4.2 presents summary statistics for several of the participants' demographic and attitudinal characteristics, separately for the treatment and control groups using the pooled sample. After the experiment, we surveyed tax morale using a question

[15]Estimating the effect of NTR rather than the tax rate is standard in both the labor supply literature (Keane 2011) and the public finance literature (Saez et al. 2012).

[16]Note that the labor decision is not censored from above, making use of estimators that allow for censoring, such as Tobit models, unnecessary. Table 4.2 depicts that the average number of correctly adjusted sliders was 18.22 (out of 48). The maximum number of correct positions was 37, suggesting that it was impossible to reach the upper bound of 48. Additionally, there is no evidence that the distribution of the number of correctly positioned sliders is skewed in any particular direction.

very similar to one used by the World Values Survey (Minkov 2012).[17] We also derived a measure of risk aversion by asking subjects to choose between a certain pay-off of $50 and a gamble that pays $100 with probability of 0.5 and $0 with probability of 0.5.

Casual observation of the data shows that randomization into the treatment states worked well. This is confirmed by non-parametric Wilcoxon rank-sum tests for differences in distributions between groups (Wilcoxon 1945; Mann and Whitney 1947);[18] we do not observe any statistically significant differences in demographic characteristics (gender, age, share of participants whose native language is German) and attitudinal variables (tax morale, risk aversion) across the two groups. While we do not explicitly measure other attitudinal variables such as social norms or preferences, randomization implies that these omitted variables are also balanced across groups and therefore do not have any effect on our results. We also find evidence that the tax evolutions are balanced with respect to individual characteristics, suggesting that randomization also worked well across tax evolutions (results available upon request).

Among all participants, approximately 46% were male , 78% indicated German to be their native language, and the average age was 24.6 years. The average tax morale is 7.11 on a scale of one to ten, with ten being the highest level of morale.

Full Sample

Non-parametric Results Panel B of Table 4.2 reports summary statistics for the two choice variables in our experiment: labor effort and evasion decision. We find that participants in the treatment group failed to report roughly 73% of their gross income on average. This level of evasion is at the high end of estimates in the tax evasion literature, which finds noncompliance rates from 30 percent to 78 percent (Fortin et al. 2007; Alm et al. 2009; Alm et al. 2010; Coricelli et al. 2010). Additionally (not reported in the Table), we find that the evasion share is positively correlated with tax rates. The simple correlation coefficient is 0.091 and the regression coefficient from regressing evasion share on the tax rate controlling for period effects is 0.002; both are statistically different from zero.

Although the mean level of effort differs slightly between the treatment group (18.51) and the control group (17.92), the difference is not statistically different from zero (p-value Wilcoxon rank-sum test: 0.238). A similar story holds for the mean change in

[17]"Please tell me for the following statement whether you think it can always be justified, never be justified, or something in between: 'Cheating on taxes if you have the chance'." This is the most frequently used question to measure tax morale in observational studies (e.g., Slemrod 2003, Alm and Torgler 2006, Halla 2012 and Doerrenberg and Peichl 2013)

[18]A difference in distributions usually implies a difference in means. The Wilcoxon rank-sum test is basically the non-parametric equivalent to a t-test for differences in means. It is the most common test in experimental economics to test for differences between treatment states.

Table 4.2: Summary Statistics

	Control		Treatment			
Variable	Mean	Std. Dev.	Mean	Std. Dev.	$p > \|z\|$	N
Panel A: Characteristics						
Male	0.46	0.50	0.47	0.50	0.942	205
Age	24.07	3.58	25.00	6.15	0.656	205
German native	0.78	0.42	0.77	0.42	0.883	205
Tax Morale	7.26	2.71	6.97	2.68	0.409	205
Risk Aversion	1.28	0.64	1.33	0.70	0.664	205
Panel B: Choice Variables						
Effort	17.92	3.85	18.51	4.13	0.238	205
Effort Change	0.98	1.01	1.01	1.03	0.992	205
Share Evaded	/	/	0.73	0.33	/	105

Means and Standard Deviations of demographic and attitudinal variables by Treatment Status. *Male* and *German native* are dummy variables. *Effort* indicates the number of correctly positioned slides. *Share evaded* is the share of gross income that has not been reported for tax purposes. *Effort change* is the difference in correctly positioned sliders between period P and the previous period, $P-1$. $p > |z|$ reports the p-value of a (non-parametric) Wilcoxon rank-sum test for differences in distributions (H0: no differences) between the two treatment groups (Wilcoxon 1945; Mann and Whitney 1947). N indicates the number of observations in both treatments.

effort, but this simple difference likely masks variation in the difference between rounds. We provide a more rigorous test for the treatment effect by regressing effort on the NTR.

Regression Results The first column of Table 4.3 presents results from estimating equation (4.3) using the full sample. We find that the net-of-tax rate has a smaller effect on effort in the treatment group than the control group. The effect of a 10-percentage point increase in the net-of-tax rate is smaller by 0.08 sliders. Although marginally economically meaningful, the effect is not statistically different from zero.

As a result of randomization, equation (4.3) identifies the causal impact of interest without any need for further adjustments. However, identifying δ using equation (4.3) implies that participants in the two treatment states make the same progress in learning how to play the effort task because tax rates and learning are marginally correlated. In order to ensure that different learning abilities across the two groups do not confound our results, we extend the baseline model to include a learning trend that takes the values 1, 2, 3 or 4 for each period. We also run specifications including the logged learning trend to account for potentially diminishing learning trend. These results, which are presented in columns two and three, respectively, of Table 4.3, show that the treatment effect is not driven by learning effects. Finally, column four of Table 4.3 shows that the estimates are not driven by individual characteristics such as age, gender, a dummy for German being the native language, risk aversion, tax morale, or a dummy for the time of day at which a participant's session was run (morning or afternoon).[19] Although the control variables are jointly significant (Wald test of joint significance: $p < 0.000$), the fact that the results of interest are not altered by controlling for these variables provides further evidence that randomization worked well and that the results are internally valid.

Table 4.3: Regression of Effort on Taxes. Full Sample

	I	II	III	IV
NTR ∗ Treat	−0.837	−0.892	−0.926	−0.890
	(0.935)	(0.890)	(0.888)	(0.892)
NTR	0.288	−0.276	0.042	0.015
	(0.703)	(0.688)	(0.705)	(0.712)
Treatment Status	1.162	1.199	1.222	1.444*
	(0.883)	(0.848)	(0.832)	(0.810)
Trend (linear)		0.987***		
		(0.068)		
Trend (log)			2.146***	2.146***
			(0.142)	(0.142)
constant	17.721***	15.634***	16.182***	21.891***
	(0.629)	(0.612)	(0.606)	(1.727)
Controls	No	No	No	Yes
Panel Obs	820	820	820	820
Indiv. Obs	205	205	205	205
Periods	4	4	4	4
R2	0.01	0.06	0.07	0.21

Random Effects Regressions for the full Sample. Dependent Variable is *Effort*. Standard Errors in parentheses are clustered by Individual. *Treated* is a dummy variable taking "1" if a participant is in the group with evasion opportunity (treated) and "0" if not. *NTR* (net-of-tax rate) is defined as (1 - Tax Rate). Specification IV includes controls for age, gender, a dummy indicating native German, risk aversion, tax morale and time of day. Significance levels: $* < 0.10$, $** < 0.05$, $*** < 0.01$.

Besides looking at the variables in levels, we also run models in which we use the logged levels of labor effort $L_{i,r,s}$ and $(1 - \tau_r)$. That is, we additionally identify the differences in labor effort elasticities with respect to the NTR across the two treatment

[19] To the extent that individuals who show up in the evening are systematically different from those who show up in the morning, one might expect that this difference is driving our treatment effect. A similar concern might also be expressed for native Germans, gender, age, tax morale and risk aversion. Note however, that even if the response to taxation is affected by these characteristics, this is only a problem for us if the treatment and control groups differ with respect to the characteristics. In other words, randomization into groups is all that is needed for identification (Charness et al. 2012).

states. The results presented in Table 4.6 of Appendix 4.8 show qualitatively similar results for the estimated elasticity; the elasticity is smaller in treatment group, but the difference not statistically different from zero.

Although pooling the data like we do here makes sense, it is possible that it masks heterogenous responses to the various evolutions of the tax rate. In particular, it is possible that the response of taxpayers depends on the order in which tax rates change across rounds. We investigate this possibility by looking at each tax evolution separately in the following sections.

Tax Evolution Analysis

Effort trend Figures 4.1, 4.2 and 4.3 depict the evolution of effort over the course of all four rounds in each of the three tax evolutions. First, the figures show that there is an upward trend in effort for both groups, which suggests a learning effect (as also found in Gill and Prowse 2012). However, the figures also show that the slopes in each graph differ between the two groups, suggesting that the adjustment to changing tax rates depends on evasion opportunities. Additionally, the visual evidence suggests that the difference between the treatment states likely varies with tax evolution.

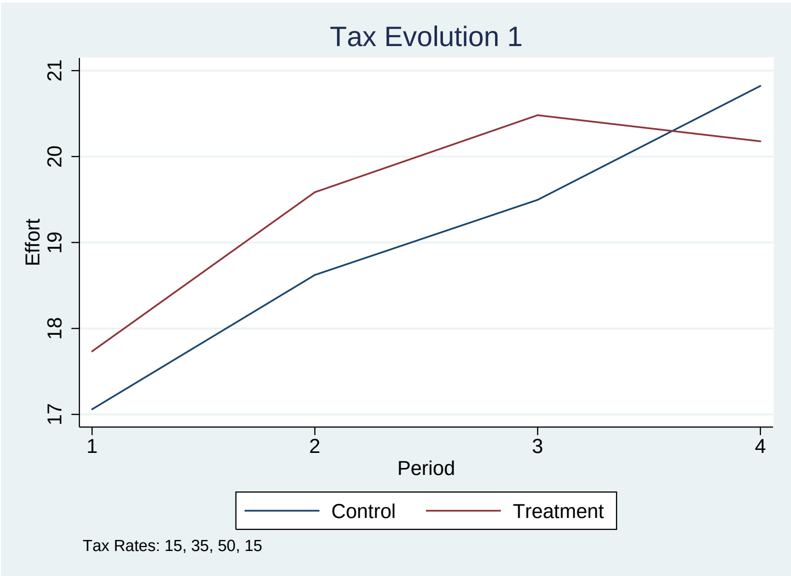


Figure 4.1: Effort by period and treatment status. Tax Rates: 15, 35, 50, 15

Non-parametric Comparison of Effort Adjustments A more careful look at the treatment effect is presented in Table 4.4 where we compare the differential changes in labor effort in response to tax changes between periods t and $t-1$ for each tax evolution and tax change. The results presented in panel A for tax evolution 1 show that the effort response to tax rate changes is significantly higher in the treatment group relative to the control group when the tax rate falls from 50% to 15%. This difference is statistically

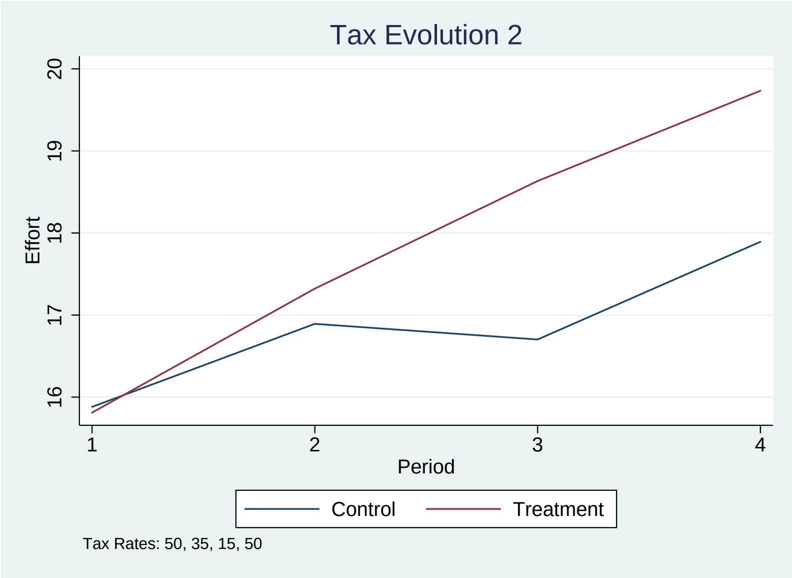


Figure 4.2: Effort by period and treatment status. Tax Rates: 50, 35, 15, 50

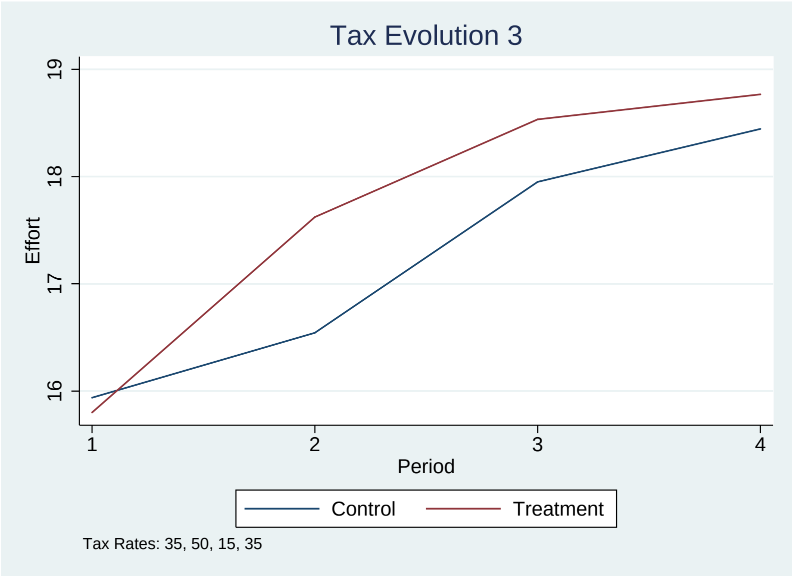


Figure 4.3: Effort by period and treatment status. Tax Rates: 35, 50, 15, 35

different from zero at the 2% level and is also visible in Figure 4.1. The results for tax evolution 2, which are presented in panel B of Table 4.4 and Figure 4.2, also show a difference in effort response between the two groups. Again, this effect is statistically different from zero at the 3% level when the tax rate falls from 35% to 15%. More importantly, though, the sign of the treatment effect is opposite of what we find in panel A. Finally, the results for tax evolution 3 in panel C show that the treatment group is more responsive when the tax rate increases from 35% to 50% and less responsive when the tax rate falls from 50% to 15%. The former effect is statistically different from zero at the 2-percent level while the latter is marginally significant at the 11-percent level. This result is demonstrated in Figure 4.3.

Table 4 also reveals three additional findings. First, we observe an asymmetric labor supply response to tax rate changes. For example, an increase in the tax rate from 15% to 35% has no effect in Panels A or C, while a decrease from 35% to 15% has a fairly large effect in Panel B. Similarly, an increase from 15% to 50% has no effect in Panel B while a decrease from 50% to 15% has a statistically significant effect in Panel A and is marginally significant in panel C. Additionally, an increase from 35% to 50% has a statistically significant effect in Panel C, while a decrease from 50% to 35% has no effect in Panel B. Second, this asymmetric labor supply response further suggests that subjects appear to be more responsive to tax decreases than tax increases. Three out of four tax decreases lead to a treatment effect that is either statistically or marginally statistically different from zero: 50% to 15% in Panels A and C, and 35% to 15% in Panel B. On the other hand, only one out of five tax increases triggered a statistically significant treatment effect: 35% to 50% in Panel C.

Finally, we observe that the direction of the treatment effect depends on the order in which tax rates changes across rounds. In other words, tax evolution and history of tax rates matter. Subjects in Panels A and C experienced tax increases before tax decreases while subjects in Panel B experienced tax decreases before tax increases. We observe that tax changes in Panel B lead to a positive treatment effect (treatment more responsive than control) while tax changes in Panels A and C lead to negative treatment effects. The treatment group in Panels A and C appears to be more responsive to tax increases while the control group is more responsive to tax decreases; the opposite is observed in Panel B.[20] The sign of the treatment effect in panel C is therefore consistent with the results in panel A in that the effort adjustment is higher (lower) in the control group when taxes go down (up). We observe the opposite result in panel B, where the control group has a smaller adjustment to a tax decrease than the treatment group (see section 4.5 for a discussion of this finding). This variation in the sign of the treatment effect across tax evolutions explains why the treatment effect is not distinguishable from zero when we pool the data (see regression results in Table 4.3).

Regression The results discussed so far are supported by a parametric analysis in which we estimate equation (4.3) separately for each tax evolution. The results in Table 4.5 confirm the existence of a negative treatment effect for tax evolutions one (panel A) and three (panel C). The coefficient of -3.056 in panel A is statistically different from zero and implies that the effect of a 10 percentage point increase in the net-of-tax rate on the supply of labor effort is 0.3 lower in the treatment group relative to the control group. The estimate in panel C is about one third that of panel A but is not distinguishable from zero. On the other hand, the estimated treatment effect is 2.714 in tax evolution two

[20] These differences are not always statistically significant; as we indicated above, the effect is more likely to be significant for tax decreases than for tax increases.

Table 4.4: Effort Adjustments

		Control		Treatment			
Period	Δ Tax	Mean	Std. Dev.	Mean	Std. Dev.	$p > \|z\|$	N
Panel A: Tax Evolution 1							
1 to 2	$\uparrow$ (15-35%)	1.56	2.61	1.85	2.37	0.651	90
2 to 3	$\uparrow$ (35-50%)	0.87	2.12	0.90	2.12	0.958	90
3 to 4	$\downarrow$ (50-15%)	1.33	2.99	-0.30	2.42	0.013	90
Panel B: Tax Evolution 2							
1 to 2	$\downarrow$ (50-35%)	1.01	2.01	1.51	2.10	0.405	58
2 to 3	$\downarrow$ (35-15%)	-0.19	2.68	1.31	2.00	0.025	58
3 to 4	$\uparrow$ (15-50%)	1.19	2.72	1.10	2.89	0.987	58
Panel C: Tax Evolution 3							
1 to 2	$\uparrow$ (35-50%)	0.61	1.85	1.82	1.47	0.011	57
2 to 3	$\downarrow$ (50-15%)	1.40	2.18	0.91	1.30	0.110	57
3 to 4	$\uparrow$ (15-35%)	0.49	2.54	0.23	2.99	0.631	57

Means and Standard Deviations for effort change between periods t and $t-1$ by Treatment Status. Δ Tax indicates the direction of the tax rate change. Tax rates in periods 1-4 are 15, 35, 50, 15% in panel A; 50, 35, 15, 50% in panel B and 35, 50, 15, 35% in panel C. $p > |z|$ reports the p-value of a (non-parametric) Wilcoxon rank-sum test for differences in distributions (H0: no differences) between the two treatment groups (Wilcoxon 1945; Mann and Whitney 1947). N indicates the number of observations in both treatments.

(panel B). These results are robust to the inclusion of period trend as well as individual demographics. Similar results are obtained when we estimate the elasticity of labor effort with respect to the NTR (see Table 4.7) in appendix 4.8.

4.5 Discussion of Results

Our empirical analysis highlights three important findings: access to evasion affects the labor supply response of workers, this effect is stronger and most obvious when tax rates fall, and the sign of the effect depends on the evolution of tax rates. This section first provides a discussion of these results (4.5.1). Because the pooled results are being driven by opposing effects of the three tax evolutions, we focus on tax evolution specific results. We then discuss how our experimental design differs from studies that focus on the effect of wages and taxes (4.5.2) as well as the external validity of our findings (4.5.3).

Table 4.5: Regression of Effort on Taxes by Tax Evolution

	I	II	III	IV
	Panel A: Tax Evolution 1			
NTR ∗ Treat	−3.056**	−3.056**	−3.056**	−3.056**
	(1.398)	(1.400)	(1.400)	(1.412)
NTR	−1.130	−0.273	1.078	1.078
	(1.082)	(1.092)	(1.119)	(1.128)
Treat	2.672*	2.672*	2.672*	2.209
	(1.535)	(1.537)	(1.537)	(1.406)
R2	0.01	0.06	0.06	0.35
	Panel B: Tax Evolution 2			
NTR ∗ Treat	2.714*	2.714*	2.714*	2.714*
	(1.596)	(1.599)	(1.599)	(1.617)
NTR	−0.501	−1.670	−2.895**	−2.895**
	(1.236)	(1.242)	(1.276)	(1.290)
Treat	−0.663	−0.663	−0.663	−0.786
	(1.084)	(1.086)	(1.086)	(1.065)
R2	0.03	0.13	0.13	0.22
	Panel C: Tax Evolution 3			
NTR ∗ Treat	−1.105	−1.105	−1.105	−1.105
	(1.386)	(1.389)	(1.389)	(1.405)
NTR	4.003***	1.431	1.563	1.563
	(1.197)	(1.225)	(1.210)	(1.224)
Treat	1.194	1.194	1.194	2.095
	(1.307)	(1.309)	(1.309)	(1.354)
R2	0.01	0.07	0.08	0.26
Trend	No	Linear	Log	Log
Controls	No	No	No	Yes

Random Effects Regressions. Dependent Variable is *Effort*. Tax rates in periods 1-4 are 15, 35, 50, 15%, in panel A; 50, 35, 15, 35% in panel B and 35, 50, 15, 35% in panel C. Number of observations is 360, 232, and 228 in panels A, B and C. Number of subjects is 90, 58 and 57 in panels A, B and C. Observations are from four periods per participant. Standard Errors in parentheses are clustered by Individual. *Treated* is a dummy variable taking "1" if a participant is in the group with evasion opportunity (treated) and "0" if not. *NTR* (net-of-tax rate) is defined as (1 - Tax Rate). Specification IV includes controls for age, gender, a dummy indicating native German, risk aversion, tax morale and time of day. Significance levels: $* < 0.10$, $** < 0.05$, $*** < 0.01$.

4.5.1 Labor Supply Responses

The theoretical model discussed in section 4.3.2 suggests two channels through which access to evasion might affect labor effort response. First, an increase in the tax rate yields the standard income and substitution effects, which together ought to have a theoretically ambiguous effect on labor effort. Although we are unable to cleanly identify the sum of these standard effects, we argue that they should be approximately the same for both groups given randomization into groups. In other words, these standard income and substitution effects cannot explain the differential effects we observe.

The second channel is related to the fact that individuals with evasion opportunities have two interrelated margins along which to adjust their income: labor effort and reported income. As described in section 4.3.2 (as well as 4.9.3 in the appendix), evaders' labor response is affected by evasion-induced income and substitution effects in addition to standard income and substitution effects. An increase in the tax rate is likely to induce an increase in evasion, which would then lead to higher net wages for subjects in the treatment group relative to those in the control group, *all else equal*. The higher net wage then leads to evasion-induced income and substitution effects. While this income effect reduces labor effort, the substitution effect increases effort.

This theoretical framework explains the results in panels A and C of Table 4.4. We find that subjects in the treatment group underreport their income by approximately 73 percent and that the share of income that is underreported increases with the tax rate (see section 4.4.2). Since the effect of the NTR (tax rate) on effort is smaller (larger) in the treatment group relative to the control group, it seems plausible to argue that the evasion-induced substitution effect outweighs the evasion-induced income effect, which then explains the observed differential effect in panels A and C of Table 4.5.

To see the evasion-induced effects more clearly, consider the impact of evasion on net wages. Recall that the gross wage of 6 ECU per correctly positioned slider is constant and identical both between and within groups. This is not the case for the net wage. Since subjects in the treatment group can underreport income, their average net wage is higher than that of subjects in the control group. For example in tax evolution 1, the control group receives a net wage of 5.10 ECU per correctly positioned slider in period one where the tax rate is 15%, while the treatment group faces an average net wage of 5.64 ECU.[21] This difference is due solely to the share of underreported income in the treatment group: approximately 60%. Now consider the effect of the tax rate increase from 15 to 35% between rounds one and two. It is straightforward to show that the net wage of the control group falls to 3.90 ECU and that the impact of the tax rate change on

[21]Net wage is calculated as $w = (1 - \tau) \times W$ in the control group and $w = (1 - (1 - \pi)\tau) \times W$ in the treatment group, where π is the share of income that is underreported on average, τ is the tax rate, w is the net wage rate and W is the gross wage.

the net wage in the treatment group depends on the evasion response. The increase in the tax rate reduces the average net wage in the treatment group to 5.16 ECU, assuming the average share of underreported income stayed at 60%. Concurrently, the higher tax rate increases underreporting to 74%, which causes the net wage to increase to 5.45 ECU.[22] We argue that this evasion-induced effect on net wages leads to income and substitution effects that explain the differential effort supply response we observe in panels A and C of Tables 4.4 and 4.5. In both cases, the treatment group is less responsive to tax increases and more responsive to tax decreases, even if the differences are not always distinguishable from zero.

The results in tax evolution two are also consistent with the theoretical framework, but only if the evasion-induced income effect outweighs the evasion-induced substitution effect or if increasing tax rates trigger less tax evasion. However, it is unlikely that income effects in our laboratory setting are large enough to outweigh the substitution effect. Additionally, we find a positive correlation between tax evasion and tax rates in all evolutions.

An alternative explanation is that the treatment effect predicted by the model depends on the evolution and history of tax rates. This conclusion is supported by two crucial features of the experimental design. First, the tax evolutions and treatment and control groups are balanced with respect to individual characteristics as demonstrated by the fact that including these covariates does not affect the estimated treatment effects. Second, the experimental design is internally valid, which implies that the results are not being driven by design features that differentially affect treatment and control groups in the various tax evolutions. We also argue that the results are not being driven by differences in the year in which the tax evolutions were implemented. Tax evolution one was implemented in June 2012, and tax evolutions two and three were implemented in the same week of July 2013. The fact that tax evolution three is consistent with tax evolution one suggests that year of implementation is not driving the results. We are therefore confident that the only difference between the three panels in Tables 4.4 and 4.5 is the evolution of tax rates.

As a result, we interpret the variation in the sign and significance of the treatment effect across tax evolutions along with the fact that the treatment effect is more obvious when tax rates fall as evidence that the institutional setting for how tax rates evolve across rounds matters for how individuals respond to tax rates. In particular, we argue that the difference in the sign of the treatment effect between panel B and panels A and C is due to the difference in the evolution of tax rates. This is evident from the discussion in section 4.4.2.

[22] Similar calculations for the other rounds show that tax rate increases have a substantially smaller effect on net wages in the treatment group relative to the control group due to evasion.

The result that the institutional setting matters for the direction of the treatment effect is consistent with Collins et al. (1992), who find that the opportunity to evade has a positive effect on labor effort when taxes are proportional or steeply progressive and a negative effect when taxes are mildly progressive.

4.5.2 Evasion vs. Net Wage and Effective Tax Rate

We acknowledge that providing one group of workers the opportunity to hide taxable income might be viewed as equivalent to a reduction in the effective tax rate of this group and therefore an increase in the net wage rate (Rosen 1976a, 1976b). This argument suggests that our experiment is eventually similar to other experimental studies examining the effect of wage rates or tax rates on supply of effort in standard real-effort tasks. However, several theoretical and empirical studies show that people respond differently to taxes than they do to wages (Koenig et al. 1995; Slemrod 2001; Fochman and Weimann 2011; Fochman et al. 2012), which suggests that the impact of evasion opportunities on effort supply in our experiment is not equivalent to the effect of the net wage rate.

We further argue that our experiment differs from experimental set-ups that examine the effect of varying effective tax rates on real-effort because our subjects have to take an explicit action, underreporting, in order to benefit from a lower tax burden. Since this action is costly due to risk and loss aversion as well as subjective moral costs of evading, it is likely to illicit different behavioral responses than would an explicit exogenous change in effective tax rates, which would not be associated with any costs.[23] The observation that people substantially overestimate risks (Tversky and Kahneman 1974) further supports our argument that (risky) evasion opportunities are not equivalent to reductions in effective tax rates or net wages.

4.5.3 External Validity

As with all economics laboratory experiments, there remains doubt about the external validity of our results.[24] One major concern is that the setting in the lab is abstract and artificial. However, subjects in our experiment must complete a "real effort" task for which they earn money dependent on their performance. While not perfectly equivalent to a naturally occurring environment, the effort task represents real economic choices and is similar to many "real world" labor tasks in the sense that it is annoying and somewhat "painful". Again, we acknowledge the need to exercise caution when generalizing labor effort decisions to more traditional labor supply decisions, such as hours worked. However,

[23] The literature shows that tax compliance decisions are indeed associated with non-pecuniary costs such as increased heart-rates (Dulleck et al. 2012) or skin conductance (Coricelli et al. 2010).

[24] See Levitt and List (2007) for a critical discussion of the generalizability of lab experiments. Falk and Heckman (2009) offer a defense of most concerns, some of which are also discussed here.

since the choices between effort and hours worked are isomorphic in many ways (Charness and Kuhn 2011) and share many characteristics, we argue that choices in the real-effort experiment can serve as a good proxy for actual labor supply. In order to make the tax evasion decision as realistic as possible we used actual tax terminology and announced to the participants that all tax revenue would be donated to the German Red Cross, a non-ideological charity organization that is usually perceived as reliable and transparent.[25]

It is also often argued that the stakes in lab experiments are too small to interpret the outcomes as realistic. This is unlikely to be true in our case because our average pay-off of EUR 11.08 roughly corresponds to five full lunch meals in the student cafeteria at Cologne University.[26] Furthermore, many experiments conducted in locations where the stakes were equivalent to more than a month's earnings find very similar results to conventional "small-stake" experiments (Slonim and Roth 1998). Additionally, it is questionable whether high stakes render a more realistic setting, since most "real-life" decisions do not involve massive amounts of money (Falk and Heckman 2009). Another concern is the reliance on university students as participants. Many experiments with non-student populations find results comparable to experiments with students (Charness and Kuhn 2011; Falk et al. 2013). Furthermore, Alm et al. (2011) compare students to non-students in tax compliance experiments and find that the reporting responses of students to policy innovations are largely the same as those of non-students in identical experiments and "real" people in non-experiment environments.

4.6 Conclusion

Motivated by the observation that some types of workers have the opportunity to hide their income while others do not have such opportunities, this paper is the first to cleanly identify the differential effect of access to evasion opportunities on labor supply elasticities with respect to tax rates. We first use a theoretical framework to describe the channels through which evasion is likely to affect labor supply elasticities. A lab experiment with 205 participants is then utilized to test the theoretical prediction that access to evasion influences the responsiveness of labor supply to tax rates.

In the experiment, all subjects first undertake a simple labor effort task and then have to pay taxes on the income they earn from the labor task. The tax rate varies across rounds and is announced before each new round. A treatment group is given the

[25]Tax morale research (Torgler 2007) finds that taxpayers are more likely to comply with tax laws if they believe that the tax revenue is spent transparently. Eckel and Grossman (1996) show that dictators share more in dictator games if the recipient is the American Red Cross. Overall, we donated EUR 541.69 to the Red Cross.

[26]The show-up fee is equivalent to one meal. The cafeteria at University of Cologne is the most popular spot for students to buy their daily lunch.

opportunity to report their income either correctly or falsely, while the control group does not have the opportunity to evade; their tax liability is simply withdrawn. We find evidence that subjects with evasion opportunities respond differently to tax rate changes than those without such opportunities. However, the sign of the difference between the two groups depends on the tax institutional setting of how tax rates evolve over time. In particular, the treatment effect is more prevalent when tax rates fall than when tax rates increase. We also find that the evolution of tax rates over time matters; the treatment effect is negative in tax evolutions where tax rates first increase then decrease and positive in tax evolutions where tax rates first decrease then increase.

Our results provide one possible explanation for why observed labor supply elasticities might be different across different types of workers. Workers with a reporting decision, i.e., the opportunity to evade taxes, have two channels through which to adjust their taxable income - labor supply and evading - whereas workers who are subject to third-party reporting are only able to resort to labor adjustments. The present paper is the first to empirically show that heterogeneous access to evasion opportunities, as found in recent empirical studies (Slemrod 2007; Kleven et al. 2011; Chetty et al. 2013), has an impact on economic decisions such as labor supply.

The findings of the paper further show that tax rate reforms may have heterogeneous effects on different types of workers. The elasticity of taxable income is significantly larger among the group with an evasion opportunity, which is due to the fact that this group is able to exploit an extra margin to adjust taxable income. In this sense, our findings are consistent with the observation that bunching around kink points in the tax schedule is more prevalent among the self-employed, who have greater access to evasion (Saez 2010; Bastani and Selin 2014; Chetty et al. 2013). Finding that the ETI is relatively higher in the treatment group is also consistent with recent work by Burns and Ziliak (2012), who find that the ETI increases with education and is greater than one for individuals with a graduate degree, as well as by Heim (2009, 2010) who shows that the ETI increases in income and is larger for the self-employed. We argue that our findings represent one possible explanation for these results since well-educated individuals with higher incomes and a higher propensity to be self-employed are more likely to have better evasion opportunities.

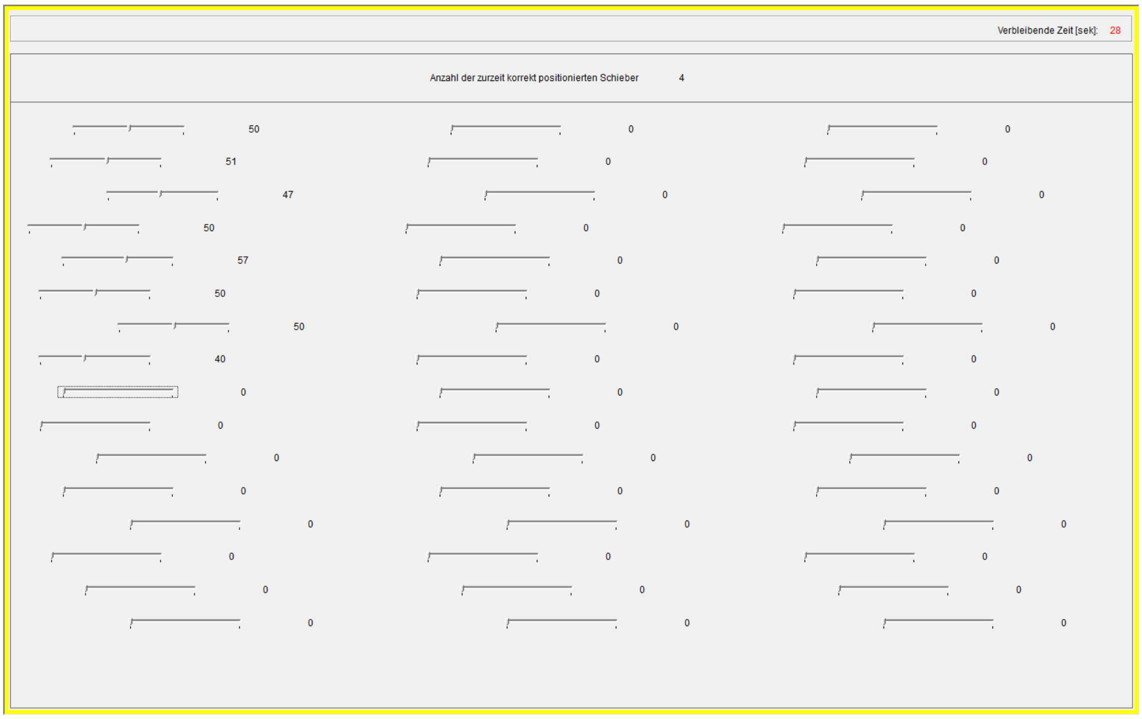
Along with previous findings in the literature, our results support the notion that tax reform evaluations ought to consider how behavioral responses to tax policy vary across populations. The results would seem to suggest that policy makers ought to adopt discriminatory tax rates; e.g., rates that vary by *access* to evasion. However, we argue that policy makers should strive to minimize evasion/avoidance loopholes instead of adopting discriminatory tax rates. Minimizing evasion and avoidance opportunities would reduce the responsiveness of the tax base and allow for more uniform tax rates, which may be higher or lower. Closing loopholes would also lead to equity gains by ensuring that the effective tax rate is equal across all types of workers *all else equal* (Alm and Finlay 2013). Additionally, closing loopholes and evasion opportunities is likely to have positive tax revenue implications as well. We find in our setting that the average tax payment per person is considerably higher in the control than in the treatment group.

Our findings also suggest that policy makers ought to consider the direction of tax rate changes when evaluating tax policy proposals. As we show, there is evidence that the evolution and history of tax rate changes matters for the behavioral impact of tax policy. Therefore, policy makers evaluating tax rate increases following years of tax rate decreases might find different results than those evaluating tax rate increases following years of tax rate increases. This is especially important given that recent fiscal deficits have lead to a number of countries considering or implementing tax rate increases after almost three decades of lowering tax rates (Sabirianova Peter et al. 2010).

An obvious caveat to our experimental findings is the fact that our experimental set-up mirrors a world where only two channels are available to adjust taxable income. It is not immediately clear what the implications are for a model with more than two response margins. We leave such considerations for future research. We further acknowledge that external validity has to be considered when extrapolating our results to the "real" world. However, just as with field experiments and quasi-experimental observational studies, we argue that our results should be interpreted as Local Average Treatment Effects (LATE) that provide causal evidence on a particular sub-population.

4.7 Appendix A: The Slider Task

Figure 4.4: Screen showing the slider task



Note: The slider task was designed by Gill and Prowse (2012). In the displayed screen, the subject positioned four sliders correctly and four falsely. She currently works on positioning the ninth slider. 28 seconds are left in this round.

4.8 Appendix B: Elasticity Estimates

Table 4.6: Elasticity Estimates. Full Sample

	I	II	III	IV
NTR (log) × Treat	−0.035	−0.037	−0.038	−0.037
	(0.038)	(0.036)	(0.036)	(0.037)
NTR (log)	0.017	−0.000	0.011	0.011
	(0.030)	(0.029)	(0.030)	(0.031)
Treatment Status	0.022	0.021	0.020	0.038
	(0.037)	(0.038)	(0.038)	(0.037)
Trend (linear)		0.057***		
		(0.004)		
Trend (log)			0.124***	0.124***
			(0.009)	(0.009)
constant	2.851***	2.701***	2.750***	3.082***
	(0.027)	(0.030)	(0.029)	(0.098)
Controls	No	No	No	Yes
Panel Obs	820	820	820	820
Indiv. Obs	205	205	205	205
Periods	4	4	4	4
R2	0.004	0.05	0.05	0.18

Random Effects Regressions. Full Sample. Dependent Variable is *Logged Effort*. Standard Errors in parentheses are clustered by Individual. *Treated* is a dummy variable taking “1” if a participant is in the group with evasion opportunity (treated) and “0” if not. *NTR* (net-of-tax rate) is defined as (1 - Tax Rate). Specification IV includes controls for age, gender, a dummy indicating native German, risk aversion, tax morale and time of day. Significance levels: $* < 0.10$, $** < 0.05$, $*** < 0.01$.

Table 4.7: Elasticity Estimates by Tax Evolution

	I	II	III	IV
Panel A: Tax Evolution 1				
NTR (log) × Treat	−0.123**	−0.123**	−0.123**	−0.123**
	(0.056)	(0.056)	(0.056)	(0.057)
NTR (log)	−0.025	0.011	0.058	0.058
	(0.045)	(0.045)	(0.046)	(0.046)
Treat	−0.021	−0.021	−0.021	−0.045
	(0.054)	(0.054)	(0.054)	(0.048)
R2	0.01	0.05	0.05	0.36
Panel B: Tax Evolution 2				
NTR (log) × Treat	0.117*	0.117*	0.117*	0.117*
	(0.066)	(0.066)	(0.066)	(0.067)
NTR (log)	−0.035	−0.076	−0.127**	−0.127**
	(0.054)	(0.054)	(0.055)	(0.056)
Treat	0.113**	0.113**	0.113**	0.109*
	(0.057)	(0.057)	(0.057)	(0.058)
R2	0.02	0.13	0.13	0.20
Panel C: Tax Evolution 3				
NTR (log) × Treat	−0.067	−0.067	−0.067	−0.067
	(0.056)	(0.056)	(0.056)	(0.057)
NTR (log)	0.169***	0.058	0.071	0.071
	(0.049)	(0.049)	(0.049)	(0.050)
Treat	0.016	0.016	0.016	0.100
	(0.086)	(0.086)	(0.086)	(0.099)
R2	0.01	0.04	0.04	0.24
Trend	No	Linear	Log	Log
Controls	No	No	No	Yes

Random Effects Regressions. Dependent Variable is *logged Effort.* Tax rates in periods 1-4 are 15, 35, 50, 15%, in panel A, 50, 35, 15, 35% in panel B, and 35, 50, 15, 35% in panel C. Number of observations is 360, 232, and 228 in panels A, B and C. Number of subjects is 90, 58, and 57 in panels A, B and C. Observations from four periods per participant. Standard Errors in parentheses are clustered by Individual. *Treated* is a dummy variable taking "1" if a participant is in the group with evasion opportunity (treated) and "0" if not. *NTR* (net-of-tax rate) is defined as (1 - Tax Rate). Specification IV includes controls for age, gender, a dummy indicating native German, risk aversion, tax morale and time of day. Significance levels: $* < 0.10$, $** < 0.05$, $*** < 0.01$.

4.9 Appendix C: Theoretical model

This section of the appendix outlines a simple theoretical framework that captures the main features of the labor supply and evasion decision that individuals face in the treatment group as defined in section 4.2. The model merges the standard neo-classical labor supply model with the Allingham and Sandmo (1972) model of tax evasion (Pencavel 1979). We acknowledge that the theoretical framework used here models labor supply as hours of work rather than effort, which is the basis of our experimental design. However, as discussed in section 4.2.6, the choices between labor effort and hours of work share many characteristics and are isomorphic in many ways so that modeling labor effort in a labor supply framework should nonetheless yield valid hypotheses for our experimental set-up. Modeling labor supply instead of effort also eases the comparison with the standard labor supply model.

Because the model describing the control group is the standard labor supply model, we simply state the results of this model where appropriate without any derivations.

4.9.1 Model

Individuals make a labor supply decision L, which yields labor income wL and non-labor income M; w is the wage rate. They then report $R \leq (wL + M)$ to the tax authority to determine their tax liability. Reported income R is audited with probability p and, because we assume all audits lead to the full discovery of true income, a fine equal to twice the evaded taxes must be paid if audited. Assuming individuals consume all of their income implies:

$$Consumption = \begin{cases} C_a & = (wL + M) - \tau R - 2\tau(wL + M - R) \\ & = (wL + M)(1 - 2\tau) + \tau R \text{ with probability p} \\ C_n & = wL + M - \tau R \text{ with probability (1-p),} \end{cases} \tag{4.4}$$

where subscripts a and n indicate audited and not audited, respectively, and τ is the proportional tax rate.

As in Pencavel (1979), we assume individuals choose L and R to maximize an expected utility function that satisfies the standard assumptions of the neo-classical labor supply model; continuous, twice differentiable, and concave. We also assume the utility function is strongly separable in consumption and labor (Pencavel 1979). With these assumptions in mind, the individual maximization problem is specified as follows

$$
\begin{aligned}
maxEU = pU(C_a, L) + (1-p)U(C_n, L) & \\
& st. \\
& C_a = (wL + M)(1 - 2\tau) + \tau R \\
& C_n = wL + M - \tau R,
\end{aligned}
\tag{4.5}
$$

Differentiating equation (4.5) with respect to R and L, respectively, yields the following first order conditions:

$$
\begin{aligned}
pU'_{C_a}\tau - (1-p)U'_{C_n}\tau = 0 \\
pU'_{C_a}w(1-2\tau) + pU'_L + (1-p)U'_{C_n}w + (1-p)U'_L = 0,
\end{aligned}
\tag{4.6}
$$

which we rewrite as

$$pU'_{C_a} = (1-p)U'_{C_n} \tag{4.7}$$

$$w = -\frac{U'_L}{p(1-2\tau)U'_{C_a} + (1-p)U'_{C_n}} \tag{4.8}$$

Assuming interior solutions exist, it can be shown from equation (4.7) that utility is maximized by the level of reported income that equalizes the weighted marginal utility of consumption in both states of the world. Similarly, equation (4.8) implies individuals choose the level of labor supply that equalizes the wage rate to the ratio of the marginal utility of labor and the expected marginal utility of consumption. This is comparable to the equilibrium condition facing our non-evader treatment in the standard neoclassical labor supply model with taxes and no tax evasion; $w = -\frac{U'_L}{(1-\tau)U'_C}$.

4.9.2 Maximum exists

This section of the appendix shows that the conditions for a maximum are satisfied. To see this we differentiate the equations in (4.6) with respect to L and R to get:

$$V = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where $A = [pw^2\gamma^2 U''_{C_a} + qw^2 U''_{C_n} + U''_L]$, $B = [pw\gamma\tau U''_{C_a} - qw\tau U''_{C_n}]$, $C = [p\gamma w U''_{C_a} - qwU''_{C_n}]$, and $D = [p\tau U''_{C_a} + q\tau U''_{C_n}]$, and $q = (1-p)$ and $\gamma = (1-2\tau)$. Maximum requires that V is negative definite; i.e.; $A < 0$ and $\Lambda = A * D - B * C > 0$. A is less than zero as long as $U''_{C_a} < 0$, $U''_{C_n} < 0$, and $U''_L < 0$. Therefore, we need to show that $\Lambda = A * D - B * C > 0$.

It can be shown that

$$\begin{aligned} A*D &= (pw\gamma U''_{C_a})^2\tau + pq(w\gamma)^2\tau U''_{C_a}U''_{C_n} + pqw^2\tau U''_{C_a}U''_{C_n} + (qw)^2\tau(U''_{C_n})^2 + p\tau U''_{C_a}U''_L + q\tau U''_{C_n}U''_L \\ &= p^2w^2\tau U''_{C_n}U''_{C_a}[\gamma^2\frac{U''_{C_a}}{U''_{C_n}} + \frac{q}{p}\gamma^2 + \frac{q}{p} + \frac{q}{p}^2\frac{U''_{C_n}}{U''_{C_a}}] + p\tau U''_L[U''_{C_a} + \frac{q}{p}U''_{C_n}] \end{aligned}$$

and that

$$\begin{aligned} B*C &= (pw\gamma U''_{C_a})^2\tau - pq\gamma w^2\tau U''_{C_a}U''_{C_n} - pqw^2\gamma\tau U''_{C_a}U''_{C_n} + (qw)^2\tau(U''_{C_n})^2 \\ &= p^2w^2\tau U''_{C_n}U''_{C_a}[\gamma^2\frac{U''_{C_a}}{U''_{C_n}} - 2\frac{q}{p}\gamma + \frac{q}{p}^2\frac{U''_{C_n}}{U''_{C_a}}] \end{aligned}$$

Therefore,

$$A*D - B*C = p^2w^2\tau U''_{C_n}U''_{C_a}(1+\gamma)^2 + p\tau U''_L[U''_{C_a} + \frac{q}{p}U''_{C_n}] > 0 \qquad Q.E.D.$$

4.9.3 Comparative Statics

This section of the appendix derives comparative static results for our theoretical model. We are particularly interested in the effect of non-labor income and the tax rate on labor supply. Start by totally differentiating the equations in (4.6) under the assumption that $d\tau = dw = 0$, $dm \neq 0$, $dL \neq 0$, and $dR \neq 0$. This yields:

$$\begin{aligned} pw\gamma U''_{C_a}(\gamma wdL + \gamma dM + \tau dR) + qwU''_{C_n}(wdL + dM - tdR) + U''_L dL = 0 \\ pU''_{C_a}(\gamma wdL + \gamma dM + \tau dR) - qU''_{C_n}(wdL + dM - tdR) = 0 \end{aligned} \tag{4.9}$$

where $q = (1-p)$ and $\gamma = (1-2\tau)$. Rewrite equation (4.13) as[27]

$$\begin{aligned} AdL + BdR + GdM = 0 \\ AdL + BdR + HdM = 0 \end{aligned} \tag{4.10}$$

and solve for the income effect to get:

$$\frac{dL}{dM} = -\frac{pqw\tau U''_{C_a}U''_{C_n}(1+\gamma)^2}{\Lambda} \leq 0 \tag{4.11}$$

This result implies that labor supply is an inferior good; i.e., an increase in non-labor income reduces labor supply.[28]

[27] A, B, C, and D are as defined earlier while $G = pw\gamma U''_{C_a} + qwU''_{C_n}$, and $H = p\gamma U''_{C_a} - qU''_{C_n}$.

[28] The effect of non-labor income on labor supply would be ambiguous as in the standard labor supply model had we not imposed the strong separability assumption.

Similarly, the effect of true income on reported income is:

$$
\begin{aligned}
\frac{dR}{dM} &= \frac{(p(1-2\tau)\frac{U''_{C_a}}{U''_{C_n}} - q)}{\Lambda} \lesseqgtr 0 \\
&= \frac{(1-p)U''_{C_a}U''_L(\frac{R^a_{cc}}{R^a_{cn}}(1-2\tau)-1)}{\Lambda},
\end{aligned}
\tag{4.12}
$$

where R^a_{ci} is the measure of absolute risk aversion in state i.

Next we derive the effect of tax rate on labor supply by totally differentiating the equations in (4.6) under the assumption that $dM = dw = 0$, $d\tau \neq 0$, $dL \neq 0$, and $dR \neq 0$. This yields:

$$
\begin{aligned}
-2pwU'_{C_a}d\tau + pw\gamma U''_{C_a}[\gamma wdL - 2Id\tau + \tau dR + Rd\tau] + qwU''_{C_n}[wdL - \tau dR - Rd\tau] + U''_L dL = 0 \\
pU''_{C_a}(\gamma wdL - 2Id\tau + \tau dR + Rd\tau) - qU''_{C_n}(wdL - \tau dR - Rd\tau) = 0
\end{aligned}
$$

which we rewrite as[29]

$$
\begin{aligned}
AdL + BdR - Ed\tau = 0 \\
CdL + DdR - Fd\tau = 0
\end{aligned}
\tag{4.13}
$$

Rewriting equation (4.13) in matrix form and solving for $\frac{dL}{dt}$ yields:

$$
\frac{dL}{d\tau} = \overbrace{\frac{p^2w\tau U'_{C_a}U''_{C_a}Z}{\Lambda}}^{1} - \overbrace{\frac{pw\tau U''_{C_a}U''_{C_n}Q}{\Lambda}}^{2} - \overbrace{\frac{dL}{dM}\frac{(wL+M)}{(1-\tau)}}^{3} \lesseqgtr 0, \tag{4.14}
$$

where[30]

$$
\begin{aligned}
Z &= [2 + 2(1-2\tau)(wL+M-R) + \frac{(1-p)}{p}R\frac{U''_{C_n}}{U''_{C_a}}] \\
Q &= [2(1-2\tau)\frac{U''_{C_a}}{U''_{C_n}}(wL+M) + (p(1-2\tau)\frac{U''_{C_a}}{U''_{C_n}} - (1-p))R + 2(1-p)] \\
\Lambda &= p^2w^2\tau U''_{C_a}U''_{C_n}\frac{(1-p)}{p}4(1-\tau)^2 + p\tau U''_L(U''_{C_a} + \frac{(1-p)}{p}U''_{C_n}),
\end{aligned}
$$

.

As in the standard model, equation (4.14) may or may not be negative. The first two terms capture substitution effects and is always negative as long as $\tau < 0.5$. In other words, an increase in the tax rate leads to lower labor supply. On the other hand, higher

[29] A, B, C, and D are as defined earlier while $E = [2pwU'_{C_a} + 2pwI\gamma U''_{C_a} - pw\gamma RU''_{C_a} + qwRU''_{C_n}]$ and $F = [2pIU''_{C_a} - pRU''_{C_a} - qRU''_{C_n}]$.

[30] Note that $Q > 0$ since $2(1-2\tau)\frac{U''_{C_a}}{U''_{C_n}} > (p(1-2\tau)\frac{U''_{C_a}}{U''_{C_n}} - (1-p))$, and $wL + M > R$. This follows from the fact that $0 < p < 1$. We also assume $(1-2\tau) > 0$, which requires $\tau < 0.5$ as in our experiment.

taxes imply lower income, which leads to higher labor supply to compensate for the lost income. This income effect is captured by the third term.

4.10 Appendix D: Instructions

4.10.1 English translation: Evasion opportunity treatment

Welcome and thank you for participating in our experiment. From now on until the end of the experiment, please refrain from communicating with other participants. If you do not abide by this rule, we will have to exclude you from the experiment.

We kindly ask you to read the instructions thoroughly. If you have any questions after reading the instructions or during the experiment, please raise your hand and one of the instructors will come to you and answer your question in person. Your payment and your decisions throughout the experiment will be treated confidentially. None of the other participants is informed, neither during nor after the experiment, about your decisions in the experiment or your payment.

You can earn money in this experiment. How much you earn depends on your decisions. During the experiment, your payments will be calculated in a virtual currency: Experimental Currency Units (ETU). **1 ECU corresponds to 0.10 Euro**. After the experiment, your pay-off will be converted to Euro and given to you in cash. Additionally, you will receive a show-up fee of 2.50 Euro.

The Experiment

Overview The experiment consists of 11 rounds. In each round, you will first complete a labor task and, depending on your performance, earn money from this labor task. You will have to pay taxes on your earned income. Therefore, after the labor task you will be faced with a tax reporting decision. Both the labor task and the tax return filing are described in more detail further below.

Payment The first of 11 rounds serves as a practice round, in which you cannot earn money. The subsequent 10 rounds are paying rounds. All rounds are independent of each other. What is more, your pay-off does not at all depend on the decisions of other participants. The pay-off, which you will be paid in cash at the end of the experiment, does not consist of the sum of the net incomes from all 10 paying rounds. Instead, after the experiment, one round will be randomly chosen to determine your payment. Practically this means that after all 10 paying rounds have been completed you will throw a 10-sided die. The number shown by the die determines the round for which you will be paid.

Flow of each round You will be told the tax rate for a given round at the beginning of that round. The tax rate may, but does not have to, vary from round to round. In the next step you will complete the **labor task** on the computer screen. After completion of this labor task, you will be shown your gross income. This is based on your performance during the labor task plus a fixed amount. Details on the labor task are described further below.

After completion of the labor task, you will be faced with a **tax reporting decision**. You choose an amount, which shall be taxed at the prevailing tax rate. The amount you chose may be as high as your earned gross income or lower. There is a random chance of 10% that your decision will be checked to see whether you reported your true gross income. If you are *not* checked, your payment for this round - the net income - will consist of your gross income less the tax payment. If you are checked *and* you have not fully reported your gross income, you will have to pay a penalty. More details are explained further below. All tax revenues paid by you and all other participants will be donated to the German Red Cross.

In Summary, each of the 11 rounds is timed as follows:

1. Information on prevailing tax rate
2. Completing the labor task
3. Information on gross income
4. Tax reporting
5. Check whether tax reporting is checked.
6. Calculation of this round's net income

The Labor Task At the beginning of each round you will first be informed about the tax rate in that round. Afterwards, you undertake a labor task on the computer screen using the computer mouse. In each round the task will last 120 seconds. During the task a screen with 48 so-called sliders appears on the screen. Each slider is initially positioned at "0" (Zero) and can be moved by you. You can move the slider to every whole number between "0" and "100" by clicking on the slider and moving the computer mouse. The current position of each slider is displayed to the right of this slider. You can readjust the position of each slider as many times as you wish. For each slider that you position exactly at "50" during the 120 seconds, you earn 6 ECU. During the 120 seconds of the labor task, on the upper right of the screen you are shown how many sliders are currently positioned at "50".

The Tax Return In the next step, you are shown on the screen how many ECU you earned in the labor task. Regardless of the ECU you earned based on correctly positioned sliders, you will receive an additional 5 ECU per round. The sum of the income from the labor task and this unconditionally given amount determines your **gross income**. This is shown on the screen. You are, once more, also shown which tax rate is in force.

You are now asked to **report income for tax purposes**. You choose an amount which shall be taxed at this round's prevailing tax rate. This chosen amount can be between zero and your gross income.

Calculation of Net Income After the completion of the tax reporting decision, one of the experimental investigators will come up to your booth with a 10-sided die. Please throw this 10-sided die. Based on the result of the die throw, there are two alternative scenarios, of which one is realized.

a) The die shows a **number between 2 and 10** (that is, a number out of 2,3,..., 9, 10)
In case the die shows a number between 2 and 10, your reporting decision will *not* be checked to determine whether you reported your full gross income for tax purposes. Your payment for this round - the net income -, in this case, consists of your gross income (earnings from labor task plus 5 ECU) less the tax payment. Thereby, the tax payment is the reported income multiplied with the prevailing tax rate. Hence:

<u>Net income</u> = gross income - (reported amount*tax rate)

b) The die shows the **number 1**:
In case the die shows the number 1 your reporting decision will be checked to determine whether you reported your full gross income for tax purposes. Depending on your previous decision, there are two different possibilities for your net income:

- If your reported income *equals* your true gross income, then your net income consists of your gross income less your tax liability. Hence:

<u>Net income</u> = gross income - (gross income*tax rate)

- If your reported income is *lower* than your gross income, then you will have to pay the tax liability based on your true gross income and additionally you will have to pay a penalty. This penalty is equal to the difference between your true gross income and your reported income multiplied by the prevailing tax rate. Hence:

Net income = gross income - (gross income*tax rate) - [(gross income - reported income) * tax rate]

Final Remarks After the completion of all 11 rounds - one practice round plus 10 paying rounds - the experiment is finished. One of the experimental investigators will come up to your booth and, once more, we ask you to throw a 10-sided die. The die throw determines the round (out of the 10 paying rounds) for which you are paid. For example, if the die shows the number "2", then your payment consists of the net income that you earned in the second paying round. In addition, you receive the show-up fee of 2,50 Euro. You will also be asked to complete a short questionnaire at the end of the experiment while we prepare the payments. All information collected through this questionnaire, just like all data gathered during the experiment, are anonymous and exclusively used for scientific purposes. After you have completed the questionnaire, please remain seated at your booth until we call you to come up front to pick up your payment.

4.10.2 English translation: No-Evasion opportunity treatment

Welcome and thank you for participating in our experiment. From now on until the end of the experiment, please refrain from communicating with other participants. If you do not abide by this rule, we will have to exclude you from the experiment. We kindly ask you to read the instructions thoroughly. If you have any questions after reading the instructions or during the experiment, please raise your hand and one of the instructors will come to you and answer your question in person. Your payment and your decisions throughout the experiment will be treated confidentially. None of the other participants is informed, neither during nor after the experiment, about your decisions in the experiment or your payment. You can earn money in this experiment. How much you earn depends on your decisions. During the experiment, your payments will be calculated in a virtual currency: Experimental Currency Units (ETU). **1 ECU corresponds to 0.10 Euro**. After the experiment, your pay-off will be converted to Euro and given to you in cash. Additionally, you will receive a show-up fee of 2.50 Euro.

The Experiment

Overview The experiment consists of 11 rounds. In each round, you will complete a labor task and, depending on your performance, earn money from this labor task. You will have to pay taxes on your earned income.

Payment The first of 11 rounds serves as a practice round, in which you cannot earn money. The subsequent 10 rounds are paying rounds. All rounds are independent of each other. What is more, your pay-off does not at all depend on the decisions of other participants. The pay-off, which you will be paid in cash at the end of the experiment, does not consist of the sum of the net incomes from all 10 paying rounds. Instead, after the experiment, one round will be randomly chosen to determine your payment. Practically this means that after all 10 paying rounds have been completed you will throw a 10-sided die. The number shown by the die determines the round for which you will be paid.

Flow of each round You will be told the tax rate for a given round at the beginning of that round. The tax rate may, but does not have to, vary from round to round. In the next step you will complete the labor task on the computer screen. After completion of this labor task, you will be shown your gross income, the prevailing tax rate and your corresponding net income. Your gross income is based on your performance during the labor task plus a fixed amount of 5 ECU. Details on the labor task and calculation of the net income are described further below. All tax revenues paid by you and all other participants will be donated to the German Red Cross.

In Summary, each of the 11 rounds is timed as follows:

a) Information on prevailing tax rate

b) Completing the labor task

c) Information on gross income

d) Calculation of this round's net income

The Labor Task At the beginning of each round you will first be informed about the tax rate in that round. Afterwards, you undertake a labor task on the computer screen using the computer mouse. In each round the task will last 120 seconds. During the task a screen with 48 so-called sliders appears on the screen. Each slider is initially positioned at "0" (Zero) and can be moved by you. You can move the slider to every whole number between "0" and "100" by clicking on the slider and moving the computer mouse. The current position of each slider is displayed to the right of this slider. You can readjust the position of each slider as many times as you wish. For each slider that you position exactly at "50" during the 120 seconds, you earn 6 ECU. During the 120 seconds of the labor task, on the upper right of the screen you are shown how many sliders are currently positioned at "50".

Calculation of Net Income After the completion of the tax reporting decision, you will be shown your gross income for this round. The gross income is based on the number of correctly positioned sliders plus a fixed amount of 5 ECU. Your gross income will be taxed at this round's prevailing tax rate. Your payment for this round - the net income -, consists of your gross income less the tax payment. Thereby, the tax payment is the gross income multiplied with the prevailing tax rate. Hence:

Net income = gross income - (gross income*tax rate)

Final Remarks After the completion of all 11 rounds - one practice round plus 10 paying rounds - the experiment is finished. One of the experimental investigators will come up to your booth and we ask you to throw a 10-sided die. The die throw determines the round (out of the 10 paying rounds) for which you are paid. For example, if the die shows the number "2", then your payment consists of the net income that you earned in the second paying round. In addition, you receive the show-up fee of 2,50 Euro. You will also be asked to complete a short questionnaire at the end of the experiment while we prepare the payments. All information collected through this questionnaire, just like all data gathered during the experiment, are anonymous and exclusively used for scientific purposes. After you have completed the questionnaire, please remain seated at your booth until we call you to come up front to pick up your payment.

Chapter 5

Nice guys finish last: Do honest taxpayers face higher tax rates?

5.1 Introduction

Tax ethics, tax honesty or tax morale[1] – the intrinsic motivation to honestly pay taxes – is widely seen as beneficial for an economy because it reduces the cost of financing the public sector. As a result, the literature on tax morale mostly focuses on the determinants of tax morale and on ways of improving it. This paper takes a different perspective and seeks to examine the tax policy implications of different levels of tax morale. Using a unique cross-country data set based on the World Values Survey and the World Tax Indicators, we find empirical support for the conclusion that tax rates increase with tax morale. To our knowledge, we are the first to investigate whether differences in tax morale affect the distribution of the tax burden across different groups of taxpayers.

Whereas the early evasion literature usually assumes that evasion decisions are driven solely by monetary considerations (Allingham and Sandmo 1972), more recent research has established that evasion behavior is *also* determined by psychological and cultural factors such as tax morale (Erard and Feinstein 1994; Andreoni et al. 1998; Alm and Torgler 2011). This is confirmed by several studies that have shown that tax morale is indeed negatively correlated with levels of tax evasion and the size of the shadow economy (Torgler and Schneider 2009; Halla 2012).[2] These findings suggest that the

[1]The literature usually uses the term 'tax morale'. This might be misleading and 'tax honesty' or 'tax ethics' might be more appropriate. However, since 'tax morale' is the commonly used terminology, we stick to it throughout this paper. In the literature, tax morale is typically defined as '*the intrinsic motivation to pay taxes which arises from the moral obligation to pay taxes as a contribution to society*' (e.g., Schwartz and Orleans 1967; Cummings et al. 2009).

[2]The standard model fails to explain why taxpayers honestly pay their taxes even in situations where detection is unlikely and penalties are low. It is now widely accepted that missing access to evasion opportunities (Kleven et al. 2011) and morale costs of evading (Frey and Feld 2002; Alm 2012) can explain high observed levels of compliance despite low audit probabilities and penalties.

prevailing level of tax morale is an important determinant of the government's ability to raise taxes and the cost of doing so (Feld and Frey 2007). As a result, many scholars argue that policy makers should design tax systems and broader political institutions so as to preserve and improve tax morale.

Unlike the existing literature which mostly examines the determinants of tax morale,[3] this study focuses on a different aspect of tax morale. We ask whether policymakers exploit the fact that their citizens have different levels of tax morale when setting tax rates. In other words, we take the level of tax morale as given and ask how tax morale affects the tax burden governments impose on different groups of taxpayers.[4]

We examine this research question using data from the World Value Survey (WVS), the European Values Survey (EVS), and detailed income tax data from the World Tax Indicators Database (WTI) (Sabirianova Peter et al. 2010). The WVS data allow us to observe levels of tax morale for different income groups in different countries, as well as various control variables. We then use the WTI database to compute average and marginal tax rates for the different income groups. The resulting unique data set allows us to study the effect of income group-level tax morale on tax rates.

Causal identification of our research question would require a (quasi-) experimental approach. Unfortunately, such an identification strategy is not available in our setting since it is difficult, if not impossible, to exploit (quasi-) random variation in either tax morale itself or another (quasi-) random variable that would be suitable to instrument for tax morale. Nor is it feasible to run a controlled field experiment that randomly assigns tax morale to individuals or income groups within countries.[5] Nonetheless, we think it is important to fully exploit the available data in order to improve our understanding of the relationship between tax morale and tax policy. Therefore, we intend to approximate the causal effect as precisely as possible by employing a (non-randomly assigned) IV and conditioning on a rich set of control variables (including country group fixed effects, income, education, trust, religiosity, patriotism and occupation).

Using this empirical strategy, we find that groups with higher levels of tax morale, *ceteris paribus*, face higher average and marginal tax rates. The results are robust to various specification checks. Further research would be helpful to confirm our suggestive

[3]The only study that we are aware of which uses tax morale as an explanatory variable (to explain subjective well-being) is Lubian and Zarri (2011).

[4]Qari et al. (2011) show that countries with higher levels of patriotism typically have higher levels of taxation as well. See Torgler (2007) for an extensive overview of the literature on the determinants of tax morale. Alm and McClellan (2012) show that the concept of tax morale is not restricted to individuals but can also be applied to firms.

[5]Slemrod and Weber (2012) survey the empirical tax evasion literature and conclude that causal identification is a major shortcoming of this strand of literature. Some studies have also used laboratory experiments to examine research questions related to tax compliance in general (Alm 2010 for a recent survey) and also tax morale (e.g., Alm et al. 1999; Feld and Tyran 2002; Wahl et al. 2010). This is, however, not an option in our context.

results and to establish a causal interpretation.

We propose three alternative mechanisms that may explain our results: *i)* The first mechanism relies on arguments from the *optimal taxation* literature. Assuming that groups with high levels of tax morale have a lower tendency to evade in response to higher taxes relative to groups with low morale, governments will impose higher taxes on the high morale group simply because taxing them creates less distortions. *ii)* The second is a *political economy* argument suggesting that groups with high levels of tax morale are less politically opposed to high tax rates. Politicians seeking reelection might then impose higher taxes on these high morale groups because they are less likely to vote against (parties in favor of) tax increases. *iii)* The third mechanism focuses on the *administrative costs* of taxation. If high morale groups are more likely to comply with taxes, politicians might seek to minimize the administrative costs of taxation by imposing higher taxes on high morale groups because it requires less enforcement costs (i.e., audits) to collect tax revenue from these groups.

The remainder of the paper is set up as follows. Section 5.2 describes the data sources and presents summary statistics. The empirical strategy and our main results as well as robustness checks are presented in Section 5.3. We discuss several mechanisms behind our empirical findings in section 5.4. Finally, section 5.5 concludes.

5.2 Data and Operationalization

In order to approach our research question empirically, we combine micro data on tax morale and other covariates from the World Values Survey (WVS) and European Values Survey (EVS) (Inglehart 2000) with information on tax rates from the World Tax Indicators (WTI) (Sabirianova Peter et al. 2010). Below we describe our approach to combine the data, discuss each data source and define our measures of tax morale and tax rates.

Tax Morale The *WVS/EVS* is the most common data source in tax morale research. It is a worldwide survey which collects comparative data on many values and attitudes using standardized questionnaires for representative national samples of at least 1000 respondents per country. The surveys are conducted by professional scientific institutions and performed through face-to-face interviews at the respondents' home and in their respective national language. We employ all five waves, which were carried out between 1981-1984, 1989-1993, 1994-1998, 1999-2004, and 2005-2008, respectively.

Our key explanatory variable, tax morale, is measured by individuals' responses to the following question:

> *Please tell me for the following statement whether you think it can always be justified, never be justified, or something in between: 'Cheating on taxes if you*

have the chance'.

The question is measured on a ten-scale index with one (1) meaning 'never justifiable' and ten (10) meaning 'always justifiable'. This is by far the most frequently used measure for tax morale (e.g., Slemrod 2003, Alm and Torgler 2006, Richardson 2006, Torgler 2006 and Halla 2012), but it is of course not free of bias. For example, Andreoni et al. (1998) argue that people might overstate their degree of morality in self-reports such as the WVS and those who have evaded might want to excuse their behavior by declaring a high tax morale. Elffers et al. (1987) find that there are significant differences between actual tax evasion and self-reported tax evasion in surveys. Nevertheless, asking about tax morale is less blunt than asking about tax evading behavior, and so the degree of honesty should be higher (Frey and Torgler 2007). Another shortcoming of the question is the fact that taxpayers might find tax evasion justifiable if tax revenue is used for, say, financing a dictator's war machine (Frey and Torgler 2007).[6]

However, previous studies show that low WVS levels of tax morale are associated with high tax evasion and vice versa (Torgler and Schneider 2009; Halla 2012). This provides evidence in our favor of the view that true tax evasion behavior can indeed be proxied with responses to questions about tax morale. As we describe below, we aggregate the WVS/EVS data on the level of income groups. This might help to cancel out incorrect reporting by respondents. Given these arguments in favor of the variable and the frequent use in the literature, we believe that it is appropriate to measure tax morale with this question.

Tax Rates Unlike tax morale, which is covered at the individual level across countries and time in the WVS/EVS, tax rates at this level are more difficult to obtain. Of course, statutory variables, such as the top marginal personal income tax rate, have very wide country-year coverage and are available from many sources. However, our analysis requires tax rates that vary across time, countries, and income groups. We rely on data from the recently published *World Tax Indicator* database to overcome these challenges. This large panel data set covers personal income tax structures at the country level in 189 countries for the period 1981 to 2005 (Sabirianova Peter et al. 2010). As it contains the complete national income tax structures, including statutory rates, tax brackets, country-specific tax formulae, standard deductions and tax credits, among others, the data allow us to compute average and marginal tax rates.[7]

[6]The question used to measure tax morale is conceptually related to another WVS question which has recently been used to study benefit morale (Heinemann 2008; Halla et al. 2010).

[7]The WTI collects tax schedule information for single tax payers only. This is not likely to have any noticeable effect on the results since very few countries tax family income (exceptions include Germany, France and the U.S.) or have tax schedules that depend on marital status (Sabirianova Peter et al. 2010).

More importantly, we are able to calculate these tax rates for any level of gross income. However, the WVS/EVS does not contain information on a respondent's personal gross income; instead of reporting their actual income level, respondents only indicate which of ten income groups (brackets) their income falls into.[8] That is, we do not know each respondent's actual income, but only her income group.

This implies that we are not able to calculate a "personal" tax rate for each respondent of the survey. We therefore use the raw WTI data to estimate average (AR) and marginal (MR) tax rates for each income group reported in the WVS/EVS for each country-year. The income level which we use to calculate the tax rate is equal to the lower bound of each country-wave-income group. Hence, the WTI are used to estimate a marginal and an average tax rate for each country-wave-income group cell which is based on the lower bound of this cell's income range.[9] Both tax rates adjust for standard deductions and credits and are calculated using country specific tax formulae.[10] Because there is no adjustment for tax evasion or avoidance, these tax rates are close to, but are not, effective tax rates. Nonetheless, they are superior to using statutory rates.

Data Structure and Summary Statistics In order to relate the calculated income group tax rates to the WVS/EVS data, all information from the WVS/EVS are aggregated (means) on the level of country-wave-income groups. We restrict the sample to employed individuals before aggregating the data in an effort to limit our analysis to individuals who potentially paid income taxes. We also exclude the lowest income group in each country-year observation from our estimations as individuals in theses groups usually do not pay income taxes; hence, we do not observe any variation in taxes within and across these groups. Finally, the aggregated WVS/EVS data are merged with the tax rates from the WTI. The unit of observation in our main analyses therefore is the country-wave-income group cell.

It seems reasonable to aggregate individual information – including tax morale – on the income group level because i) it is very unlikely that policy makers have individual level information on tax morale and ii) even if they do, they could not tax each person individually. In addition, grouping can alleviate measurement error in the covariates.

We use the tax morale question to define two measures of tax morale for our em-

[8] The provided income steps are adjusted to the respective national income distributions, but they do not reflect income deciles.

[9] The downloadable WVS/EVS data do not contain information about the income levels which correspond to each country-wave income group. We therefore had to retrieve this information from the questionnaires used in each country and wave. Estimation is restricted to 52 countries that recorded gross income in the questionnaires. For a sensitivity check (see Section 5.3.5), we also estimate lead tax rates to analyze the impact of tax morale in year t on tax rates in $t+1$.

[10] We are not able to adjust for deductions and credits that vary by individual characteristics (e.g., child credits). See Sabirianova Peter et al. (2010) for a more detail description of the WTI and the tax rates.

pirical analyses which, in the light of our data structure, are defined as follows: i) 'tax morale (ten)' represents income-group averages of the original 10-scale variable as reported in the survey; ii) 'tax morale (bi)' is based on a dummy variable that is equal to 1 for individuals who report the highest level of tax morale (1 on the original scale) and '0' for individuals who report a value greater than 1. It follows that 'tax morale (bi)' is the share of individuals in each income-group that report the highest possible level of tax morale – an operationalization commonly used (see, e.g., Alm and Torgler 2006). Both variables are coded such that a higher value implies a higher level of tax morale.

Tables 5.5 and 5.6 in the Appendix provide summary statistics for all relevant variables. The standard deviations reported in Table 5.6 reveal that there is variation across income-groups within a given country-wave for tax morale and tax rates. We report summary statistics by income group in Table 5.7 and find that while the relationship between tax rates and income is approximately linear, the relationship between tax morale and income has an inverted-U shape.

5.3 Empirical Strategy and Results

5.3.1 Empirical Model

We are interested in examining the relationship between tax rates and tax morale, which we wish to do by estimating the following model:

$$\overline{t_{ijt}} = \alpha + \beta\overline{m_{ijt}} + \overline{X}\delta + C\phi + \upsilon_i + \theta_t + \lambda_j + \epsilon_{ijt}, \tag{5.1}$$

where subscripts indicate an income group i in country j surveyed in the t-th wave of the WVS. Our dependent variable $\overline{t_{ijt}}$ is either the average or marginal tax rate. The explanatory variable of interest is $\overline{m_{ijt}}$, which is one of our tax morale measures (tax morale (ten) or tax morale (bi))[11] and ϵ_{ijt} is an *iid* error term. We further include survey wave fixed effects θ_t to capture time specific effects and income group dummies υ_i. The latter is crucial in light of the progressive structure of almost all tax systems around the world and the literature in the field of tax morale which finds that income has an effect on tax morale (e.g., Torgler 2007). λ_j indicates a set of country fixed effects.

The vectors $\overline{X}$ and C control for income group and country-level variables, respectively. $\overline{X}$ includes several confounding variables on the income-group level which are known determinants of both tax morale *and* tax rates: marital status, number of children, education, religiosity, patriotism, trust, and employment type (full time vs part time). At first glance, it might not appear obvious why patriotism and trust in other

[11] Recall that 'tax morale (bi)' is the share of individuals in each income group that reports the highest level of tax morale whereas 'tax morale (ten)' is the average of the 10-point scale value.

people are potential determinants of both tax rates and tax morale.

However, Qari et al. (2011) show that patriotism affects tax rates because patriotic citizens are less likely to move to a foreign country in response to higher taxation. In another study, Konrad and Qari (2012) show that tax morale and patriotism are related. Trust in other people is likely to confound our result of interest since trust affects tax compliance and tax morale (Scholz and Lubell 1998; Slemrod 2003). Trust also affects the size of the government via its impact on preferences for redistribution and the demand for social insurance and public goods (Luttmer 2001; Bjornskov and Svendsen 2012). Since evasion opportunities vary by occupation (e.g., Kleven et al. 2011) and possibly affect tax morale, we also control for the share of self-employed individuals in an income group.[12] We additionally include age, age squared and gender in $\overline{X}$. All variables in $\overline{X}$ are obtained from the WVS/EVS and averaged over income groups. We further include per capita GDP (in PPP), GDP growth rate, and foreign direct investments (FDI) in vector C in order to account for confounding country-level variables. All country-level variables are taken from the Worldbank's World Development Indicators (World Bank 2010).

Ignoring unobserved country-specific factors that affect both tax rates and tax morale will potentially lead to biased estimates. For example, countries such as Norway and the United States are not comparable because the former has a culture of high taxation, whereas the latter is known for low taxes. Since these two countries also have different levels of tax morale due to reasons other than the level of tax rates, relying on variation between them would bias our estimates.

While controlling for country-level variables mitigates this bias, estimating the model with a full set of country fixed effects is problematic given our data constraints: First, the fixed effects are highly collinear with tax morale, which inflates the standard errors and thus reduces the precision of the estimates. Second, controlling for country fixed effects implies including a large set of dummy variables, which reduces our degrees of freedom: this, too, leads to inflated standard errors because our number of observations is relatively small. Finally, the sample is highly unbalanced with many gaps in the data because many countries did not participate in every wave of the WVS/EVS: for example, some countries participated in the first and last waves while others participated in only one wave (see Table 5.6 in the Appendix for an overview). This makes the problem of collinearity even more serious. We find a positive relationship between tax rates and tax morale when we estimate equation (5.1) with country fixed effects (and excluding the country-level variables). However, due to the above reasons the results are only partly statistically different from zero.

[12]Robustness checks in which we restrict the analysis to self-employed individuals who have better evasion opportunities reveal a positive relationship between tax morale and tax rates and are hence similar to our baseline results. These results are available upon request.

In order to control for cultural effects and estimate precise standard errors, we form 9 groups of countries which are homogenous with respect to their tax system and include dummies for each of the 9 groups in our regressions (e.g., Helliwell 2003 employs a similar approach using WVS data). That is, we replace the country fixed effects λ_j in equation 5.1 with country group fixed effects. This approach mitigates the discussed problems of country fixed effects while still controlling for heterogeneity in the culture of taxation and tax morale.[13] The groups are defined as: 1) English-speaking countries (Anglo-Saxon plus Australia and New Zealand), 2) Mid-central Europe plus Israel, 3) Southern Europe, 4) Scandinavia, 5) Eastern Central Europe, 6) former Soviet countries, 7) Latin America, 8) Asia, and 9) (other) Developing Countries (see Table 5.6 for an overview of countries in each country groups). To the extent that country-level time-invariant factors are common across countries within a region, this approach should address any bias resulting from omitted cultural fixed effects, even if imperfectly so.

Regarding the source of variation, the use of income-group and country-group dummies eventually implies that our empirical analysis exploits variation in tax morale between the same income groups within a country-group, thereby conditioning on a wide set of control variables as well as wave fixed effects.

5.3.2 Ordinary Least Squares

Estimation Results We start examining the data by estimating the country group fixed effects model with Ordinary Least Squares methods in order to learn about the relationship between tax morale and tax rates.[14] The OLS results are displayed in table 5.1: Specifications I and II include the effect of tax morale – measured on the ten and two point scale, respectively – on average tax rates, while III and IV show the effect on marginal tax rates. All four coefficients are positive but the effect is only statistically distinguishable from zero when tax morale (bi) is the explanatory variable of interest; we observe a clear significant and positive relationship between tax morale and tax rates in specifications II and IV.

Bias in OLS estimates While the OLS results provide a useful first glance at the data, we suspect they are biased for two main reasons. First, previous research has argued that there is a feedback effect from tax rates (or the general system of taxation) to the level of tax morale, which implies reverse causality bias. That is, taxes are not only influenced by the level of tax morale, but they also affect tax morale. Second, we are

[13] We present robustness checks in section 5.3.5 in which we conduct the analysis on the individual person level (rather than the income group level) in the presence of actual country fixed effects.

[14] All subsequent estimations display panel-adjusted standard errors that account for clustering effects of a certain country's income groups and are robust to the presence of heteroscedasticity.

Table 5.1: Effect of tax morale on tax rates: OLS regressions

Model	I	II	III	IV
Dependent Variable	AR	AR	MR	MR
Tax Morale (ten)	0.939		0.779	
	(0.718)		(0.750)	
Tax Morale (bi)		11.172***		10.626***
		(3.946)		(4.012)
Constant	−10.970	−5.672	−4.477	−0.182
	(31.382)	(30.471)	(31.265)	(30.843)
N	576	576	575	575
R2	0.529	0.534	0.548	0.553
Country Group FE	yes	yes	yes	yes

[1] Dependent variables are average (AR) and marginal (MR) tax rates [2] Tax Morale (ten) is based on the original ten-point scale. Tax Morale (bi) is based on a dummy variable indicating the highest level of tax morale. [3] Estimation is by OLS with clustered standard errors [4] All specifications condition on full set of control variables (see Table 5.8 in the appendix for the full specification) [5] * < 0.10, **< 0.05, *** < 0.01

not able to control for the actual extent of tax evasion or the size of the shadow economy – except for the part captured by the country group fixed effects. Since both variables are known determinants of tax rates and tax morale, we must also contend with omitted variable bias. In order to evaluate the presented OLS results, it is important to discuss the direction of the bias. In the following, we argue that omitting measures of the shadow economy and tax evasion is expected to induce a downward bias in the OLS estimates.

Empirical evidence suggests that tax rates and shadow economy/evasion are positively related (e.g., Gorodnichenko et al. 2009 and Schneider et al. 2010 for observational studies and Alm 2012 for experimental work). The reason is that in countries with high levels of evasion and shadow economy, the government can only rely on a smaller tax base and has to use higher rates in order to finance (a given level of) public expenditures – under the assumption that governments consider the level of evasion as given. At the same time, empirical evidence shows that the shadow economy is negatively correlated with tax morale because people are less willing to pay their taxes if many of the fellow citizens engage in tax evasion (e.g., Frey and Torgler 2007). Hence, the omitted shadow economy variable is likely to be positively correlated with tax rates and negatively correlated with tax morale; resulting in a downward bias of the estimated (positive) OLS effect of tax morale on tax rates.

Additionally, the bias generated by reverse causality is likely to be directed downward as well because higher tax rates presumably have a negative effect on tax morale.

Although the tax morale literature is mostly silent in this respect, the scarce empirical evidence suggests that the effect is negative (e.g., Torgler et al. 2010). Indeed, it seems reasonable to expect a negative relationship for at least two reasons. First, higher tax rates make evasion more profitable, *all else equal*, and therefore increase the temptation to evade. Second, if the perception is such that the government keeps too much of an individual's gross income, it is likely that a negative attitude towards the government is triggered.

Summing up, the positive effect of the shadow economy on tax rates and its negative effect on tax morale as well as the negative effect of tax rates on tax morale are likely to create downward bias in the OLS estimates. This possibly explains why we cannot distinguish the effect of tax morale (ten) on tax rates from zero in Table 5.1. The biased OLS estimations hence require a different empirical strategy which we put forward in the next section.

5.3.3 Instrumental Variables

In order to approximate the true effect of tax morale on tax rates more precisely, we require variation that is exogenous conditional on the country-group fixed effects and all other control variables in equation 5.1. In other words, we require an instrumental variable (IV) z_{ijt} that is independent of the error term in equation 5.1. Most importantly, the IV should overcome the two main threats to identification discussed in the previous section: it needs to be uncorrelated with *i)* tax rates in order to overcome the problems stemming from reverse causality, and *ii)* the size of the shadow economy to mitigate potential omitted variable bias.

Unfortunately, it is not possible to obtain (quasi) experimental evidence to answer our research question as this requires either a controlled field experiment or an IV based on random assignment. Although both alternatives are unavailable in our setting with many countries over time, we argue that the research question at hand is sufficiently relevant so that, even in the absence of randomization based evidence, an attempt ought to be made to pursue an empirical analysis (see the discussion in the Introduction).

Instrument

Finding suitable instruments is generally a difficult task. We require a variable that is related to income group tax morale, but does not have any direct link to the same income group's tax rate. Given the structure of our dataset, we also require a variable with sufficient variation across income groups. Our chosen IV is based on a question asked in the WVS/EVS. In this regard, our strategy is similar to the approach by Lubian and Zarri (2011) who also use tax morale as an explanatory variable (to explain happiness) and instrument it with another question from the same survey.[15]

The IV we employ in the subsequent analyses exploits the answer to the question '*Tell me whether you think it can always be justified, never be justified or something in between, to avoid a fare on public transportation*'. Respondents are asked to respond to the question on a 10 point scale where 1 implies never justified. We first reverse the scale of the variable so that higher values indicate higher 'dodging-fares-morale' and then use the income group average as our IV. We presume that many individuals who have high tax morale also develop a high level of 'dodging-fares-morale' because of general moral attitudes driving both types of morality. Therefore, this instrumental variable allows us to exploit variation in general morale attitudes that is not related to other variables affecting both tax rates and tax morale (see below for a thorough discussion of the exclusion condition). A valid IV in our set-up needs to be sufficiently correlated with tax morale and orthogonal to the error term in the equation of interest. We discuss both conditions below.

Instrument Relevance Instrument relevance is testable and requires examining the effect of the instrument on the explanatory variable of interest conditional on all control variables. In our set-up this first-stage equation reads:

$$\overline{m_{ijgt}} = \gamma_0 + \gamma_1 z_{ijgt} + \overline{X}\delta^1 + C\phi^1 + \upsilon_i^1 + \theta_t^1 + \lambda_g^1 + \varsigma_{ijgt}^1, \tag{5.2}$$

where subscript g now indicates one of nine country groups and all other subscripts are defined as in equation (5.1). z_{ijgt} is our instrument and all other variables are as described above.

Our first-stage results are depicted in Table 5.2 (as well as Table 5.9 in the appendix where the effects of all covariates are displayed). The results indicate that our chosen instrument is strongly correlated with the variable of interest, tax morale. Models I and III show the effect of the instrument "Cheat" on tax morale (ten) and II and IV show the effect on the tax morale (bi). The coefficient of interest is positive and highly significant in all four models suggesting that the effect is as expected a-priori and that the instrument is relevant. The observation of a positive correlation between our IV and tax morale is not surprising since individuals who report high tax morale are also likely to develop a high level of 'dodging-fares-morale'. The instrument's relevance is further supported by F-statistics of excluded instruments which are considerably larger than the critical value of 10 (Staiger and Stock 1997). The F-statistics in our set-up vary between 33 and 55 thus putting us in the 'safe zone'.[16]

[15] Lubian and Zarri (2011), however, use a different dataset (for Italy) and their instrument is not available in our data.

[16] Note that the number of observations in these first-stage regressions varies because they are estimated

Table 5.2: First Stage Regressions of tax morale on 'dodging-fares-morale'

Model	I	II	III	IV
Dependent Variable	TM10	TM2	TM10	TM2
Cheat Transport	0.412***	0.059***	0.418***	0.060***
	(0.056)	(0.010)	(0.056)	(0.010)
Constant	4.678*	−0.227	4.695*	−0.226
	(2.612)	(0.387)	(2.615)	(0.388)
N	504	504	503	503
R2	0.670	0.703	0.673	0.703
Country Group FE	yes	yes	yes	yes
F Statistic	54.95	33.86	55.52	33.80

[1] Dependent variables is Tax Morale (ten) in *I* and *III* and Tax Morale (bi) in *II* and *IV*. Tax Morale (ten) is based on the original ten-point scale. Tax Morale (bi) is based on a dummy variable indicating the highest level of tax morale. [2] First Stage Results corresponding to models *I* to *IV* in equation 5.3 [3] Instrument: Cheating on Public Transportation [4] All specifications condition on full set of control variables (see Table 5.9 in the appendix for the full specification) [5] *F-statistic* indicates F-statistic of excluded instruments [6] The corresponding second-stage results are displayed in Table 5.3. [7] * < 0.10, **< 0.05, *** < 0.01

Exclusion Restriction The second condition of valid instruments, exclusion restriction, is not testable and can only be defended on the basis of economic theory and intuitive reasoning. Our main concerns for identification are reverse causality bias and omitted variable bias. Although the latter is primarily due to the omission of the shadow economy, it is also possible that our IV introduces additional sources of omitted variable bias. We argue that our IV and model specification are able to overcome some of these concerns, even though we are not able to establish a clean causal identification. In other words, although our IV lacks random assignment, we argue below that it generates income-group-level variation in tax morale that is exogenous *conditional* on the vector of covariates in equation (5.1). In particular, the IV generates variation that is not driven by either the tax system or the shadow economy.

First, we argue that it is reasonable to assume that individuals do not directly link public transportation to tax legislation. The willingness to pay for public transportation is instead likely to be shaped by the quality of the public transportation system (e.g. coverage, prices, timeliness). Additionally, Algan and Cahuc (2009), show that 'civic attitudes' (i.e., 'benefit morale' from WVS in their case) do not change systematically in response to changing institutions. Therefore, it is unlikely that 'dodging fares' is *directly*

on the same sample as the corresponding second-stage regressions.

driven by tax rates. Nonetheless, there might be an *indirect* link between tax rates and our IV via trust and levels of patriotism (see previous discussion). For example, patriotic citizens might be more willing to pay for public transportation in "their" country and therefore have a higher "doging-fare-morale". We control for this indirect link between tax rates and our IV by including patriotism and trust in our regressions.

Second, we argue that our IV helps to prevent potential omitted variable bias due to the omission of measures of the shadow economy. Using a similar argument as before, it is unlikely that variation in the shadow economy is related to the system of public transportation and the willingness to pay for it. Variation in the level of shadow economic activity is unlikely to have any effect on the willingness to pay for public transportation which, as argued before, is rather shaped by the quality of public transportation itself. It is therefore unlikely that the level of shadow economic activity *directly* affects our IV. Still, it is possible that the shadow economy *indirectly* affects our IV through levels of trust and patriotism. For example, individuals might lose their trust in other people once they realize that many people work in the shadow economy. Similarly, people could lose pride for a country that is characterized by high levels of shadow economic activity. However, this is not a problem for our analysis since we control for both trust and patriotism.

Third, because the IV is not based on random assignment, it is likely to be correlated with other variables that affect tax rates. For example, high income individuals might have higher 'dodging-fares-morale' because bus and train fare is a relatively smaller component of their income. Since high income individuals also have higher tax rates, our 2SLS estimates would be biased if we do not control for income groups.[17] A similar argument can be constructed for education. Educated individuals acknowledge the value of public transportation better and therefore dodge less fares. Since the well-educated are also richer on average and therefore face higher tax rates, the estimates would be biased if we did not condition on the level of education in an income group. The omission of cultural differences in tax rates, tax morale, and the IV (the previously discussed example with Norway and the United States illustrates this) may also lead to biased estimates. We address this problem by including country-group fixed effects, where each group represents a set of homogeneous countries.

An additional identification issue arises for countries where public transportation is operated by the government and where taxes and public transportation are financed from the same budget. If, in those countries, many people dodge fares, the transportation sector might have to be cross-subsidized by tax revenue, which implies higher taxes, *ceteris paribus*. However, public transportation is privately or semi-privately run in most countries. Moreover, semi-privately run transportation companies usually operate on a

[17] In our data set the relationship between 'dodging-fares-morale' and income rather has an inverted-U shape.

different budget than the tax legislating government.

5.3.4 Estimation Results

This section describes the second stage results, which are obtained by estimating the following second stage regressions:

$$\overline{t_{ijgt}} = \alpha + \beta\overline{m_{ijgt}} + \overline{X}\delta^2 + C\phi^2 + \upsilon_i^2 + \theta_t^2 + \lambda_g^2 + \varsigma_{ijgt}^2 \tag{5.3}$$

where m_{ijgt} represents the fitted values of tax morale from the first-stage regressions (Equation 5.2) and all other variables and subscripts are defined as defined before. Our coefficient of interest is β.

The second stage IV results are displayed in Table 5.3 (the full specification is presented in Table 5.10 in the appendix). First, note that the coefficients from the 2SLS estimations are larger than those of the OLS regressions reported in Table 5.1. This is consistent with our expectations on the OLS bias; OLS estimates are downward biased due to the omission of a measure of the shadow economy and reverse causality. The larger coefficients also suggest that our IV accounts for some of the endogeneity concerns in OLS estimations.

Models I and III in Table 5.3 depict the effect of tax morale as measured on the ten point scale on average and marginal rates, respectively. Accordingly, models II and IV measure the impact of tax morale (bi) on both types of tax rates. The coefficient on tax morale is positive and highly significant in all four specifications. Additionally, the coefficient on tax morale (bi) is considerably larger than that on the ten-point-scale tax morale variable. This is intuitive because the former tells us the change in average tax rates if a group's average tax morale increases to the highest level, whereas the former indicates the effect of an increase in average tax morale in an income group by 1 on the ten point scale.

We observe that a one standard deviation increase in tax morale (ten) increases the average rate by 0.659 ($= 12.441 \times 0.806/15.208$) and the marginal rate by 0.583 standard deviations. A one standard deviation increase in 'tax morale (bi)' increases the average and marginal tax rate by 0.970 and 0.865 standard deviations, respectively.[18]

Regarding the effect of confounding variables (see Table 5.10 in the appendix), we mostly observe theoretically expected effects: We include dummies for each income

[18]Recall that tax morale (bi) is the share of individuals in an income group with the highest level of tax morale. Therefore, an alternative interpretation of our coefficients is that, relative to groups where subjects have levels of tax morale between 1 and 9 on the 10 point scale, average tax rate is 86.239 percentage points higher for income groups where everyone has the highest level (10 on the scale) of tax morale. However, since none of the groups in our sample has this extreme value of high average tax morale, it makes more sense to use the standard deviations interpretation of the results. The same reasoning applies to 'tax morale (ten)'.

Table 5.3: Effect of tax morale on tax rates: Second Stage IV Regressions

Model	I	II	III	IV
Dependent Variable	AR	AR	MR	MR
Tax Morale (ten)	12.441***		11.193***	
	(2.698)		(2.743)	
Tax Morale (bi)		86.239***		78.254***
		(18.814)		(19.260)
Constant	−86.744	−9.013	−76.794	−6.593
	(56.847)	(47.236)	(54.439)	(46.063)
N	504	504	503	503
R2	0.359	0.294	0.422	0.380
Country Group FE	yes	yes	yes	yes
F Statistic	54.95	33.86	55.52	33.80

[1] Dependent variables are average (AR) and marginal (MR) tax rates [2] Tax Morale (ten) is based on the original ten-point scale. Tax Morale (bi) is based on a dummy variable indicating the highest level of tax morale. [3] IV regressions with clustered standard errors. The corresponding first stage results are displayed in table 5.2 [4] Instrument: Cheating on Public Transportation [5] All specifications condition on full set of control variables (see Table 5.10 in the appendix for the full specification) [6] *F-statistic* indicates F-statistic of excluded instruments [7] * < 0.10, **< 0.05, *** < 0.01

group in the analysis in order to capture non-linear effects of income. Reflecting the progressivity of tax systems in almost all countries in the analysis, we clearly observe that higher income groups pay higher tax rates. Dummy variables for each survey wave indicate that there is a trend in time towards lower levels of taxation – a trend that is widely recognized (Sabirianova Peter et al. 2010). We further observe that tax rates are higher in rich countries and, all else equal, a higher level of foreign direct investments (FDI, as a proxy for openness) yields lower rates. We also see from the data that groups with high average numbers of children face lower rates (for reasons of clear arrangement, the displayed result tables do not show the coefficients of included sociodemographic control variables).

Overall, we find evidence for a positive effect of tax morale on tax rates. It seems to be the case that income groups with high levels of tax morale, ceteris paribus, face systematically higher average and marginal tax rates.

5.3.5 Robustness Checks

In this section, we present several robustness checks in order to evaluate the sensitivity of our results. In particular, we first present evidence that our results are robust to the inclusion of a full set of country fixed effects (as opposed to the country group fixed effects on which we condition in our previous estimations). In a second step, we show that our results are robust to: i) the inclusion of a control variable on institutional quality, ii) regressions that are weighted with the number of observations in each income group, iii) using lead tax rates as dependent variables, and iv) estimating the effect for the more homogeneous set of OECD countries.

Country Fixed Effects in Individual Level Regressions

The results presented so far rely on within country group variation where each country group represents a set of countries that are homogeneous with respect to their tax system and culture of taxation. As we discussed earlier, controlling for a full set of country fixed effects is difficult and leads to upward biased standard errors. However, since controlling for country fixed effects might yet be important in order to ensure that the results are not driven by cultural differences within the country groups, we explore individual level data as a robustness check; this allows us to include the full set of country fixed effects.

This approach provides greater within country variation than our previous estimations based on averages of individual values within income groups. The challenge in implementing this approach is that we do not have individual level data on income. Instead, the WVS provides data on each respondent's income group. We take several approaches in order to overcome this challenge and estimate our models on the individual level using country fixed effects. Running the regressions on the individual level will yield coefficients that are an order of magnitude smaller than the income group estimates reported in Table 5.3. This can be explained by the increased variation in the explanatory variables, which results in smaller coefficients on the individual level compared to the aggregate level.[19] Additionally, whereas before our coefficients represented the effect of a one-unit increase in a *group*'s average tax morale, in the individual regressions they represent a one-unit change in an *individual*'s tax morale. This also yields smaller coefficients.

We use three separate estimation strategies to implement the analysis on the individual level: (1) 'inverse multilevel analysis' with dependent variable on the income group-level and independent variables on the individual level, (2) interval regression, and (3) imputation of individual income levels and corresponding tax rates. The three ap-

[19]See, e.g., Wooldridge (2006). For intuition, consider a simple univariate OLS regression of y on x where the slope coefficient is given by the covariance of x and y over the variance of x. Estimation at the individual level allows us to use the maximum available variation in each variable (see Table 5.11 in the Appendix for Summary Statistics of the used sample). Because the variation on the individual level is greater than on the group level, the slope coefficient decreases c.p. (the actual change also depends on the change in the covariance). In addition, the smaller coefficients could be an indicator of attenuation bias due to measurement error in the explanatory variables, which may be alleviated by grouping the covariates. Therefore, we prefer to keep the income group level regressions as our baseline specifications.

Table 5.4: Effect of tax morale on tax rates: Individual Level Fixed Effect Estimations

Expl. Variable	Tax Morale (ten)		Tax Morale (bi)	
Dependent Variable	AR	MR	AR	MR
Panel A: Inverse Multilevel Analysis				
Country Group Fixed Effects:				
Tax Morale	0.866***	0.761***	4.851***	4.265***
	(0.144)	(0.149)	(0.790)	(0.820)
Country Fixed Effects:				
Tax Morale	0.168***	0.157***	0.937***	0.871***
	(0.051)	(0.054)	(0.279)	(0.299)
Observations	30,024	30,024	30,024	30,024
Panel B: Interval Regression				
Country Group Fixed Effects:				
Tax Morale	0.861***	0.701***	4.830***	3.935***
	(0.141)	(0.134)	(0.787)	(0.753)
Country Fixed Effects:				
Tax Morale	0.167***	0.151***	0.929***	0.844***
	(0.050)	(0.048)	(0.276)	(0.265)
Observations	29,816	29,816	29,816	29,816
Panel C: Multiple Imputation Approach				
Country Group Fixed Effects:				
Tax Morale	0.865***	0.723***	4.850***	4.056***
	(0.143)	(0.142)	(0.787)	(0.784)
Country Fixed Effects:				
Tax Morale	0.171***	0.163***	0.953***	0.906***
	(0.050)	(0.051)	(0.274)	(0.283)
Observations	29,834	29,834	29,834	29,834

[1] 2SLS IV regressions [2] Instrument: cheating on public transportation [2] Tax Morale (ten) is based on the original ten-point scale. Tax Morale (bi) is based on a dummy variable indicating the highest level of tax morale. [4] Individual person level [5] All specifications include the same control variables as the previous, but exclude country-level variables [6] Groups as defined before (see Appendix) [7] Multiple Imputation Standard Errors are calculated following Reiter (2003) [8] * < 0.10, **< 0.05, *** < 0.01

proaches differ in how they treat the (unobserved) variation in the dependent variable. In the first approach, we neglect this variation by assigning each individual the reported (midpoint) income and corresponding tax rate (i.e. AR or MR) of his/her income group. In the second approach, we assign each individual her income group's lower and upper bound tax rates and run an interval regression, i.e. a generalized Tobit regression for interval censored data (Wooldridge 2006). In the third approach, we employ a multiple imputation procedure, which is comparable to the approach proposed by Jenkins et al. (2011), and described in more detail in Appendix 5.7, in order to generate even more variation on the left-hand side of our estimations. For each individual, we randomly draw an income (and corresponding AR and MR) lying between the lower and upper bounds of her income group and then estimate the model. This is repeated 500 times and the average coefficients are reported. We employ the combination method proposed by Reiter (2003), and applied by Jenkins et al. (2011), to calculate standard-errors and levels of significance, taking into account the finite number of imputations. Note that randomly assigning tax rates within each bracket generates noise and renders the relationship between tax morale and tax rates less strong. This might lead to smaller coefficients – hence, it is a conservative approach for the expected positive relationship.

Our analysis includes the following control variables; i) the same individual-level controls as in the baseline, ii) survey wave fixed effects, iii) income group fixed effects, and iv) country fixed effects. We exclude country-level variables from our estimations in order to use the full sample of countries, including those that are only part of one WVS wave. Results from these regressions are presented in Table 5.4.

The results are very similar for the three approaches – despite the different treatment of the variation in the dependent variable. As expected, the individual level estimations are an order of magnitude smaller than the income group estimates. Nonetheless, the coefficients – for both tax morale (ten) and tax morale (bi) – are positive and statistically different from zero. This is not only true for the country group fixed effects but also when including actual country fixed effects. Of course, the estimated coefficients become smaller since country fixed effects take out more variation. However, as all country specific effects are captured and controlled for, the positive coefficients can be attributed solely to within-country variation in tax morale. This provides further evidence that our results are not only driven by heterogeneity between countries within a country group. For example, using the inverse multilevel analysis including the full set of country fixed effects, we find that a one standard deviation rise in an individual's tax morale (ten) increases her group's marginal tax rate by 0.023 ($= 0.157 \times 2.336/15.642$) standard deviations.

Further Robustness Checks

For our set of sensitivity checks, we go back to income-group level estimations with country group fixed effects. We first include a country level measure of bureaucratic quality (ICRG 2011) – a variable found to be a possible determinant of tax morale (Barone and Mocetti 2011) – in order to control for the possibility that countries with less efficient governments have higher tax rates (Brennan and Buchanan 1980). Unfortunately, this variable is not available for all country-year observations which is why we only include it in a robustness check. We also run regressions in which we weight the income groups by the number of their members. Additionally, we restrict the analysis to OECD countries in order to gain insights for a more homogeneous set of countries. Finally, we employ lead tax rates where tax rates in year $t + 1$ are related to tax morale in year t. Table 5.12 in the Appendix summarizes these sensitivity checks.

We are able to confirm our baseline results in all sensitivity checks. Tax rates, ceteris paribus, depend positively on the level of tax morale and the results are mostly significantly different from zero. In most specifications, the size of the tax morale point estimates is roughly similar to the sizes in the baseline. Interestingly, when the sample is restricted to OECD countries, we find significant point estimates that are slightly larger than in the baseline scenario. Additionally, we confirm Brennan and Buchanan's (1980) argument of a negative relationship between tax rates and the quality of bureaucracy.

5.4 Discussion and Potential Explanations

Our empirical analysis provides suggestive but robust evidence for a positive effect of tax morale on tax rates – a relationship that has not been examined in the literature before. While we think that we have fully exploited the data to approximate the causal effect as precisely as possible, we are not able to rely on random-based evidence that would be able to establish causality. We therefore find it important to discuss potential mechanisms that might explain our finding that tax rates are highest on the most honest. This section describes three such mechanisms. *i)* an inverse elasticity argument where governments seek to minimize economic distortions, *ii)* a political economy argument where governments take voting behavior into account, and *iii)* an administrative costs argument where taxing high morale groups is more cost efficient. While the latter arguments do not make any assumptions about government behavior, the first shows that it is even possible to derive such a tax setting from a model with a benevolent welfare maximizing government based on optimal tax considerations.

Inverse Elasticity Rule Individuals with different levels of tax morale most likely respond differently to increases in tax rates: individuals with low tax morale, and therefore low subjective costs of evading, are likely to respond by evading more relative to subjects with high levels of tax morale, who feel an intrinsic motivation to pay their taxes. This argument implies that individuals with high levels of tax morale may have lower elasticities of evading. Following an inverse elasticity rule reasoning, governments setting tax rates that minimize distortions created by the tax system will account for heterogeneous levels of evading elasticities by imposing higher tax rates on groups with high average levels of tax morale, *ceteris paribus*. This potential driver behind our findings is based on a simple Ramsey-type argument: groups with less elastic responses to tax rate changes should be taxed at higher rates than groups with elastic responses. It is straight-forward to formalize this story using utility maximizing individuals who have heterogeneous levels of subjective evasion costs and a benevolent welfare maximizing government (see Appendix 5.8).

This argument of course implies that governments are able to observe the average levels of tax morale in different income groups. While politicians usually do not consult surveys like the WVS/EVS before passing tax reforms, they might learn about tax morale levels from history over time. Algan and Cahuc (2009), for example, show that "civic attitudes", which are likely to contain tax morale, tend to be very persistent over time. Politicians are therefore likely to have an impression of different income or occupation groups' tax morale.

Although the idea that governments seek to minimize distortions is appealing to economists, it is not an empirical fact that politicians indeed set tax rates following the laws of optimal taxation. One can interpret our paper as one of the first empirical papers supporting this view. The lack of evidence in the literature, along with a long line of political economy findings suggesting that governments do not only seek to maximize welfare, could nevertheless also indicate that politicians set higher tax rates for high morale groups not *only* to minimize distortions but also for other reasons.

Maximizing Votes An alternative explanation for our findings is that groups with high levels of tax morale might also be less politically opposed to high tax rates. This is supported by the literature on tax morale research finding that the perception of the tax system and the government are positively correlated with tax morale (see Torgler 2007 or Doerrenberg and Peichl 2013 for overviews). Politicians facing strict revenue requirements and reelection might then impose higher taxes on these high morale groups because they are less likely to vote against (parties in favor of) tax increases. For example, a Gallup poll in 2011 showed that 71% of Democrats in the US are in favor of higher taxes (on the rich) whereas 69% of Republican voters were against it. If this heterogeneity in attitudes towards higher taxes is correlated with tax morale – which, based on the tax morale literature, we consider to be reasonable – then our results might be driven by such a political economy argument.

Administrative Costs Governments might also seek to minimize the administrative costs of taxation and its enforcement. If high morale groups are more likely to comply with taxes, politicians might be inclined to impose higher taxes on these groups simply because it is more cost efficient: it requires less enforcement costs (i.e., audits) to collect tax revenue from groups with high tax morale. This argument is obviously related to the previous "inverse elasticity" argument, but it is based on a different assumption regarding the intentions of politicians. They do not intend to minimize distortions but instead try to gather as much tax revenue as possible with a given administrative budget.

Other Mechanisms We acknowledge the possibility that no single argument explains our results, but that our findings are driven by a combination of different mechanisms. We also do not claim that our list of possible mechanisms is exhaustive. However, there are at least two seemingly plausible mechanisms that we argue do not drive our results. For example, one potential mechanism is the following: some countries have both high tax rates and efficient tax enforcement, which leaves little scope for tax evasion. On the other hand, other countries have lower tax rates and less efficient tax enforcement and thus a lot of room for evasion. If tax morale is positively related with the strength of enforcement, then governments in the "low tax low enforcement" countries will have to impose lower taxes on factors that are susceptible to evasion or avoidance by necessity.

We argue that this mechanism neglects the fact that the tax morale question in the survey ask respondents about a hypothetical situation: do you think evasion is justifiable if the opportunity to evade exists? The ability to evade does not imply that an individual will find it justifiable to evade. The story is also inconsistent with previous research on tax morale that finds no robust association between tax morale and tax enforcement parameters such as the fine rate or audit probability (Torgler 2005; Torgler and Schneider 2007). More importantly, our empirical analysis established the positive relationship between tax morale and tax rates conditional on country-group and country fixed effects, i.e. exploiting within country (group) variation. In other words, cross country differences are not driving our findings.

Another possible mechanism is the following: Suppose two individuals have the same true gross income. One individual with high tax morale reports the full amount for tax purposes, whereas the other person has low morale and therefore reports less. In a system with progressive taxation, the high morale subject will then be taxed at a higher rate. Our results would only be driven by this mechanical effect if individuals with low morale indeed underreported their true gross income in the WVS/EVS data, just as they do on their tax returns. We argue that this is likely not the case: First, the fact that respondents are asked to indicate the income group in which they fall instead of their exact income alleviates the concern of systematic underreporting by tax evaders in the survey. Second, since the WVS/EVS is an international academic survey, we consider it unlikely that individuals underreport their income class in fear of any connection between the tax authorities and the conductors of the survey. We do not wish to neglect that income might be misreported in some cases, but we do not find it plausible that the misreporting is more prevalent among low morale respondents relative to high morale respondents.[20]

5.5 Conclusion

In this paper, we construct a unique dataset of (average and marginal) tax rates and tax morale parameters in order to provide evidence of the relationship between tax morale and the tax burden imposed on different income groups. Combining data from the EVS/WVS and the WTI, we find empirical support showing that 'nice guys finish last', i.e., groups with higher tax morale have to bear a higher tax burden. In order to explain our results, we propose three different mechanisms which could be drivers of our results: governments may tax high morale groups higher because of intentions *i)* to minimize distortions caused by the tax system, *ii)* to get reelected, and/or *iii)* to minimize the administrative costs of taxation.

One should note that the empirical strategy used in this paper has limitations. Unfortunately, a (quasi-) experimental set-up is not available to investigate the impact of tax morale on the distribution of tax burdens. As a result, we rely on a non-random IV strategy and conditioning on a rich set of control variables including country group fixed effects to approximate the true causal effect. Although we believe this is an important first step in answering the question posed here, it would be interesting to test our results with data that allow for cleaner identification in future research.

Our findings shed new light on the growing literature on tax morale. So far, scholars have mostly argued that a high general level of tax morale is advantageous for a society because it increases the efficiency of a tax system. Many empirical studies have worked out possible determinants of tax morale and derived the policy implication that strengthening these determinants helps to increase tax morale and therefore the efficiency of raising taxes. While we do not contradict this view, we show that governments already seem to exploit high relative levels of tax morale among particular groups and, *ceteris paribus*, tax them higher than low morale groups in the same country.

The welfare implications of our findings are, however, less clear. Although high tax morale groups are taxed more heavily, they may still benefit from this policy if they receive some kind of 'warm glow' due to the intrinsic satisfaction of doing the right thing

[20] Recall that we do not relate a person's actual tax rate to his tax morale. Instead, we relate the average morale for each income group to the income group's tax rate using the income group's gross income class.

when complying with the tax law. This is also in line with recent observations from the US, where rich individuals like Warren Buffet claim they would like to pay higher taxes, provided that other rich people also face higher taxes. When deriving policy implications from our findings, it is important to account for endogenous tax morale (see, e.g., Traxler 2010, for endogenous tax morale in the standard model of tax evasion). A tax policy as sketched in our study could be self defeating in the long run if it created incentives to develop a lower level of tax morale.

5.6 Appendix A: Tables and Results

5.6.1 Summary Statistics: Income Group Level

Table 5.5: Summary statistics

Variable	Mean	Std. Dev.	N
Average Tax Rate (AR)	23.311	15.208	576
Marginal Tax Rate (MR)	30.147	15.475	575
AR lead	23.27	15.626	549
MR lead	29.924	15.736	548
Tax Morale Ten	8.49	0.806	583
Tax Morale Bi	0.575	0.171	583
Cheat public transp	8.547	0.671	511
Full time	0.722	0.151	583
Part time	0.126	0.094	583
Self employed	0.152	0.136	583
Single	0.296	0.136	583
Married	0.631	0.168	583
Divorced	0.052	0.075	583
Widowed	0.021	0.033	583
Age	39.071	3.787	583
Age squared / 100	15.409	2.983	583
Male	0.607	0.123	583
Numb children	1.764	0.560	583
Education	19.375	3.274	583
Church once month	0.347	0.244	583
Patriotism	3.32	0.374	583
Trust most people	0.358	0.187	583
GDP per cap ppp	19.725	11.344	583
GDP growth	2.714	3.292	583
FDI net inflows	12.796	69.635	583
Institutional Quality	3.148	0.879	542

Table 5.6: Means and Standard Deviations of Key Variables by country and year

Country	Group	Year	TM ten	TM bi	AR	MR	cheat
Albania	East Eur	2002	9.079	0.587	17.672	21.972	8.583
			0.185	0.049	5.449	6.337	0.250
Australia	Anglo	1995	8.616	0.563	44.611	46.441	8.819

continues on next page

continued from previous page

			0.209	0.059	0.912	1.480	0.187
		2005	8.701	0.546	23.058	33.248	8.361
			0.164	0.089	7.708	12.012	0.498
Austria	Mid Eur	1999	8.687	0.553	21.224	33.229	8.509
			0.184	0.039	7.578	11.275	0.415
Belgium	Mid Eur	1990	6.290	0.261	24.841	39.387	8.217
			0.528	0.056	8.187	9.122	0.375
		1999	6.941	0.310	23.078	36.633	8.538
			0.524	0.063	9.618	13.581	0.267
Brazil	Latin	1990	7.789	0.531	24.751	24.885	8.341
			0.540	0.055	0.304	0.325	0.467
Bulgaria	East Eur	1999	8.804	0.626	14.997	24.356	.
			0.429	0.097	5.269	7.095	.
Canada	Anglo	1982	8.920	0.704	17.103	26.267	8.687
			0.724	0.164	6.643	9.177	0.627
		1990	8.437	0.546	24.224	31.524	8.764
			0.500	0.090	6.121	9.489	0.325
		2000	8.886	0.657	21.856	30.253	8.662
			0.272	0.088	7.053	9.844	0.177
		2006	9.116	0.655	20.900	27.254	8.786
			0.246	0.053	6.807	9.239	0.323
Chile	Latin	1990	9.121	0.748	14.285	24.712	8.040
			0.183	0.048	11.747	17.849	0.432
		1996	8.818	0.634	17.509	28.563	8.582
			0.416	0.081	11.651	15.678	0.400
		2000	8.726	0.697	22.826	34.549	7.546
			0.550	0.068	12.185	13.709	0.739
Croatia	East Eur	1999	8.187	0.578	19.322	26.484	7.785
			0.526	0.079	9.609	12.704	0.487
Czech Republic	East Eur	1999	8.785	0.527	13.666	19.210	7.863
			0.190	0.080	2.644	4.723	0.409
Denmark	Scand	1981	8.436	0.618	35.985	48.582	9.162
			0.417	0.079	8.459	14.175	0.226
		1999	8.795	0.601	42.805	56.731	9.161
			0.342	0.035	7.640	7.391	0.243
Estonia	Soviet	1999	7.605	0.357	22.220	24.845	.
			0.490	0.084	2.962	3.464	.
Finland	Scand	1990	7.542	0.315	11.171	23.508	8.704
			1.321	0.189	8.603	13.530	0.470
		2005	8.715	0.543	14.719	25.566	8.515
			0.339	0.095	9.016	13.124	0.324
France	Mid Eur	1999	7.599	0.421	12.016	20.705	8.195
			0.443	0.055	6.031	9.120	0.238
Germany	Mid Eur	1990	8.267	0.464	23.754	32.703	8.879
			0.196	0.040	4.318	8.297	0.181
Hungary	East Eur	1991	7.682	0.482	19.412	30.257	7.712
			0.301	0.075	12.321	15.603	0.456
Iceland	Scand	1984	8.570	0.590	26.081	30.974	8.908
			0.449	0.130	5.355	9.122	0.250
India	Asia	1990	9.439	0.801	8.686	17.930	9.592
			0.152	0.050	10.441	17.471	0.259
		2001	9.126	0.819	8.868	12.917	9.282
			0.588	0.113	10.753	14.706	0.616
Ireland	Anglo	1990	8.045	0.444	23.787	38.492	8.740
			0.302	0.076	13.700	18.184	0.361

continues on next page

continued from previous page

Italy	South Eur	1990	7.647	0.393	49.877	49.943	8.704
			0.986	0.228	0.160	0.170	1.301
		1999	8.572	0.538	45.457	45.491	8.732
			0.244	0.055	0.019	0.026	0.111
		2005	8.415	0.522	27.996	33.774	9.065
			0.586	0.090	3.772	6.395	0.464
Japan	Asia	1990	9.541	0.814	64.896	64.973	9.529
			0.119	0.041	0.063	0.081	0.108
		1995	9.504	0.811	21.246	32.636	9.461
			0.124	0.019	5.958	10.724	0.124
		2000	9.527	0.812	49.944	49.988	9.367
			0.134	0.047	0.027	0.037	0.159
		2005	9.390	0.778	49.944	49.988	9.358
			0.260	0.074	0.027	0.037	0.249
Latvia	Soviet	1999	7.946	0.467	21.977	23.856	.
			0.863	0.154	3.096	3.431	.
Lithuania	Soviet	1999	7.278	0.379	32.319	32.782	7.754
			0.578	0.079	0.583	0.654	0.548
Luxembourg	Mid Eur	1999	7.792	0.447	11.922	25.227	8.475
			0.584	0.097	6.641	12.221	0.609
Malta	Mid Eur	1999	9.238	0.742	7.209	16.946	.
			0.222	0.054	6.657	13.908	.
Mexico	Latin	2000	8.666	0.669	4.133	8.498	7.339
			0.366	0.076	5.373	9.989	0.259
		2005	8.746	0.681	7.513	15.243	7.013
			0.345	0.083	6.379	11.805	0.894
Morocco	Develop	2001	9.578	0.930	18.459	24.182	8.801
			0.383	0.048	18.422	20.526	0.406
Netherlands	Mid Eur	1981	7.701	0.371	12.933	24.974	8.340
			0.685	0.155	5.542	10.088	0.306
		1990	7.771	0.381	32.546	40.010	8.348
			0.321	0.082	4.417	8.678	0.500
Nigeria	Develop	1990	8.683	0.614	37.634	43.428	8.476
			0.264	0.048	6.734	7.278	0.255
Norway	Scand	1996	8.196	0.458	25.411	34.634	9.134
			0.290	0.068	5.999	9.786	0.447
Peru	Latin	1996	8.759	0.625	5.046	8.316	8.083
			0.498	0.074	5.332	7.458	0.559
		2001	8.900	0.673	3.781	6.563	8.477
			0.316	0.080	5.020	7.433	0.553
Portugal	South Eur	1990	7.403	0.375	39.944	39.980	8.325
			0.344	0.046	0.051	0.059	0.302
Slovakia	East Eur	1999	8.898	0.602	15.539	22.375	.
			0.191	0.064	3.222	6.365	.
Slovenia	East Eur	1999	8.515	0.568	22.756	30.034	.
			0.360	0.063	4.840	6.787	.
South Africa	Develop	1996	8.889	0.706	22.405	33.542	9.095
			0.309	0.075	12.893	16.196	0.330
		2001	8.755	0.598	14.575	21.494	8.774
			0.150	0.056	14.041	18.432	0.304
Spain	South Eur	1981	7.858	0.445	40.000	40.000	8.478
			0.266	0.053	0.000	0.000	0.260
		1990	7.898	0.475	14.195	22.927	8.408
			0.699	0.118	8.710	11.496	0.283
		1995	8.652	0.654	55.872	55.957	8.703

continues on next page

continued from previous page

			0.655	0.073	0.117	0.128	0.743
Sweden	Scand	1996	8.297	0.459	8.695	19.398	8.025
			0.289	0.083	8.185	15.971	0.290
		1999	8.423	0.469	6.889	17.085	.
			0.308	0.097	6.954	15.250	.
		2006	8.466	0.478	4.884	15.316	8.058
			0.351	0.098	5.512	15.119	0.272
Turkey	Develop	2001	9.630	0.882	39.998	39.999	.
			0.458	0.151	0.002	0.002	.
Uganda	Develop	2001	7.802	0.611	29.961	29.982	8.439
			1.241	0.257	0.039	0.045	0.438
United Kingdom	Anglo	1990	8.231	0.471	15.011	21.664	8.670
			0.236	0.073	7.723	8.292	0.466
United States	Anglo	1990	9.010	0.654	13.390	19.146	8.688
			0.499	0.148	5.006	8.304	0.556
		2000	8.661	0.591	15.049	22.008	8.375
			0.304	0.087	5.648	8.656	0.205
Venezuela	Latin	2000	9.291	0.738	33.973	33.994	8.460
			0.244	0.056	0.014	0.018	0.448

For each country-year: first row indicates the mean; second row indicates standard deviation. Abbreviations: TM ten: Tax morale (ten), TM (bi): Tax morale (bi), AR: Average tax rate, MR: marginal tax rate, cheat: Cheating on public transportation, Anglo: Anglo-Saxon plus AUS and NZ, Mid Eur: Mid Central Europe plus Israel, South Eur: Southern Europe, Scand: Scandinavia, East Eur: Eastern Central Europe, Soviet: Former Soviet countries, Latin: Latin America, Asia: Asia, Develop: Developing countries

Table 5.7: Means and Standard Deviations of Key Variables by Income Group

Group	**TM ten**	**TM bi**	**AR**	**MR**	**Cheat**
2	8.450	0.586	14.999	14.947	8.330
	0.866	0.181	16.948	17.025	0.800
3	8.562	0.589	17.345	22.280	8.435
	0.813	0.166	16.248	15.941	0.649
4	8.617	0.590	19.045	24.640	8.544
	0.705	0.157	15.881	15.641	0.634
5	8.592	0.601	21.087	28.262	8.596
	0.664	0.151	15.178	14.500	0.608
6	8.585	0.597	23.445	32.349	8.647
	0.698	0.152	14.043	11.853	0.536
7	8.472	0.561	25.919	35.231	8.613
	0.840	0.176	12.983	11.177	0.640
8	8.451	0.560	28.122	37.167	8.658
	0.763	0.160	12.514	11.261	0.523
9	8.409	0.554	29.612	38.327	8.636
	0.914	0.201	11.731	10.373	0.636
10	8.242	0.534	32.307	40.398	8.468
	0.943	0.188	11.074	9.358	0.912

Means and Standard Deviations by Income Groups. For each income group, the first row indicates the mean and the second row is the standard deviation. Abbreviations: TM ten: tax morale (ten), TM bi: tax morale (bi), AR: Average tax rate, MR: marginal tax rate, Cheat: Cheating on public transportation

5.6.2 OLS results displaying covariates

Table 5.8: OLS Estimations of Tax Morale on Tax Rates

Model	I	II	III	IV
Dependent Variable	AR	AR	MR	MR
Tax Morale (ten)	0.939		0.779	
	(0.718)		(0.750)	
Tax Morale (bi)		11.172***		10.626***
		(3.946)		(4.012)
Income Group 3	2.780	2.884	6.790**	6.878***
	(2.311)	(2.310)	(2.626)	(2.618)
Income Group 4	4.173*	4.364*	8.680***	8.850***
	(2.293)	(2.284)	(2.620)	(2.601)
Income Group 5	5.845**	5.952**	11.431***	11.532***
	(2.444)	(2.451)	(2.717)	(2.723)
Income Group 6	8.284***	8.425***	15.384***	15.521***
	(2.385)	(2.381)	(2.435)	(2.432)
Income Group 7	10.747***	11.242***	17.743***	18.236***
	(2.332)	(2.323)	(2.481)	(2.476)
Income Group 8	12.811***	13.158***	19.600***	19.945***
	(2.263)	(2.246)	(2.502)	(2.489)
Income Group 9	14.541***	14.921***	21.115***	21.505***
	(2.345)	(2.302)	(2.557)	(2.524)
Income Group 10	16.893***	17.244***	22.536***	22.910***
	(2.568)	(2.513)	(2.752)	(2.706)
Part Time	4.673	4.751	−5.331	−5.265
	(7.094)	(7.107)	(7.512)	(7.506)
Self-employed	4.191	4.884	−0.098	0.626
	(4.254)	(4.249)	(4.006)	(4.013)
Married	0.736	−0.193	6.816	5.899
	(5.296)	(5.173)	(5.936)	(5.840)
Divorced	−22.077***	−22.138***	−24.088***	−24.082***
	(8.213)	(8.450)	(8.782)	(9.064)
Widowed	−0.367	−1.601	8.010	6.907
	(17.193)	(16.980)	(16.580)	(16.495)
Age	1.625	1.442	1.861	1.687
	(1.525)	(1.513)	(1.570)	(1.576)
Age (squared)	−2.265	−2.026	−2.277	−2.053
	(1.947)	(1.935)	(2.018)	(2.030)
Male	−1.461	−0.586	−5.395	−4.565
	(4.802)	(4.830)	(4.794)	(4.779)
Number Children	0.278	0.847	−0.363	0.202
	(1.353)	(1.329)	(1.394)	(1.355)
Education	−0.300*	−0.266*	−0.359**	−0.325**
	(0.157)	(0.157)	(0.167)	(0.164)
Church	−9.628***	−9.547***	−8.327***	−8.234***
	(2.900)	(2.927)	(3.018)	(3.048)
Patriotism	−5.330***	−6.201***	−6.621***	−7.524***
	(1.793)	(1.895)	(1.873)	(1.966)
Trust	−0.089	−0.834	3.592	2.763
	(4.334)	(4.196)	(4.321)	(4.152)
wave 2	−2.241	−0.749	−3.245	−1.732
	(2.329)	(2.344)	(2.332)	(2.302)
wave 3	−6.435*	−5.152	−6.146*	−4.858
	(3.771)	(3.723)	(3.579)	(3.515)
wave 4	−5.942**	−5.004*	−7.950***	−7.008***
	(2.587)	(2.580)	(2.645)	(2.627)
wave 5	−11.046***	−10.156***	−11.239***	−10.337***
	(2.998)	(2.974)	(3.209)	(3.197)
GDP per capita	0.771***	0.781***	0.589***	0.598***
	(0.125)	(0.127)	(0.132)	(0.134)
GDP growth	0.123	0.131	0.296*	0.301*

continues on next page

continued from previous page

	(0.157)	(0.155)	(0.170)	(0.167)
FDI	−0.055***	−0.055***	−0.039***	−0.039***
	(0.007)	(0.007)	(0.008)	(0.008)
constant	−10.970	−5.672	−4.477	−0.182
	(31.382)	(30.471)	(31.265)	(30.843)
N	576.000	576.000	575.000	575.000
R2	0.529	0.534	0.548	0.553
Country Group FE	yes	yes	yes	yes

[1] Dependent variables are average (AR) and marginal (MR) tax rates [2] Estimation is by OLS with clustered standard errors [3] Income group 2, wave 1, full time, single and female are reference categories [4] * < 0.10, ** < 0.05, *** < 0.01

5.6.3 Baseline IV Results displaying covariates

First Stage Results

Table 5.9: IV Estimations of Tax Morale on Tax Rates: First Stage

Model	I	II	III	IV
Dependent Variable	TM10	TM2	TM10	TM2
Cheat Transport	0.412***	0.059***	0.418***	0.060***
	(0.056)	(0.010)	(0.056)	(0.010)
Income Group 3	0.079	−0.005	0.056	−0.006
	(0.109)	(0.022)	(0.110)	(0.023)
Income Group 4	0.036	−0.016	0.012	−0.018
	(0.107)	(0.022)	(0.107)	(0.022)
Income Group 5	−0.085	−0.021	−0.110	−0.022
	(0.107)	(0.023)	(0.107)	(0.023)
Income Group 6	−0.138	−0.030	−0.164	−0.032
	(0.115)	(0.024)	(0.115)	(0.025)
Income Group 7	−0.266*	−0.068**	−0.291**	−0.069**
	(0.137)	(0.027)	(0.138)	(0.028)
Income Group 8	−0.239*	−0.055**	−0.262**	−0.056**
	(0.126)	(0.025)	(0.126)	(0.025)
Income Group 9	−0.366***	−0.071**	−0.390***	−0.072**
	(0.134)	(0.028)	(0.134)	(0.028)
Income Group 10	−0.379***	−0.062**	−0.402***	−0.063**
	(0.142)	(0.027)	(0.143)	(0.027)
Part Time	0.210	0.039	0.152	0.036
	(0.375)	(0.081)	(0.374)	(0.082)
Self-employed	−0.804**	−0.111	−0.831**	−0.113
	(0.343)	(0.077)	(0.339)	(0.078)
Married	−0.104	0.023	−0.115	0.022
	(0.364)	(0.066)	(0.367)	(0.066)
Divorced	−0.752	−0.107	−0.646	−0.101
	(0.969)	(0.165)	(1.004)	(0.169)
Widowed	−1.030	−0.007	−1.133	−0.013
	(1.077)	(0.250)	(1.084)	(0.251)
Age	−0.055	0.009	−0.059	0.009
	(0.126)	(0.018)	(0.126)	(0.018)
Age (squared)	0.061	−0.014	0.066	−0.014
	(0.154)	(0.022)	(0.154)	(0.022)
Male	0.083	−0.084	0.086	−0.084
	(0.440)	(0.077)	(0.441)	(0.077)
Number Children	−0.038	−0.039**	−0.043	−0.039**
	(0.080)	(0.019)	(0.080)	(0.019)
Education	0.009	0.000	0.008	0.000
	(0.011)	(0.002)	(0.011)	(0.002)

continues on next page

continued from previous page

Church	−0.489**	−0.097***	−0.461**	−0.096**
	(0.191)	(0.037)	(0.190)	(0.037)
Patriotism	0.394**	0.100***	0.403**	0.101***
	(0.174)	(0.031)	(0.174)	(0.031)
Trust	1.211***	0.175***	1.274***	0.178***
	(0.302)	(0.055)	(0.302)	(0.056)
wave 2	−0.657***	−0.172***	−0.667***	−0.172***
	(0.138)	(0.032)	(0.139)	(0.032)
wave 3	−0.562***	−0.162***	−0.572***	−0.163***
	(0.162)	(0.037)	(0.163)	(0.037)
wave 4	−0.399**	−0.120***	−0.398**	−0.120***
	(0.158)	(0.034)	(0.158)	(0.034)
wave 5	−0.275	−0.096**	−0.292	−0.097**
	(0.199)	(0.044)	(0.197)	(0.044)
GDP per capita	0.011*	0.001	0.011*	0.001
	(0.006)	(0.001)	(0.006)	(0.001)
GDP growth	0.055***	0.006***	0.055***	0.006***
	(0.010)	(0.002)	(0.010)	(0.002)
FDI	−0.001	−0.000	−0.001	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	4.678*	−0.227	4.695*	−0.226
	(2.612)	(0.387)	(2.615)	(0.388)
N	504.000	504.000	503.000	503.000
R2	0.670	0.703	0.673	0.703
Country Group FE	yes	yes	yes	yes
F Statistic	54.95	33.86	55.52	33.80

[1] Dependent variables is tax morale (ten) in I and III and tax morale (bi) in II and IV [2] First Stage Results corresponding to models I to IV in equation 5.10 [3] Instrument: Cheating on Public Transportation [4] Income group 2, wave 1, full time, single and female are reference categories [5] *F-statistic* indicates F-statistic of excluded instruments [6] * < 0.10, **< 0.05, *** < 0.01

Second Stage Results

Table 5.10: IV Estimations of Tax Morale on Tax Rates: Second Stage

Model	I	II	III	IV
Dependent Variable	AR	AR	MR	MR
Tax Morale (ten)	12.441***		11.193***	
	(2.698)		(2.743)	
Tax Morale (bi)		86.239***		78.254***
		(18.814)		(19.260)
Income Group 3	1.857	3.290	5.957**	7.092**
	(2.873)	(3.071)	(3.019)	(3.135)
Income Group 4	3.245	5.088*	7.951***	9.458***
	(2.831)	(3.004)	(2.980)	(3.109)
Income Group 5	6.552**	7.292**	12.121***	12.634***
	(2.914)	(3.224)	(3.070)	(3.324)
Income Group 6	9.417***	10.305***	16.577***	17.223***
	(2.900)	(3.154)	(2.848)	(3.089)
Income Group 7	13.944***	16.486***	20.712***	22.874***
	(3.005)	(3.228)	(2.968)	(3.157)
Income Group 8	15.545***	17.325***	22.059***	23.543***
	(2.763)	(2.948)	(2.872)	(3.040)
Income Group 9	18.670***	20.239***	24.981***	26.275***
	(2.744)	(2.887)	(2.865)	(3.007)
Income Group 10	22.611***	23.238***	27.609***	28.060***
	(3.008)	(3.007)	(3.089)	(3.072)
Part Time	5.049	4.275	−3.972	−5.085

continues on next page

continued from previous page

	(8.433)	(10.129)	(8.402)	(9.968)
Self-employed	7.847	7.443	4.157	3.683
	(6.726)	(7.901)	(6.195)	(7.361)
Married	−4.807	−8.068	3.821	0.797
	(7.457)	(7.460)	(7.572)	(7.485)
Divorced	−19.312	−19.409	−20.145	−19.450
	(17.501)	(18.583)	(16.831)	(17.950)
Widowed	0.363	−11.807	12.524	0.885
	(22.669)	(25.866)	(21.569)	(23.916)
Age	1.866	0.418	2.323	0.988
	(2.598)	(2.348)	(2.507)	(2.317)
Age (squared)	−2.521	−0.516	−2.832	−0.985
	(3.218)	(2.966)	(3.137)	(2.962)
Male	−1.638	6.627	−4.729	2.788
	(7.631)	(8.246)	(6.892)	(7.421)
Number Children	1.998	4.898**	0.945	3.543*
	(1.640)	(2.095)	(1.571)	(1.955)
Education	−0.307	−0.212	−0.366*	−0.288
	(0.202)	(0.229)	(0.193)	(0.210)
Church	−4.632	−2.305	−3.329	−0.983
	(4.591)	(5.179)	(4.472)	(5.090)
Patriotism	−12.793***	−16.520***	−13.960***	−17.317***
	(3.138)	(3.958)	(3.111)	(3.904)
Trust	−15.318**	−15.309**	−10.122	−9.803
	(6.558)	(7.588)	(6.409)	(7.224)
wave 2	7.312**	13.930***	5.521*	11.521**
	(3.418)	(4.755)	(3.221)	(4.492)
wave 3	1.336	8.340	1.277	7.622
	(4.276)	(5.186)	(3.998)	(4.831)
wave 4	1.267	6.697	−0.960	4.008
	(3.426)	(4.451)	(3.334)	(4.261)
wave 5	−4.098	0.794	−4.450	−0.096
	(4.108)	(5.189)	(4.153)	(5.125)
GDP per capita	0.565***	0.626***	0.377**	0.432**
	(0.161)	(0.184)	(0.158)	(0.178)
GDP growth	−0.166	−0.005	−0.044	0.096
	(0.280)	(0.257)	(0.284)	(0.256)
FDI	−0.044***	−0.050***	−0.028***	−0.033***
	(0.010)	(0.012)	(0.011)	(0.011)
constant	−86.744	−9.013	−76.794	−6.593
	(56.847)	(47.236)	(54.439)	(46.063)
N	504.000	504.000	503.000	503.000
R2	0.359	0.294	0.422	0.380
Country Group FE	yes	yes	yes	yes
F Statistic	54.95	33.86	55.52	33.80

[1] Dependent variables are average (AR) and marginal (MR) tax rates [2] IV regressions with clustered standard errors. Second Stage Results [3] Instrument: Cheating on Public Transportation [4] Income group 2, wave 1, full time, single and female are reference categories [5] *F-statistic* indicates F-statistic of excluded instruments [6] $^{*} < 0.10$, $^{**} < 0.05$, $^{***} < 0.01$

5.6.4 Summary Statistics: Individual level

Table 5.11: Summary Statistics: Individual Level

Variable	Mean	Std. Dev.
Average Tax Rate (AR)	23.125	14.883
Marginal Tax Rate (MR)	30.274	15.642
Tax Morale (ten)	8.538	2.336
tax Morale (bi)	0.581	0.493
Cheat public transp	8.632	2.207
N	30044	

5.6.5 Robustness Checks

Table 5.12: Robustness Checks

Model	I	II	III	IV
Dependent Variable	AR	AR	MR	MR
Baseline				
Tax Morale (ten)	12.441***		11.193***	
	(2.698)		(2.743)	
Tax Morale (bi)		86.239***		78.254***
		(18.814)		(19.260)
Inst Quality	0.956	-2.610	2.389	-0.982
	(2.345)	(2.296)	(2.288)	(2.254)
N	504	504	503	503
R2	0.359	0.294	0.422	0.380
Institutional Quality				
Tax Morale (ten)	16.675***		15.784***	
	(4.145)		(4.039)	
Tax Morale (bi)		109.836***		105.415***
		(27.237)		(27.034)
N	464	464	463	463
R2	0.299	0.202	0.368	0.280
Weighted by number of observations in income group				
Tax Morale (ten)	14.731***		12.340***	
	(2.913)		(2.964)	
Tax Morale (bi)		106.555***		89.465***
		(20.536)		(20.191)
N	502	502	501	501

continues on next page

continued from previous page

R2	0.280	0.184	0.403	0.365
OECD Countries				
Tax Morale (ten)	12.774***		13.855***	
	(3.283)		(3.417)	
Tax Morale (bi)		77.558***		85.327***
		(18.689)		(19.996)
N	385	385	384	384
R2	0.469	0.463	0.420	0.414
Lead Tax Rates				
Tax Morale (ten)	8.828***		9.071***	
	(2.644)		(2.745)	
Tax Morale (bi)		61.902***		64.084***
		(18.192)		(19.242)
N	477	477	476	476
R2	0.404	0.392	0.429	0.420

[1] Dependent variables are average (AR) and marginal (MR) tax rates [2] IV regressions with clustered standard errors. [3] Instrument: Cheating on Public Transportation [4] All estimations include same control variables as in baseline specifications including country group fixed effects. [5] Lead tax rates estimations are IV regressions of tax rates in year $t+1$ on tax morale in t. [6] Weighted IV regressions weight income groups with the number of individuals in each respective group [7] * < 0.10, **< 0.05, *** < 0.01

5.7 Appendix B: Multiple Imputation Procedure

In order to tackle the problem of not having individual level data on income, among others, we employ a multiple imputation procedure, which is comparable to the approach proposed by Jenkins et al. (2011). The basic steps are as follows:

a) For each country-year, divide each of the 10 income groups of the WVS data into 100 equally wide segments.[21]

b) For each of the resulting 100 incomes within each income group, calculate the corresponding (average and marginal) tax rates using the WTI data.

[21] In principle, it is possible to randomly assign each individual a random income between the lower and upper bounds for the respective income brackets. However, we would not gain more variation in tax rates, whereas the computational procedure would be slightly more difficult.

c) Randomly assign one of the 100 incomes (within an income group) to each individual in that income group. This step also automatically assigns corresponding tax rates to individuals. Since some groups have more than 100 individuals, it is possible for two or more individuals within a given group to have the same income.

d) Run individual level IV-regressions.

e) Repeat steps (3) and (4) 500 times.

The average (mean) coefficients of these 500 replications is used to derive our coefficient of interest. We employ the combination method proposed by Reiter (2003), and applied by Jenkins et al. (2011), to calculate standard-errors and levels of significance, taking into account the finite number of imputations.

5.8 Appendix C: An Optimal Tax Model with Tax Morale

This section describes a Ramsey-type model to describe the relationship between tax morale and tax rate.[22] Income groups have heterogeneous levels of tax morale and maximize their utility with respect to their labor supply and evasion decisions. Governments may tax different income groups differently. The optimal policy maximizes an objective function which may be interpreted as a welfare function or a function reflecting political influence. The optimal tax rates set by the government then depend, among other things, on the level of tax morale of the different groups. Groups with a high level of tax morale are taxed more heavily because, other things equal, their reported income reacts less elastically to tax rate changes than the reported income of groups with lower tax morale. While this explanation is intuitive, one should bear in mind that it is based on the assumption that different income groups can be taxed independently of one another.

5.8.1 Households

Consider an economy with n groups of households, $i = 1, ..., n$. For simplicity, we normalize group size to unity, i.e., there is one household representing each group. We model tax morale as follows. We assume that households can evade taxes, but doing so gives rise to a subjective cost. The cost function is given by the function $\beta(e_i, m_i)$, where e_i is undeclared income of group i and m_i is a parameter which captures differences in tax

[22] An alternative approach to modeling the relationship of interest is to incorporate tax morale in the standard Mirrleesian model of optimal income taxation. The challenge with this approach is that the resulting relationship between tax burdens and tax morale is ambiguous. This is yet another reason to exploit the available data to examine the relationship between tax morale and tax rates.

morale—the intrinsic motivation to honestly pay taxes—across groups. The cost function $\beta(e_i, m_i)$ can be interpreted broadly as also including possible real costs of evading or being detected and fined. The crucial point for our analysis is that it also includes differences in tax morale.[23] The utility function of household i is given by

$$u_i = c_i - \alpha(l_i) - \beta(e_i, m_i), \tag{5.4}$$

where c_i is consumption, l_i is labor supply and $\alpha(l_i)$ is a strictly convex function which represents the disutility of work. The assumption of quasi-linearity in consumption simplifies notation by allowing us to abstract from income effects of taxation on labor supply and evasion. We assume that the tax evasion cost function has the following properties:

Assumption 1: $\beta_e, \beta_{ee} > 0, \beta_{em} > 0, \beta_{eem} > 0$.

Assumption 1 implies that the marginal cost of evading taxes is strictly convex in the amount of evaded income e_i. In addition, the marginal cost of undeclared income is higher for households with higher tax morale ($\beta_{em} > 0$) and this marginal cost also increases more quickly ($\beta_{eem} > 0$).

Following Feldstein (1999), it is convenient to express utility in terms of earned income $y_i = l_i w_i$, where w_i is the wage rate of household i and declared income $d_i = y_i - e_i$. This yields for our utility function:

$$u_i = c_i - \alpha(\frac{y_i}{w_i}) - \beta(y_i - d_i, m_i). \tag{5.5}$$

The household's budget constraint is

$$c_i = y_i - t_i d_i, \tag{5.6}$$

where t_i is the income tax rate. The household chooses earned income y_i and declared income d_i to maximize utility, subject to the budget constraint. This yields the following first order conditions:

$$\begin{aligned} 1 - \alpha_1(\frac{y_i^*}{w_i})\frac{1}{w_i} - \beta_e(y_i^* - d_i^*, m_i) &= 0, && (5.7) \\ -t_i + \beta_e(y_i^* - d_i^*, m_i) &= 0. && (5.8) \end{aligned}$$

Note that (5.7) and (5.8) implicitly define the household's optimal choices, which

[23]Note that our approach can also be interpreted in terms of a 'warm glow' effect, i.e., the intrinsic satisfaction of doing the right thing when complying with the tax law. Similar theoretical results can also be derived from political economy models of rentier states or government capture, though yielding different welfare implications.

can be expressed by the functions $y_i^*(t_i, m_i)$, $d_i^*(t_i, m_i)$. For later use note that

$$\frac{\partial d_i}{\partial t_i} = -\left[\frac{1}{\beta_{ee}} + \frac{w_i^2}{\alpha_{11}}\right] < 0, \frac{\partial d_i}{\partial m_i} = \frac{\beta_{em}}{\beta_{ee}} > 0. \tag{5.9}$$

An increase in the tax rate on household i reduces declared income for two reasons. Firstly, labor supply and, hence, earned income declines. Secondly, the higher tax rate induce the household to declare a smaller part of the income earned, which means that tax evasion increases. Moreover, declared income is increasing in the level of tax morale, as one would expect.

5.8.2 The Government

The government finances a given revenue target R using a wage tax, which may differ across household groups. The government budget constraint is given by

$$R = \sum t_i d_i. \tag{5.10}$$

In the decision making process of the government, each group i is given a weight b_i, which is assumed to be independent of tax morale[24] and may be interpreted either as a welfare weight or as a parameter reflecting relative political influence. Given the optimal choices of the households, the government maximizes the objective function

$$W = \sum b_i u_i \tag{5.11}$$

subject to (5.10).[25] The optimal tax rate levied on group i is given by the formula

$$t_i^* = -\frac{(\eta - b_i)}{\eta} d_i \left[\frac{\partial d_i}{\partial t_i}\right]^{-1}, \tag{5.12}$$

where η is the marginal cost of public funds. This formula implies that, for a given cost of public funds, the optimal tax rate levied on household i is increasing in the household's tax morale parameter m_i if $\beta_{em} > 0$ (which implies $\partial d_i / \partial m_i > 0$) and $\beta_{eem} > 0$, as assumed above in assumption 1. The economic explanation is as follows. If $\beta_{em} > 0$, a higher tax

[24]If the 'social planner' systematically attributed a higher weight to the wellbeing of groups with high tax morale, or if these groups had more influence in the political process than other groups, this would of course imply that high tax morale would not necessarily lead to higher tax burdens.

[25]This objective function implies that the government takes into account that some groups benefit more from tax evasion than others because their tax morale is lower. In addition, the expected payment of fines in case of detection represents redistribution from the taxpayer to the government (Chetty 2009). One could argue that governments would rather maximize a function based on a group specific utility of the type $W = \sum b_i u_i^*$ where $u_i^* = c_i - \alpha(\frac{y_i}{w_i})$. However, it is problematic to assume that the government pursues objectives which are not based on individual utilities. In any case, it is easy to check that the optimal tax policy would be the same.

morale increases the tax base, holding everything else constant, so that it is optimal to levy a higher tax rate on this group. $\beta_{eem}(e_i, m_i) > 0$ implies that the decline in the tax base caused by a higher tax rate is smaller for groups with higher tax morale. This is because their marginal cost of evading taxes increases quickly as evasion increases. Both effects imply that the elasticity of declared income with respect to the tax base declines if tax morale increases. The presence of one of these effects is sufficient for the result that a higher tax morale leads to a higher optimal tax rate. Clearly, this is well in line with standard results of optimal income tax theory following Ramsey (1927) or Mirrlees (1971): Groups with a higher responsiveness to taxation should be taxed lower than groups with low levels of responsiveness, i.e., a low elasticity of taxable income. Hence, our empirical analysis also provides an empirical verification of the inverse elasticity rule of optimal taxation, which, to the best of our knowledge, has not been tested for income taxation before.

Note also that, by assuming separability between the disutility of work and the subjective cost of evading taxes, we have assumed that tax morale does not affect the elasticity of labor supply. Yet, group specific differences may also be driven by differences in labor supply elasticities. We assume that tax morale and the elasticity of labor supply are unrelated in our empirical analysis. While it is straightforward to generalize our model in various aspects, we opt for a simple model which includes the features and channels we can explore empirically. For instance, we do not build in detection probabilities and penalties. We assume that these are implicitly captured in the model parameters on evasion and tax morale. In principle, it would also be possible to construct a richer model, which allows for tax avoidance (e.g., income shifting) in addition to evasion and that would allow for sorting of individuals into tax brackets. As evasion implies a tax rate of zero while avoidance or income shifting yields a positive (but smaller) tax rate, our main results would not change in that case. A richer model could also consider the interplay of both the evasion and labor supply margins (e.g., Pencavel 1979 and Doerrenberg and Duncan 2014b). The impact of tax morale on labor supply and evasion responses is ambiguous in such a model. However, labor supply elasticities are typically rather small (Saez et al. 2012; Bargain et al. 2014). It is also worth noting that Simonovits (2011) builds on an earlier version of our paper and extends our model by introducing redistributive concerns, but does not derive different conclusions in his numerical simulations. Additionally, Traxler (2010) allows tax morale to be endogenous and incorporates it into the seminal tax evasion framework of Allingham and Sandmo (1972).

References

Aaberge, R. and U. Colombino (2013). Using a microeconometric model of household labour supply to design optimal income taxes. *Scandinavian Journal of Economics 115*(2), 449–475.

Abeler, J. and S. Jaeger (2014). Complex Tax Incentives. *American Economic Journal: Economic Policy*. Forthcoming.

Algan, Y. and P. Cahuc (2009). Civic virtue and labor market institutions. *American Economic Journal: Macroeconomics 1*(1), 111–145.

Algan, Y., P. Cahuc, and M. Sangnier (2014). Efficient and Inefficient Welfare States. *Economic Journal*. Forthcoming.

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

Alm, J. (2010). Testing behavioral public economics theories in the laboratory. *National Tax Journal 63*(4), 635–658.

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International Tax and Public Finance 19*(1), 54–77.

Alm, J., K. M. Bloomquist, and M. McKee (2011). Comparing student and non-student reporting behavior in tax compliance experiments. In A. Plumley (Ed.), *IRS Research Bulletin (Publication 1500)*, pp. 93–99. Washington D.C.: International Revenue Service.

Alm, J., T. Cherry, M. Jones, and M. McKee (2010). Taxpayer information assistance services and tax compliance behavior. *Journal of Economic Psychology 31*(4), 577–586.

Alm, J. and K. Finlay (2013). Who benefits from tax evasion? *Economic Analysis and Policy*. forthcoming.

Alm, J., B. Jackson, and M. McKee (1992a). Institutional uncertainty and taxpayer compliance. *American Economic Review 82*(4), 1018–1026.

Alm, J. and B. R. Jackson (1993). Fiscal exchange, collective decision institutions, and tax compliance. *Journal of Economic Behavior and Organization 22*, 285 – 303.

Alm, J., B. R. Jackson, and M. McKee (1992b). Estimating the determinants of taxpayer compliance with experimental data. *National Tax Journal 45*(1), 107 – 114.

Alm, J., B. R. Jackson, and M. McKee (2009). Getting the word out: Enforcement information dissemination and compliance behavior. *Journal of Public Economics 93*(3-4), 392–402.

Alm, J. and S. Jacobson (2007). Using Laboratory Experiments in Public Economics. *National Tax Journal 60*(1), 129–152.

Alm, J. and C. McClellan (2012). Tax morale and tax compliance from the firm's perspective. *Kyklos 65*(1), 1–17.

Alm, J., G. H. McClelland, and W. D. Schulze (1992). Why do people pay taxes? *Journal of Public Economics 48*(1), 21–38.

Alm, J., G. H. McClelland, and W. D. Schulze (1999). Changing the social norm of tax compliance by voting. *Kyklos 52*(2), 141–171.

Alm, J., E. Sennoga, and M. Skidmore (2009). Perfect competition, urbanization, and tax incidence in the retail gasoline market. *Economic Inquiry 47*(1), 118–134.

Alm, J. and E. B. Sennoga (2010). Mobility, competition, and the distributional effects of tax evasion. *National Tax Journal 63*(4), 1055–84.

Alm, J. and B. Torgler (2006). Culture differences and tax morale in the United States and in Europe. *Journal of Economic Psychology 27*(2), 224–246.

Alm, J. and B. Torgler (2011). Do ethics matter? Tax compliance and morality. *Journal of Business Ethics 101*(4), 635–651.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal 100*(401), 464 – 477.

Andreoni, J., B. Erard, and J. Feinstein (1998). Tax compliance. *Journal of Economic Literature 36*(2), 818–860.

Balafoutas, L., A. Beck, R. Kerschbamer, and M. Sutter (2014). The hidden costs of tax evasion: collaborative tax evasion in markets for expert services. University of Innsbruck Working Papers in Economics and Statistics 2014-01.

Bargain, O., K. Orsini, and A. Peichl (2014). Comparing labor supply elasticities in Europe and the US: New results. *Journal of Human Resources*. Forthcoming.

Barone, G. and S. Mocetti (2011). Tax morale and public spending inefficiency. *International Tax and Public Finance*. DOI: 10.1007/s10797-011-9174-z.

Bastani, S. and H. Selin (2014). Bunching and non-bunching at kink points of the Swedish tax schedule. *Journal of Public Economics*. Forthcoming.

Bayer, R. and F. Cowell (2009). Tax compliance and firms' strategic interdependence. *Journal of Public Economics 93*(11-12), 1131–1143.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76*(2), 169–217.

Becker, W., H.-H. Buechner, and S. Sleeking (1987). The impact of public transfer expenditures on tax evasion. *Journal of Public Economics 37*, 243 – 252.

Bjornskov, C. and G. T. Svendsen (2012). Does social trust determine the size of the welfare state? Evidence using historical identification. *Public Choice*. DOI: 10.1007/s11127-012-9944-x.

Blumenthal, M., C. Christian, and J. Slemrod (2001). Do normative appeals affect tax compliance? evidence from a controlled experiment in minnesota. *National Tax Journal 54*(1), 125 – 138.

Blumkin, T., B. J. Ruffle, and Y. Ganun (2012). Are income and consumption taxes ever really equivalent? evidence from a real-effort experiment with real goods. *European Economic Review 56*(6), 1200–1219.

Bo, E. E., J. Slemrod, and T. O. Thoresen (2014). Taxes on the internet. Deterrence effects of public disclosure. *American Economic Journal: Economic Policy*. forthcoming.

Borck, R., D. Engelmann, W. Mueller, and H.-T. Normann (2002). Tax liability-side equivalence in experimental posted-offer markets. *Southern Economic Journal 68*(3), 672–682.

Brennan, G. and J. M. Buchanan (1980). *The Power to Tax*. Cambridge University Press.

Breusch, T. (2005a). The canadian underground economy: An examination of giles and tedds. *Canadian Tax Journal 53*(2), 367.

Breusch, T. (2005b). Estimating the Underground Economy using MIMIC Models. Australian National University Working Paper. Available online: `http://128.118.178.162/eps/em/papers/0507/0507003.pdf`.

Burns, S. K. and J. P. Ziliak (2012). Identifying the elasticity of taxable income. mimeo, University of Kentucky.

Carpenter, J. and C. K. Myers (2010). Why volunteer? evidence on the role of altruism, image, and incentives. *Journal of Public Economics 94*(11-12), 911 – 920.

Carrillo, P., D. D. Pomeranz, and M. Singhal (2013). Tax me if you can: Evidence on firm misreporting behavior and evasion substitution. mimeo.

Cettolin, E. and A. Riedl (2013). Justice under uncertainty. Maastricht University Working Paper.

Charness, G., U. Gneezy, and M. A. Kuhn (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization 81*(1), 1 – 8.

Charness, G. and P. Kuhn (2011). Lab labor: What can labor economists learn from the lab? In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics Vol 4A*, pp. 229–330. Amsterdam: North Holland.

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

Chetty, R., J. N. Friedman, and E. Saez (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings. *American Economic Review 103*(7), 2683–2721.

Chetty, R. and E. Saez (2013). Teaching the tax code: Earnings responses to an experiment with EITC recipients. *American Economic Journal: Applied Economics 5*(1), 1–31.

Chetty, R. and E. Saez (2014). How can we increase pro-social behavior? An experiment with referees at the journal of public economics. *Journal of Economic Perspectives*. forthcoming.

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

Collins, J., D. Murphy, and D. Plumlee (1992). The taxpayer's dilemma: How hard to work and what to report? *Advances in Taxation 4*, 31–53.

Coricelli, G., M. Joffily, C. Montmarquette, and M. Villeval (2010). Cheating, emotions, and rationality: an experiment on tax evasion. *Experimental Economics 13*(2), 226–247.

Cowell, F. A. (1985). Tax evasion with labour income. *Journal of Public Economics 26*(1), 19–34.

Cowell, F. A. and J. P. Gordon (1988). Tax evasion and public good provision. *Journal of Public Economics 36*, 305 – 321.

Cox, J. C., M. Rider, and A. Sen (2012). Tax incidence: Do institutions matter? an experimental study. Experimental Economics Center Working Paper Series No 2012-17.

Cramer, J. S., J. Hartog, N. Jonker, and C. M. Van Praag (2002). Low risk aversion encourages the choice for entrepreneurship: an empirical test of a truism. *Journal of Economic Behavior & Organization 48*(1), 29–36.

Cremer, H. and F. Gahvari (1993). Tax evasion and optimal commodity taxation. *Journal of Public Economics 50*(2), 261 – 275.

Cummings, R. G., J. Martinez-Vazquez, M. McKee, and B. Torgler (2009). Tax morale affects tax compliance: Evidence from surveys and an artefactual field experiment. *Journal of Economic Behavior & Organization 70*(3), 447 – 457.

Dickinson, D. L. (1999). An experimental examination of labor supply and work intensities. *Journal of Labor Economics 17*(4), 638–670.

Djawadi, B. M. and R. Fahr (2012). The impact of tax knowledge and budget spending influence on tax compliance. mimeo, University of Paderborn.

Doerrenberg, P. and D. Duncan (2014a). Distributional implications of tax evasion: Evidence from the lab. *Public Finance Review*. Forthcoming.

Doerrenberg, P. and D. Duncan (2014b). Experimental evidence on the relationship between tax evasion opportunities and labor supply. *European Economic Review 68*, 48 – 70.

Doerrenberg, P. and D. Duncan (2014c). Tax incidence in the presence of tax evasion. IZA Discussion Paper No. 8137.

Doerrenberg, P., D. R. Duncan, C. Fuest, and A. Peichl (2014). Nice guys finish last: Do honest taxpayers face higher tax rates? *Kyklos 67*(21), 29–53.

Doerrenberg, P. and A. Peichl (2013). Progressive taxation and tax morale. *Public Choice 155*(3-4), 293–316.

Dufwenberg, M., T. Lindqvist, and E. Moore (2005). Bubbles and experience: An experiment. *American Economic Review 95*(5), 1731–1737.

Dulleck, U., J. Fooken, C. Newton, A. Ristl, M. Schaffner, and B. Torgler (2012). Tax compliance and psychic costs: Behavioral experimental evidence using a physiological marker. CREMA Working Paper Series 2012-19.

Duncan, D. (2014). Behavioral responses and the distributional effects of the Rusian 'flat' tax. *Journal of Policy Modelling 36*, 226–242.

Duncan, D. and J. Graham (2013). Road user fees instead of fuel taxes: The quest for political acceptability. *Public Administration Review 73*(3), 415–426.

Duncan, D. and K. Sabirianova-Peter (2014). Switching on the lights: Do higher income taxes push economic activity into the shade? *National Tax Journal 67*(2), 321–350.

Dwenger, N., H. Kleven, I. Rasul, and J. Rincke (2014). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. Working paper. MPI Munich.

Eckel, C. C. and P. J. Grossman (1996). Altruism in anonymous dictator games. *Games and Economic Behavior 16*(2), 181–191.

Elffers, H., R. H. Weigel, and D. J. Hessing (1987). The consequences of different strategies for measuring tax evasion behavior. *Journal of Economic Psychology 8*(3), 311 – 337.

Erard, B. and J. S. Feinstein (1994). The role of moral sentiment and audit perceptions in tax compliance. *Public Finance 49*(Supplement), 70 – 89.

Evans, W., J. Ringel, and D. Stech (1999). Tobacco taxes and public policy to discourage smoking. In J. Poterba (Ed.), *Tax Policy and the Economy*, Volume 13. Cambridge, USA: MIT Press.

Falk, A. and E. Fehr (2003). Why labour market experiments? *Labour Economics 10*(4), 399–406.

Falk, A. and J. J. Heckman (2009). Lab experiments are a major source of knowledge in the social sciences. *Science 326*(5952), 535–538.

Falk, A., S. Meier, and C. Zehnder (2013). Do lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of the European Economic Association 11*(4), 839–852.

Feige, E. L. (1990). Defining and estimating underground and informal economies: The new institutional economics approach. *World Development 18*(7), 989 – 1002.

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

Feld, L. P. and B. S. Frey (2007). Tax compliance as the result of a psychological tax contract: The role of incentives and responsive regulation. *Law & Policy 29*(1), 102 – 120.

Feld, L. P. and J.-R. Tyran (2002). Tax evasion and voting: An experimental analysis. *Kyklos 55*(2), 197–221.

Feldman, N. E. and J. Slemrod (2007). Estimating tax noncompliance with evidence from unaudited tax returns. *Economic Journal 117*(March), 327–352.

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

Finkelstein, A. (2009). E-ztax: Tax salience and tax rates. *Quarterly Journal of Economics 124*(3), 969–1010.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Fochman, M. and J. Weimann (2011). The effects of tax salience and tax experience on individual work efforts in a framed field experiment. FEMM Working Paper 20.

Fochman, M., J. Weimann, K. Blaufus, J. Hundsdoerfer, and D. Kiesewetter (2012). Net wage illusion in a real effort experiment. *Scandinavian Journal of Economics*. Forthcoming.

Fong, C. M. (2007). Evidence from an experiment on charity to welfare recipients: Reciprocity, altruism and the empathic responsiveness hypothesis. *The Economic Journal 117*(522), 1008 – 1024.

Fong, C. M. and E. F. Luttmer (2011). Do fairness and race matter in generosity? evidence from a nationally representative charity experiment. *Journal of Public Economics 95*(5-6), 372 – 394.

Fong, C. M. and E. F. P. Luttmer (2009). What determines giving to hurricane katrina victims? experimental evidence on racial group loyalty. *American Economic Journal: Applied Economics 1*(2), 64–87.

Fortin, B., G. Lacroix, and M.-C. Villeval (2007). Tax evasion and social interactions. *Journal of Public Economics 91*(11-12), 2089–2112.

Frederiksen, A., E. K. Graversen, and N. Smith (2005). Tax evasion and work in the underground sector. *Labour Economics 12*(5), 613–628.

Frey, B. and H. Weck-Hanneman (1984). The hidden economy as an 'unobserved' variable. *European Economic Review 26*(1-2), 33 – 53.

Frey, B. S. and L. P. Feld (2002). Deterrence and morale in taxation. CESifo Working Paper No. 760.

Frey, B. S. and B. Torgler (2007). Tax morale and conditional cooperation. *Journal of Comparative Economics 35*(1), 136 – 159.

Fuest, C., A. Peichl, and S. Siegloch (2013). Do higher corporate taxes reduce wages? Micro evidence from Germany. IZA Discussion Papers 7390, Institute for the Study of Labor (IZA).

Gallagher, K. S. and E. Muehlegger (2011). Giving green to get green? incentives and consumer adoption of hybrid vehicle technology. *Journal of Environmental Economics and Management 61*(1), 1 – 15.

GAO (2000). Sales taxes: Electronic commerce growth presents challenges; revenue losses are uncertain. US Government Accounting Office (GAO): Report to Congressional Requesters No. GAO/GGD/OCE-00-165, Washington D.C.

Gill, D. and V. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review 102*(1), 469–503.

Gill, D. and V. Prowse (2013). A novel computerized real effort task based on sliders. Oxford Department of Economics Discussion Paper 435.

Gill, D., V. Prowse, and M. Vlassopoulos (2013). Cheating in the workplace: An experimental study of the impact of bonuses and productivity. *Journal of Economic Behavior and Organization 96*, 120–134.

Gill, D. and V. L. Prowse (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*. Forthcoming.

Gorodnichenko, Y., J. Martinez-Vazquez, and K. S. Peter (2009). Myth and reality of flat tax reform: Micro estimates of tax evasion response and welfare effects in russia. *Journal of Political Economy 117*, 504–554.

Graetz, M. J., J. F. Reinganum, and L. L. Wilde (1986). The tax compliance game: Toward an interactive theory of law enforcement. *Journal of Law, Economics and Organization 2*(1), 1–32.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer and V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63*, pp. 79–93. Goettingen: Ges. fuer Wiss. Datenverarbeitung.

Grosser, J. and E. Reuben (2013). Redistribution and market efficiency: An experimental study. *Journal of Public Economics 101*(May), 39 – 52.

Grossman, P. J. and C. C. Eckel (2012). Giving versus Taking: A "Real Donation" Comparison of Warm Glow and Cold Prickle in a Context-Rich Environment. Monash University Economics Working Papers 20-12.

Gruber, J. and B. Koszegi (2004). Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes. *Journal of Public Economics 88*(9-10), 1959–1987.

Halla, M. (2012). Tax morale and compliance behavior: First evidence on a causal link. *The B.E. Journal of Economic Analysis & Policy 12*(1).

Halla, M., M. Lackner, and F. G. Schneider (2010). An empirical analysis of the dynamics of the welfare state: The case of benefit morale. *Kyklos 63*(1), 55–74.

Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev (2014). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. National Bureau of Economic Research (NBER) Working Paper 20007.

Hammermann, A., A. Mohnen, and P. Nieken (2012). Whom to choose as a team mate? A lab experiment about in-group favouritism. IZA Discussion Papers 6286.

Heim, B. T. (2009). The effect of recent tax changes on taxable income: Evidence from a new panel of tax returns. *Journal of Policy Analysis and Management 28*(1), 147–163.

Heim, B. T. (2010). The responsiveness of self-employment income to tax rate changes. *Labour Economics 17*(6), 940–950.

Heinemann, F. (2008). Is the welfare state self-destructive? A study of government benefit morale. *Kyklos 61*(2), 237–257.

Helliwell, J. F. (2003). How's life? Combining individual and national variables to explain subjective well-being. *Economic Modelling 20*(2), 331 – 360.

HMRC (2012). Measuring tax gaps 2012. Tax gap estimates for 2010-11. United Kingdom HM Revenue & Customs. Available online: `http://www.hmrc.gov.uk/statistics/tax-gaps/mtg-2012.pdf` (accessed June 2014).

Holt, C. A. (1995). Industrial organization: A survey of laboratory research. In J. H. Kagel and A. E. Roth (Eds.), *The handbook of experimental economics*, pp. 349 – 443. Priceton, USA: Prince- ton University Press.

Hsu, L. (2008). Experimental evidence on tax compliance and voluntary public good provision. *National Tax Journal 61*(2), 205–223.

ICRG (2011). International country risk guide, table 3b: Political risk components. Available online: `http://www.prsgroup.com/prsgroup_shoppingcart/p-75-icrg-historical-data.aspx` (accessed September 2011).

Inglehart, R. (2000). *Codebook for World Values Survey.* Ann Arbor, USA: Institute for Social Research.

IRS (2012). Tax gap for tax year 2006. United States Internal Revenue Service. Available online: `http://www.irs.gov/pub/newsroom/overview_tax_gap_2006.pdf` (accessed June 2014).

Iyer, G. S., P. M. Reckers, and D. L. Sanders (2010). Increasing tax compliance in Washington state: A field experiment. *National Tax Journal 63*(1), 7–32.

Jenkins, S., R. V. Burkhauser, S. Feng, and J. Larrimore (2011). Measuring inequality using censored data: a multiple-imputation approach to estimation and inference. *Journal Of The Royal Statistical Society Series A 174*(1), 63 – 81.

Johns, A. and J. Slemrod (2010). The distribution of income tax noncompliance. *National Tax Journal 63*, 397–418.

Johnson, S., D. Kaufmann, and A. Shleifer (1997). The unofficial economy in transition. *Brookings Papers on Economic Activity 2*, 159–239.

Kachelmeier, S. J., S. T. Limberg, and M. S. Schadewald (1994). Experimental evidence of market reactions to new consumption taxes. *Contemporary Accounting Research 10*(2), 505–545.

Kaufmann, D. and A. Kaliberda (1996). Integrating the unofficial economy into the dynamics of post socialist economies: A framework of analyses and evidence. Worldbank policy research paper 1691.

Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature 49*(4), 961–1075.

Kerschbamer, R. and G. Kirchsteiger (2000). Theoretically robust but empirically invalid? an experimental investigation into tax equivalence. *Economic Theory 16*(3), 719–734.

Kessler, J. and L. Vesterlund (2014). The external validity of laboratory experiments: Qualitative rather than quantitative effects. In G. Frechette and A. Schotter (Eds.), *Methods of Modern Experimental Economics*. Oxford: Oxford University Press.

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

Koenig, H., F. Laisney, M. Lechner, and W. Pohlmeier (1995). Tax illusion and labour supply of married women: Evidence from German data. *Kyklos 48*(3), 347–368.

Konrad, K. A. and S. Qari (2012). The last refuge of a scoundrel? Patriotism and tax compliance. *Economica 79*(15), 516 – 533.

Kopczuk, W., J. Marion, E. Muehlegger, and J. Slemrod (2013). Do the laws of tax incidence hold? point of collection and the pass-through of state diesel taxes. NBER working paper 19410.

Lamberton, C., J.-E. D. Neve, and M. I. Norton (2014). Eliciting taxpayer preferences increases tax compliance. Harvard Business School Working Paper 14-106.

Lefebvre, M., P. Pestieau, A. Riedl, and M. Villeval (2014). Tax evasion and social information: an experiment in belgium, france, and the netherlands. *International Tax and Public Finance*. forthcoming.

Lemieux, T., B. Fortin, and P. Frechette (1994). The effect of taxes on labor supply in the underground economy. *American Economic Review 84*(1), 231–54.

Levitt, S. D. and J. A. List (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives 21*(2), 153–174.

Li, S. X., C. C. Eckel, P. J. Grossman, and T. L. Brown (2011). Giving to government: Voluntary taxation in the lab. *Journal of Public Economics 95*(9-10), 1190 – 1201.

Lubian, D. and L. Zarri (2011). Happiness and tax morale: An empirical analysis. *Journal of Economic Behavior & Organization 80*(1), 223 – 243.

Luttmer, E. F. P. (2001). Group loyalty and the taste for redistribution. *Journal of Political Economy 109*(3), 500–528.

Mann, H. B. and D. R. Whitney (1947). On a test whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics 18*, 50–60.

Marion, J. and E. Muehlegger (2011). Fuel tax incidence and supply conditions. *Journal of Public Economics 95*(9-10), 1202 – 1212.

Marrelli, M. (1984). On indirect tax evasion. *Journal of Public Economics 25*(1-2), 181–196.

Marwell, G. and R. E. Ames (1981). Economists free ride, does anyone else?: Experiments on the provision of public goods. *Journal of Public Economics 15*(3), 295 – 310.

Mettler, S. (2011). *The submerged state*. Chicago, USA: University of Chicago Press.

Mikesell, J. (2014). Misconceptions about value-added and retail sales taxes: are they barriers to sensible tax policy? *Public Budgeting & Finance*, Forthcoming.

Minkov, M. (2012). World values survey. In G. Ritzer (Ed.), *The Wiley-Blackwell Encyclopedia of Globalization*. Oxford: Blackwell Publishing Ltd.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies 38*(2), 175 – 208.

Okun, A. M. (1975). *Equity and efficiency: The big tradeoff*. Washington D.C., USA: Brookings Institution Press.

Paetzold, J. and H. Winner (2014). Unwilling, unable or uninformed to cheat. tax evasion via (quasi-) self-reporting in austria. Unpublished working paper.

Pencavel, J. H. (1979). A note on income tax evasion, labor supply, and nonlinear tax schedules. *Journal of Public Economics 12*(1), 115–124.

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

Pomeranz, D. (2013). No taxation without information: Deterrence and self-enforcement in the value added tax. Harvard Business School Working Paper 13-057.

Qari, S., K. A. Konrad, and B. Geys (2011). Patriotism, taxation and international mobility. *Public Choice*. DOI 10.1007/s11127-011-9765-3.

Quirmbach, H. C., C. W. Swenson, and C. C. Vines (1996). An experimental examination of general equilibrium tax incidence. *Journal of Public Economics 61*(3), 337–358.

Ramsey, F. P. (1927). A contribution to the theory of taxation. *The Economic Journal 37*(145), 47 – 61.

Reiter, J. (2003). Inference for partially synthetic, public use microdatasets. *Survey Methodology 29*(2), 181 – 188.

Richardson, G. (2006). Determinants of tax evasion: A cross-country investigation. *Journal of International Accounting, Auditing and Taxation 15*(2), 150 – 169.

Riedl, A. (2010). Behavioral and experimental economics do inform public policy. *FinanzArchiv: Public Finance Analysis 66*(1), 65–95.

Riedl, A. and J.-R. Tyran (2005). Tax liability side equivalence in gift-exchange labor markets. *Journal of Public Economics 89*(11-12), 2369–2382.

Riener, G. and S. Wiederhold (2011). On social identity, subjective expectations, and the costs of control. Jena Economic Research Papers 2011-035.

Roed, K. and S. Strom (2002). Progressive taxes and the labour market: Is the trade-off between equality and efficiency inevitable? *Journal of Economic Surveys 16*(1), 77–110.

Roine, J. (2006). The political economics of not paying taxes. *Public Choice 126*, 107–134.

Rosen, H. S. (1976a). Tax illusion and the labor supply of married women. *The Review of Economics and Statistics 58*(2), 167–172.

Rosen, H. S. (1976b). Taxes in a labor supply model with joint wage-hours determination. *Econometrica 44*(3), 485–507.

Rothstein, J. (2010). Is the EITC as good as an NIT? Conditional cash transfers and tax incidence. *American Economic Journal: Economic Policy 2*(1), 177–208.

Ruffle, B. J. (2005). Tax and subsidy incidence equivalence theories: experimental evidence from competitive markets. *Journal of Public Economics 89*(8), 1519–1542.

Sabirianova Peter, K., S. Buttrick, and D. Duncan (2010). Global reform of personal income taxation, 1981-2005: Evidence from 189 countries. *National Tax Journa 63l*(3), 447–78.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy 2*(3), 180–212.

Saez, E., M. Matsaganis, and P. Tsakloglou (2012). Earnings determination and taxes: Evidence from a cohort-based payroll tax reform in greece. *The Quarterly Journal of Economics 127*(1), 493–533.

Saez, E., J. Slemrod, and S. H. Giertz (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature 50*(1), 3–50.

Sandmo, A. (1981). Income tax evasion, labour supply, and the equity-efficiency trade-off. *Journal of Public Economics 16*(3), 265–288.

Sandmo, A. (2005). The theory of tax evasion: A retrospective view. *National Tax Journal 58*(4), 643–663.

Sausgruber, R. and J.-R. Tyran (2005). Testing the mill hypothesis of fiscal illusion. *Public Choice 122*(1), 39–68.

Schneider, F., A. Buehn, and C. Montenegro (2010). New estimates for the shadow economies all over the world. *International Economic Journal 24*(4), 443–461.

Schneider, F. and D. H. Enste (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature 38*(1), 77 – 114.

Scholz, J. T. and M. Lubell (1998). Trust and taxpaying: Testing the heuristic approach to collective action. *American Journal of Political Science 42*(2), 398 – 417.

Scholz, J. T. and N. Pinney (1995). Duty, fear, and tax compliance: The heuristic basis of citizenship behavior. *American Journal of Political Science 39*(2), 490–512.

Schwartz, R. D. and S. Orleans (1967). On legal sanctions. *University of Chicago Law review 34*, 274 – 300.

Simonovits, A. (2011). Higher tax morale implies a higher optimal income tax rate. Hungarian Academy of Science, Institute of Economics, Working Paper 2011/37.

Slemrod, J. (1994). On the high-income laffer curve. In J. Slemrod (Ed.), *Tax Progressivity and Income Inequality*, pp. 177–210. Cambridge, New York and Melbourne: Cambridge University Press.

Slemrod, J. (1995). Income creation or income shifting? behavioral responses to the tax reform act of 1986. *The American Economic Review 85*(2), pp. 175–180.

Slemrod, J. (2001). A general model of the behavioral response to taxation. *International Tax and Public Finance 8*, 119–128.

Slemrod, J. (2003). Trust in public finance. In S. Cnossen and H.-W. Sinn (Eds.), *Public Finance and Public Policy in the New Century*, pp. 49–88. Cambridge, USA: The MIT Press.

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

Slemrod, J. (2008). Does it matter who writes the check to the government? the economics of tax remittance. *National Tax Journal 61*, 251–75.

Slemrod, J., M. Blumenthal, and C. Christian (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in minnesota. *Journal of Public Economics 79*(3), 455 – 483.

Slemrod, J. and C. Weber (2012). Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance 19*, 25–53.

Slemrod, J. and S. Yitzhaki (2002). Tax avoidance, evasion, and administration. In A. J. Auerbach and M. Feldstein (Eds.), *n.a.*, Volume 3 of *Handbook of Public Economics*, pp. 1423 – 1470. Amsterdam, Netherlands: Elsevier.

Slonim, R. and A. E. Roth (1998). Learning in high stakes ultimatum games: An experiment in the slovak republic. *Econometrica 66*(3).

Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy 70*, 322.

SNTA (2008). Tax gap map for sweden. Swedish National Tax Agency. Available online: `http://www.skatteverket.se/download/18.225c96e811ae46c823f800014872/Report_2008_1B.pdf` (accessed June 2014).

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557 – 586.

Tanzi, V. (1983). The underground economy in the united states: Annual estimates. *IMF Staff Papers 30*(02), 283 – 305.

Torgler, B. (2002). Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. *Journal of Economic Surveys 16*(5), 657–683.

Torgler, B. (2005). Tax morale and direct democracy. *European Journal of Political Economy 21*(2), 525 – 531.

Torgler, B. (2006). The importance of faith: Tax morale and religiosity. *Journal of Economic Behavior & Organization 61*(1), 81 – 109.

Torgler, B. (2007). *Tax Compliance and Tax Morale: A Theoretical and Empirical Analysis*. Cheltenham, UK: Edward Elgar.

Torgler, B. and F. Schneider (2007). What shapes attitudes toward paying taxes? Evidence from multicultural European countries. *Social Science Quarterly 88*(2), 443 – 470.

Torgler, B. and F. Schneider (2009). The impact of tax morale and institutional quality on the shadow economy. *Journal of Economic Psychology 30*(2), 228 – 245.

Torgler, B., F. Schneider, and C. Schaltegger (2010). Local autonomy, tax morale, and the shadow economy. *Public Choice 144*, 293–321.

Tran, A. and N. Nguyen (2014). The darker side of private ownership: Tax manipulation in vietnamese privatized firms. Indiana university working paper.

Traxler, C. (2010). Social norms and conditional cooperative taxpayers. *European Journal of Political Economy 26*(1), 89 – 103.

Traxler, C. (2012). The distribution of income tax noncompliance. *Journal of Public Economics 96*, 1–9.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*, 1124–1131.

Vesterlund, L. (2014). Voluntary giving to public goods: Moving beyond the linear VCM. In J. H. Kagel and A. E. Roth (Eds.), *Handbook of Experimental Economics Volume 2*, pp. 229–330. Princeton: Princeton University Press.

Wahl, I., S. Muehlbacher, and E. Kirchler (2010). The impact of voting on tax payments. *Kyklos 63*(1), 144–158.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics 1*, 80–83.

Wooldridge, J. M. (2006). *Introductory Econometrics. A Modern Approach* (3rd ed.). Cincinnati, USA: South-Western College Publishing.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2 ed.). The MIT Press.

World Bank (2010). The Word Development Indicators. Available online: `http://databank.worldbank.org/` (accessed June 2011).

Yaniv, G. (1995). A note on the tax-evading firm. *National Tax Journal 48*(1), 113–120.

Yitzhaki, S. (1974). A note on income tax evasion: A theoretical analysis. *Journal of Public Economics 3*(2), 201–202.

Curriculum Vitae

Contact Information

ZEW Mannheim
L 7, 1
D-68161 Mannheim, Germany

Email: doerrenberg@zew.de
Phone: +49 (0)621 1235 162
Web: http://www.zew.de/

Personal Details

Born	10 November 1984 in Ratingen (Germany)
Citizenship	German
Marital Status	Single

Current Positions

since 09.2013	Researcher, Centre for European Economic Research (ZEW), Mannheim
since 01.2012	Research Fellow (since 07.2014) and Affiliate (01.2012-07.2014), Institute for the Study of Labor (IZA), Bonn

Previous Academic Positions

10.2010—08.2013	Fellow and Scholarship Holder, Cologne Graduate School in Economics, Management and Social Sciences (CGS), University of Cologne
04.2011—08.2013	Resident Research Affiliate, Institute for the Study of Labor (IZA), Bonn
01.2012—04.2012	Visiting Scholar, School of Public and Environmental Affairs (SPEA), Indiana University, Bloomington, IN
11.2007—07.2010	Student Research Assistant, FiFo Institute for Public Economics at University of Cologne

Education

since 10.2010	Doctoral Studies in Economics, University of Cologne, Advisor: Prof. Dr. Clemens Fuest
08.2010	Diplom-Volkswirt (equiv. M.A. Economics), University of Cologne, Grade: 1.3 (top of class),
10.2005—08.2010	Studies in Economics (major) and Political Science, University of Cologne and Trinity College Dublin (09.2008—03.2009)
1995—2004	High School: Kopernikus Gymnasium Ratingen, Germany and Wyomissing Area High School, Wyomissing, PA, USA (2001-2002)

Refereed Journal Publications

"Experimental Evidence on the Relationship between Tax Evasion Opportunities and Labor Supply", 2014, *European Economic Review* 68(May), pages 48-70 (with D. Duncan)

"Is soccer good for you? The motivational impact of big sporting events on the unemployed", 2014, *Economics Letters* 123(1), pages 66-69 (with S. Siegloch)

"Distributional Implications of Tax Evasion: Evidence from the Lab", forthcoming 2014, *Public Finance Review* (with D. Duncan)

"Nice Guys Finish Last: Do Honest Taxpayers Face Higher Tax Rates?", 2014, *Kyklos* 67(1), pages 29-53 (with D. Duncan, C. Fuest and A. Peichl) (News coverage, *Handelsblatt*, 03.2012)

"The impact of redistributive policies on inequality in OECD countries", 2014, *Applied Economics* 46(17), pages 2066-2086 (with A. Peichl)

"Progressive Taxation and Tax Morale", 2013, *Public Choice* 155(3-4), pages 293-316 (with A. Peichl)

Under Review at Academic Journals / Working Paper

"Circumstantial Risk: Impact of future tax evasion and labor supply opportunities on risk exposure", revise and resubmit, *Journal of Economic Behavior & Organization* (with D. Duncan and C. Zeppenfeld)

"Tax Incidence and Tax Evasion", *submitted* (with D. Duncan)

Other Publications

"Finanzwissenschaftliches Gutachten zum Länderfinanzausgleich", Expertise on behalf of the Ministry of Economics and Finance at the federal state of Baden-Wuerttemberg, 2014 (with F. Heinemann)

"Biofuels – At what cost? Government support for ethanol and biodiesel in the European Union", *Research report for the Global Subsidies Initiative (GSI)*, Geneva, Switzerland, 2010 (with A. Jung, A. Rauch and M. Thoene)

"Green Buildings – A niche becomes mainstream", *DB Research. Current Issues*, 2010 (with A. Nelson and O. Rakau)

Work in Progress

"The effect of tax rate changes on tax deductions" (with A. Peichl and S. Siegloch)

"Google me this, Google me that: What can Google tell us about tax evasion?" (with J. Alm and D. Duncan)

"Elasticities of business and property taxation: Empirical evidence from German municipalities" (with A. Rauch)

"Pre-Populated Income Tax Forms: A Nudge in the Wrong Direction?" (with N. Brodnax and D. Duncan)

"Does the usage of tax revenue matter for compliance behavior? Experimental Evidence"

Teaching Experience

Fall 13/Spring 14	"Seminar on International Taxation" (undergraduate level), University of Mannheim, Teaching Assistant for Professor C. Fuest
Spring 2013	"Empirical Public Economics" (graduate level), University of Cologne, Center for Macroeconomic Research (CMR), Teaching Assistant for Dr. A. Peichl
since 2012	(Co-) supervision of several BA-theses and term papers, University of Mannheim and University of Cologne
Spring 2010	"Introductory Microeconomics" (undergraduate level), University of Cologne, Teaching Assistant for Professor O. Guertler

---

Conference Presentations and Seminar Talks

2014	International Institute of Public Finance (IIPF) Annual Conference, Lugano (scheduled); ZEW Mannheim (scheduled)
2013	ZEW Mannheim (09.2013); European Commission, Brussels (invited talk, 09.2013); German Economic Association (Verein fuer Socialpolitik), Duesseldorf (09.2013); International Institute of Public Finance (IIPF) Annual Conference, Taormina/Sicily (08.2012); Society of Labor Economists (SOLE) Annual Meeting, Boston (05.2013); University of Frankfurt (invited talk, 02.2013)
2012	European Commission, Brussels (invited talk, 12.2012); CESifo Political Economy Conference, Dresden (12.2012); IZA Bonn (11.2012); SPEA at Indiana University, Bloomington (invited talk, 10.2012); International Institute of Public Finance (IIPF) Annual Conference, Dresden (08.2012); European Meeting of the International Microsimulation Association, Dublin (05.2012); SPEA at Indiana University, Bloomington (04.2012)
2011	CESifo Political Economy Conference, Dresden (12.2011); IZA Bonn (09.2011); shadow2011: The Shadow Economy, Tax Evasion and Money Laundering, Muenster (07.2011)
2010—2013	Several presentations at University of Cologne (2010—2013)

Awards, Grants and Scholarships

Full Scholarship for Ph.D. Studies, granted by the German Federal State NRW (10.2010—08.2013)

Scholarship for Research Visit to the USA, granted by the German Academic Exchange Service (DAAD) (01.2012—04.2012)

Dermot McAleese Medal (Award—including prize money—for "Best Overall Essay" in the Student Economic Review 2009 at Trinity College Dublin)

Erasmus Scholarship for Studies Abroad (09.2008—03.2009)

Funding for "Risk Taking, Labor Supply and Tax Evasion: Theory and Experimental Evidence" (EURO 2650, with D. Duncan and C. Zeppenfeld), granted by "KölnAlumni – Freunde und Förderer der Universität zu Köln" (07.2013)

Further Research Activities

MaTax	Coordinator (jointly with K. Finke), Leibniz Research Cluster "Science Campus Mannheim Taxation" (MaTax), University of Mannheim and ZEW Mannheim
Refereeing	Economics of Governance, Business Administration Review
Organizer	MaTax Annual Conference 2014, Keynote: Joel Slemrod (jointly with K. Finke)

Memberships	International Institute of Public Finance, Society of Labor Economists, Verein für Socialpolitik, European Economic Association

Non-Academic Professional Experience

2005-2009	Deutsche Bank Research, Deutsche Bank AG, Frankfurt a.M. (Internship, 08—10.2009); Ministry of Economics at the German Federal State of Rhineland-Palatinate, Mainz (Internship, 10.2007); State Parliament at the German Federal State North Rhine-Westphalia, Duesseldorf (Internship, 07.2005—09.2005); School for Mentally Disabled Children: Helen-Keller-Schule, Ratingen (Civilian Service, 09.2004—05.2005)

Miscellaneous

Languages

- German (native), English (fluent)

Computer

- Stata, MS Office, zTree, LaTeX

Mannheim, August 11, 2014