Essays in Development and Public Economics: Savings, Information and Formalization

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

Albrecht Bohne

August 2018

Abteilungssprecher:	Prof. Dr. Jochen Streb
Referent:	Prof. Dr. Markus Frölich
Korreferent:	Prof. Dr. Andreas Peichl

Tag der Verteidigung: 07. November 2018

Für Vero & Felix

Acknowledgements

I would like to thank my advisors Markus Frölich and Andreas Peichl for their support and valuable feedback throughout the whole process of writing this thesis. I especially value Markus Frölich's exemplary drive for finding solutions to issues affecting the poorest and most vulnerable individuals on our planet. Explicit thanks go to Andreas Peichl for supporting my endeavors with the Ecuadorian tax authority from the very beginning and providing crucial guidance in shaping research questions into academic papers. I extend my thanks to Katja Kaufmann for numerous discussions about my research papers.

I am indebted to the Ecuadorian tax authority *Servicio de Rentas Internas* for providing data access and indirectly to all Ecuadorian taxpayers for the opportunity to study their tax system in such detail. I especially thank Néstor Villacrese, José Almeida, María Andrade and the whole team at the Centro de Estudios Fiscales for hosting me during numerous research stays and providing invaluable help by discussing details of the tax system.

Moreover, I would like to thank my co-authors on the first chapter of this thesis, Markus Frölich and Alexandra Avdeenko for the opportunity to work on all stages of a randomized control trial. I will never forget the numerous discussions on how to design the experiment, organize the fieldwork, lead the budget negotiations, get a shipment of moneyboxes through Ethiopian customs and of course analyze the data and write the paper.

A very sincere thank you goes to my friend Jan Nimczik for the numerous and stimulating discussions we had about economic research, our joint paper on taxes in Ecuador (chapter 2) and life in general during the whole PhD program, numerous trips to conferences, and during our joint research stay in Ecuador.

Additionally, I thank numerous fellow researchers and colleagues at the University of Mannheim and the graduate school for fruitful and extensive discussions. Special thanks go to Torben Fischer, Jasper Haller, Sebastian Panthöfer, and Maria Isabel Santana-Penczynski.

Lastly, I would like to thank my whole family, and in particular my parents, Friedgund and Thomas Bohne, for their continuous support during the very long course of my studies. Most importantly, I thank my wife Verónica for her immense love, unwavering support and tolerance for the ups and downs associated with doing a PhD. I am extremely grateful for the wonderful family we have together with our son Felix Mateo.

Contents

List of Figures

List of Tables

Preface

Improving the livelihoods of all individuals around the globe can be considered as one of the key aims of modern economic and political endeavors. A crucial challenges is how to set up structures and frameworks which help individuals in developing countries make lasting improvements to their economic prospects and overall well-being. These structures and frameworks need to address a multitude of issues at varying levels such as individual behavior, interactions between individuals and nation-wide effects governing the structure of the economy.

This thesis addresses these fundamental issues at various levels and in differing contexts. Chapter 1 studies how to provide individuals, in this case smallholder farmers in Ethiopia, the necessary tools to overcome behavioral biases and economic restrictions due to extremely seasonal income patterns. Chapters 2 and 3 look into how to efficiently design tax and transfer systems in developing countries. While chapter 2 deals with information frictions in the tax system and how these can be alleviated through interactions between individuals, chapter 3 analyzes how a specific reform can shift business activity from the informal to the formal economy and thereby change the structure of the economy.

The chapters of this thesis also touch on various aspects of how to analyze and design effective public policies in general. Chapter 1 is an example of how to leverage behavioral insights to design innovative policies providing individuals with the tools to improve their outcomes. Chapter 2 looks at the roll-out of a policy where information frictions prevent universal take-up and analyzes how these information frictions can be overcome. Finally, chapter 3 studies how individual policy responses can be harnessed to create spillovers in a different economic sector with improvements for the whole economy. When designing effective policies in any setting, all three of these aspects need to be taken into account: how to make a policy effective at the individual level, how to roll out the policy to the whole economy, and what its spillover effects are on other sectors of the economy. In many cases, policy conclusions are reached without giving sufficient consideration to the latter two issues.

Methodologically, the chapters in this thesis use a range of approaches depending on the type of question at hand. Chapter 1 implements a randomized control trial and thereby provides very robust answers to a specific question about individual behavior while abstracting

from almost all other confounders. However, this gold standard cannot be applied to answer all questions. Analyzing the questions addressed in chapter 2 about information flows within a whole society would be extremely costly to address with a randomized control trial. Therefore, this thesis draws on various sources of novel and extremely rich administrative data with employer-employee linkages and details on the universe of personal income tax records. This extensive data provides unique quasi-experimental variation in the information environment individuals face along various dimensions, providing the basis of the analysis on information frictions in chapter 2. Chapter 3 exploits plausibly exogenous variation in the usage of a policy to estimate its spillovers towards a different sector of the economy.

Chapter 1: Savings Behavior Chapter 1 is based on joint work with Markus Frölich and Alexandra Avdeenko. In this chapter, we analyze behavioral constraints to savings among smallholder farmers in Ethiopia. These farmers face extremely seasonal income patterns, with a short harvest season with high income and much lower or no income generation during the rest of the year. However, expenditures including vital investments must often be made outside of the harvest season. Even in settings with more regular income patterns, previous literature has documented positive effects of increased savings on a number of economic outcomes in areas such as health, education and agricultural investment.

Due to these positive effects, a large literature in development economics focuses on finding innovative ways to increase savings among individuals in low-income economies. Along with providing individuals with access to savings accounts or savings technologies, many of these studies focus on overcoming behavioral biases hindering individuals from reaching their full savings potential. Key behavioral biases studied include present-biasedness, inattention, as well as those arising from non-standard risk preferences. Most interventions revolve around the provision of commitment devices restricting individual future choice sets.

This paper proposes a previously unexplored behavioral bias to savings, namely underconfidence. We find a strong empirical link between confidence levels and savings behavior, even when controlling for important behavioral traits such as present-biasedness and risk preferences. Individuals with low confidence levels save significantly less than individuals with high confidence levels. Within an intervention enabling individuals to save more through the provision of moneyboxes as a simple savings technology, we experimentally implement an innovative feedback loop in the form of recommendations to change previously self-stated saving goals. Individuals receiving the additional feedback save significantly more than individuals not receiving this feedback, and this effect is especially strong for underconfident individuals. We rule out a number of alternative explanations due to other behavioral biases and crowding-out behavior into other forms of saving.

Chapter 2: Information Frictions and Learning Dynamics Chapter 2 is joint work with Jan Sebastian Nimczik. This chapter analyzes how dynamic learning processes reduce information frictions hindering individuals from responding optimally to policies. In many settings, policies and changes in regulation are implemented with the underlying idea that all economic actors with potential benefits will immediately participate and gain from a given policy change. We look at a setting in which many individuals, even though they are potential beneficiaries of a new policy, do not take it up. In our setting, the main obstacle seems to be information frictions and we thoroughly study how these can be overcome through experience and dynamic learning processes. In particular, we provide causal evidence on the exact mechanisms of information transmission by studying various patterns of (exogenous) job mobility.

We look at these questions in the context of legal tax avoidance opportunities in a country with a recent unprecedented growth in the size of its formal economy, Ecuador. By using generous deduction opportunities in the personal income tax system, taxpayers can significantly reduce their tax burden and often completely avoid paying taxes. Drawing on novel administrative data based on the universe of personal income tax declarations, we document that individuals are more likely to avoid paying taxes both as they personally gain experience in the formal economy but also as their firm gains experience in the formal economy. By studying individuals changing their jobs for exogenous reasons we provide evidence for how the firm-level information environment causally affects individual tax avoidance behavior.

Moreover, we identify the underlying channels of information transmission by exploiting further changes in the information environment individuals face through changes in their co-worker composition and accountant switches. We find peers in the form of new co-workers joining a firm to be important in affecting the tax avoidance behavior of incumbent workers. If these new co-workers were previously avoiding tax payments, they increase the likelihood that the incumbent co-workers also avoid paying taxes. Likewise, experts play an important role in driving individual tax avoidance behavior. Accountants previously working for a firm in which employees avoided paying taxes bring this knowledge to a new firm and increase the probability that employees at their new firm will avoid paying taxes.

Chapter 3: Formalization Chapter 3 studies how individual-level responses to incentives by consumers can be harnessed to improve tax compliance of business owners with economy-wide implications for the size of the formal sector. Increasing tax compliance is an especially important goal for developing economies, which are generally plagued by very large informal sectors (ILO, 2014). Getting all economic actors to participate in the formal sector economy is often a prerequisite for improving the quantity and quality of public spending and for ensuring that the tax and transfer system is perceived as just by all sectors of the population. However, although a number of recent efforts have been aimed

at formalizing developing economies, some sectors of the economy are particularly difficult to formalize. Among these are self-employed business owners, who also in more developed economies are very prone to tax avoidance and evasion (Chetty et al., 2011; Bastani and Selin, 2014). This effect is further exacerbated in the absence of the self-enforcing properties of the VAT system due to frequent exemptions.

This chapter evaluates a novel reform in Ecuador based on enlisting all individuals paying personal income tax to create paper trails and third-party information about business activity in sectors of the economy largely exempt from VAT payments. These taxpayers can deduct expenses in health, education, housing, clothing and food (up to certain limits) from their tax liability. Using administrative data based on the universe of self-employed personal income tax declarations, I estimate the effects of this increased demand for receipts on reported business profits of self-employed business owners. The main identifying variation stems from the fact that personal income taxes are levied at relatively high levels in Ecuador and therefore only a small share of the population uses these deductions. Moreover, there is substantial regional heterogeneity in the density of these high-income individuals and therefore business owners in some regions are exposed to a high demand for receipts while business owners in other regions to a low demand for receipts. I exploit this variation in a difference-in-differences framework and find large effects of the reform on aggregated economic activity: Reported regional business profits of self-employed increase by up to 33% per inhabitant living in a given region.

At the individual self-employed level, I exploit an additional source of variation due to the fact that the deductions only affect certain self-employed based on their professions. Drawing on data from the civil registry, I identify self-employed particularly affected by the reform, such as doctors, and those not affected, such journalists and economists. This second treatment layer allows to conduct triple difference regressions. I corroborate my previous findings at the aggregate level and estimate a treatment effect of the reform in which self-employed subject to a high demand for receipts increase their reported profits by almost 100%.

Chapter 1

Linking Savings Behavior, Confidence and Individual Feedback: A Field Experiment in Ethiopia[1]

with Markus Frölich and Alexandra Avdeenko

1.1 Introduction

Recent research has highlighted the importance of savings for individuals in developing countries. People in these countries are often exposed to potentially large idiosyncratic shocks while facing seasonal income patterns and lacking access to social insurance schemes. Finding ways to increase savings among these households has attracted considerable attention from economists. In fact, a broad literature has shown positive effects of increased savings on a range of development outcomes in areas ranging from agricultural investments (Brune et al., 2016) to health (Dupas and Robinson, 2013) and education (Karlan and Linden, 2017).[2]

In devising strategies to increase savings, previous research has put a strong focus on overcoming behavioral biases, especially through the use of innovative commitment devices (Ashraf et al., 2006; Dupas and Robinson, 2013; John, 2017).[3] This paper addresses a previ-

[1] An earlier version of this chapter was circulated under the name "Underconfidence and the Use of Persuasive Messages in the Attainment of Savings Goals". This study obtained IRB approval from the University of Mannheim "Ethikkommission" on April 22, 2015 and is registered at the AEA RCT Registry under # AEARCTR-0000613. Special thanks go to Niels Kemper for valuable discussions and inputs during the design of the intervention.

[2] Please refer to Karlan et al. (2014) for a comprehensive overview.

[3] The literature has also proposed various other barriers to savings, including transaction costs, information asymmetries, lack of access to financial services, and social constraints, to name just a few.

ously unexplored behavioral bias in determining savings behavior, namely confidence. Our first contribution to the literature is that we establish the link between confidence and savings behavior drawing on a detailed survey of microfinance clients in rural Ethiopia. We document that confidence is an independent predictor for baseline savings behavior, and at least as important as other behavioral characteristics such as risk-aversion and present-biasedness. Individuals with low levels of confidence, whom we label as underconfident, exhibit significantly lower levels of savings even when controlling for a range of socio-demographic and economically relevant variables. We classify individuals as underconfident if they underestimate their ability to find correct answers to a standard World Bank (1998) financial literacy module.

We further propose and experimentally test an innovative feedback loop enabling individuals to overcome their underconfidence and reach higher savings outcomes. To do so, we conduct a randomized controlled trial in 94 rural villages in Northern Ethiopia. Smallholder farmers in this region provide the ideal study population to analyze savings patterns due to the high variability in their income reflecting seasonal patterns in agriculture. We distribute moneyboxes along with individualized, self-set savings plans to around 600 randomly selected farmers during peak income season. Therewith, the farmers receive individual feedback on their reported savings goals in the form of two types of recommendations to save either more or less than their originally stated plan. We find that individuals who receive this individualized feedback to reflect and reconsider their original savings goal save more (an increase of 181 Birr or 36 percent), and this effect is especially strong for underconfident farmers. Even though respondents mostly change their goal in the intended direction (upwards or downwards), the direction of the recommendation we provided had no impact on actual savings outcomes. We take this as confirmation that the additional request to reconsider and revise the original savings goal, disregarding the direction of the revision, is crucial in helping underconfident individuals reach their full savings potential. We confirm the robustness of our results by ruling out various other behavioral mechanisms such as risk-aversion and present-biasedness. Moreover, we do not find any crowding-out behavior with respect to alternative savings mechanisms.[4]

Our paper relates to the growing literature on savings, especially in developing countries. Even though individuals in these countries may be poor, they still have (considerable) scope in deciding what to spend their money on (Banerjee and Duflo, 2006). In settings without access to formal savings mechanisms, people often save in informal mechanisms like "under-the-mattress", informal saving groups or investments into livestock (Karlan et al.,

[4] We are aware that some of the respondents in our sample may have simultaneous debt at potentially high interest rates and are therefore unable to make direct statements about the welfare effects of these increases in savings.

2014). Numerous studies have shown the importance of savings for a range of development outcomes for individuals in all income ranges (Brune et al., 2016; Dupas and Robinson, 2013; Karlan and Linden, 2017). Therefore, the literature has focused on identifying a set of constraints on savings, such as lack of financial knowledge (Perry and Morris, 2005; Berg and Zia, 2013; Bayer et al., 2009; Karlan et al., 2014), market frictions and reduced access to financial services (Dupas et al., forthcoming; Karlan et al., 2014). Moreover, numerous behavioral frictions have been put forward such as time-inconsistent preferences (Ashraf et al., 2006; Dupas and Robinson, 2013; John, 2017), inattention problems (Karlan et al., 2016; Kast et al., 2018), intra-household barriers (Dupas and Robinson, 2013), and procrastination in financial decision making (Becchetti et al., 2015; Bisin and Hyndman, 2014; Brown and Previtero, 2014; Duflo et al., 2009; Thaler and Benartzi, 2004; Linardi and Tanaka, 2013). We draw special attention to studies highlighting the importance of reminders and feedback in savings behavior. Karlan et al. (2016) show that reminders to save increased total bank savings and savings goal attainment by 6%. Kast et al. (2018) show the relevance of feedback and follow-up text messages in self-help saving groups. Carvalho et al. (2016) created bank accounts and had weekly visits by deposit collectors whereby the treatment group had the chance to make a considered saving decision. They conclude that the treatment led to a higher accumulation of wealth. Lastly, our paper contributes to the literature showing the effectiveness of lockboxes as a simple savings technology to increase savings (Dupas and Robinson, 2013; Shipton et al., 1990).

The second broad literature our paper connects to is one linking confidence and financial decision making. Overconfident individuals are known to take higher risks in financial markets (Kirchler and Maciejovsky 2002, Caliendo and Huang 2007, Doerr et al. 2011), trade too much, too aggressive and earn lower net returns (Barber and Odean 2001, Gervais and Odean 2001, Barberis and Thaler 2003). More closely related to our argument are the studies by Malmendier and Tate (2005) and Garrard and Robinson (2015). The authors argue that overconfident managers over-invest and take lower quality decisions when firms have abundant internal funds. Further studies have focused on the behavior of consumers and have shown that overconfident consumers overpay overestimating the benefit of the product or service (Grubb 2009, Grubb 2015 and Li et al. 2016). However, within the literature on confidence and financial decision making there are only few studies relating underconfidence to economic outcomes, especially savings behavior. Perhaps closest to this question is a study in the field of psychology by Tang and Baker (2016), who state that financial actions might be "intimidating, and short-term failures or distractions can undermine responsible long-term financial behaviors". They suggest that self-esteem relates to financial behavior both directly and indirectly through subjective financial knowledge. In the same vein, Ameriks et al. (2007) find that a group of people with a higher negative difference between

expected consumption and ideal consumption, the so-called underconsumers, actually accumulates more wealth being less tempted to consume. Likewise, a range of studies relate measurable personality traits such as locus of control and perceived control to financial decisions. In fact, high levels of perceived control seem to be key to savings decisions (Rotter 1966, Ajzen 2002, Perry and Morris 2005, Cobb-Clark et al. 2013, Fouarge et al. 2013). The widely used personality trait locus of control is also relevant, since self-confidence is one of its defining elements. Cobb-Clark et al. (2013) find that households with internal locus of control (believing to be in full control of their lives) save more than households with an external locus of control (believing lives are controlled by external factors). Moreover, Chatterjee et al. (2011) analyze the relationship between self-efficacy and savings behavior. They find that individuals with higher self-efficacy accumulate more wealth.

Our study combines both strands of literature by focusing on a behavioral constraint to savings so far only analyzed for developed countries, namely confidence. In particular, we contribute by complementing the prior focus on overconfidence and taking a closer look at underconfident individuals and their saving behavior. We establish a clear link between the two. Our last contribution is the design and implementation of a targeted encouragement experiment to overcome this behavioral constraint.

The rest of the paper is organized as follows: In Section 1.2 we describe the background and experimental design. Section 1.3 presents our data and analyzes the relationship between confidence and baseline saving levels. In Section 1.4 we present our results, followed by robustness checks in Section 1.5 Section 1.6 concludes.

1.2 Background and Experimental Design

The study was conducted in rural Northern Ethiopia (Tigray region). The experiment took place in December 2014 together with smallholder farmers in 94 rural villages. Participants were randomly selected out of lists of current and former microfinance clients.

1.2.1 Background

Economic activity in Northern rural Ethiopia is almost entirely dependent on agriculture. More than 90% of the households in our study directly engage in agriculture as their main source of income, and almost all of this is generated through crop production. Agricultural activity in Northern Ethiopia is heavily dependent on rainfall patterns, creating three seasons that are of relevance to smallholder farmers: *Belg* (March to June), associated with

little agricultural activity, *Kiremti* (July to September), where cultivation and heavy rainfalls take place, and *Kewie* (October to February), which is the season in which farmers harvest, sell or store their goods. Naturally, seasonal patterns in household income and cultural activities are a direct consequence of these agricultural seasons. During the harvest period *Kewie* (October to February), households typically generate relatively high income streams due to irregular sales of their goods at local markets, wage work as harvest helpers and a general uptake in all other economic activities. The subsequent season associated with little agricultural activities (*Belg*, March to June) typically entails several religious festivities and weddings, for which households are socially expected to make non-negligible expenditures. Agricultural investments, such as fertilizer or seed acquisition, usually tend to take place in *Kiremti* (July to September), the planting and cultivation season. Obviously, the lag between income generation and investment causes difficulties for our study population.

Our partner organization, a local microfinance institution (MFI), has long tried to devise strategies to increase savings and investment among their client population. Smallholder farmers in our regions of Tigray usually have access to several savings technologies: cash savings in their home, savings with the MFI, or savings in informal savings arrangements, most importantly *Equb* (savings society) and *Iddir* (funeral society). Only roughly 5% of the study population has a formal bank account, and in the subsequent paper we will refer to “bank savings" as savings in the local MFI. Savings at the local microfinance institution, however, also incur sizable transaction costs, since clients usually need to travel to the next larger-sized village. In our sample mean travel time to the next MFI branch is slightly above 60 minutes. Especially during harvest season, it may be difficult for farmers to find this additional time.

Apart from savings, however, farmers in Northern Ethiopia do have a considerable demand for further microfinance products. Almost 90% of our sample has taken out a loan from the MFI at some point in time. The most frequent months for these loans are June and July, right before cultivation begins and agricultural investments need to be made. Due to this, the MFI has often undergone attempts to promote savings behavior among the local population. We chose the introduction of our intervention and provision of savings technology to coincide with the seasonality patterns of agriculture and income generation. Therefore, we conducted the intervention in December, the middle of the harvest season *Kewie*, where most households are selling crops or otherwise in the middle of their income generating activities.

1.2.2 Experimental Design

We visited all sampled households and asked to talk to the household head along with a further adult household member (generally the spouse). After a few survey questions, we started with a discussion on the general importance of savings and proceeded to prepare a detailed and individualized savings plan with the respondents. Our treatment was then composed of two elements: First, we offered smallholder farmers a new savings technology, namely moneyboxes with individual savings plans. Secondly, we provided individualized savings feedbacks.

Moneyboxes with Savings Plans. An emerging consensus in the literature states that savers have a demand for commitment and that softer commitments may be more effective in inducing behavioral change with respect to savings than harder commitments (Karlan and Linden, 2017; John, 2017). Thus, we choose a soft commitment - a moneybox provided to a subset of randomly chosen farmers - to leverage savings. Moneyboxes induce a certain amount of commitment since the cash inside the box is earmarked for a certain goal and using it for something else may induce unease for the owner. However, as opposed to savings in banks, MFIs, or commitment savings accounts, the money held inside moneyboxes is available at all times in case of emergency and does not entail any constraints on the individual's future choice sets.[5] Moreover, the households in our sample dwell in remote areas. Traveling to the next microfinance branch is time-consuming and visits in the village by the branches' savings officers are infrequent. Using a moneybox allows for saving at high, even daily, frequencies at virtually no transaction costs. This is especially important for the detailed saving plans designed together with the participants.

On every moneybox we distributed we also fixed an individualized savings plan. To do so, we asked respondents to formulate and talk about the most important savings goal they would like to save for. Examples for these goals include livestock (cows, goats), school books, and fertilizer. We proceeded by asking respondents how much they want to save for this goal and in how much time they want to accumulate this amount. In order to guarantee comparability, we had preset the timeline to be between 8 and 24 week.[6] For those households randomly selected to the pure control group not receiving a moneybox, the savings discussion ended at this point and was followed by a few more survey questions.[7]

[5] The moneyboxes we employed are round cylindrical plastic boxes with a slit at the top to facilitate the introduction of cash. The bottom of the moneybox has a hole with a small lock. Two keys were handed out to each household.

[6] If the costs of the goal (say, a cow) exceed the amount an individual believes to be able to save in 24 weeks, we asked that person to state the amount of money they would like to save during those 24 weeks towards their goal.

[7] We wanted to ensure that also households not receiving the moneybox were given the same general discussion about savings as treatment households. 300 individuals belong to this pure control group.

Feedback to Reconsider Original Savings Goal. After the respondents stated the initial goal amount they decided to save for, a subset of farmers was asked to reconsider this amount. The enumerator read out one of the following recommendations (randomized assignment):

I. "Our experience shows that people are more likely to reach their savings goal if they have higher goal amounts. Do you want to increase the amount of your savings goal to... [*initial goal amount* $\times$ *1.4*]?"

II. "Our experience shows that people are more likely to reach their savings goal if they have higher goal amounts. Do you want to increase the amount of your savings goal to... [*initial goal amount* $\times$ *1.2*]?"

III. "Our experience shows that people are more likely to reach their savings goal if they have lower goal amounts. Do you want to decrease the amount of your savings goal to... [*initial goal amount* $\times$ *0.8*]?"

IV. "Our experience shows that people are more likely to reach their savings goal if they have lower goal amounts. Do you want to decrease the amount of your savings goal to... [*initial goal amount* $\times$ *0.6*]?"

The control group receives a confirmation for their initially stated savings goal. In this case the interviewers read out the following statement: "Our experience shows that people are likely to reach their savings goal."

Following these recommendations, we ask participants whether they would like to revise their originally stated goal amount and what this new goal amount should be. Given this possibly revised savings amount, the enumerator calculates the regular (daily or weekly) savings installment necessary to reach the goal by the self-set deadline.[8]

At the end of the interview we put 30 *Birr* into the moneybox.[9] We asked participants to write down the following information on the label of the box: (1) the savings goal (written or drawn, e.g. cow), (2) the possibly revised goal amount in *Birr*, (3) the regular savings installment in *Birr*, and (4) the savings end date. At the end of the visit, the interviewers let households know that they would be visited again at an unspecified future date to check on their progress towards reaching the savings goal.

Sampling and Randomization. We sampled a total of 940 households in 94 village clusters. The village clusters basically represented a census of the zone in which the study was

[8] Whether the installments were daily or weekly was also randomized. [9] 30 *Birr* $\approx$ 1.5 USD. The pure control group also received the payment.

conducted. In each village cluster, 10 households were randomly chosen from lists of current and former clients at our partner MFI.

The sample of 940 households was split into random subsamples as follows: 640 households received a moneybox and 300 households belong to the pure control group. Among the 640 households receiving the moneybox, 128 were randomized to one of the four recommendation treatments. Randomization was done at the individual household level. [10]

For the randomization, we stratified the different blocks according to a range of baseline characteristics. These included important financial and economic measures, specifically savings in cash (amount in *Birr*), current savings with the local MFI (amount in *Birr*), whether household is a member of an *Iddir*, an informal funeral insurance, or *Equb*, an informal savings group (both binary indicators), the total land area under cultivation (in hectare), the total revenues from crop sales (in *Birr*) and the per-capita consumption expenditures (in *Birr*). Furthermore, we considered the demographic composition of the household (measured by the number of household members between 0-5, 6-11, 11-17, 18-65 and more than 65 years of age), whether the household head is female or married (both binary indicators) and the years of education of the head. We chose this combination of economic and socio-demographic variables to reflect those aspects of our baseline data which we believe are most important in determining savings outcomes and for which we sought balance.

Randomization was conducted in December 2014 before treatment implementation and after collection of the first baseline survey. Household-level treatment status was randomly assigned while balancing the Mahalanobis distances in the above-mentioned baseline variables used for stratification. The 940 households were assigned to either pure control group or one of the treatment arms. Thereafter the Mahalanobis distance with respect to the baseline covariates given above was calculated. Mahalanobis distance was calculated pairwise between any two treatment arms and also for each treatment arm relative to control. The random assignment was only accepted if the maximum distance to control was smaller than 0.605 and the maximum distance between any of the treatment arms was smaller than 1.322. These numbers corresponded to the 0.01 percentile of the empirical distribution of these distances. The treatments in the study were assigned according to the first random assignment vector that passed the balance thresholds.

[10] We also cross-randomized two further treatments with 320 households in each group: the frequency of the savings installment (daily or weekly) and the transparency of the moneybox. Results are available upon request.

1.3 Data and Background

1.3.1 Timeline, Data and Attrition

We use data from three rounds of comprehensive household-level surveys. An extensive baseline survey was administered in November/December 2013 on all sample households.[11] The implementation of the treatments was combined with a short survey to allow for a clean experimental procedure and to enable gathering further data on key baseline values. This setup effectively gives us two baseline surveys. For important variables available at both points in time, we can average them across both waves to create cleaner measures less prone to measurement error.

We administered the endline survey in January/February 2015 right at the end of the harvest season and the beginning of the subsequent season with little agricultural activity for two main reasons. First, this point marks the end of the high-income period for agricultural activity and is the ideal timing to measure the stock of savings households were able to accumulate during this period of relative prosperity. Second, in the design of the experiment, we asked households to save towards self-set saving goals within the next two to six months. The unannounced endline survey was at most two months after treatment implementation. This allows us to check whether households are on track towards reaching their savings goal right before the first households are scheduled to reach their goal date.

Our baseline and endline surveys encompass a large range of economically and behaviorally relevant variables and measures. We put a special focus into gathering comprehensive savings data. Given that savings (especially non-bank savings) are typically difficult to observe and measure correctly, we exploit the fact that the moneyboxes let us accurately observe how much cash is inside them at at a given point in time, thereby reducing measurement error compared to conventional survey questions considerably. We consider total monetary savings, which are savings in the moneybox (opened and directly counted by our enumerators), *plus* savings in cash the household head holds during the interview. In most cases, enumerators also physically counted further cash holdings together with the respondents during the interview.[12] Therefore, our measure of monetary savings is based on observations without measurement error and does not, as is often the case, represent self-reported

[11] We timed the savings intervention in December 2014 to coincide with the middle of the harvest season, when our study population enjoys high levels of income and has the opportunity to save (refer to Section 1.2.1 for details).

[12] This was achieved by asking respondents to report the composition of their further cash holdings in bills and coins, which typically led to enumerators counting the money together with respondents and thereby considerably reduce measurement error and ensure truthful reporting.

savings.

Moreover, to assess possible crowing-out behavior, we cover a range of non-monetary saving measures, including savings at the local MFI ("bank")[13] and contributions to informal savings arrangements, specifically *Iddir* (funeral society) and *Equb* (savings society). These measures will be introduced and discussed in more detail in Section 1.5 on the robustness of our results.

A key aspect of our investigation is associated with our measures of confidence. We assess whether individuals are (1) overconfident, (2) underconfident or (3) correctly confident using the following standard World Bank financial literacy questions:

I. Let's assume that you deposited 1000 *Birr* in a bank account at 5 percent annual interest rate. How much money will you have in your account in 2 years if you do not withdraw from or add to this account any money?

II. Let's assume that in 2015 your income is twice as now, and the food prices also grow twofold. Do you think that in 2015 you will be able to buy more, less, or the same amount of goods and services as today?

III. Let's assume that you saw a mobile phone of the same model offered from two different sellers. The initial retail price of it was *Birr* 1000. One seller offered a discount of *Birr* 150, while the other one offered a 10 percent discount. Which one is a better bargain – a discount of *Birr* 150 or 10 percent?

Every question has three possible answers but only one of them is correct. After individuals answer these questions we ask how many questions they believe they have answered correctly.[14] We classify as underconfident those individuals who believe to answer less of the financial literacy questions correctly than they actually did. Conversely, an individual that overestimates the amount of correct answers is classified as overconfident. Individuals that neither over- nor underestimate their performance are classified as correctly confident. Moreover, we ask individuals how many questions they believe all other people in their village would answer correctly on average.

Sample Size and Attrition The study draws on three waves of extensive surveys with relatively low levels of attrition. In our endline survey, we were able to re-locate all individuals

[13] To reduce measurement error, we also ask respondents to show us their MFI savings passbooks so that the enumerators can copy the last entries.

[14] The exact wording of the question is "What do you think: how many of the last 3 questions did you answer correctly?"

who received a moneybox and thereby have an attrition rate of 0% between treatment and the measurement of our results (endline). However, there was a minor amount of attrition between the first baseline survey and the implementation of the treatment. Out of the 640 households from our baseline randomly assigned to receive a moneybox, we were able to locate 614 for the treatment implementation, which translates to a modest attrition rate of about 4%. After deleting households with missing information in key outcome variables, the final sample throughout our main analysis amounts to 599 households.[15]

1.3.2 Descriptive Statistics

This section starts by describing key features of our data with a special focus on comparing individuals according to their confidence levels. We continue by presenting detailed balance statistics for our main experiment.

Table 1.1 reports summary statistics (means and standard deviations) on a range of variables for the full sample, as well as the subsamples classified as underconfident, overconfident and correctly confident (neither over- nor underconfident). Out of 599 individuals in the full sample, 16 percent are characterized as underconfident (94 individuals), 47 percent as overconfident (283 individuals) and the remaining 37 percent are correctly confident (222 individuals). The majority of household heads are married (74 percent) and 22 percent of households are led by women. Household members are on average relatively young, with the average household having more members below 18 years of age (3.05 individuals on average) than above 18 years (2.6). The household head has on average 3 years of education. Comparing the characteristics across the three types of confidence levels, we find that underconfident individuals are as likely to live in a female-headed household as others. Their household head is also as likely to be married. Yet, at baseline underconfident individuals do seem to differ to overconfident individuals in a number of characteristics: the household head has about one year less in education, land size is slightly bigger, debt is considerably smaller, and revenue from business activities is smaller.

The variation in baseline savings is big. However, this is partly due to possible measurement error in our reported savings variables, which is eliminated in our endline survey due to our improved measurement process based on direct observations. The baseline average of cash savings is about 4430 *Birr*.[16] However, our baseline savings measures show striking

[15] The complete baseline (considering also individuals without a moneybox) consists of 940 individuals, 905 of which were revisited for the second baseline coupled with treatment implementation. At the endline we were able to find and interview 899 individuals. Attrition rates are also small in the full sample, below 4% between first baseline and treatment implementation and less than 0.7% between treatment implementation and endline.

[16] 1 USD $\approx$ 20 *Birr*.

Table 1.1 – Summary Statistics

	(I) All		(II) Underconfident		(III) Overconfident		(IV) Correctly Confident	
	(1) Mean	(2) SD	(3) Mean	(4) SD	(5) Mean	(6) SD	(7) Mean	(8) SD
Baseline								
Female household head	0.22	0.41	0.23	0.43	0.20	0.40	0.22	0.42
Married household head	0.74	0.44	0.73	0.44	0.77	0.42	0.70	0.46
Education household head	3.28	3.75	2.35	3.37	3.57	3.63	3.30	3.99
Present-biased	0.28	0.45	0.32	0.47	0.29	0.45	0.26	0.44
Risk-loving	0.49	0.50	0.55	0.50	0.46	0.50	0.50	0.50
Cash savings	4429.93	12407.02	2816.49	3605.52	5632.93	15975.10	3579.55	9040.21
Bank savings	2289.85	9414.75	1149.94	1707.37	3174.63	12581.11	1644.63	5888.07
Iddir member	0.44	0.50	0.56	0.50	0.39	0.49	0.45	0.50
Equb member	0.30	0.46	0.29	0.45	0.32	0.47	0.27	0.45
Land size	4.58	5.49	5.45	8.31	4.37	4.88	4.47	4.67
Revenue	3951.39	12978.67	2938.40	4113.87	5186.33	18163.36	2806.04	4898.77
Debt	2974.74	4461.47	1923.40	3110.10	3479.88	4863.12	2775.97	4333.01
Treatments & Confidence Measures								
Positive recommendation	0.41	0.49	0.39	0.49	0.39	0.49	0.44	0.50
Negative recommendation	0.40	0.49	0.38	0.49	0.40	0.49	0.40	0.49
Initial goal amount (*Birr*)	1975.75	5234.63	1593.62	3778.63	1799.22	4520.26	2366.09	6476.24
Revised goal amount (*Birr*)	1988.40	5251.24	1650.04	3893.76	1776.68	4466.38	2403.43	6524.62
Investment goal	0.73	0.45	0.71	0.45	0.71	0.46	0.76	0.43
Subjective probability to reach goal	82.89	20.59	80.85	21.49	82.17	20.45	84.67	20.33
No perceived problem in reaching goal	0.45	0.50	0.34	0.48	0.46	0.50	0.47	0.50
Financial literacy score	1.57	0.92	2.47	0.62	1.17	0.75	1.68	0.93
Expected financial literacy score	1.99	0.90	1.32	0.59	2.46	0.70	1.68	0.93
Underconfident	0.16	0.36						
Overconfident	0.47	0.50						
Endline								
Does not possess moneybox anymore	0.03	0.17	0.02	0.15	0.04	0.19	0.02	0.13
Total amount in moneybox	239.01	468.03	175.17	240.91	270.59	567.78	226.54	395.49
No money in moneybox	0.04	0.19	0.03	0.18	0.03	0.17	0.05	0.23
Took money out	0.16	0.37	0.15	0.36	0.13	0.34	0.20	0.40
Amount taken out last time	462.31	1810.20	138.44	184.41	365.52	1159.98	651.02	2448.84
N	599		94		283		222	

Note: Column (1) reports means for all individuals in the sample, followed by the standard errors (SD) in column (2). The remaining columns follow the same pattern for underconfident, overconfident, and correctly confident individuals respectively. All monetary variables are given in local currency *Birr*, exchange rate at time of treatment (Dec 2014) was 1 USD $\approx$ 20 *Birr*.

differences according to an individual's confidence level: underconfident individuals save 2816 *Birr* in cash, while overconfident ones save twice as much (5633 *Birr*). Bank savings are at considerably lower overall levels (2290 *Birr* on average) and are again smaller for underconfident individuals (1150 *Birr*) than for overconfident individuals (3175 *Birr*). Furthermore, underconfident individuals seem to save slightly more in traditional saving groups (*Iddir* membership 12 percentage points above sample average). Our baseline measures for time and risk preferences (present-biasedness and risk-lovingness) show no systematic differences between individuals with varying levels of confidence.

Table 1.1 further reports summary statistics of key factors related to the introduction of the treatments and the simultaneous measurement of our confidence measures. The self-set savings plans and individual responses to the recommendations strongly differ between under- and overconfident individuals. Underconfident individuals set lower savings goals and are more likely to revise them. While the average original goal amount in the sample is 1976 *Birr*, the corresponding amount for underconfident individuals is about 19 percent lower. After being encouraged by interviewers to reconsider their savings goal, underconfident individuals change their goals upwards by 56 *Birr* on average. This is especially remarkable considering the fact that respondents received recommendations to increase and decrease saving goals in equal proportions. Overconfident individuals, on the contrary, decreased their saving goals by 23 *Birr* on average. Most respondents report saving towards some form of investment goal (73 percent).

Additionally, we collected information on individuals' beliefs about their ability to save and their expected obstacles. 66 percent of the underconfident respondents name specific obstacles in reaching their goals, whereas almost half of the overconfident individuals do not expect any problems in reaching their goal. Likewise, underconfident individuals report slightly lower probabilities in reaching their goal (on average 81 percent compared to 83 percent for overconfident individuals). The financial literacy score is the number of correctly answered questions (out of three) of the World Bank survey module and the expected financial literacy score are the number of questions respondents believe to answer correctly. These scores show the drivers behind the classification of individuals into different groups based on their confidence levels. It is remarkable that while underconfident individuals believe to have lower financial literacy scores than overconfident respondents, in reality their scores are higher than those of individuals with high confidence levels.

In our endline survey, we document that moneyboxes have been a useful and frequently employed savings technology. Only 3 percent of our sample did not possess the box at the endline survey, mostly because they were given away as presents or were damaged. Nearly

all individuals used the moneybox for saving purposes, and only 4 percent of the respondents had no money in the box. Every sixth person had taken out money at least once. On average individuals reported to have taken 462 *Birr* out of their savings device. Interestingly, underconfident individuals, although their probability of having taken out money is comparable to overconfident individuals, reported taking out lower amounts (only 184 *Birr*). Given the small amount of time respondents had access to this new savings devise, they used it considerably. The average amount in the moneybox was 239 *Birr*, with underconfident respondents having lower amounts (just like with overall savings) than overconfident individuals.

Table 1.2 reports the summary statistics subdivided by treatment group. Column (1) describes the characteristics of the treatment group while column (2) does the same for the control group. Column (3) presents the difference and its standard deviation. We present the variables on which we randomized and our central measure of confidence. All variables are well balanced with the difference in means being insignificant in all but one case. Only participation in the traditional savings group *Equb* manifests a statistically significant difference at the 10 percent level. Moreover, in a test for joint significance of all baseline variables (all those reported in Table 1.2 with the expectation of the confidence levels), we cannot reject the null that these baseline values do not explain treatment status (p-value = 0.8).

While we did not randomize with respect to the individuals confidence levels, we observe that the share of those who received a recommendation and those who did not is comparable across both groups. 15 percent of those who received an encouragement were underconfident, while for the group of those who did not receive one the share is 18 percent. The difference of 2.8 percent is statistically insignificant. Similarly, 46 percent of those who received a recommendation were overconfident. Among those who did not receive a recommendation, 51 percent were classified as overconfident. The difference between these two groups is statistically insignificant as well.

1.4 Results

In this section we present the main results of the paper, including the effects of our recommendation treatments and the behavioral factors driving the results. To estimate the effects of the treatments on our outcomes of interest, we use standard methods from the analysis of randomized control trials and estimate OLS regressions with treatment indicators and baseline controls. Our preferred regression specification is

Table 1.2 – Balance Statistics

	(1) Recommendation	(2) No Recommendation	(3) Difference
Cash savings	4171.175	5495.923	-1324.748
	(11652.667)	(5135.869)	(1278.609)
Bank savings	2160.774	2821.607	-660.833
	(8153.065)	(13451.993)	(970.735)
Iddir member	0.434	0.462	-0.028
	(0.496)	(0.501)	(0.051)
Equb member	0.280	0.368	-0.087*
	(0.450)	(0.484)	(0.047)
Household members aged 0-5	0.873	0.752	0.121
	(0.778)	(0.765)	(0.080)
Household members aged 6-11	1.089	1.034	0.055
	(0.890)	(0.830)	(0.091)
Household members aged 11-17	1.098	1.214	-0.116
	(0.919)	(0.954)	(0.095)
Household members aged 18-64	2.541	2.573	-0.031
	(1.043)	(1.037)	(0.107)
Household members aged 65+	0.046	0.060	-0.014
	(0.219)	(0.238)	(0.023)
Female household head	0.216	0.214	0.002
	(0.412)	(0.412)	(0.042)
Married household head	0.730	0.761	-0.030
	(0.444)	(0.429)	(0.045)
Eduaction household head	3.259	3.368	-0.108
	(3.765)	(3.697)	(0.387)
Land size baseline	4.495	4.917	-0.422
	(4.970)	(7.282)	(0.566)
Revenue baseline	4023.166	3655.684	367.482
	(14075.977)	(6832.953)	(1338.638)
Debt baseline	2932.512	3148.718	-216.206
	(4448.389)	(4530.065)	(460.106)
Consumption expenditure baseline	747.178	693.190	53.987
	(973.479)	(1010.565)	(101.082)
Underconfident	0.151	0.179	-0.028
	(0.359)	(0.385)	(0.038)
Overconfident	0.463	0.513	-0.050
	(0.499)	(0.502)	(0.051)
F-Test			0.700
p-value			0.804

Note: N=599. Standard deviations in parenthesis. Columns (1) and (2) denote the average value of the relevant variable depending on treatment status. Column (3) denotes the difference between treatment and control, its significance is given as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All variables (except the last two) are measured at baseline before randomization. Underconfidence and overconfidence are measured as described in Section 1.3. For a test of joint significance, we estimate a regression with a binary indicator for treatment status on the left-hand side and all the balance variables used for stratification during randomization in the table on the right-hand side. The F-statistic and corresponding p-value for the null of joint zero effects are denoted in the last two rows.

$$Y_i = \alpha + \beta \mathbf{T}_i + \gamma \mathbf{X}_{i0} + \delta_1 Y_{i0} + \delta_2 lit_i + \zeta \mathbf{I}_i + \epsilon_{it}, \tag{1.1}$$

with Y_i being the outcome of interest of household i at time t and $\mathbf{T}_i$ the vector of treatment dummies. $\mathbf{X}_{i0}$ represents the vector of baseline control variables: savings in cash and at the local MFI, membership in the informal savings societies Iddir and Equb, land area, revenue from crop sales, per-capita consumption expenditures, number of household members by age, gender, marital status and education of the household head (please refer to Section 1.2.2 for details). We further control for the baseline values of the outcome variables Y_{i0} in order to improve precision.[17] We explicitly control for lit_i, an indicator for above average financial literacy score, in order to rule out any possible mechanical effects of financial literacy on confidence. Although the enumerators followed a detailed protocol and received intensive training, we include a vector of enumerator dummies $\mathbf{I}_i$ at treatment to control for any possible enumerator-specific effects in administrating the treatments. ϵ_{it} denotes the remaining error term. Throughout the rest of this section, we will keep this estimation framework and vary the definition of the treatment vector $\mathbf{T}_i$. We will also introduce several interactions along behavioral dimensions, especially for the individual degrees of confidence.

In understanding these results, it is crucial to first note that the moneyboxes themselves are a very effective method to increase savings among smallholder farmers in Northern Ethiopia. When comparing those (randomly selected) individuals who received a moneybox to those in the pure control group without a moneybox, savings increased substantially. Cash savings plus savings in the moneybox increased by about 117 *Birr* (22 percent increase compared to control group). Even a broader measure of savings, including savings in the bank (local MFI) plus savings in the informal arrangements *Iddir* (savings club) and *Equb* (funeral society), is larger for individuals who received a moneybox, with a significant increase of about 0.6 log points. These results stem from simple OLS regressions following our standard estimation approach set out in equation (1.1) and are presented in detail in Table A.1 in the Appendix. These results are in line with a large literature documenting the effectiveness of moneyboxes in increasing savings (among others Dupas and Robinson 2013). The moneyboxes thus helped to initiate larger savings and further permit us to measure cash savings without measurement error (through direct counting of the money in the box). The rest of our paper goes beyond this observation and focuses only on the 599 individuals who received a moneybox and analyzes the effects of the main behavioral treatments among this

[17] Our small survey just before treatment implementation effectively provides us with two baseline surveys on a select number of measures. For these measures we take the average between both pre-treatment values to reduce volatility and measurement error.

subsample.

1.4.1 Savings and Confidence

This section provides evidence on the importance of the link between confidence levels and savings behavior. For our study population, confidence levels are an independent factor predicting the level of savings individuals generate even *before* the implementation of any behavioral treatments.

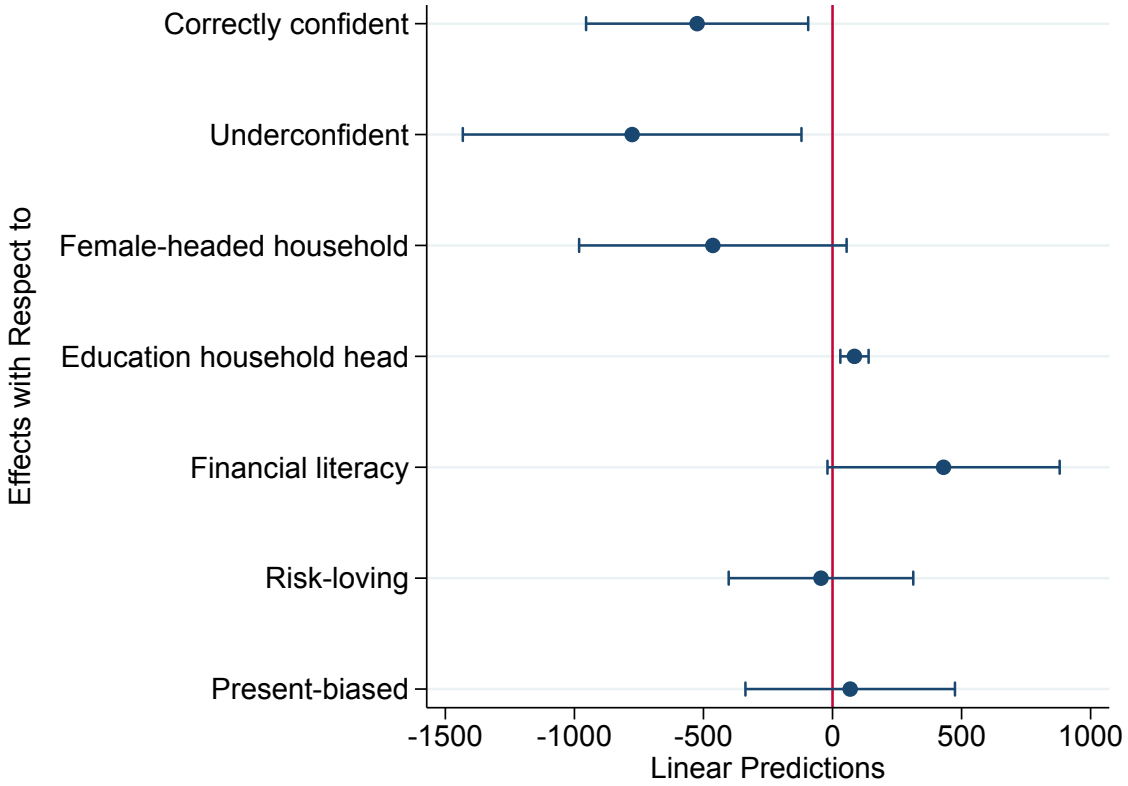
We construct a comprehensive measure of baseline savings as the sum of cash holdings, bank savings, and money in informal saving arrangements. To this end, we run a simple OLS regression with our measure of baseline savings as the outcome and indicators for our central behavioral measure of underconfident and correctly confident individuals. We control for important socio-demographic observables and behavioral traits, including whether a household is female-headed, the years of education of the household head, our score for financial literacy, and indicators for risk-loving and present-biased individuals. Additionally, we also control for a range of socio-demographic and economically relevant variables: indicators for the age structure of the household, marital status of the household head, land ownings, revenues, outstanding debt, per capita consumption expenditures, and lastly enumerator indicators. We are interesting in quantifying the relative importance of various individual characteristics and behavioral traits in explaining savings behavior. Thus, we compare confidence to other key characteristics introduced above. Figure 1.1 shows the coefficients from this OLS regression along with 95% confidence intervals comparing confidence to an indicator for female-headed households, the education of the household head, financial literacy, risk and time preferences[18].

The figure clearly depicts the strong association between the importance of confidence levels and savings behavior. Being underconfident is a statistically significant predictor for holding less savings at baseline, even when controlling for a range of socio-demographic and economic variables. Underconfident individuals save more than one third less than overconfident individuals. Interestingly, the confidence levels are more important in predicting savings behavior than other behavioral traits frequently associated with financial outcomes such as risk-lovingness and present-biasedness. The coefficients of the indicator for financial literacy and our measure for years of education of the household head are positively associated with saving levels, as would be expected. We interpret these results as underlining the importance of confidence measures in being strongly associated with financial

[18] Risk and time preferences were incentivized measures we discuss further in Section 1.5

outcomes and will investigate the effectiveness of our behavioral treatments especially with respect to our measure of confidence.

Figure 1.1 – Savings Behavior Before Treatment



Note: N=599. This figure depicts the coefficients from a simple OLS regression with a broad measure of savings at baseline (cash + bank + informal arrangements) as the dependent variable. Additional control variables include all baseline variables used for stratification during randomization excluding savings outcomes (socio-demographics, land and consumption variables) as well as enumerator indicators. The savings measure is taken *before* treatment implementation. The variables depicted in the figure are measured as indicator variables with the exception of education, which is measured in years of schooling. 95% confidence intervals are depicted around the point estimates.

1.4.2 Savings Recommendations

Recommendations. We start by analyzing the overall effect of providing smallholder farmers with individually-tailored feedback in the form of recommendations on their self-set saving goals. To this end we compare those who received a recommendation to reconsider the originally stated savings goal to those who did not receive such a recommendation.[19] Table 1.3 reports these main results. The outcome variable measures savings in the moneybox

[19] As individuals who received a recommendation we combine all those with the treatments I–IV presented in Section 1.2.2

plus cash savings the household head carries at the time of the interview.[20] The estimation strategy follows the exact setting laid out in equation (1.1). Panel A of Table 1.3 shows that receiving a recommendation leads to an increase in savings of 181.5 *Birr*, statistically significant at the 5 percent level. This corresponds to a sizable increase of 36 percent compared to the control group (receiving no recommendation).

The sizable increase in savings following a recommendation shows concisely that a simple message motivating smallholder farmers to reconsider their original savings goal leads these individuals to save more.[21] Even though there could be a multitude of behavioral and cognitive factors at play, we rationalize this main result as follows: The recommendations the enumerators give effectively provide instantaneous and individually-tailored feedback to the respondents. This individualized feedback loop lends importance to the fact that the original savings plans might benefit from a reassessment. The additional reflection may lead farmers to take further relevant factors into account when deciding on the amount they wish to save for and this may in turn increase the likelihood of reaching the goal. The visibility of the moneybox and the revised goal depicted on it further enforces the updated individualized savings plan.

Confidence. We observe strong heterogeneity in the positive effect of our recommendations on individual savings behavior along behavioral characteristics. Motivated by our results in Section 1.4.1 documenting the general importance of confidence levels in determining baseline savings behavior, we are especially interested in whether underconfident individuals, who save less at baseline, may be differentially affected by the recommendations. Panel B of Table 1.3 introduces the indicator variable underconfidence (refer to Section 1.3 for details on the measurement of behavioral characteristics) and its interaction with the recommendation treatment. Just as at baseline, underconfidence leads individuals to save significantly less (on average 440 *Birr* less). Interestingly, the recommendation treatment is especially helpful in increasing savings for underconfident individuals. The estimate of the interaction term between both corresponds to an increase of 409 *Birr*. We explain this by noting that underconfident individuals may need additional encouragement to reach their self-set goals. Our control group also receives a generic feedback in the form of a statement that in general people are likely to reach their savings goal. However, it seems that the individualized component of the recommendations and the subsequent revision of the

[20] As detailed in Section 1.3, this measure is extremely resistant against measurement error since our enumerators open the moneybox to count the money inside and ask respondents to detail their additional cash savings by naming the exact composition in bills and coins.

[21] It is important to note that this happens irrespective of the direction the recommendations take: in some cases we asked respondents to save more, in other cases we asked them to save less. The paper provides more details on these directions and mechanisms in further sections.

Table 1.3 – Effect of Recommendations on Savings

	(1) Mean Control Group	(2) Recommendation	(3) Financial Literacy	(4) Recommendation × Underconfidence	(5) Underconfidence
Panel A: Recommendations					
Savings	496.009	181.463**	-6.082		
		(80.733)	(76.722)		
Panel B: Interaction Underconfidence					
Savings	496.009	106.141	23.315	409.802**	-440.383***
		(92.405)	(83.985)	(162.633)	(138.329)

Note: N=599. The dependent variable savings is calculated as the value of cash savings the household head carries at the time of the interview plus savings in the moneybox. Panel A reports results ignoring the effect of underconfidence. Panel B introduces the effect of underconfidence, measured by an indicator equal to 1 for individuals believing to answer fewer questions on financial literacy correctly than they actually do. We interact underconfidence with the recommendations. Mean refers to the average value of the outcome variable in the control group (no recommendation). All variables are winsorized at the 95% percentile. Additional control variables include all baseline variables used for stratification during randomization (savings in cash, bank and informal arrangements, socio-demographics, land and consumption variables) as well as enumerator indicators. Robust standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

self-set savings goals matters especially for underconfident individuals. Throughout these estimations, financial literacy has no independent effect on savings, thereby ruling out any mechanical effects accruing to the way we measure confidence levels. Interestingly, we replicate the results using an interaction term for overconfidence and find that the individualized recommendations have no differential impact on savings for these individuals (results available upon request).

1.4.3 Content of Individual Feedback

Direction and Intensity of Recommendation. The previous analysis has shown a clear causal increase in savings induced by the recommendations, and this effect is especially strong for underconfident individuals. In the following we further explore the effects of the contents of the recommendations. To do so, we look at a set of intermediate outcomes. In essence the following analysis allows us to test whether the messages were well delivered by the enumerators, understood by the farmers, and consequently actually triggered a change in savings goals in the recommended direction.

Table 1.4 presents these intermediate results and addresses whether individuals followed

Table 1.4 – Intermediate Results

	(1) Revised − Initial Goal	(2) Revised − Initial Goal (% of Initial)	(3) Share of HH Changed Goal	(4) Share of HH with Recommended Change
		Panel A: Direction of Recommendation		
Positive recommendation	218.969***	0.102***	0.286***	0.346***
	(53.338)	(0.034)	(0.038)	(0.033)
Negative recommendation	-220.270**	-0.037	0.180***	0.142***
	(100.622)	(0.032)	(0.038)	(0.028)
N	597	597	597	597
		Panel B: Intensity of Recommendation		
Positive recommendation + 40%	248.721***	0.105***	0.257***	0.318***
	(81.470)	(0.029)	(0.049)	(0.045)
Positive recommendation + 20%	195.715***	0.100**	0.312***	0.372***
	(67.020)	(0.051)	(0.047)	(0.042)
Negative recommendation – 20%	-116.177	-0.006	0.189***	0.137***
	(129.461)	(0.041)	(0.049)	(0.036)
Negative recommendation – 40%	-324.837**	-0.067**	0.169***	0.145***
	(144.722)	(0.034)	(0.044)	(0.034)
N	597	597	597	597
		Panel C: Underconfident Individuals Only		
Positive recommendation	110.735*	0.067**	0.323***	0.323***
	(63.462)	(0.028)	(0.100)	(0.100)
Negative recommendation	-71.466	-0.034	0.135	0.135
	(75.656)	(0.028)	(0.112)	(0.112)
N	94	94	94	94

Note: This table shows the effect of the recommendations on intermediate outcomes. The outcome variable in column (1) is the difference between the revised and initial savings goal. In column (2) the outcome is this difference as a fraction of the initial goal amount, more specifically: (revised − initial goal)/initial goal. The dependent variable in column (3) is an indicator for changing the goal at all and in column (4) we take an indicator for changing the goal in the recommended direction. Panel A depicts results for recommendations grouped by their direction, Panel B disaggregates the recommendations by their intensity and Panel C focuses on the subsample of underconfident indivinduals. 2 observations are missing: one without data on the initial goal level, another without data on any goal level. In 24 cases we had missing data on the *revised goal* and assumed the *revised goal* to be equal to the *initial goal*. Further control variables include all baseline variables used for stratification in the randomization process (savings in cash, bank and informal arrangements, socio-demographics, land and consumption variables) as well as enumerator controls. Robust standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

the randomly assigned savings recommendations given to them. As outcome variables we construct various measures of the extent to which individuals revised their self-set goals.[22] In column (1) we look at the difference between the revised and initial savings goal. For example, an individual planned to save 1000 *Birr* for a cow, but we recommended saving 1400. The respondent could then change the goal in any way and the revised amount was noted down on the moneybox. Column (2) reports this change as a percentage of the original amount.[23] The outcome in column (3) is an indicator whether an individual has changed the goal at all and column (4) is an indicator whether an individual has changed the goal in the recommended direction.

For Panel A of Table 1.4 we group the four different types of recommendations (please refer to Section 1.2.2 for details) into "positive" and "negative" ones, i.e. messages that encourage to save more or less, respectively. As we see from the first column of Table 1.4, the direction of the recommendation clearly has an effect on how the goals are revised. While positive recommendations increase savings goals, negative recommendations induce respondents to decrease their goals. These differences are statistically significant and the magnitudes of positive and negative recommendations are almost identical and correspond to roughly an 11 percent change in the goal amount on average. Columns (2) - (4) support these findings. Both individuals receiving positive and negative recommendations change their goals, but it seems that the probability of revising the goal is slightly stronger for recommendations to save more.

Panel B analyzes the intensity of the recommended change, whereby we distinguish between recommended revisions of plus/minus 40 percent and plus/minus 20 percent. In general, stronger recommendations are accompanied by higher changes in goal amounts. For example, being recommended to save 40 percent more leads to an upward revision of the initial savings goal by 249 *Birr* compared to individuals who did not receive any recommendation. On the other hand, the recommendation to increase the savings goal by only 20 percent results in a slightly lower change of only 196 *Birr*. We see a similar pattern for the intensity of the negative recommendations. As before, we find some evidence that positive recommendations seem to be more effective at influencing the process of savings goal formulation.

Panel C focuses on the subsample of underconfident individuals, as their savings behavior seems to be particularly affected by the recommendations. At the stage of formulating

[22] Note, we lose two observations due to missing information on intermediate outcomes in at least one of the 4 regressions. [23] The variable is calculated as (revised amount - initial amount)/ initial amount.

their savings goals, underconfident individuals also respond to the recommendations in the suggested directions. Their increase in saving goals following a positive recommendation is comparable to that of the whole sample in relative terms, as the coefficient of 111 *Birr* corresponds to a 7 percent increase in their goal amount.[24] However, underconfident individuals respond significantly more strongly to positive and encouraging recommendations as compared to negative recommendations. While the effect of positive recommendations is significant in all specifications, the effect of the negative recommendations is insignificant throughout. This is in line with our expectation that farmers with low confidence levels set low goals and might be more likely to revise these upwards when prompted.

Savings Attainment. In this section we have so far described the effects of the content of the recommendations on the goal setting behavior. In Table 1.5, we now analyze actual savings achievement. Different to Table 1.3, here we distinguish explicitly between positive and negative recommendations. Table 1.5 provides similar results as before: the effect of the recommendations on savings behavior is driven by underconfident individuals. Interestingly, both positive as well as negative recommendations increase savings for this group. Moreover, the coefficients on both of the interactions between underconfidence and positive as well as negative recommendations are almost identical. We take this as evidence that although the recommendations have the intended effect on the formation of the savings goal (as documented in Table 1.4), the direction of the recommendation becomes irrelevant for the actual savings behavior. We therefore conclude that the effects of recommendations on savings are driven mainly through the additional attention individuals receive for their savings plans. This additional attention comes in the form of individualized feedback together with the opportunity to reflect and revise the savings goal. We conclude that through this additional attention, especially underconfident households develop a stronger attachment to their savings plan and hence were able to achieve higher saving levels.

1.5 Robustness

This section presents a range of robustness checks. A number of behavioral factors are possibly related to confidence levels and could partly explain some of the observed results. This section addresses such alternative mechanisms and furthermore rules out crowding-out behavior of savings into other savings vehicles.

[24] Underconfident individuals have lower initial goal amounts.

Table 1.5 – Direction of Recommendations and Underconfidence

	(1) Savings
Positive Recommendation	98.610
	(102.346)
Negative Recommendation	114.151
	(102.251)
Positive Recommendation x Underconfident	419.866**
	(187.637)
Negative Recommendation x Underconfident	400.443*
	(205.225)
Underconfident	-440.564***
	(138.597)
Financial literacy	23.459
	(84.606)
N	599
Mean	496.009

Note: The outcome variable savings is measured as cash savings plus savings in the moneybox. This table disaggregates the effect of recommendations on savings by their direction (positive or negative) and interacts these with our measure of underconfidence. Further control variables include all baseline variables used for stratification during randomization (savings in cash, bank and informal arrangements, sociodemographics, land and consumption variables) as well as enumerator controls. The last row gives the mean value of the outcome variable in the control group (no recommendation). Robust standard errors in parentheses, * p < 0.1, ** p < 0.05, *** p < 0.01.

Financial Literacy, Perceived Obstacles, and Preferences. Respondents in our sample might differ in their ability or willingness to anticipate obstacles hindering them from achieving their savings goal. This may in turn have implications for the effectiveness of our recommendation treatments and might be an alternative mechanism to general confidence levels. In our detailed survey data, we ask respondents to explicitly state any problems they anticipate in reaching their savings goal. About 55 percent of the sample stated one or several problems, the remaining 45 percent did not see any problem hindering them from reaching their savings goal. In Panel A of Table A.2 in the Appendix, we include an indicator for expecting at least one problem and interact this with our recommendation treatment. The recommendations did not work differentially for those individuals. We conclude that the ability or willingness to anticipate obstacles does not drive the observed results.

In Panel B of Table A.2 we take a closer look at financial literacy. Throughout our regressions we always controlled for financial literacy to eliminate any mechanical effects on

confidence levels as well as any direct effects on savings behavior. A remaining issue might be that the ability to properly apply the recommendations also differs by financial literacy. To test this hypothesis, we interact an indicator for having above average financial literacy with the recommendation treatment in Panel B. We find no statistically significant effect and conclude that financial literacy seems not to be an independent driver of our results.

Risk and Time Preferences. A further open question is whether the results are due to risk and time preferences. These personality traits have been shown to be highly correlated with financial decisions, including savings behavior.

In our baseline survey we elicit time and risk preferences of household heads. Risk preferences are elicited with a standard incentivized framework going back to Holt and Laury (2002). We present 7 distinct lotteries to participants, each consisting of two alternatives which will materialize with a 50% probability. The first lottery has a payout of 3 *Birr* in both states of the world. The following lotteries incrementally increase the expected payout but also the variance. The respondents are asked to choose which of these lotteries they would like to participate in.

In a further section of the baseline survey, we elicit time preferences. To this end, we ask participants whether they would like to receive a certain amount of money tomorrow or within one month.[25] In five subsequent questions, the payoff occurring in one month is incrementally increased. The switching point provides us with a measure for the underlying discount rate that an individual applies. In a later part of the survey, we repeat the elicitation with questions comparing payoffs in one year to those in one year and one month. If an individuals discounts values faster in the present than in one year, they are classified as present-biased.

Individuals were informed at the beginning of the survey that one of the questions would be randomly chosen and the payouts provided. Future payouts were provided through the branch office of the local MFI.

Panels C and D of Table A.2 show that in our setting present-biasedness and risk-lovingness cannot explain differences in the reaction to our feedback mechanism.

[25] By setting the time for receiving the first payment to the day after the survey instead of the day of the survey, we guarantee that individuals make the decision purely based on their time preferences and are not affected by other considerations like the perceived credibility of actually receiving the payment.

Crowding-Out Effects. We have shown that the recommendations increase savings, in particular for underconfident individuals. In doing so, we have measured savings as the sum of savings in the moneybox and cash savings. A key concern is crowding out of other forms of savings. If budget constraints are binding, an increase in one form of savings should lead to a reduction in other saving vehicles.[26] Table A.3 in the Appendix shows that we observe no crowding-out behavior. Column (1) shows the effects on bank savings (at the local MFI), column (2) and (3) holdings in the informal savings arrangements *Equb* (savings group) and *Iddir* (funeral society). In column (4) we use a "wide" definition of savings, defined as the sum of our previous measure (cash + moneybox) plus bank savings, *Equb* and *Iddir* holdings. Our recommendation treatment has no effect on all four alternative saving measures. Significant coefficients are mostly related to the baseline values of these saving measures. We conclude that we find no evidence for any crowding-out into alternative savings vehicles.

1.6 Conclusion

In this paper we study the link between confidence and saving behavior. Controlling for a set of key individual- and household-level characteristics we find that in particular underconfidence is strongly associated with lower savings. In fact, the relationship is stronger than for other important and well-studied determinants. This finding in itself is innovative and contributes to a growing literature that attempts to understand behavioral constraints to savings.

We analyze a large-scale randomized control trial in rural Ethiopia. In our experiment, we first encourage savings in a moneybox with individualized savings plans. We expected to nudge higher savings and initiate a habit formation. In fact, we are able to put forward additional supportive evidence for a well-established finding in the literature: Soft commitment devices with individualized savings plans increase savings by 22 percent compared to the control group without moneyboxes and savings plans. Taking a closer look at intermediate outcomes, we show that our soft commitment devices were frequently used and well-kept by our targeted population. Additionally, farmers were on track with their personal savings plans that recorded and even visualized the individualized savings goals on the boxes (overall final savings amount and goal targeted).

Originally hypothesizing that underconfident individuals might need an additional level of encouragement, we test a further set of hypotheses. We encourage a randomly selected

[26] Note, however, that Brune et al. (2016) observe the opposite: an increase in one type of saving increases the demand for other forms of savings.

subset of our individuals to save more than they originally planned. We mirror this approach by proposing a further random group of individuals to set lower, more realistic saving goals. Additionally, we vary the intensity of our encouraging and discouraging messages. Our approach seems to have convinced the farmers: The more encouraging our treatment message was, the higher the deviation from the original savings plan. The same is true for the discouraging messages. As expected, the individuals we classified as underconfident reacted in particular to the encouraging messages. The revised savings goal was noted down on the moneybox.

Revisiting the farmers a few months later, we make an interesting and rather unexpected observation: First, with the feedback mechanism we employed in our experiment, we successfully increase savings by 36 percent. This effect is especially strong among underconfident individuals, almost nullifying their behavioral constraint. Surprisingly, however, we find that the underconfident farmers reacted as strongly to the messages that encouraged to save more as to the messages that encouraged to opt for lower, but easier to achieve savings amounts. Thus, we believe that the observed change in saving behavior is a reaction to the additional attention and reflection upon the original savings goal. In essence, our treatment boils down to not only underlining the importance of savings, but also taking individual needs and wishes explicitly into account.

As such our results have immediate policy implications: Given the importance of savings for a range of development outcomes, it is important to provide households with the opportunity to save more. The feedback mechanism employed in our experiment is a cost-efficient and simple procedure, capable of doing so, without crowding out other forms of savings.

Moreover, following up on the observations we made in this study, we believe it would be worthwhile adjusting savings plans to the confidence levels of the savers. We conclude so given the robust descriptive evidence of differences in savings behavior. In contrast to the prior literature which mainly focuses on overconfident individuals, we believe that the group of underconfident savers is at least as important. Further research could explore the exact behavioral processes involved in the savings processes, which were not within the scope of this paper. In particular, could further personalization of saving messages in general trigger even greater savings? Or is an even greater feedback and reflection of original savings goals helpful? Finally, how can savers with varying confidence levels develop and reach their full savings potential?

Chapter 2

Information Frictions and Learning Dynamics: Evidence from Tax Avoidance in Ecuador[1]

with Jan Sebastian Nimczik

2.1 Introduction

Formalization of developing economies is a key policy goal. Informal employment eluding government control represents a large portion of economies in low and middle income countries, estimated to be almost 50 percent in Latin America (ILO, 2014). A primary barrier to formalization is the lack of information about the functioning of government programs according to survey evidence from the World Bank (Perry et al., 2007). While a growing literature looks into the determinants of formalization and its impact on key economic areas (Gerard and Gonzaga, 2016; Naritomi, 2016; Jensen, 2016; Pomeranz, 2015), little is known about the dynamic processes shaping the responses of economic agents that adapt to the formal system and try to learn about its incentives. This is particularly relevant in the context of behavioral responses to – often complicated – tax incentives. Previous work has extensively explored the role of adjustment frictions in constraining responses to the tax system (Chetty et al., 2011).[2] We are the first to thoroughly study the role of information frictions and how dynamic learning processes remove these obstacles. A number of studies have explored general spillovers between taxpayers (Chetty et al., 2013; Paetzold and Winner, 2016), but there is no clear consensus on how information frictions can be overcome.

[1] An earlier version of this chapter was circulated under the name "Learning Dynamics in Tax Bunching at the Kink: Evidence from Ecuador"

[2] Moreover, an emerging literature highlights the influence of behavioral biases on responses to tax incentives (Bhargava and Manoli, 2015; Taubinsky and Rees-Jones, forthcoming; Benzarti, 2017).

While a specific intervention teaching the tax code to EITC recipients in the US has proven to be rather ineffective (Chetty and Saez, 2013), we show that dynamic adjustments and learning processes lead to substantial changes in reported taxable income. Worker mobility is the most important driver of information transmission. We identify co-workers and accountants as specific channels of information transmission and show that information about tax adjustment opportunities spreads through top-down learning processes induced by job switches of managers and accountants.

We draw on novel administrative data on personal income tax (PIT) returns in Ecuador to assess how workers and firms learn about tax avoidance opportunities in a developing country. Ecuador's rapidly formalizing economy with a steady inflow of new workers and firms to the tax system provides a unique setting to study dynamic information flows between taxpayers. We make four main contributions: First, we document dynamic developments of individual tax avoidance. With increasing tenure in the formal sector, individuals are more likely to avoid paying taxes. Second, we exploit exogenous job mobility to show that the increase in tax avoidance is causally affected by changes in the information environment individuals face. Third, we show that knowledge about tax avoidance opportunities spreads across firms and document that firms with more experience in the formal sector are more likely to have employees who avoid tax payments. Fourth, we identify specific channels of information transmission: peers (co-workers) and experts (accountants). In particular, we show that the learning process within firms is driven by top-down information transmission. Incoming co-workers in the top decile of a firm's wage distribution have a lasting effect on the tax avoidance behavior of their new co-workers. Likewise, introducing a knowledgeable accountant into a firm increases the tax adjustment behavior of the firm's employees.

Tax avoidance in Ecuador is mostly achieved by filing deductions for personal expenses in housing, health, nutrition, education, and clothing. Generous deduction possibilities are one of the government's main policies to induce an increase in formalization stimulating the demand for formal receipts. Strikingly, however, many individuals do not capitalize on the deduction possibilities. Among those workers who could use the deductions to completely avoid paying taxes, 60 percent still pay some taxes (this share is decreasing over time and reaches just above 50 percent in 2015). 65 percent of those remaining taxpayers earn gross income in a range where they could even avoid paying taxes without actually having to hand in any receipts to the tax authority.[3] This low usage of easily accessible tax adjustment opportunities speaks to the presence of information frictions.

Our main measure of tax avoidance is the extent to which workers use deductions to lower their reported taxable income just below the income tax exemption threshold ("bunching"). We find a large and pronounced spike in the distribution of taxable income at the tax exemp-

[3] Only if the value of deductions exceeds a certain reporting threshold are taxpayers obliged to hand in the receipts to the tax authority. More details in Section 2.2.

tion threshold while the distribution of gross income (before using deductions) is smooth around all discontinuities in the marginal tax schedule. In extensive robustness analyses, we replicate all our results for alternative measures of tax avoidance without substantial change in the results.[4]

To document dynamic adjustments and learning processes, we begin by focusing on individual taxpayers' adaptation to the incentives of the formal sector. We estimate the prevalence of tax avoidance among cohorts of taxpayers by their year of entry into the formal sector. We find clear evidence of individual-level learning: across all cohorts, tax avoidance becomes stronger as individuals gain experience in the formal sector. We approximate the effect of experience through flexible polynomials and find strong initial increases in tax avoidance which level off after about five years in the formal system. We conclude that, with tenure in the formal sector, workers in Ecuador learn how to avoid paying taxes. The correlation between experience and avoidance remains strong and unchanged when controlling for a broad range of observable characteristics and unobserved heterogeneity.

However, it is unclear exactly *how* workers learn about the tax system. We provide causal evidence on how the information environment in firms drives individual learning processes. Exploiting the matched employer-employee component of our data, we identify asymmetric responses to exogenous changes in an individual's knowledge environment due to job transitions. Individuals moving into a firm with high levels of tax avoidance are more likely to avoid paying taxes themselves while individuals moving into a firm with low levels are just as likely to avoid paying taxes as before. These findings are robust to several identification strategies and can be interpreted as causal evidence that confirms the hypothesis of learning and memory in the literature (Chetty et al., 2013; Paetzold and Winner, 2016).

The importance of the firm environment in shaping individual learning processes motivates our interest in the firm-level dynamics of expanding the formal sector. We show that firms themselves are more likely to employ workers who avoid tax payments as they gain experience in the formal economy. When looking at firm cohorts by their year of entry into the formal sector, we document a strong rise in the prevalence of tax avoidance. However, once a firm engages in tax avoidance the *level* of tax avoidance within the firm remains relatively stable over time. We conclude that for firms information about tax avoidance practices is either available or it is not.

What are the determinants of a firm's information environment? We identify and quantify two specific information transmission mechanisms between firms: Peers and experts. To characterize the peers channel, we study co-workers coming into a firm and the knowledge

[4] The first alternative measure tracks taxpayers lowering their taxable income to any value below the income tax exemption threshold while having gross income above the threshold, and the second measure indicates taxpayers using deductions at any position in the income distribution. Please refer to the supplemental online Appendix B.3 for replications of our whole analysis.

they bring about tax avoidance due to their behavior in the previous job. The experts channel is characterized by knowledgeable accountants who were previously working for a firm that was employing tax avoiders. We identify these effects through changes in the co-worker composition and switches of accountants. Both the peers and experts channels are sizeable, leading to average increases in firm-level avoidance by 21 and 13 percent respectively. We corroborate our findings in an alternative identification strategy based on event studies in subsamples with plausible control groups for both channels. Incumbent employees in firms with new co-workers that were previously avoiding are significantly more likely to avoid tax payments than incumbents among firms with new workers that were previously not avoiding. Likewise, firms with new accountants previously at a firm with no tax avoidance activity are less likely to avoid than those with new accountants with tax avoidance at their previous firm.

Our findings are highly policy relevant since they give indications for tax authorities in designing audit strategies and deciding who should be targeted. Moreover, in settings where a policy instrument is only partially used by economic agents, slow adjustments can have distributional implications. In our setting, the usage of the deduction opportunities is strongly related to advantaged demographic characteristics and firms in particular sectors. This increases inequality compared to a scenario with full adoption. A flexible labor market mitigates these information frictions by enhancing information transmission through job mobility.

Literature Our main contribution is towards the small but growing literature on knowledge diffusion and spillover effects in taxation (Chetty et al., 2013; Paetzold and Winner, 2016). These papers analyze the effects of moving into high or low information environments (regions and firms) and emphasize the role of learning. In contrast to these papers, however, we provide extensive evidence that the effects are not driven by selection into specific knowledge environments but are indeed causal. We establish causality of the knowledge environment by exploiting exogenous job mobility through firm closures, controlling for a broad range of observed and unobserved confounders, and additionally creating a balanced control group by matching on observables. Moreover, we extend this literature by exploring the dynamic learning processes and by identifying specific channels of information transmission. In a recent contribution, Aghion et al. (2017) show that sluggish adjustments to newly introduced tax regimes are also present in a developed country, France.

Our results are embedded in a broader literature that has established the importance of job mobility for the transmission of information and innovation, and, hence, for firm performance. Using worker transitions from particularly productive firms, a number of recent studies show that mobility substantially contributes to the diffusion of human capital and helps increase productivity (Song et al., 2003; Balsvik, 2011; Parrotta and Pozzoli, 2012;

Stoyanov and Zubanov, 2012; Poole, 2013; Serafinelli, forthcoming). In particular, mobility of managers plays a crucial role for firm productivity, confirming parallel results in our paper (Mion and Opromolla, 2014; Bender et al., 2016).

The paper further contributes to the literature on bunching at kinks and notches in the tax schedule started by Saez (2010) and Chetty et al. (2011). The method was refined and expanded to estimate further behavioral parameters influencing bunching behavior like frictions, fixed adjustment costs, and reference dependencies (Kleven and Waseem, 2013; Gelber et al., 2017; Seibold, 2017).[5] We provide novel evidence on the dynamics of bunching by tracking economic agents over time. We exploit changes in the bunching estimate for workers with different exposure to the formal system to quantify the learning process. Moreover, bunching in personal income taxes has been mostly found in developed countries and for subgroups with easy adjustment opportunities such as self-employed workers (Chetty et al., 2011; Bastani and Selin, 2014).[6] We look at bunching among wage earners in a development setting and find strong reactions to a very small kink.[7]

Moreover, we contribute towards a growing literature on the determinants of formalization of developing economies (Gerard and Gonzaga, 2016; Naritomi, 2016; Pomeranz, 2015; Brockmeyer et al., 2018). We provide detailed evidence on the dynamics of individual and firm-level adjustments to the formal sector. Most importantly, we document the importance of experience and tenure in the formal economy for explaining the use of tax avoidance opportunities.

More generally, our paper relates to the literature on taxation and development. The relevance of our study is underscored by recent work showing the rising importance of personal income taxes as countries develop (Besley and Persson, 2013; Jensen, 2016). A number of studies have shown how tax systems in low enforcement settings can differ to those in more developed economies (Gordon and Li, 2009; Best et al., 2015; Keen and Slemrod, 2017). Corporate taxation and firm behavior in a development context (Asatryan and Peichl, 2017; Bachas and Soto, 2017) and in Ecuador in particular (Carrillo et al., 2012, 2017) have been studied extensively. The role of firms in driving tax avoidance and evasion opportunities has been put forward recently (Best, 2014; Kumler et al., 2015; Kleven et al., 2016). We specifically investigate the dynamics and determinants of the information environment at the firm level.

Finally, we contribute to the literature on the role of accountants and tax preparers in fa-

[5] For a comprehensive review, please refer to Kleven (2016).

[6] A notable exception is Kleven and Waseem (2013) who look at bunching of wage earners at notch points in Pakistan.

[7] The first kink (income tax exemption threshold) in the Ecuadorian tax schedule is very salient. The change in marginal tax rates from zero to five percent, however, is very small in international comparison. In line with the literature on the role of deduction opportunities in personal income taxation (Doerrenberg et al., 2017; Matikka, forthcoming), strong bunching responses at this first kink are driven by reporting effects using deductions and not real labor supply responses.

cilitating tax avoidance behavior (Kopczuk and Pop-Eleches, 2007; Chetty and Saez, 2013; Mahon and Zwick, 2017). We provide evidence of the importance of a firm's accountant in driving tax avoidance behavior not of the firm itself but of its employees.

The remainder of the paper is organized as follows. Section 2.2 provides information on the institutional background in Ecuador and describes the PIT system in detail. Section 2.3 gives detailed information on the various data sources employed in our study. In Section 2.4 we present the results on the drivers of individual and firm dynamics. Section 2.5 concludes.

2.2 Institutional Background

Ecuador is a middle-income country with a large but shrinking informal sector.[8] In the past years the government has implemented a range of economic and political reforms aimed at expanding social programs and public service delivery. While a surge in oil revenues facilitated some of this increased spending, the tax administration has also pushed wide-ranging reforms of the tax system and tax collection policies. As a result, tax revenue as well as the tax base have grown substantially over the past years. Between 2006 and 2015, central government tax revenues have increased from about 10% to almost 14% of GDP and have more than doubled in real terms. Taxation in Ecuador can be broadly categorized into personal income taxes (PIT), a value-added tax (VAT) of 12 % (food and some other goods are exempt), corporate taxes (22% of profits since 2013), a tax on foreign money transfers, and special consumption taxes. One of the main reasons for higher tax revenue is an increase in formalization of the economy induced by the tax administration's wide-ranging efforts to increase tax compliance.

The most relevant policy is the introduction of extensive deduction possibilities in income tax, substantially increasing the demand for formal receipts.[9] The receipts handed in to the authorities are used to cross-check the sales of businesses and fight tax fraud, especially with respect to VAT reporting behavior. From a firm's perspective, emitting receipts is not only linked to paying more VAT but also to taking part in other aspects of the formal economy such as withholding income tax and social security contributions for employees.

Apart from a general hike in tax revenue, these formalization efforts induced a strong increase in the number of taxpayers subject to personal income taxation. Between 2006 and

[8] According to a survey in 2006, about 70 percent of the labor force was employed in the informal sector (Canelas, 2015).

[9] Sellers of goods and services are obliged to offer two different types of receipts. The standard receipt ("nota de venta") includes information on goods and prices, while the enhanced version ("factura") contains additional information about the client's name and unique identification number. Only these detailed receipts issued to the taxpayer or his/her dependents can be presented to the tax authority. This policy guarantees a paper trail and impedes illegal sale of receipts. Further policies to increase tax compliance include improved information sharing between government agencies.

2015, the total number of tax declarations submitted for private sector employees increased from 1 million to about 2.5 million.

Personal Income Taxes (PIT) Ecuador has a unified PIT schedule which is levied on almost all regular sources of wage and self-employed income.[10] Tax liability in Ecuador is individually determined (i.e., no family taxation).[11] The PIT liability is calculated progressively with numerous small jumps in the marginal tax rate, starting at 5% and going up to 35%. The cutoff income levels change yearly according to inflation.[12]

PIT in Ecuador starts being levied only at relatively high levels. In 2013, the exemption threshold was set such that income tax was not charged on annual income below 10,180 USD.[13] For the same year, the monthly minimum wage was set at 318 USD, corresponding to yearly taxable income of 3,816 USD, well below the exemption threshold. The minimum wage is estimated to be slightly above the median wage and slightly below the average wage in Ecuador for 2008 to 2012 (Canelas, 2014). Therefore, PIT is only applicable to individuals in the top 10 % of the distribution of formal sector income.

The Ecuadorian tax system is unique in its generous deduction allowances for personal expenses in education, health, food, clothing and housing introduced in 2008 (Villacreses, 2014). The total deductible amount of personal expenses is limited to the smaller of 50% of individual income or 1.3 times the exemption threshold (in 2013 this was $1.3 \times 10{,}180 = 13{,}234$ USD).[14] Ecuadorian taxpayers are legally obliged to keep the receipts of all of their deductions. However, only if individuals claim deductions above a specific reporting threshold (50% of the tax free amount, or 5090 USD, in 2013[15]), must they submit the receipts of all of the claimed deductions to the tax authority via an online annex.

The mechanism by which tax declarations and deductions are submitted in Ecuador deserves some special attention and is key to understanding the findings in our analysis. PIT is primarily filed on a firm-reported tax form (F107, see figure B.1 in the Appendix). This form can only be submitted to the tax authority by the employer and includes the level of deductions in personal expenses. In March of each year, wage earners fill out a form with

[10] Notable exceptions include all forms of payments from the social security system (pension payments, educational stipends, disability benefits, etc.), severance payments, interest on savings accounts, occasional capital gains, returns from investment funds or long-term deposits as well as certain additional wage benefits mandatory under labor market regulations.

[11] Additional to PIT, employees in the private sector pay 9.45% of their wage income in social security contributions and the employer pays 11.15%. Paying these social security contributions entitles people to a range of benefits including pensions, health insurance, disability insurance and unemployment benefits.

[12] The rate used for inflation adjustments is the yearly change in consumer price index for urban areas published by Ecuador's National Statistics Institute INEC on November 30 of a given year. Exact nominal values since 2006 are displayed in Table B.1 in the appendix. In 2008, the government enacted a series of tax system reforms, including increasing the top marginal tax rate from 25% to 35%.

[13] The Ecuadorian economy was completely dollarized in 2000 following extreme hyperinflation.

[14] Each category is individually capped at 0.325 times the exemption threshold, except for health expenditures, which have an upper limit of 1.3 times the exemption threshold.[15] Until 2010 this limit was set at 7500 USD.

their *projected* expenses in health, education, food, clothing and housing for that whole year and submit it to their employer. Based on these figures, the employer computes the level of the withholding tax for the following year. Workers are given the opportunity to update their information on deductions in October. While the ultimate responsibility for the overall correctness of these deductions lies solely with the employee, this system induces a weak form of third-party reporting of deductions. Recent literature shows that third-party information reporting by firms is a key driver for sustaining high levels of taxation (Kleven et al., 2016).

For the vast majority of employees (87% of our observations), taxes and personal deductions are only reported by the employer. The remaining 13% of all observations additionally submit a self-reported tax declaration (form F102). The primary purpose of this self-reported tax declaration form is to report self-employment income. However, individuals can also use it to update the employer-reported information.

2.3 Data and Descriptives

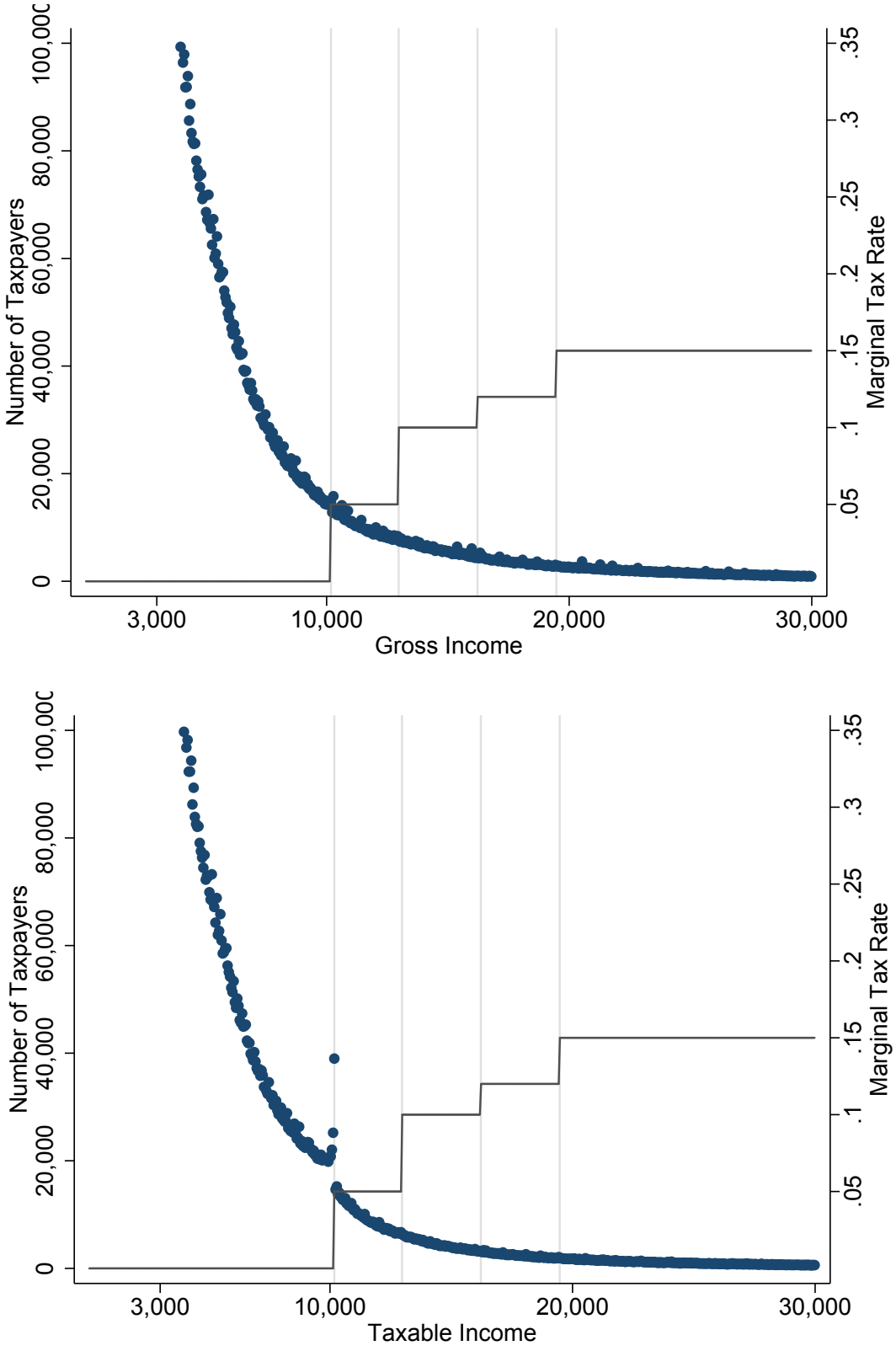
Our data combines several administrative datasets in Ecuador administered by the Ecuadorian tax authority *Servicio de Rentas Internas* (SRI). The core data consist of the universe of firm-reported PIT returns of regular employees (tax form F107) for the years 2006-2015.

We augment these tax records by three additional datasets. First, we use unique individual identifiers to merge the data to the Ecuadorian civil registry (*Registro Civil*). This register data provides a range of socio-demographic variables, including the year of birth, highest level of education, and gender. Second, we merge the tax returns to the central firm registry in Ecuador (*Catastro de RUC*). This registry contains firm-level data on industry affiliation, sector (public or private), time of formation of the firm, and place of registry. Lastly, for the subset of corporate firms we draw on their corporate tax declarations to identify the accountant working at the firm.[16] We end up with detailed matched employer-employee data that allows us to track taxpayers, firms, and co-workers over time.

A significant fraction of wage earners has various employers throughout a given calender year and therefore multiple tax declarations. We sum up the different income values to compute a unified measure of yearly individual income. Moreover, we consider the spell with the highest earnings as the main employer. We deflate all earnings to real 2013 USD values using the same consumer price index that is employed by the SRI to adjust the tax brackets annually (cf. footnote 12). Thereby the tax brackets, even though they change yearly in nominal values, remain unchanged in real terms.

[16] Firms are obliged to file a corporate tax declaration if their annual gross income exceeds 100,000 USD. Firms can have several corporate tax declarations and accountants per year. Here we take all accountants given in any of a firm's corporate tax declarations as being at the firm in a given year. Likewise, some accountants work for several firms in a given year. This is exactly the source of variation we are exploiting in Section 2.4.2.

Figure 2.1 – Income Distribution in Ecuador



This figure shows binned scatterplots of the distribution of gross income (upper panel) and taxable income (lower panel) in Ecuador. We restrict the sample to individuals who earn at least 12 times the monthly minimum wage and at most 30,000 USD. The income distribution is contrasted with the marginal tax schedule (right y-axis) and vertical lines mark the location of kink points in the marginal tax rate.

Throughout our analysis, we exclude all individuals employed in the public sector and only focus on private sector employees. About one quarter of the formal sector employees are in the public sector.[17]

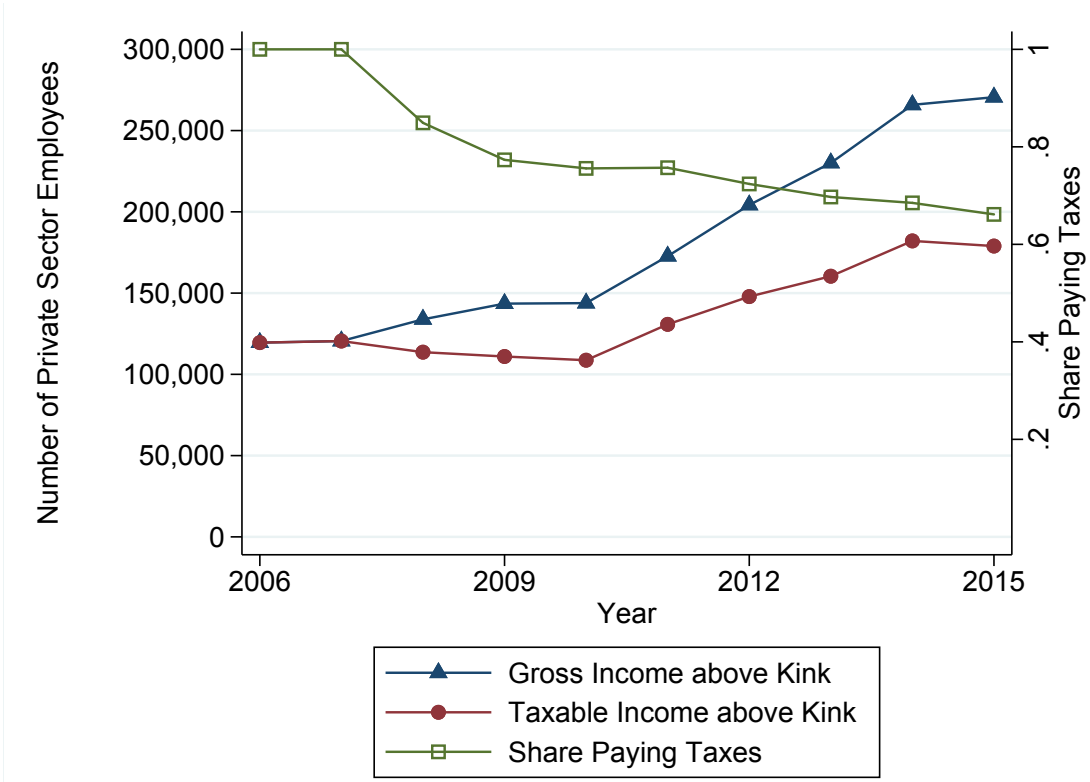
Figure 2.1 displays the reported income distribution in Ecuador pooling all observations in our sample from 2006 to 2015. We concentrate on workers who earn at least twelve times the monthly Ecuadorian minimum wage (yearly earnings of $12 \times 318 = 3{,}816$ USD in 2013) and those who earn less than 30,000 USD. The individual data is compressed into bins of 50 USD and plotted as bin frequencies for each bin. In general, the distribution of gross income in the upper panel is downward sloping, with the most frequent points around the minimum wage. The graph contrasts the income distribution with the marginal tax schedule, as given by the step function with values on the right vertical axis. The gross income distribution is smooth around all kink points of the marginal tax schedule depicted in the figure. The distribution of *taxable* income (gross income minus any deductions) in the lower panel, however, looks different. There is a pronounced spike in the distribution just before the exemption threshold. The difference between gross and taxable income indicates that tax avoidance is driven by reporting effects rather than real labor supply responses.

Our main measure of tax avoidance is the amount of individuals adjusting their income such that they locate just below the tax exemption threshold ("bunching"). In online Appendix B.3, we conduct our entire analyses using two alternative measures of tax avoidance, the amount of individuals reducing their taxable income to any value below the first kink and the amount of individuals with deductions with very similar results.

While bunching is strong at the exemption threshold, we do not observe any bunching at subsequent kink points of the marginal tax schedule. The exemption threshold, even though it is associated with a very modest increase in the marginal tax rate of only 5%, is arguably the most salient aspect of the tax schedule. Behavioral biases may make the disutility associated with the first dollar of tax payments discretely higher than any other subsequent increases in the tax liability. Moreover, individuals may perceive a discontinuity in audit probabilities at the exemption threshold and prefer to stay under the radar of the tax authority. Lastly, the marginal returns to filing more deductions vanish once taxpayers have successfully reduced their taxable income below the exemption threshold.

[17] We exclude public sector employees for three main reasons. First, public sector employees face different incentives than private sector employees, and their pay is often regulated by predetermined government pay scales. Second, the main drive in formalization of the past years was being carried out in the private sector as the public sector was already formal by definition. Third, private sector employees might have better opportunities to adjust their taxable income by bargaining with their employer about wages, and employers in the private sector might provide more support in filing the deductions.

Figure 2.2 – Number of Employees



This figure displays the number of tax declaration of employees with gross and taxable income above the tax exemption threshold over time. The green squares indicate the share of individuals with taxable income above the kink among those with gross income above the kink (right y-axis).

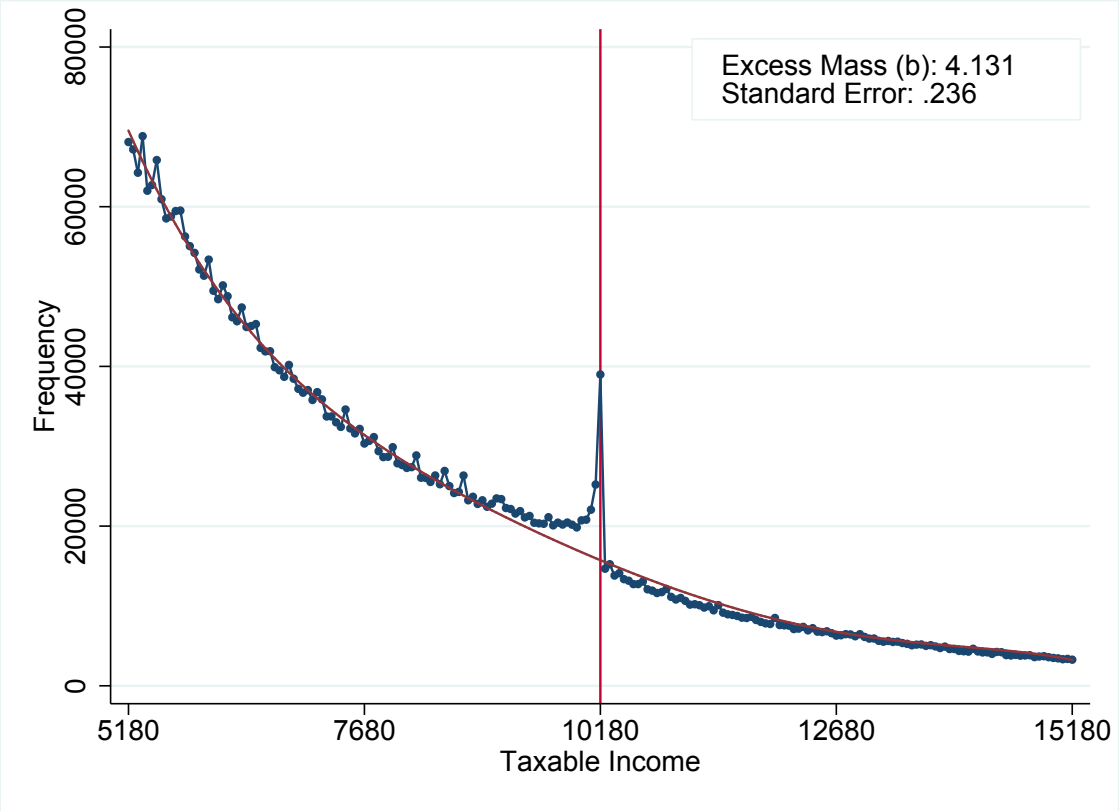
The relevance of dynamic aspects in driving tax adjustment behavior becomes especially pronounced when tracking the number of taxpayers over time. Figure 2.2 indicates a strong 2.5-fold increase in the number of private sector employees with tax-liable gross income between 2006 and 2015 (blue triangles).[18] After the introduction of generous deduction possibilities in 2008, however, a substantial and increasing share of employees reduced their reported taxable income below the exemption threshold (red dots). The growing wedge between gross income and taxable income results in a decreasing share of individuals that actually pay taxes (green squares on right hand axis) and reflects the growth in tax avoidance over time. The main part of our analysis examines the learning processes driving this dynamic increase in tax avoidance.

[18] The increase in the overall number of private sector employees is proportional but about an order of magnitude larger: The number increases from about 1 million to 2.5 million.

2.4 Results

In this section we present empirical results from our analysis of learning dynamics about avoidance opportunities in personal income taxes. The first part explores the dynamics of individual learning and exploits a sample of job switchers to identify firms as the driving environment for individual learning. The second part documents firm-level dynamics in tax adjustment behavior and identifies peers and experts as the main drivers of information transmission on tax avoidance opportunities. Throughout this section, our measure of tax avoidance is bunching just below the income tax exemption threshold. All of our results, however, are robust to using two alternative measures for tax avoidance: reducing taxable income to any value below the exemption threshold and an indicator for using deductions. Please refer to online Appendix B.3 for all graphs and tables using these alternative definitions.

Figure 2.3 – Bunching Estimates Taxable Income



This figure shows the actual distribution of taxable income around the tax exempt threshold as a binned scatterplot with 50 USD bin width. The red line shows a polynomial fit (of degree 5) to the distribution leaving out bins in a window around the kink (1000 USD to the left and 100 USD to the right). The vertical line indicates the location of the kink point.

To quantify the amount of bunching at the exemption threshold, we draw on the methods laid out in Saez (2010) and Chetty et al. (2011). Using binned income data (50 USD bin size)

and leaving out a window around the kink (1000 USD to the left and 50 USD to the right), we estimate a counterfactual density (polynomial of degree 5) around the kink that would prevail in its absence. The difference between the observed density and the counterfactual is used to compute the excess mass as multiples of the counterfactual.[19] Figure 2.3 displays the distribution of taxable income around the kink. The empirical density is represented by the blue dots and the estimated counterfactual is represented by the red line. The estimate for the excess mass is highly significant and very large, indicating that more than four times as many individuals are located around the kink compared to the expected mass under the counterfactual of no kink.[20]

2.4.1 Individual Dynamics

In this section we explore the dynamics in the usage of tax adjustment opportunities among individual workers. First, we document strong increases in tax avoidance as individuals gain experience in the formal sector. Second, we provide causal evidence for the influence of the firm information environment on individual learning processes.

Individual Learning

The massive expansion in the number of taxpayers in Ecuador allows us to follow cohorts of individuals who entered the formal sector at various points in time. Hence, we compare bunching levels among the same set of individuals depending on their tenure in the formal system. To hold the sample composition constant within cohorts, we restrict the sample to individuals that are observed without interruption once they entered the formal economy.

Table 2.1 displays bunching estimates over time for different cohorts. Each row corresponds to one of the cohorts that entered the formal sector between 2007 and 2014. The columns indicate how the level of bunching changes over time for these cohorts. For each cohort, there is a clear increase in the amount of bunching in taxable income as experience in the formal sector increases. Moreover, the estimates become more precise over time, indicating less heterogeneity within cohorts over years. Individuals entering the formal economy in 2010 for instance had a modest (and insignificant) excess mass of 0.62 in their first year which increased to 5.56 in 2015. We observe this steep increase throughout all cohorts.

[19] Standard errors are obtained using a non-parametric bootstrap procedure. Our results are robust to sensitivity checks varying the bin width, the parametric form of the polynomial and the bunching window left out in the estimation of the counterfactual density (available on request).

[20] When using these estimates to calculate elasticities we find extremely large values. However, we do not believe these to be very informative about the underlying labor supply elasticity or elasticity of taxable income for a variety of reasons (see also Blomquist and Newey (2017)). First, as discussed in Section 2.3, there are number of factors exacerbating bunching at this first kink. Second, recent research has shown that in the presence of deduction possibilities it becomes difficult to structurally interpret inferred elasticities (Doerrenberg et al., 2017).

Table 2.1 – Bunching estimates over time by cohort

Cohort	2007	2008	2009	2010	2011	2012	2013	2014	2015	Observations
A. Taxable Income										
2007	2.59*	2.95***	2.89***	3.08***	4.25***	4.98***	4.31***	4.93***	6.65***	48,570
	(1.50)	(1.08)	(1.08)	(0.77)	(0.74)	(0.70)	(0.58)	(0.60)	(0.65)	
2008		3.44**	-0.57	2.90***	2.64***	4.78***	3.08***	4.72***	3.83***	79,785
		(1.59)	(0.92)	(0.75)	(0.65)	(0.68)	(0.56)	(0.51)	(0.52)	
2009			0.26	0.75	2.26**	5.74***	4.34***	5.67***	5.61***	59,427
			(0.66)	(1.60)	(1.02)	(1.02)	(1.03)	(0.70)	(0.79)	
2010				0.62	2.16	3.94***	4.75***	5.45***	5.56***	67,024
				(0.98)	(1.74)	(1.21)	(1.19)	(1.00)	(0.82)	
2011					1.18	3.72*	6.05***	6.15***	7.19***	108,496
					(0.97)	(2.15)	(1.61)	(1.15)	(1.04)	
2012						2.91	4.64*	5.69***	5.49***	140,777
						(3.23)	(2.57)	(1.35)	(0.96)	
2013							5.21	4.08*	6.25***	168,952
							(3.43)	(2.19)	(1.38)	
2014								3.73	7.38***	219,543
								(3.07)	(1.78)	
B. Gross Income										
2007	2.56*	1.68	1.15	1.81*	1.59**	0.72	0.64	0.15	0.58	48,570
	(1.50)	(1.11)	(1.14)	(0.94)	(0.86)	(0.80)	(0.77)	(0.71)	(0.79)	
2008		1.85	-2.24**	1.06	0.98	1.50**	0.04	0.63	-0.46	79,785
		(1.68)	(1.03)	(0.79)	(0.76)	(0.76)	(0.63)	(0.64)	(0.61)	
2009			1.25	-1.54	-0.73	2.30**	0.05	0.55	-0.02	59,427
			(3.67)	(1.57)	(1.09)	(1.04)	(1.14)	(0.83)	(0.87)	
2010				1.28	-1.06	1.27	0.47	0.43	0.18	67,024
				(3.43)	(1.75)	(1.30)	(1.27)	(1.08)	(0.94)	
2011					0.20	-1.19	-0.87	-0.08	0.41	108,496
					(3.33)	(2.19)	(1.69)	(1.30)	(1.10)	
2012						-2.05	-0.78	-0.46	-1.06	140,777
						(3.28)	(2.65)	(1.52)	(1.13)	
2013							-2.57	-2.39	-0.89	168,952
							(3.36)	(2.36)	(1.35)	
2014								-3.72	-1.18	219,543
								(3.10)	(1.91)	

Note: This table reports bunching estimates for taxable and gross income by year conditioned on the cohort of entry into the formal economy. Sample restricted to individuals observed without interruption after entering the formal economy. The estimates are based on binned income data (50$ bin size) and a counterfactual density using a polynomial of degree 5. Bootstrapped standard errors reported in parentheses, significance levels are given by * < 0.1, ** < 0.05, and *** < 0.01.

Learning did not only occur within cohorts but also across cohorts as individuals entering the formal economy in later years tend to start at higher degrees of bunching.[21] Bunching in gross income (Panel B of Table 2.1), in contrast, stays relatively low and does not increase as individuals gain experience in the formal system.

One major concern in comparing bunching estimates according to tenure and experience in the formal system is that workers might sort into firm environments where it is more common to bunch. Hence, factors like wage growth and selection on (un)observables may confound our results. These factors are already mitigated to a large extent by the fact that the bunching estimator is a local estimator measuring the excess mass for a given subsample and in the vicinity of the kink. Moreover, by holding constant the individuals within a cohort, we abstract from a range of selection effects. To address any remaining selection issues, we regress an indicator whether a worker bunches on flexible functions of experience in the formal sector while controlling for a broad range of observable characteristics and unobserved heterogeneity.[22] Table 2.2 presents results from various specifications of a simple linear probability model. The first three columns show regression results for a linear, quadratic, and cubic polynomial in years of experience. In all specifications, we include year fixed effects to control for general time trends. The estimates show strong initial increases in the probability to bunch which level off after 4 to 5 years. In Column (4), we add individual-level control variables such as an age polynomial, gender, education, marital status, nationality and the number of jobs a worker holds in the given year, but do not observe any change in the impact of experience on bunching. Women and married individuals are more likely to bunch and tax avoidance increases with age. Higher education levels tend to increase the likelihood of bunching. Having multiple jobs within a year makes it more difficult to adjust income and deductions and therefore reduces the probability of bunching. In order to take care of income dynamics as potential confounders, Column (5) additionally controls for (log) gross income and income growth. The coefficients on experience are slightly smaller in magnitude but still strongly significant and indicate the same pattern of diminishing effects as in previous specifications. In Column (6) we add firm-level characteristics such as firm age, firm size, an indicator for corporate firm status, and industry (14 broad categories) and region (24 provinces) fixed effects with no change in the main effects. The identifying variation hence derives from differences in experience within industry and within region cells, holding fixed observable characteristics and general time

[21] Notable exceptions to this are the 2007 and 2008 cohorts, which start at relatively high levels. The 2007 cohort has the same amount of (not very significant) bunching in *gross* income levels in 2007, indicating other mechanisms at work than the tax avoidance mechanisms studied in this paper. The 2008 cohort might be inherently different to the other cohorts as these are the very first individuals affected by the government's drive to formalize the economy.

[22] We define bunching as having taxable income within the range of 1000 USD to the left of the exemption threshold and restrict the sample to individuals in the years 2006-2015 with gross income above the exemption threshold but still within the relevant range for bunching using the deduction possibilities.

trends. Finally, incorporating worker fixed effects in Columns (7) indicates that the relation between experience and bunching behavior remains stable when the effects are identified by within-individual variation in experience.

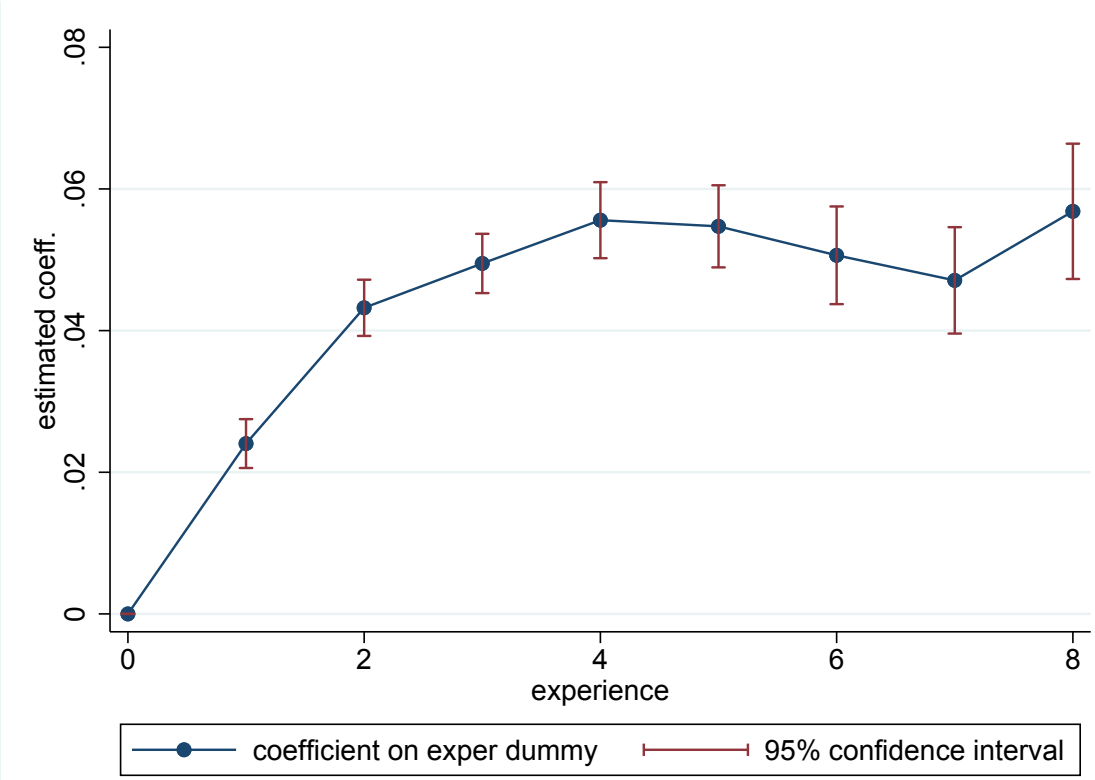
Table 2.2 – Bunching Individuals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Experience	0.0066***	0.040***	0.085***	0.088***	0.065***	0.064***	0.048***
	(0.00058)	(0.0012)	(0.0030)	(0.0031)	(0.0063)	(0.0062)	(0.0065)
Experience^2		-0.0038***	-0.015***	-0.016***	-0.012***	-0.012***	-0.0077***
		(0.00012)	(0.00078)	(0.00079)	(0.0013)	(0.0013)	(0.0014)
Experience^3			0.00083***	0.00089***	0.00065***	0.00062***	0.00041***
			(0.000056)	(0.000056)	(0.000083)	(0.000081)	(0.000089)
Married				0.0055***	0.0038***	0.0025*	
				(0.0011)	(0.0013)	(0.0013)	
Age				0.0049***	0.0033***	0.0029***	
				(0.00049)	(0.00061)	(0.00056)	
Age^2				-0.000043***	-0.000026***	-0.000028***	
				(0.0000061)	(0.0000072)	(0.0000067)	
Female				0.015***	0.019***	0.011***	
				(0.0017)	(0.0021)	(0.0019)	
Secondary Education				0.033**	0.022	0.015	
				(0.014)	(0.017)	(0.017)	
Tertiary Education				0.032**	0.015	0.0042	
				(0.015)	(0.017)	(0.017)	
Foreign				-0.0040	-0.011*	-0.015***	
				(0.0047)	(0.0059)	(0.0053)	
Number of Jobs				-0.043***	-0.046***	-0.046***	-0.027***
				(0.0013)	(0.0015)	(0.0014)	(0.0014)
Log Gross Income					0.032***	0.038***	0.060***
					(0.0033)	(0.0028)	(0.0021)
Gross Income Growth					0.0062***	0.0048***	-0.0016
					(0.0015)	(0.0014)	(0.0014)
Corporate Firm						-0.0077**	0.011***
						(0.0039)	(0.0040)
Firm Age						-0.00012	-0.00018**
						(0.000097)	(0.000070)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	No	Yes	Yes
Industry FE	No	No	No	No	No	Yes	Yes
Worker FE	No	No	No	No	No	No	Yes
R^2	0.016	0.019	0.020	0.028	0.021	0.027	0.280
Observations	618,356	618,356	618,356	618,356	508,417	508,417	508,417

The table shows results from linear regressions with a binary indicator for bunching individuals as dependent variable. The sample is restricted to potential bunchers in 2008 to 2015. Further (unreported) control variables include firmsize, firm age. Standard errors (in parentheses) are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Overall, the learning process can be described well by a polynomial in years of experience. We find strong initial increases in bunching activity: Between the first and the second year in the formal sector, experience leads to an increase in the bunching probability of 3 to 6 percentage points. The increase becomes less steep over time and levels off completely after four to five years of experience. The development of the effects is clearly presented in Figure 2.4 which displays coefficients in a specification that controls for worker fixed effects as Column (7) but includes separate dummy variables for each year of experience in the formal sector.[23]

Figure 2.4 – Coefficients on experience dummies



This figure depicts estimated coefficients from a linear regression of a bunching indicator on dummy variables for each year of experience in the formal sector. We control for individual fixed effects, income dynamics and a broad range of firm characteristics.

The evidence presented in this subsection strongly supports the hypothesis of individual learning dynamics in tax bunching. We provide robust evidence of individuals increasing their bunching activity as they gain experience in the formal sector – even when controlling for income dynamics and other potential confounding factors. The next subsection turns to the question of how learning takes place and investigates how individuals react to changes in their information environment.

[23] The estimates are interpreted relative to the first year in the formal sector (with no previous experience).

Job Switchers

To gain insights into the impacts of the firm environment on tax avoidance behavior, we draw on a sample of job switchers and exploit variation in the information environment individuals face. Following Chetty et al. (2013), we compare tax avoidance behavior of workers moving into a high-avoidance environment to those moving into a low-avoidance environment. In contrast to that paper, however, we examine job-to-job transitions between firms directly (instead of regional mobility). Due to several identification strategies addressing possible strategic job mobility patterns, we are able to make statements about the causal effect of information environments on individual tax avoidance behavior.

We draw on the universe of formal sector job transitions in Ecuador. To keep sample composition fixed across years, we only consider job transitions where we observe at least two consecutive years before and after the job switch. Moreover, we only consider job switches of the main employer[24] and only an individual's first job transition.[25] Hence, we end up with a sample of 152,617 job transitions that occurred between 2010 and 2014.

We characterize the job switchers' information environments by assigning their origin and destination firms to quintiles based on the share of co-workers who are bunching.[26] Table 2.3 reports summary statistics for our sample of job switchers. We concentrate on workers who work in the medium quintile and move to the bottom, medium, or high quintile. Average characteristics of these workers are displayed in Column (1). Demographic characteristics as well as income before and after the job transition differ substantially between workers with different destination quintiles. Column (2) reports characteristics for switchers to the bottom quintile and Column (3) indicates significant differences to workers who switch to another firm in the medium quintile. Similarly, Columns (5) reports characteristics for those switching to the high quintile and Column (6) provides significant differences to those switching to the mid quintile. We therefore employ a broad range of identification strategies that address the potential selection of workers into specific knowledge environments. The main challenge is that transitions into higher knowledge quintiles are also associated with higher wage increases. We first provide graphical evidence of bunching shares around the job transition based on raw data before we address selection using event study regressions with (1) a broad range of control variables including wage growth and unobservable worker heterogeneity, (2) a matched control group with excellent balancing properties, and (3) the subsample of workers who switch their job due to exogenous job displacement.

[24] The main employer is the one with the highest annual earnings. Job switches are by definition to a firm the individual has not worked at before.

[25] In unreported robustness checks we consider the subsample of individuals who switched jobs only once with no change in the results.

[26] For every year, we compute the distribution of the share of co-workers who bunch and split the sample into quintiles. As before, we define bunching as reporting taxable income of 1000 USD to the left of the exemption threshold. To abstract from individuals too far away from the exemption threshold, we draw on the full sample of private sector employees with gross earnings between 5000 and 25000 USD.

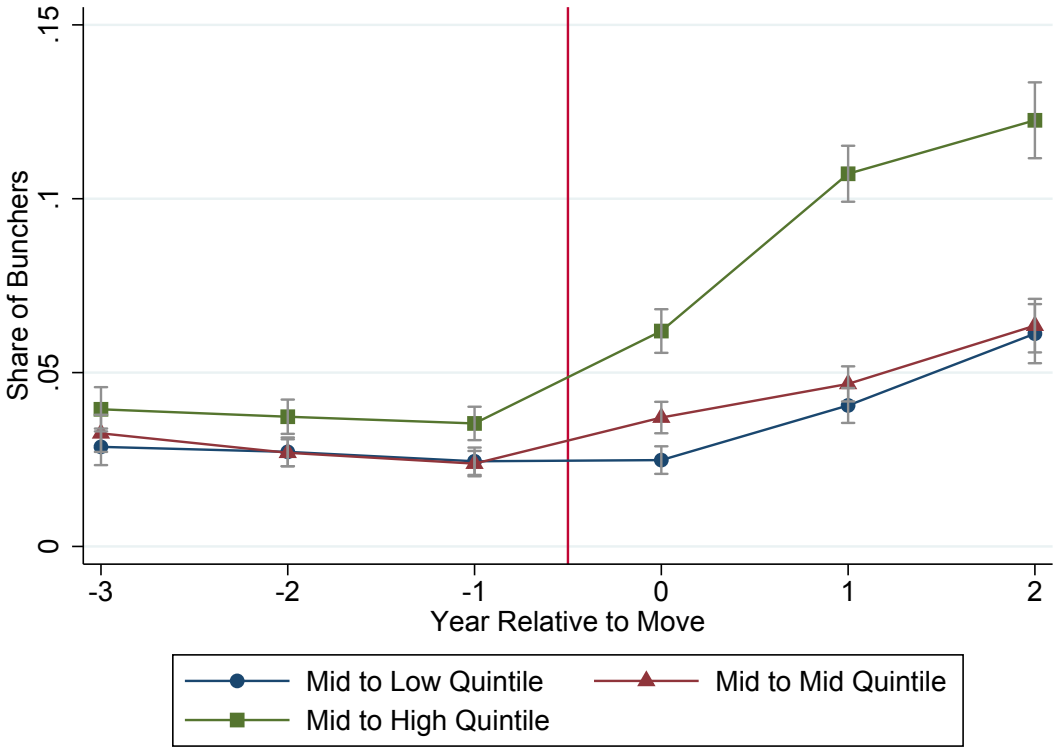
Table 2.3 – Job Switchers - Descriptives

	Descriptive Statistics						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	From Mid	Mid to Low	Diff	Matched Diff	Mid to High	Diff	Matched Diff
Demographics							
Age	31.75	33.27	2.00	-0.03	30.75	-0.52	0.14
	(8.84)	(9.57)	(0.16)	(0.18)	(8.16)	(0.15)	(0.15)
Married	0.46	0.47	0.02	-0.01	0.46	0.01	-0.01
	(0.50)	(0.50)	(0.01)	(0.01)	(0.50)	(0.01)	(0.01)
Female	0.29	0.30	0.03	-0.00	0.31	0.03	0.01
	(0.46)	(0.46)	(0.01)	(0.01)	(0.46)	(0.01)	(0.01)
Tertiary Education	0.23	0.23	0.02	-0.02	0.27	0.07	0.00
	(0.42)	(0.42)	(0.01)	(0.01)	(0.45)	(0.01)	(0.01)
Pre-Switch							
Gross Income	6278.08	5868.99	-409.32	140.90	6703.97	425.65	189.97
	(6305.17)	(6033.08)	(110.82)	(107.29)	(6471.90)	(115.69)	(120.25)
Taxable Income	5849.38	5493.57	-345.23	144.32	6232.53	393.73	199.15
	(5272.79)	(5083.61)	(92.94)	(89.49)	(5378.45)	(96.42)	(99.59)
Share Deduction Filers	0.08	0.07	-0.00	0.00	0.08	0.01	0.01
	(0.26)	(0.25)	(0.00)	(0.00)	(0.28)	(0.00)	(0.01)
Buncher	0.03	0.02	0.00	0.00	0.04	0.01	0.00
	(0.16)	(0.15)	(0.00)	(0.00)	(0.18)	(0.00)	(0.00)
Post-Switch							
Gross Income	6544.62	5115.60	-1921.69	-110.45	7450.82	413.52	171.57
	(6145.82)	(4953.44)	(109.62)	(87.80)	(5871.09)	(117.66)	(115.47)
Taxable Income	6039.29	4854.24	-1629.36	-104.66	6748.53	264.93	84.80
	(5064.40)	(4204.85)	(91.68)	(74.37)	(4656.59)	(96.13)	(91.92)
Share Deduction Filers	0.10	0.06	-0.04	0.00	0.14	0.05	0.04
	(0.30)	(0.23)	(0.00)	(0.00)	(0.35)	(0.01)	(0.01)
Buncher	0.04	0.02	-0.01	0.00	0.06	0.02	0.03
	(0.20)	(0.16)	(0.00)	(0.00)	(0.24)	(0.00)	(0.00)
Observations	18,318	5,919			5,682	.	.

Notes: This table reports summary statistics for the job switcher sample, consisting of all individuals who switch their job between 2010 and 2014 (regarding only their first move) and for whom it is possible to observe at least two consecutive years before and after the move. Pre-move gives mean values in the two years before the move, post-move the respective values in the first two years at the destination firm. Individuals are grouped into quintiles depending on their co-worker bunching shares for any given year. Columns (2) to (4) represent individuals starting in the mid (third) quintile of the bunching distribution in the year before the move and moving to a firm in the low (first), mid (third) or high (fifth) quintile.

Graphical Evidence Using an event study graph, we observe the dynamic adjustment process of individuals depending on the quintile they are moving towards. Figure 2.5 plots the share of bunchers among workers starting from a firm in the mid-quintile of the bunching distribution. The horizontal axis indicates the year relative to the move with year zero being the first year at the destination firm. The data show an asymmetric pattern of adjustment. The share of bunchers among workers switching to a high-bunching firm sharply increases after the transition, resulting in the bunching share more than doubling its pre-switch level after three years. In contrast, even though we observe a moderate overall upward trend, bunching probabilities remain relatively unchanged for job transitions into a mid- or low-bunching environment.[27]

Figure 2.5 – Event Study Job Switchers



This figure shows an event study with bunching shares of job switchers around the time of the job transition. The vertical line indicates the time of the transition. We observe bunching among individuals who come from a firm in the medium quintile of the distribution of co-worker bunching shares and differentiate between those who switch to a firm in the bottom, medium, and top quintile.

[27] Table B.4 in the appendix depicts the same event-study graph for individuals starting in the low or high quintile of the bunching distribution. In both alternative samples we also find a much stronger increase in the share of bunchers among individuals transitioning to the top quintile than among those moving to the mid or low quintile.

Figure 2.5 indicates parallel and stable pre-switch trends between individuals moving to firms in different parts of the bunching share distribution. While this lends credibility to standard parallel trends assumptions, the descriptive analysis has shown selection in terms of income dynamics between these groups of taxpayers. To address potential selection effects, we employ a range of identification strategies that control for unobserved heterogeneity and observed characteristics such as earnings and wage growth before and after the job switch.

Controlling for observed and unobserved heterogeneity In our first strategy, we estimate several regression-based versions of the event study design that control for a broad range of observable worker and firm characteristics and allow for unobserved heterogeneity across workers by incorporating individual worker fixed effects. Hence, the effect of the job switch on bunching is identified by the time variation within individuals. We run the following regression on the subsample of individuals starting in the medium quintile of the bunching distribution:

$$Y_{it} = \beta_0 + \delta post_{it} \times quintile_i + \theta X_{it} + \alpha_i + \lambda_t + \sum_{k=-2}^{k=2} \gamma_k D_{it}^k + \epsilon_{it}. \tag{2.1}$$

The dependent variable Y_{it} measures tax avoidance as an indicator for individual i having taxable income within a 1000 USD window to the left of the exemption threshold in year t. The indicator variable $post_{it}$ takes on the value of one in the years after the job switch and $quintile_i$ indicates if an individual moved to the high quintile. Accordingly, δ is our main coefficient of interest measuring the overall effect of moving to a high- or low-avoidance firm. We control for time-varying individual and firm characteristics X_{it} including gross income, wage growth, age squared, firm size, industry classification (18 broad industries), firm location (24 provinces), and corporate firm status. Last, we account for various sources of unobserved heterogeneity by including individual fixed effects (α_i), year fixed effects (λ_t) and fixed effects in event time (γ_k). We run a parallel analysis for individuals switching from a firm in the mid to the low quintile with $quintile_i$ being an indicator for the low quintile.

The estimates are displayed in Panel A of Table 2.4. Columns (1) and (5) are without and columns (2) and (6) with the controls X_{it}. The results confirm the importance of the firm environment in driving individual tax adjustment behavior: moving to a high quintile firm increases bunching by about 3 percentage points while moving to the low quintile has no significant effect.[28]

[28] In various sensitivity checks, we estimate this same regression without individual fixed effects but instead a wide range of individual specific demographic controls (age, gender, education) and find no substantial difference in the results. We furthermore estimate the same regression without the fixed effects in event time D_{it}^k and find no substantial change in the results.

Table 2.4 – Job Switchers

	Mid to Low				Mid to High			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full Sample		Matching	Displaced	Full Sample		Matching	Displaced
A. Overall Effect								
After event year	-0.007*	0.001	0.010*	-0.002	0.034***	0.028***	0.028***	0.025
	(0.004)	(0.004)	(0.006)	(0.014)	(0.005)	(0.005)	(0.007)	(0.018)
B. Effects by Relative Year								
Anticipatory Effects								
Event year - 3	-0.006	-0.009	-0.012	-0.007	-0.003	-0.001	0.002	0.025
	(0.007)	(0.007)	(0.009)	(0.020)	(0.008)	(0.007)	(0.011)	(0.031)
Event year - 2	-0.002	-0.005	-0.019**	-0.001	0.000	0.002	0.008	0.025
	(0.005)	(0.005)	(0.009)	(0.014)	(0.006)	(0.006)	(0.010)	(0.019)
Post Treatment Effects								
Event year	-0.013***	-0.006	0.003	-0.008	0.013**	0.010	0.021**	-0.002
	(0.005)	(0.005)	(0.006)	(0.016)	(0.006)	(0.006)	(0.008)	(0.018)
Event year + 1	-0.007	-0.000	0.002	-0.009	0.049***	0.043***	0.039***	0.058**
	(0.006)	(0.006)	(0.009)	(0.019)	(0.007)	(0.007)	(0.011)	(0.026)
Event year + 2	-0.004	0.000	-0.000	0.009	0.044***	0.037***	0.036***	0.069*
	(0.008)	(0.008)	(0.011)	(0.025)	(0.010)	(0.010)	(0.013)	(0.035)
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	60,040	60,040	53,686	6,211	59,367	59,367	52,215	4,999
C. Timing								
Event year - 2	0.005	-0.002		-0.010	0.006	0.001		-0.006
	(0.004)	(0.004)		(0.017)	(0.004)	(0.004)		(0.045)
Event year	-0.000	0.005		0.012	0.026***	0.022***		0.059
	(0.003)	(0.005)		(0.013)	(0.005)	(0.007)		(0.040)
Event year + 1	0.015***	0.016*		0.030	0.071***	0.059***		0.151*
	(0.004)	(0.008)		(0.025)	(0.006)	(0.011)		(0.077)
Controls	No	Yes		Yes	No	Yes		Yes
Observations	25,048	25,048		2,403	23,947	23,947		1,479

The panels of this table denote the results from regression equations (2.1), (2.2) and (2.3) respectively. Standard errors (in parentheses) are clustered at the destination firm by year level. Significance levels are given by * < 0.1, ** < 0.05, and *** < 0.01.

Explicitly looking at the timing of the effects, we modify the regression equation

$$Y_{it} = \beta_0 + \sum_{k=-2}^{k=2} \delta_k D_{it}^k \times quintile_i + \theta X_{it} + \alpha_i + \lambda_t + \sum_{k=-2}^{k=2} \gamma_k D_{it}^k + \epsilon_{it} \tag{2.2}$$

to include the coefficients δ_k measuring the anticipatory and post treatment effects separately for each year reported in Panel B of Table 2.4.[29] We find no evidence of anticipatory effects before the event. Switching into a high quintile firm leads to a persistent increase in bunching strongest in the second year after the move. In contrast, job transitions to a low avoidance environment are not associated with significant effects.

In a third specification, we restrict the sample to those individuals who switched to a high or low bunching environment and identify the effects only through the timing of the move. Specifically, we estimate

$$Y_{it} = \beta_0 + \sum_{k=-1}^{k=2} \gamma_k D_{it}^k + \theta X_{it} + \alpha_i + \lambda_t + \epsilon_{it} \tag{2.3}$$

with the variables as defined above.[30] Our coefficients of interest γ_k are reported in Panel C of Table 2.4. We find very similar results to before, emphasizing the robustness of our findings.

Matched control group In a second identification strategy, we define the comparison group for movers into a low and high knowledge environment by matching workers from the mid to mid group based on similar propensities to switch to the same environment. The matching algorithm is based on exact matches with regard to the industry and region in the period before the job switch and estimates propensity scores by a probit regression controlling for age, marital status, gender, education, and gross income in the years before and after the job transition. For each worker with a destination firm in the high (or low) quintile, we then select the comparison worker with the closest propensity score among those switching to a firm in the medium quintile. Columns (4) and (7) of Table 2.3 show that worker characteristics are now nicely balanced between the groups of analysis, even for characteristics that were not part of the matching algorithm, such as taxable income and bunching status in the pre-switch period.

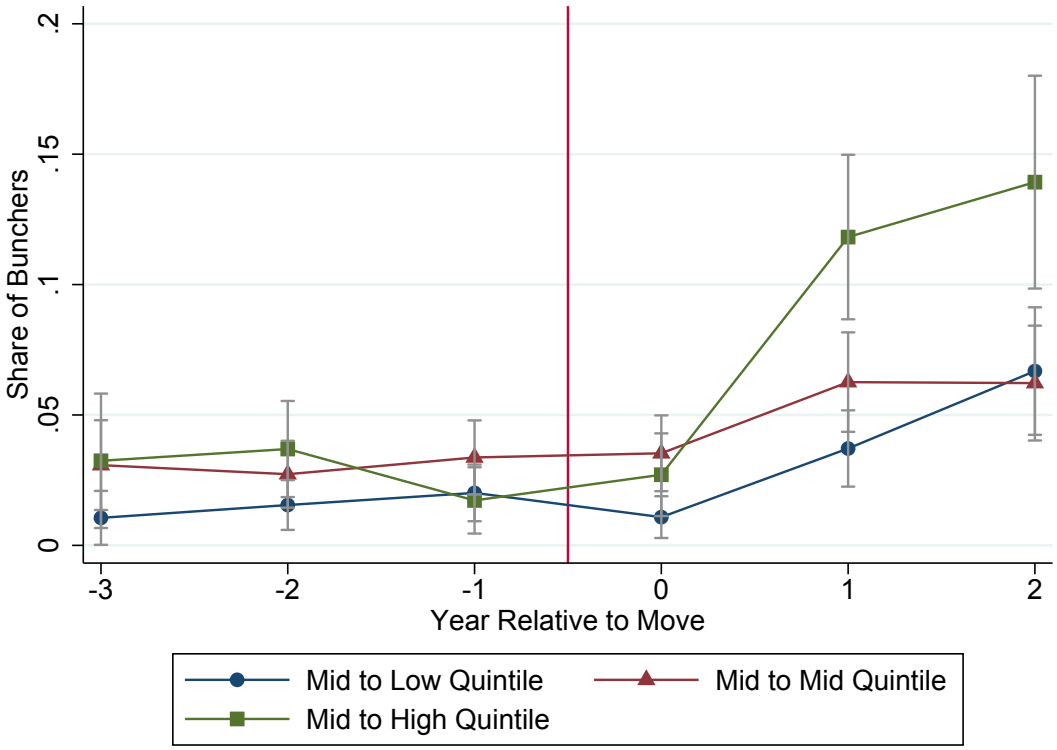
Columns (3) and (7) of Table 2.4 indicate that estimating equations (2.1) and (2.2) on the matched sample does not change the results. While moving to a low-bunching environment still does not result in a reduction of bunching, the point estimates for moving to a high-

[29] As is standard in the literature, we compare all effects to the year before the event.

[30] In order to rule out any compositional effects, we furthermore restrict the sample in this regression to only include observations from the two years before and after the move for which we have a perfectly balanced panel.

bunching environment are remarkable stable. Exposure to a high-bunching firm still leads to significant increases in tax avoidance by about 3 percentage points.

Figure 2.6 – Event Study Job Switchers – Sample of Displaced Workers



This figure shows an event study with bunching shares of job switchers around the time of the job transition in the subsample of workers who exogenously loose their job due to a firm closure.

Sample of displaced workers In our final identification strategy, we rule out strategic job mobility by restricting the sample to the subset of workers that switch their job due to a firm closure. In the spirit of Jacobson et al. (1993), this extracts the exogenous part of job mobility through job displacement. The event study graph in the sample of displaced workers (with a remaining 23,988 job transitions) is shown in Figure 2.6. It looks very similar to the full sample of job switchers. Intuitively, however, the effect is slightly delayed since displaced workers need longer to find new employment. Columns (4) and (8) of Table 2.4 report results for the same regressions as in the full sample of job switchers. Despite the much smaller sample size, results for the subsample of exogenously displaced workers are remarkably similar to the full sample, indicating that strategic mobility to specific bunching environments does not play a major role.

Our results provide robust evidence for asymmetric adjustment patterns consistent with learning and memory as have been found among self-employed in the US (Chetty et al., 2013) and commuters in Austria (Paetzold and Winner, 2016). The firm environment is

crucial in driving individual learning on bunching opportunities. Consistent results using various different identification strategies lead us to the conclusion that there is a causal relationship between the firm-level knowledge environment and individual tax avoidance. In the following section, we therefore examine dynamic learning processes on the firm level.

2.4.2 Firm Dynamics

The importance of the firm environment for individual tax avoidance behavior as well as the institutional setting in which firms directly submit tax declarations on behalf of their employees motivate a detailed study of firm dynamics. We document a strong increase in the likelihood to have bunchers in the workforce as firms gain experience in the formal sector (Section 2.4.2). Moreover, we identify two key mechanisms of information transmission between and within firms: peers and experts (Section 2.4.2).

Table 2.5 – Extensive Margin of Firm-level Bunching over time by cohort

	2008	2009	2010	2011	2012	2013	2014	2015	Obs
Cohort									
2008	0.20	0.31	0.38	0.41	0.53	0.61	0.63	0.67	489
	(0.40)	(0.46)	(0.49)	(0.49)	(0.50)	(0.49)	(0.48)	(0.47)	
2009		0.23	0.33	0.41	0.47	0.53	0.59	0.61	528
		(0.42)	(0.47)	(0.49)	(0.50)	(0.50)	(0.49)	(0.49)	
2010			0.21	0.31	0.43	0.51	0.56	0.54	555
			(0.41)	(0.46)	(0.50)	(0.50)	(0.50)	(0.50)	
2011				0.26	0.38	0.45	0.50	0.55	1100
				(0.44)	(0.49)	(0.50)	(0.50)	(0.50)	
2012					0.31	0.41	0.50	0.49	1657
					(0.46)	(0.49)	(0.50)	(0.50)	
2013						0.37	0.46	0.48	2203
						(0.48)	(0.50)	(0.50)	
2014							0.38	0.44	3280
							(0.48)	(0.50)	
2015								0.36	4847
								(0.48)	

Note: Share of firms in given cohort with at least one buncher. Cohorts conditioned on the firm's year of entry into the formal sector. Further conditioned on employing potential bunchers in all subsequent years. Standard deviations given in parentheses.

Cohort Analysis

This subsection analyzes bunching behavior through the lens of the firm by focusing on firms' experience in the formal sector. We document a strong impact of the availability of information on tax avoidance at the firm level.

We measure firm-level information on tax adjustment opportunities by looking at the number of employees bunching at a given firm. To do so, we define *potential bunchers* as individuals with gross earnings in a range allowing them to lower their taxable income below the exemption threshold by using deductions. In 2013 real USD, this is gross earnings between 10180 and 20360 USD. Analogously to the individual level cohort analysis in section 2.4.1, we follow cohorts of firms after they first appeared in the formal sector.[31] Table 2.5 reports the share of firms with at least one buncher among the potential bunchers for each year and cohort. Evidently, there is a strong increase in the share of firms that employ bunchers over time for each of the cohorts. Moreover, new cohorts start at higher bunching levels than previous cohorts. Lastly, within a given year, firms which entered the formal sector earlier exhibit higher bunching levels. We interpret this as evidence that the increase in bunching activity at the firm level is driven by experience and knowledge acquired in the formal sector and is not just a result of the general increase in bunching activity over time.

Table 2.6 focuses on the share of bunchers within a firm *conditional* on the firm having at least one buncher. This share is calculated as the number of bunchers relative to the number of potential bunchers.[32] As before, we group these firms by cohorts of entry into the formal sector. In general, the share of bunchers conditional on any bunching at the firm lies between 25 and 35 %. Notably, this share does not increase considerably with experience.

In summary, the increase in overall bunching levels is primarily driven by new firms entering the set of bunching firms. Experience of the firm in the formal sector leads to a higher probability to engage in bunching at the firm level. Given that a firm has taken the decision to allow for bunching, a relatively stable fraction of workers (around 30 percent) makes use of tax avoidance opportunities. In order to gain a more detailed understanding into what drives these firm-level decisions to start bunching, the following section analyzes how information spreads between and within firms.

Channels of Information Transmission

In this section, we characterize the channels of information transmission underlying the information flows between workers and firms. We focus on two specific channels we can identify in the data: Peers and experts. The peers channel, specifically information transmission from new co-workers towards incumbent workers, represents an important aspect of changes in the information environment at a given firm. We hypothesize that co-workers who were bunching in their previous firm induce their new colleagues to engage in bunching themselves. The experts channel focuses on the role of accountants. Here we hypothesize

[31] We restrict our sample to firms that employed potential bunchers throughout all years since their first appearance in the formal sector.

[32] We restrict the analysis to firms with at least five potential bunchers such that the share is not driven by a large number of firms with very few potential bunchers.

Table 2.6 – Intensive Margin of Firm-level Bunching over time by firm cohort

		2008	2009	2010	2011	2012	2013	2014	2015
Cohort									
2008	Share	0.23	0.25	0.28	0.27	0.28	0.31	0.31	0.33
	SD	(0.20)	(0.20)	(0.25)	(0.20)	(0.22)	(0.23)	(0.23)	(0.24)
	Obs	21	58	86	100	142	165	195	187
2009	Share		0.26	0.26	0.24	0.29	0.29	0.28	0.27
	SD		(0.23)	(0.21)	(0.20)	(0.23)	(0.21)	(0.22)	(0.22)
	Obs		32	66	92	107	126	154	147
2010	Share			0.26	0.30	0.28	0.30	0.32	0.32
	SD			(0.14)	(0.22)	(0.23)	(0.22)	(0.25)	(0.24)
	Obs			23	60	74	109	134	127
2011	Share				0.32	0.30	0.30	0.33	0.34
	SD				(0.24)	(0.23)	(0.21)	(0.24)	(0.24)
	Obs				45	100	149	196	208
2012	Share					0.29	0.29	0.32	0.31
	SD					(0.22)	(0.21)	(0.23)	(0.24)
	Obs					60	124	209	224
2013	Share						0.34	0.34	0.37
	SD						(0.26)	(0.25)	(0.27)
	Obs						71	170	194
2014	Share							0.38	0.36
	SD							(0.27)	(0.27)
	Obs							99	165
2015	Share								0.36
	SD								(0.26)

Note: Share of bunchers among potential bunchers in given cohort, conditional on firms employing at least one buncher. Cohorts conditioned on the firm's year of entry into formal sector and having potential bunchers in all subsequent years. Further conditioned on firms employing at least 5 potential bunchers in given year. The number of observations varies between year of observation since the conditioning on having at least one buncher leads to yearly changing compositions of the cohort. Standard deviations given in parentheses.

that accountants previously working for a firm with bunching activity might bring knowledge about tax avoidance opportunities to their new firm. We identify the effect of these channels through changes in the co-worker environment and accountant switches. Moreover, we shed light on the anatomy of information flows within a firm by differentiating incoming co-workers according to their relative position within their destination firm's distribution of wages.

We draw on the same panel of firms used in the cohort analysis in Section 2.4.2. However, we restrict ourselves to the subsample for which we have data on the corporate tax

declarations and thereby an identifier for the accountant.[33] We quantify the effect of the information transmission channels by estimating various linear probability models where we regress our measure of tax avoidance at the firm level on indicators whether the firm employs knowledgeable co-workers and/or accountants. In particular, we estimate variants of the following regression equation:

$$Y_{jt} = \beta_0 + \beta_1 \textit{co-worker bunch}_{jt} + \beta_2 \textit{co-worker bunch} \times \textit{above p90}_{jt} + \beta_3 \textit{accountant bunch}_{jt} + \gamma X_{jt} + \alpha_j + \lambda_t + \epsilon_{jt} \quad (2.4)$$

The outcome variable Y_{jt} is an indicator for firm j capturing whether one or more of its employees is bunching at time t. The variable $\textit{co-worker bunch}_{jt}$ is an indicator for a firm having an employee who was bunching at the previous employer.[34] The incoming buncher variable is interacted with an indicator, $\textit{above p90}_{jt}$, that is equal to one in case the incoming worker earns a wage in the 90th percentile of the destination firm's wages distribution. The indicator variable $\textit{accountant bunch}_{jt}$ takes on the value of one whenever a firm's accountant was working for a different firm with bunching activity in the periods prior to the current one.[35]

Throughout these regressions, we include year fixed effects (λ_t) and control for a range of time-varying firm level variables X_{it}. These include demographic employee characteristics like average age, share of married employees, share of female workers and share of workers with tertiary education. We also control for average gross income levels at a firm, indicators for fixed groups of firm size, industry and region (province) indicators, and an indicator for whether a given firm has employed bunchers in previous years.

Table 2.7 reports the results on the information transmission channels. Columns (1) through (3) quantify the effects of information transmission through peers and columns (4) and (5) consider information transmission through accountants. The remaining five columns represent the same specifications, but additionally include firm fixed effects (α_j), thereby controlling for unobserved firm-level heterogeneity. In these specifications the identifying variation derives from *changes* in the peer composition and *switches* in the accountants of a given firm.

Having an incoming employee who was bunching previously is associated to an increase

[33] About one fifth of the firms used in this panel do not need to file corporate tax declarations and thereby do not have official accountants. These are generally smaller firms for which it would in any case be more difficult to have enough variation to identify the channels of information transmission.

[34] We only consider incoming co-workers who were bunching in the year before joining their current firm and had gross income in the range for potential bunchers. Moreover, the $\textit{co-worker bunch}_{jt}$ indicator is equal to one in all periods in which this incoming buncher remains at the destination firm.

[35] Note that, as explained in Section 2.3, accountants can work for several firms at the same time. In this case even a single accountant at a given firm can differ over time in terms of his knowledge about bunching.

Table 2.7 – Information Transmission: Extensive Margin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Incoming Buncher	0.095***	0.052***	0.0051		0.050***	0.089***	0.040	0.029		0.041
	(0.013)	(0.019)	(0.025)		(0.019)	(0.024)	(0.032)	(0.040)		(0.032)
Incoming Buncher above p90		0.084***	0.12***		0.082***		0.11***	0.12**		0.11**
		(0.024)	(0.028)		(0.024)		(0.043)	(0.048)		(0.043)
Incoming Buncher between p50 and p90			0.088***					0.023		
			(0.029)					(0.055)		
Knowledgeable Accountant				0.054***	0.053***				0.054***	0.054***
				(0.0060)	(0.0060)				(0.0096)	(0.0096)
Avg. Age	-0.0021***	-0.0021***	-0.0021***	-0.0021***	-0.0020***	-0.0031*	-0.0030*	-0.0030*	-0.0031**	-0.0030*
	(0.00054)	(0.00054)	(0.00054)	(0.00054)	(0.00054)	(0.0016)	(0.0016)	(0.0016)	(0.0016)	(0.0016)
Share Female	0.027**	0.027**	0.027**	0.028**	0.028**	-0.048	-0.047	-0.047	-0.050	-0.046
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)
Between 25 and 250 Employees	0.071***	0.072***	0.071***	0.079***	0.073***	0.078***	0.079***	0.079***	0.080***	0.079***
	(0.0067)	(0.0067)	(0.0067)	(0.0066)	(0.0066)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
More than 250 Employees	0.20***	0.19***	0.19***	0.22***	0.19***	0.26***	0.25***	0.25***	0.27***	0.26***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
Manufacturing	0.046***	0.045***	0.045***	0.048***	0.044***					
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)					
Construction	0.037**	0.037**	0.037**	0.038**	0.037**					
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)					
Trade; Repairing	0.055***	0.055***	0.055***	0.056***	0.055***					
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)					
Hotel and Restaurant	0.042*	0.041*	0.042*	0.039*	0.038					
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)					
Transport, Storage, Communication	0.032**	0.032**	0.032**	0.036**	0.033**					
	(0.016)	(0.016)	(0.016)	(0.015)	(0.015)					
Financial Sector	0.081***	0.081***	0.081***	0.085***	0.084***					
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)					
Real Estate, Business and Renting	0.047***	0.047***	0.047***	0.047***	0.046***					
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)					
Education	0.036*	0.036*	0.036*	0.037*	0.037*					
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)					
Health and Social Services	0.032*	0.030*	0.031*	0.033*	0.031*					
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)					
Other	0.040***	0.040***	0.040***	0.043***	0.041***					
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)					
Firm FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	35025	35025	35025	35025	35025	35025	35025	35025	35025	35025

Note: The outcome variable is a binary indicator for a firm having at least one buncher. Further unreported controls: share married, share with tertiary education, average gross income, lagged bunching behavior, as well as year and province fixed effects. Standard errors clustered at firm level, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

in the probability that any of a firm's employees bunch by about 9 percentage points. With on average 42.9 percent of the firms in this sample employing bunchers, this is a strong effect corresponding to an increase in bunching activity by about 21 percent. Most of this effect is driven by employees taking up jobs relatively high in the destination firm's wage distribution. When including an indicator for incoming bunchers above the 90th percentile, the overall effect of incoming bunchers is strongly diminished and becomes insignificant, but the interaction with high-wage earning incomers is strong and significant. Bunchers joining their new firm between the 50th and the 90th percentile also have a positive impact on their coworker's bunching behavior, however, this effect is weaker than for incoming bunchers in the 90th percentile. This lets us conclude that the spread of information within a firm can be characterized through a "top-down" learning process.

Top-down learning also determines the effects of the experts in charge of accounting at a firm. Periods in which a firm has a knowledgeable accountant are associated with increases in bunching of about 5.4 percentage points, which corresponds to an increase in firm-level bunching by about 13 percent. The effects remain virtually unchanged when including both types of knowledge flows (peers and experts) simultaneously in columns (5) and (10) indicating that these are two separate mechanisms.

In order to get a grasp of what types of firms help their employees bunch, we draw on our rich firm-registry data to characterize bunching firms according to observables. The first five columns of Table 2.7 show the effects of time-varying and time-invariant observables at the firm level on bunching behavior. Firms with younger and more female workers are more likely to engage in bunching. Larger firms' employees are also more likely to bunch. Industry affiliation seems to play an important role in determining a firm's bunching activity. It is remarkable that the strongest positive coefficient belongs to firms in the financial sector, as we expect their employees to be most knowledgeable about tax adjustment opportunities.

Table 2.8 shows results from a similar set of regression using the subset of firms with at least one buncher. As outcome variable we now use the share of bunchers among potential bunchers. This outcome is thereby conditional on bunching already happening at the firm and is our previously introduced measure of the firm-level intensive margin bunching behavior.[36] Especially in our robust specifications including firm fixed effects, almost all of our estimates of the channels are very small and insignificant. We take this as evidence that neither peers nor experts have an effect on the intensive margin bunching level conditional on a firm already employing bunchers. This is in line with our results in Section 2.4.2 showing that the strong overall increases in bunching can be attributed to firms joining the group of bunching firms and not to an increase in the intensity of bunching at firms already employing some bunchers.

[36] Akin to the sample restrictions in Table 2.6, we focus on firms with at least 5 potential bunchers in order to abstract from very small firms with high variability in bunching shares conditional on bunching.

Table 2.8 – Information Transmission: Intensive Margin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Incoming Buncher	-0.027***	-0.011	-0.0042		-0.011	-0.018	-0.013	0.0052		-0.013
	(0.0085)	(0.013)	(0.016)		(0.013)	(0.012)	(0.017)	(0.021)		(0.017)
Incoming Buncher above p90		-0.029*	-0.035**		-0.030*		-0.010	-0.025		-0.010
		(0.015)	(0.017)		(0.015)		(0.023)	(0.025)		(0.023)
Incoming Buncher between p50 and p90			-0.012					-0.042*		
			(0.018)					(0.022)		
Knowledgeable Accountant				0.0079	0.0089				-0.00012	0.00026
				(0.0064)	(0.0064)				(0.0089)	(0.0088)
Avg. Age	-0.0011	-0.0011	-0.0011	-0.00088	-0.0011	0.00096	0.00092	0.00084	0.0011	0.00092
	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0030)	(0.0030)	(0.0030)	(0.0030)	(0.0030)
Share Female	0.063***	0.063***	0.063***	0.065***	0.063***	0.042	0.041	0.042	0.046	0.041
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)
Between 25 and 250 Employees	-0.089***	-0.089***	-0.089***	-0.092***	-0.089***	-0.011	-0.012	-0.011	-0.011	-0.012
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
More than 250 Employees	-0.19***	-0.19***	-0.19***	-0.20***	-0.19***	-0.037	-0.037	-0.036	-0.039	-0.037
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Manufacturing	-0.013	-0.013	-0.013	-0.015	-0.013					
	(0.016)	(0.016)	(0.017)	(0.017)	(0.016)					
Construction	0.013	0.013	0.013	0.012	0.013					
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)					
Trade; Repairing	0.024	0.025	0.025	0.024	0.025					
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)					
Hotel and Restaurant	0.0062	0.0069	0.0066	0.0051	0.0062					
	(0.028)	(0.027)	(0.027)	(0.028)	(0.027)					
Transport, Storage, Communication	0.0099	0.0100	0.0098	0.0091	0.0098					
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)					
Financial Sector	0.036	0.035	0.035	0.037	0.036					
	(0.025)	(0.025)	(0.026)	(0.026)	(0.025)					
Real Estate, Business and Renting	0.015	0.015	0.015	0.014	0.015					
	(0.016)	(0.016)	(0.016)	(0.017)	(0.016)					
Education	-0.0097	-0.010	-0.010	-0.011	-0.011					
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)					
Health and Social Services	-0.036*	-0.035	-0.035	-0.037*	-0.035					
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)					
Other	-0.014	-0.014	-0.014	-0.015	-0.014					
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)					
Firm FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	7264	7264	7264	7264	7264	7264	7264	7264	7264	7264

The outcome variable is the share of bunchers among potential bunchers. Further controls: lagged bunching behavior, share married, share with tertiary education, average gross income at firm, as well as year and province fixed effects. Standard errors clustered at firm level, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

To summarize, peers and experts play a crucial role in transmitting information between firms and are a key factor in explaining the rise in firm-level extensive margin bunching shares. Moreover, anatomy of information flows within firms shows that information is passed by managers and accountants in a “top-down" manner.

Peers This section causally identifies the peers channel by looking at individuals with recent changes to their co-worker composition. We compare a treatment group of firms with incoming bunchers to a suitable control group and analyze how knowledgeable co-workers affect the behavior at their new firm. We find strong spillover effects of new co-workers on the probability that incumbent co-workers will bunch.

Specifically, we construct a sample of firms with incoming employees who were potential bunchers due to their gross income in the year before joining the new firm. We only consider firms hiring new workers once in the years 2010-2014 and in which we can observe at least two years before and two years after the event. These restrictions provide a sample balanced in event time and allow us to abstract from various treatments happening sequentially. Among the firms with incoming potential bunchers, we divide the new employees into those that reduced their taxable income to just below the exemption threshold (“bunchers")[37] and those that did not in the year *before* joining the new firm. We use this distinction to classify firms into “treatment" (receiving bunchers) and “control" (receiving non-bunchers) groups.

Table 2.9 provides descriptive statistics for the workers in this sample of firms. Along key demographic variables, the full sample (all firms receiving incoming co-workers) is very similar to the treated group. However, as shown in Column (3), there are significant differences between the treatment and control group in terms of gross and taxable income both before and after the incoming event. To account for these differences, we create a matched control group, to which the differences disappear (Column 4).[38]

Using a similar event study methodology as in Section 2.4.1, we plot the share of firms with bunchers among their incumbent workers in both treatment and control group relative to the year of hiring the new co-worker. By focusing only on the incumbent workers, we effectively calculate the “leave-out" version of our previous firm-level probability to bunch. This indicator disregards the incoming co-worker and focuses only on the employees already working at a given firm. The results in Figure 2.7 suggest that incoming workers have a strong effect on the tax adjustment behavior of their co-workers. Firms in the treat-

[37] We again take an interval of 1000 USD to the left of the first kink.

[38] The matching algorithm employed here is a mirrored version of the one employed in Section 2.4.1 but at the firm level. The algorithm uses exact matches regarding industry by region cells in the year before the event and estimates propensity scores for being in the treatment group based on the non-outcome variables average age, share married, share female, share tertiary educated, firmsize, corporate status of the firm and average gross income pre and post event. Balance is excellent.

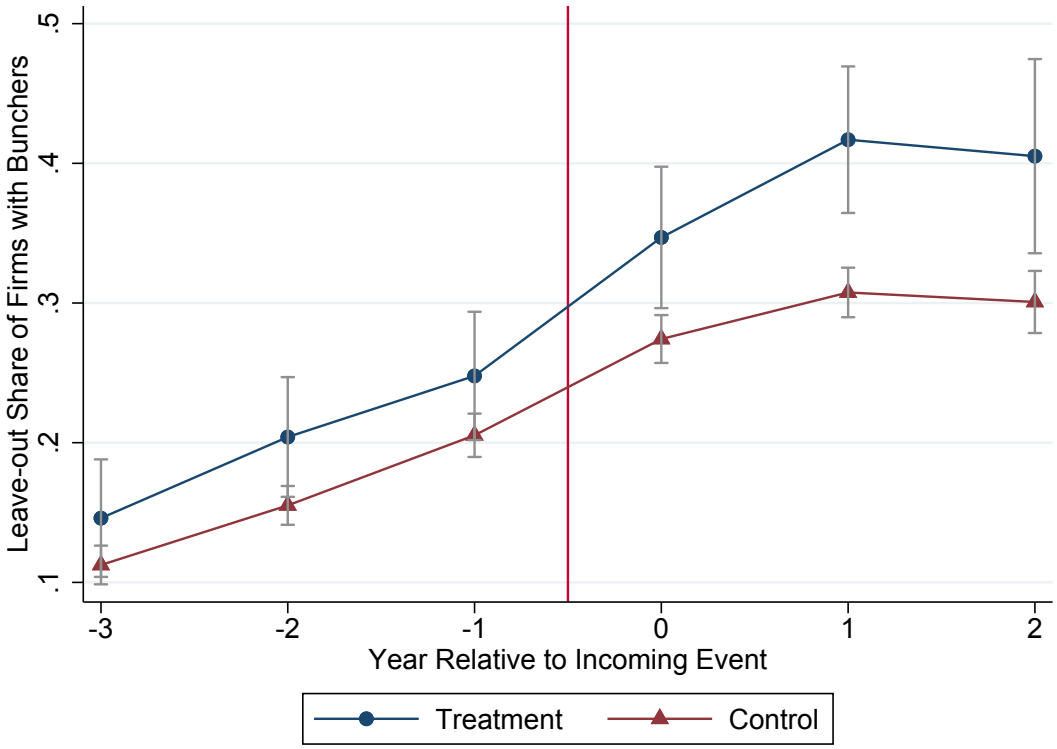
Table 2.9 – Peer Learning Event Study - Descriptives

	Descriptive Statistics			
	(1) Full Sample	(2) Treated	(3) Diff	(4) Matched Diff
Demographics				
Avg Age	36.02	36.21	0.21	-0.01
	(6.03)	(5.54)	(0.35)	(0.44)
Share Married	0.52	0.53	0.01	-0.00
	(0.24)	(0.22)	(0.01)	(0.02)
Share Female	0.37	0.40	0.03	-0.02
	(0.27)	(0.27)	(0.02)	(0.02)
Share Tertiary Education	0.32	0.33	0.01	0.02
	(0.26)	(0.26)	(0.02)	(0.02)
Pre-Event				
Firmsize	50.74	51.24	0.57	1.20
	(120.13)	(108.95)	(6.90)	(7.86)
Corporate Firm	0.85	0.88	0.04	-0.01
	(0.36)	(0.32)	(0.02)	(0.02)
Avg Gross Income	6903.01	7748.11	956.12	302.87
	(4052.51)	(4918.06)	(232.12)	(383.09)
Avg Taxable Income	6231.00	6902.77	760.02	197.43
	(3177.46)	(3722.10)	(181.98)	(300.44)
Share with Bunchers	0.21	0.25	0.04	0.03
	(0.41)	(0.43)	(0.02)	(0.03)
Post-Event				
Avg Gross Income	7761.76	8330.30	643.22	338.10
	(3949.91)	(4244.95)	(226.58)	(336.69)
Avg Taxable Income	6925.06	7258.80	377.58	96.83
	(3073.59)	(3032.26)	(176.42)	(253.22)
Share with Bunchers	0.28	0.35	0.07	0.08
	(0.45)	(0.48)	(0.03)	(0.04)
Observations	2,954	343		

Notes: This table shows descriptive statistics for the sample of firms used in the event study quantifying the peer learning channel. The sample consists of all firms receiving one incoming employee between 2010 and 2014 and for which it is possible to observe at least two consecutive years before and after the event. Treated refers to firms receiving incoming potential bunchers that bunched prior to joining their new firm. Column (3) displays the difference between treated and control and column (4) this same difference for the matched sample. Matching was done on average age, share married, female and tertiary educated, firmsize, corporate status of firm and average gross income pre and post event. Pre-event refers to the year before the arrival of new co-workers and post-event to the first year after the arrival of the new coworkers.

ment group are much more likely to have bunchers among their incumbent employees after receiving a new co-worker.

Figure 2.7 – Peer Learning Event Study



This figure shows an event study on the firm level with the share of firms employing bunchers around the hiring of a new co-worker (leaving out the new worker from the calculation). The vertical line denotes the arrival of the new worker. The treatment group is formed by firms that receive a new co-worker who was bunching in her previous firm while the control group is formed by firms with a new co-worker who was not bunching (despite being a potential buncher with gross income in the range above the kink).

Table 2.10 provides regression results for the previous graphic evidence. With the aim of addressing possible selection issues and quantifying the magnitude of the effects, we mirror the identification strategies employed in Section 2.4.1. Specifically, we estimate

$$Y_{jt} = \beta_0 + \delta post_{jt} \times treat_j + \theta X_{ij} + \alpha_j + \lambda_t + \sum_{k=-2}^{k=2} \gamma_k D_{jt}^k + \epsilon_{jt}. \tag{2.5}$$

where Y_{jt} is an indicator for whether there is bunching activity among incumbent workers, $post_{jt}$ is an indicator for observations after the new co-worker joined the firm, $treat_j$ is an indicator for a firm receiving an incoming buncher. We include fixed effects at the firm (α_j), time (λ_t) and event-time (D_{jt}^k) level and in X_{jt} we control for firm-level characteristics (firmsize, average gross income, corporate status, and industry and province dummies)

as well as employee characteristics (average income, share tertiary educated, average age, share married, and share female).

Table 2.10 – Peer Learning - Regression Results

	(1)	(2)	(3)
	Full Sample		Matching
A. Overall Effect			
DiD estimate	0.052**	0.047**	0.069**
	(0.022)	(0.022)	(0.029)
B. Effects by Relative Year			
Anticipatory Effects			
Event year - 2	0.010	0.012	0.003
	(0.024)	(0.024)	(0.033)
Post Treatment Effects			
Event year	0.035	0.031	0.039
	(0.028)	(0.028)	(0.037)
Event year + 1	0.072**	0.067**	0.105***
	(0.031)	(0.030)	(0.040)
Event year + 2	0.065*	0.061	0.063
	(0.039)	(0.039)	(0.051)
Controls	No	Yes	Yes
Observations	15,913	15,913	3,696

Notes: The table reports results from the event-study regression equations (2.5) and (**??**) at the firm level. Outcome variable is the leave-out firm bunching decision and event year refers to the year of incoming employees. Event year - 1 is excluded and serves as the base category. Firm and year fixed effects are included throughout. Columns (1) and (2) refer to the full sample, and column (3) uses matching on observables. We control for average gross income, average age, share married, share female, share tertiary educated, firmsize, corporate status of firm, as well as industry and province dummies and dummies for the year of the incoming event. Standard errors (in parentheses) are clustered at the firm level. Significance levels are given by * < 0.1, ** < 0.05, and *** < 0.01.

The results are displayed in Panel A of Table 2.10. Parallel to the analysis in Section 2.4.1, we examine effects separately for each year relative to the job transition in Panel B. Even when controlling for unobserved heterogeneity and a rich set of observables, the peer learning channel is strong and pronounced (Column 2). Moreover, when using the matched

control group, we find even stronger effects (Column 3). An incoming buncher increases the probability that at least one of the incumbent co-workers bunches by about 5 to 7 percent. The effects are strongest in the second year after the incoming event, consistent with the idea that it takes some time for incoming co-workers to spread the information to the new firm environment.[39] In the appendix we conduct a heterogeneity analysis by firm size. Figure B.6 depicts the same event study separately for small, medium and large firms. Intuitively, we find the effect of co-workers on their peers to be largest for small firms and to become smaller for larger firms.

Experts We now focus our attention on the accountant channel. In a similar event study design exploiting variation in the knowledge of accountants, we find causal evidence for the effect of accountants on firm-level bunching behavior.

We assess whether firm-level bunching behavior changes after firms receive new accountants. Like a new co-worker, a new accountant changes the information environment at a firm. Firms that receive a knowledgeable accountant who was previously working at a firm with bunchers constitute the treatment group. Firms in the control group also receive a new accountant, but this new accountant was previously working for firms without bunchers even though those firms had employees with gross income in the relevant range for bunching (potential bunchers).

We extract the universe of accountant switches observed in the corporate tax declarations. We then analyze how accountant switches have an impact on whether a firm engages in bunching activity.[40] Table 2.11 shows descriptive statistics for the firms in the experts event study. Treatment and control firms are similar along key demographic variables but show significant differences in income variables before and after the accountant switch. Using our matched algorithm, however, we achieve nearly perfect balance.[41]

Figure 2.8 graphically depicts the experts event study. The vertical axis denotes the average firm-level bunching share among treatment and control group respectively. The horizontal axis denotes event time relative to the year of the incoming accountant (year 0). We observe stable pre-trends between treatment and control group before the new accountant enters the firm. In the first year after the accountant switch we observe a clear

[39] In unreported results we additionally identify the peer channel within the sample of treated firms purely through the timing of the effect akin to the regression strategy in equation (2.3) and find very similar results.

[40] Much like in our previous event study analyses, we make a number of restrictions to guarantee tractability and credibility of the results. We exclude cases were firms simultaneously received knowledgeable and non-knowledgeable accountants. We further restrict our analysis to firms where we can observe at least two consecutive years before and after the accountant switch. Moreover, we focus on switches happening in 2010 or later so that in both years before the switch bunching was a viable option. This leaves us with a sample of 16,389 accountant switches.

[41] The matching algorithm in the experts event study is exactly like in the peer learning event study: exact matching on industry and region cells, with ensuing propensity score matching using average age, share married, share female, share tertiary educated, firmsize, and average gross income pre and post event.

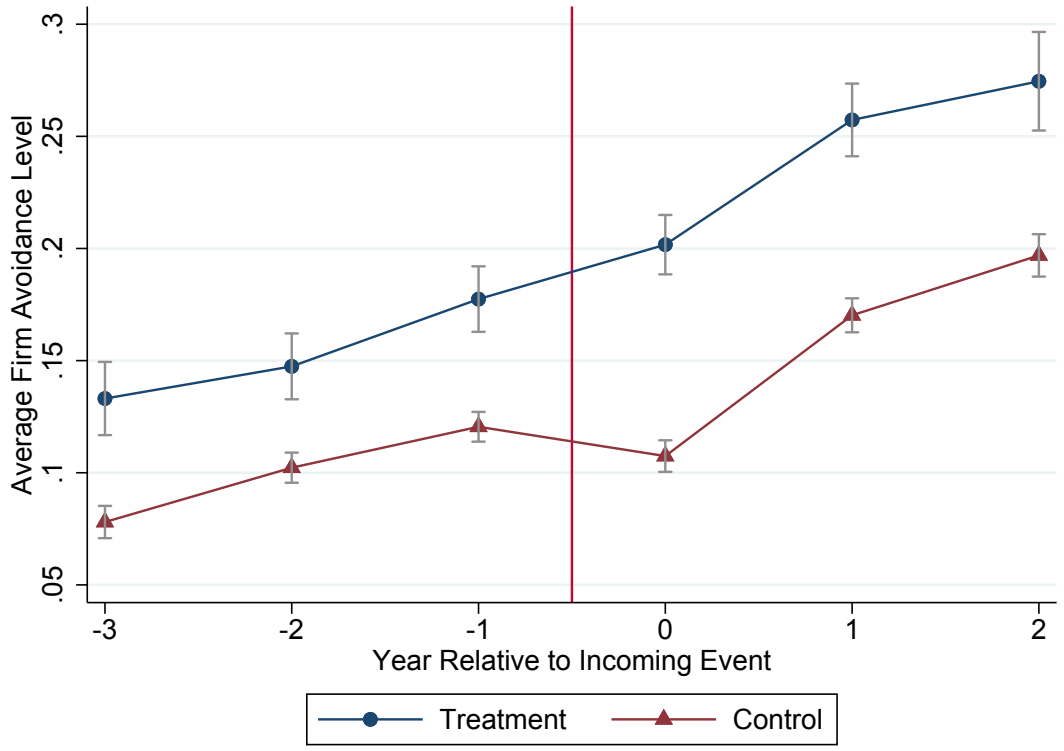
Table 2.11 – Experts Event Study - Descriptives

	Descriptive Statistics			
	(1)	(2)	(3)	(4)
	Full Sample	Treated	Diff	Matched Diff
Demographics				
Avg Age	36.00	36.03	-0.07	0.19
	(7.40)	(7.25)	(0.16)	(0.20)
Share Married	0.47	0.47	-0.01	0.01
	(0.28)	(0.27)	(0.01)	(0.01)
Share Female	0.41	0.41	0.00	-0.01
	(0.30)	(0.30)	(0.01)	(0.01)
Share Tertiary Education	0.28	0.28	-0.00	0.01
	(0.28)	(0.27)	(0.01)	(0.01)
Pre-Event				
Firmsize	43.99	53.73	13.25	9.28
	(178.40)	(248.55)	(3.91)	(5.76)
Avg Gross Income	5217.38	5932.17	894.75	-9.81
	(4529.22)	(5038.55)	(100.07)	(144.85)
Avg Taxable Income	4766.13	5316.95	683.17	-23.41
	(3671.01)	(4066.28)	(81.31)	(117.07)
Share with Bunchers	0.13	0.18	0.06	0.02
	(0.34)	(0.38)	(0.01)	(0.01)
Post-Event				
Avg Gross Income	5227.81	6026.00	1025.94	29.23
	(4668.04)	(5393.24)	(91.34)	(146.31)
Avg Taxable Income	4770.90	5370.95	773.96	6.58
	(3837.20)	(4298.05)	(75.20)	(117.86)
Share with Bunchers	0.14	0.20	0.08	0.05
	(0.35)	(0.40)	(0.01)	(0.01)
Observations	16,389	3,337		

Notes: This table shows descriptive statistics for the sample of firms used in the event study quantifying the experts channel. The sample is based on the universe of accountant switches between 2010 and 2014 for which it is possible to observe at least two consecutive years before and after the event and the firms employ potential bunchers throughout. Cases in which firms simultaneously received knowledgeable and non-knowledgeable accountants were excluded. Treated refers to firms receiving new accountants previously working at a firm with bunching employees. Columns (3) displays the difference to a control group consisting of firms receiving an accountant previously working at a firm with potential bunchers but with zero bunching employees. Column (4) displays the difference to the matched sample. Matching was done on average age, share married, female and tertiary educated, firmsize, and average gross income pre and post event. Pre-event refers to the year before the arrival of the new accountants and post-event to the first year after the arrival of the new accountants.

difference between treatment and control firms. Control firms seem to have a significantly lower propensity to employ bunchers. However, in the second and third year at the new firm this effect is harder to distinguish.

Figure 2.8 – Experts Event Study



This figure shows an event study on the firm level with the average share of bunchers around the entry of a new accountant into the firm. The vertical line denotes the arrival of the new accountant. The treatment group is formed by firms that receive an accountant who was previously working for a firm with bunchers while the control group is formed by firms with a new accountant who was working for a firm without any bunchers.

Table 2.12 denotes regression results from event-study type regressions analogous to those in the previous section. The notable exception is that the outcome variable is now the firm-level bunching decision and the treatment indicator $treat_j$ indicates firm j receiving a knowledgeable accountant. Switching towards a knowledgeable accountant is clearly associated to a strong increase in the amount of bunching at a firm. Receiving a knowledgeable accountant increases firm-level bunching by about 2.5 percent, even when including extensive control variables and using the matched control group.

Table 2.12 – Experts Event Study - Regression Results

	(1)	(2)	(3)
	Full Sample		Matching
A. Overall Effect			
DiD estimate	0.103***	0.024***	0.025*
	(0.007)	(0.008)	(0.013)
B. Effects by Relative Year			
Anticipatory Effects			
Event year - 2	-0.018**	-0.010	0.001
	(0.008)	(0.009)	(0.015)
Post Treatment Effects			
Event year	0.069***	0.021**	0.028*
	(0.008)	(0.009)	(0.016)
Event year + 1	0.115***	0.024**	0.023
	(0.010)	(0.011)	(0.018)
Event year + 2	0.133***	0.013	0.023
	(0.012)	(0.013)	(0.022)
Controls	No	Yes	Yes
Observations	60,483	60,483	22,485

Notes: The table reports results from the event study regressions quantifying the experts channel detailed in Section (2.4.2). Outcome variable is the firm bunching decision and event year refers to the year of the incoming accountant. Event year - 1 is excluded and serves as the base category. Firm and year fixed effects are included throughout. Columns (1) and (2) refer to the full sample, and column (3) uses matching on observables. We control for average gross income, average age, share married, share female, share tertiary educated, firmsize, as well as industry and province dummies and dummies for the year of the accountant switch. Standard errors (in parentheses) are clustered at the firm level. Significance levels are given by * $<$ 0.1, ** $<$ 0.05, and *** $<$ 0.01.

2.5 Conclusion

We analyze tax avoidance behavior using new administrative data on personal income taxes from Ecuador. Learning plays an important role in determining individual tax adjustments: as taxpayers gain experience in the formal sector, they are more likely to avoid paying taxes. Tax avoidance is driven through reporting behavior based on generous deduction possibilities. By exploiting matched employer-employee data and a research design based on exogenous job mobility we find the firm information environment to have a causal effect on individual learning processes about tax avoidance opportunities.

Furthermore, this paper exploits the strong rise in the size of the Ecuadorian formal sector to provide evidence for the importance of firm-level dynamics in tax avoidance behavior. We show that the knowledge environment at the firm-level can be characterized by a binary pattern: either a firm has knowledge about bunching opportunities or it does not. Conditional on tax avoidance at the firm level, the share of employees avoiding taxes remains relatively stable over time. The paper identifies and quantifies two specific channels of information transmission that explain the rise in firm-level knowledge on tax avoidance activity. We quantify the effects of peers and experts by exploiting changes in the co-worker composition of firms and accountant switches. Furthermore, we provide evidence for "top-down" information flows within firms.

From a policy perspective, these findings on how taxpayers in a low-enforcement setting learn about tax adjustment and avoidance opportunities are highly relevant. A range of developing and middle-income countries have recently undergone numerous reforms aiming towards the formalization of the economy. While designing these reforms it is important to take into account how and when they will translate into actual behavior, especially in a dynamically growing setting. Due to partial usage only by individuals in an advantageous knowledge environment, such reforms can also (at least in the short and medium run) increase inequality. Moreover, our analysis has shown the importance of firms and firm-level environments in driving the usage of tax avoidance opportunities. This observation is important when designing strategies to combat tax avoidance and setting up auditing targets. A flexible labor market with worker and job mobility is crucial for the spread of information and helps to reduce information frictions.

In future research on behavioral responses to public policies, we think it is important to focus more strongly on dynamic aspects. Especially in settings with a growing number of affected parties or beneficiaries, these economic agents do not respond to incentives immediately and take time to understand and learn about the system. Moreover, identifying the channels of information transmission underlying learning processes can be informative for the design of optimal policies and to guide policymakers in improving existing ones.

Chapter 3

Harnessing Deductions to Increase Tax Compliance and Formalization

3.1 Introduction

Increasing tax compliance and ensuring that an economy operates in the formal system is vital towards guaranteeing a well-functioning tax and transfer system. For many developing economies with large informal sectors (ILO, 2014), key policy goals often center on getting economic actors to participate in the formal economy. Increasing formality and tax compliance is important not only for revenue generation to finance crucial public expenditures, but also to guarantee a just and equitable tax and transfer system implementing the socially decided levels of redistribution. However, it is often unclear exactly how to improve tax compliance. Recent studies have highlighted the importance of generating paper trails and automatic third party reporting as driving factors behind high levels of tax compliance (Kleven et al., 2016). Especially in developing countries, efforts to increase and improve paper trails in VAT have been crucial in recent formalization efforts (Pomeranz, 2015; Naritomi, 2016). However, due to frequent exemptions based on firm size or industry, substantial parts of economies often operate outside of the VAT system. It is especially unclear how to put in place incentives inducing third party reporting and paper trail generation for these sectors of the economy.

This paper evaluates a unique reform aimed at radically increasing the demand for receipts by final consumers, thereby creating paper trails for sectors of the economy previously marred by tax avoidance and evasion. This demand for receipts is stimulated through extensive deduction possibilities in personal income tax (PIT), particularly in sectors exempt from VAT. Drawing on detailed administrative data from Ecuador and exploiting heterogeneity in the effectiveness of the reform both between regions and sectors, I estimate its impact on the size of the formal sector economy. According to a simple difference-in-differences esti-

mate, reported self-employment professional profits aggregated at the regional level increase between 18 and 33% per inhabitant. In triple difference regressions at the individual self-employed level additionally differentiating between professionals affected and unaffected by the reform, I estimate the reform to increase reported profits by almost 100%.

Ecuador has recently experienced a large increase in formalization as documented by a strong increase in the number of taxpayers (Bohne and Nimczik, 2018) and an increase in the tax-to-gdp ratio from 10 to 21% between 2000 and 2015 (Modica et al., 2018). One of the key drivers are extensive deduction opportunities from personal income tax (PIT) implemented in 2008. These allow taxpayers to deduct expenses[1] in health, education, housing, food and clothing, almost all of which (except the latter) are mostly exempt from VAT payments.[2] However, PIT is levied at relatively high levels above 10,000 USD. To estimate the effect of this policy on formalization, I exploit substantial regional variation in the density of high-income individuals and the actual amount of deductions claimed within 219 administrative regions within Ecuador.

In this paper, I draw on a unique combination of several administrative data sources based primarily on the universe of personal income tax records of self-employed business owners. For an estimate of the overall economic impacts of this reform, I aggregate the outcome variables at the regional level and consider per-capita (inhabitant) values. My main outcome variable is the total amount of professional profits reported within an administrative region per inhabitant. I define treated regions as those exhibiting a demand for receipts above the mean or above the 90th percentile. To validate the identifying assumption in the difference-in-differences framework, I plot two years of remarkably parallel and extremely low pre-trends. Trends in both treatment and control regions increase greatly after the reform, but the increase is much stronger in regions with a high demand for receipts. The difference-in-differences estimator corresponds to increases of 45 to 82 USD reported professional profits per inhabitant of a given region, or 18 to 33%. Remarkably, this effect is not existent when looking at profits of formalized medium to large businesses arguably less affected by the reform (sole proprietorships above certain size thresholds).

Having established the importance of the reform for aggregated regional outcomes, I find comparable and sizeable results in similar difference-in-differences regressions at the individual level while controlling for individual unobserved heterogeneity. Importantly, combining the tax returns with data on registered professions from the civil registry provides an additional level of heterogeneity depending on whether or not the services of an individual profession are included in the deduction categories and hence face increased demand for

[1] Limits apply. Please refer to Section 3.2 for details.

[2] In housing, only rental payments are exempt from VAT. Likewise, in the food category, only basic food items are exempt from VAT.

receipts. This further treatment layer allows to conduct triple difference regressions controlling for possible changes in the underlying trends between treatment and control regions arising from potential time-varying region-specific shocks. As a result, the changes in reported profits between doctors (strongly affected) and other professionals are about 5,000 USD higher in regions with a strong demand for receipts than those same changes in regions with a low demand for receipts. In placebo triple-difference regressions, I find no effects for evidently unaffected professions such as veterinary physicians, economists, business administrators or journalists.[3]

The findings of this paper are highly policy relevant in terms of uncovering innovative ways to increase tax compliance and formalization, both for developing and more developed economies. The very extensive and generous deduction policy was instrumental in getting a whole sector — professional services not subject to VAT — to participate in the formal economy. In this setting, since the self-employed themselves use the same deductions to lower their taxable income, it is unclear whether this reform actually leads to professional service providers paying more taxes. However, getting them to participate in the formal economy and creating paper trails of their business activities has additional long-term advantageous. Lastly, these deductions arguably have strong spillovers on the reporting of VAT and VAT revenue. For a comprehensive cost-benefit analysis of the reform (foregone PIT due to deductions versus increased revenue), the impact on VAT revenue would need to be factored in and this is outside the scope of this study.

This paper contributes towards several strands of the literature. In general, it is embedded into a growing literature on taxation and development. Key parts of this literature focus on explaining how and why tax systems in developing countries differ to those of more developed economies (Gordon and Li, 2009; Best et al., 2015; Keen and Slemrod, 2017). Recently, a number of papers have highlighted the rising importance of both PIT (Besley and Persson, 2013; Jensen, 2016) and VAT (de Paula and Scheinkman, 2010; Pomeranz, 2015; Waseem, 2018) for developing countries. Likewise, many papers have focused on corporate taxation and firm behavior in explaining tax avoidance and evasion in developing countries (Best, 2014; Carrillo et al., 2017; Asatryan and Peichl, 2017; Bachas and Soto, 2017). I combine these two research areas by focusing on the smallest firms, namely self-employed individuals, and looking at their behavior with respect to PIT liability.

More specifically, this paper fits to recent work looking into formalization of developing economies (Pomeranz, 2015; Gerard and Gonzaga, 2016; Naritomi, 2016; Brockmeyer et al., 2018; Rocha et al., 2018). For more developed economies, there is a consensus on the importance of third party information in sustaining high levels of taxation (Kleven et al.,

[3] I address general income and demand effects for health services by looking at a broader definition of health professionals excluding doctors who face different incentives mainly due to the fact that their clients may not be final consumers. For this group of professionals, I do not find an effect of the reform.

2011, 2016). This paper fits into a small group of studies recently attempting to increase the quantity and quality of third-party information in developing economies (Kumler et al., 2015; Pomeranz, 2015). A number of countries, including Brazil, China, Italy, and Taiwan, have recently implemented policies aimed specifically at getting consumers to report business revenues and submit receipts. The benefit to consumers usually comes in the form of lottery ticket reward policies. Fabbri (2015) looks at these from a theoretical point of view and Marchese (2009) evaluates a specific reform in China. Most closely related is the study by Naritomi (2016), evaluating a program also giving monetary rewards to consumers for demanding receipts and quantifying the compliance effects for firms. However, the monetary rewards for consumers were mostly a direct function of the VAT liability of firms. In contrast, in this paper consumers are specifically given monetary rewards for transactions largely exempt from VAT payments. And while the above papers focus on either regionally confined policies or monetary rewards through lottery schemes, this paper evaluates a nation-wide reform with strong monetary rewards for everyone paying PIT. In general, most of the above formalization efforts focus on improving VAT compliance and paper trails, whereas this reform specifically targets economic sectors exempt from VAT.

Additionally, this paper contributes to a small literature studying deductions within the personal income tax system (Doerrenberg et al., 2017; Matikka, forthcoming). Especially Hamilton (2018) looks at the optimal level of deductions in personal income taxes by decomposing observed bunching behavior into a gross income and a deductions response. In his setting, however, deductions do not have any effect on the formalization of the economy.

More generally, this paper relates to the general literature on tax compliance and evasion (Allingham and Sandmo, 1972; Slemrod, 2007; Kleven et al., 2011; Dwenger et al., 2016). Withing this literature, we focus on the subgroup of self-employed individuals who usually make strong adjustments to their reported income even in high-enforcement environments (Chetty et al., 2011; Bastani and Selin, 2014; Kleven and Waseem, 2013). On an even broader level, I contribute to a number of recent studies looking into general equilibrium or macro effects of micro-oriented tax policies. Saez et al. (2017) look into the effects of a tax rebate on aggregate employment rates and Alstadsaeter et al. (2017) look into the impacts of tax evasion on global inequality. The policy studied in this paper analyzes how individual incentives can be harnessed to drive formalization of a whole sector of the economy.

The remainder of this paper is organized as follows. Section 3.2 gives an overview of the specific deductions policy and provides information on the taxation of self-employed in Ecuador. Section 3.3 introduces the data sources and provides descriptive statistics. Section 3.4 provides details on the research design and discusses the results at the aggregated regional and individual self-employed level. Section 3.5 concludes.

3.2 Deductions Policy

Ecuador, as many countries in Latin America, has recently put a large focus on reducing tax evasion and increasing formality. Most of the Ecuadorian reforms were carried out in 2008 along with a series of policies aimed at expanding social programs, improving public service delivery, and making vital infrastructure investments. These policies were largely unchanged up to 2014 due to a long period of political stability, sustained economic growth and high oil revenues. The fiscal results can be clearly seen in terms of overall government tax revenue, which increased from 10.3% of GDP in 2000 to 21.1% in 2015 (Modica et al., 2018). Tax revenue in Ecuador can be broadly categorized into a value-added tax (VAT) of 12%, personal income taxes (PIT) levied on wage and self-employment income, corporate taxes (22% of profits), and a number of special taxes ranging from certain consumption goods to transferring money abroad.

Deductions Policy One of the strongest policies to formalize the economy are extensive deduction policies within the personal income tax (PIT) system. The PIT exemption threshold is relatively high at 10,180 USD in 2013[4], leading to a relatively small fraction of the population with income above this threshold (refer to Section 3.3 for descriptive statistics). Individuals liable to pay PIT can deduct expenses in health, education, food, clothing and housing for themselves or their dependents and thereby reduce their taxable income (Villacreses, 2014). The total deductible amount is limited to either 50% of gross income or 1.3 times the exemption threshold, whichever is smaller. Additionally, the limit of 0.35 times the exemption threshold applies to each specific deductions category, except health expenses, which are only capped in case the overall limit is reached. In order for taxpayers to claim these deductions, they are legally obliged to gather and keep the receipts used. The receipts can only be used in case they include information on the name and personal identification number of the taxpayer or his/her dependents.[5] While taxpayers are legally obliged to keep all receipts, they only need to actively present them to the tax authority in case the overall value of their deductions exceeds a reporting threshold set at 0.5 times the exemption threshold (corresponding to 5,090 USD in 2013).[6]

[4] The exemption threshold is adjusted to inflation in every year according to the yearly change in consumer price index in urban areas published by the national statistical agency INEC on every November 30. Throughout this paper, all nominal values are deflated according to this index and the exemption threshold thereby remains constant in real terms.

[5] In practice, sellers ask whether clients need a standard receipt (“nota de venta") without their name and id number or the enhanced version necessary to present to the tax authority (“factura"). [6] In the years up to and including 2010, this reporting threshold was set at the fixed amount of 7,500 USD.

Taxation of Self-Employed In general, all sources of individual income in Ecuador are subject to a unified personal income tax (PIT) scheme.[7] This includes regular wage income as well as profits from self-employed business activities. The tax scheme is progressive, with numerous tax brackets, ranging from 5% marginal tax rate for income immediately above the exemption threshold and increasing up to 35% for top incomes above 103,810 USD.

Micro-businesses in Ecuador benefit from a simplified tax regime (*RISE - Régimen Impositivo Simplificado*) replacing both personal income taxes and VAT payments through monthly lump-sum transfers depending on the reported yearly income and industry. The policy objective is to reduce compliance costs and facilitate formalization for very small enterprises. This is possible if overall yearly business income does not exceed 60,000 USD and a range of other restrictions apply. Most importantly for this analysis, all sorts of business activities arising from professional services in need of a university degree (doctors, lawyers, journalists, etc.) are excluded from using this simplified tax regime.

All individuals with self-employed business activities not using this simplified tax regime are obliged to pay PIT on their full business profits. In order to assess their tax liability, they must submit a self-employed tax declaration form. These forms include information on the different types of income and their immediate deductibles generated by the relevant business activity. Immediate deductibles are expenses made to obtain the relevant income, e.g. the costs of maintaining a doctor's practice can be deducted from the revenue a doctor generates. Throughout this paper, I look at profits generated by self-employed as the difference between income and deductible costs. The income categories reported on the tax form include business profits, professional income, rental income, agricultural income, dividends, financial returns, foreign income sources as well as income from the banana sector and the exploitation of natural resources. Crucial for this analysis will be the professional income category,[8] which contains income from all professional service providers like self-employed doctors, journalists, lawyers, etc. Additionally, the incomes and deductible costs in each of these fields are aggregated and reported as total self-employment profits generated by an individual.

The exact tax declaration form self-employed fill in depends on the size of their business activities. The extensive tax declaration form F102 is the default and contains comprehensive accounting information underlying the first subcategory "business profits" referred to above. The accounting information corresponds to a whole set of accounting figures includ-

[7] Important exceptions include payments from the social security system (pensions, stipends, disability benefits, etc.) severance payments, certain mandatory end-of-year wage benefits, and a range of very specific capital income including interests on certain saving accounts, occasional capital gains and returns from investment funds or long-term deposits (conditions apply).

[8] On the tax form, there are actually two relevant categories: Income from "libre ejercicio profesional" and income from "ocupación liberal". Since their boundaries can be unclear in some cases, for the purposes of this study, I aggregate the two income fields as "professional profits".

ing a revenue and expense statement according to Ecuadorian accounting rules. Additionally, this form needs to be co-signed by an official accountant. However, self-employed with capital below 60,000 USD and annual income below 100,000 USD can fill in the simpler tax declaration form F102A. This form effectively summarizes the full accounting information with one field named “business profits" and taxpayers are not required to provide full accounting books nor need a signature from an official accountant. The other types of self-employment income introduced above (importantly also the professional income category) remain unchanged in this simpler form F102A. Figure C.1 in the Appendix depicts the full form 102A and Figure C.2 zooms in on the relevant income and profit variables.

Apart from paying PIT, business owners in Ecuador are obliged to remit value-added tax (VAT) at a rate of 12% on all products and services they sell. As standard in a VAT system, the VAT they paid on their inputs serves as a credit towards the VAT they need to remit, effectively only charging VAT on the value-added by their enterprise. However, a number of exemptions to VAT apply. First, businesses operating under the simplified tax regime *RISE* introduced above (revenue below 60,000 USD) make a lump-sum payment liberating them from any PIT and VAT obligations — and effectively forfeiting their VAT credit. Second, a number of goods and services are completely exempt from VAT payments. This includes basic food items, rental payments, and importantly for the context of this study, any health and education services. Even though in terms of overall tax revenue VAT is much more important for Ecuador than PIT, these exemptions effectively exclude a large part of the economy from the self-enforcing paper trail mechanisms of VAT (Kopczuk and Slemrod, 2006). The deductions policy is an attempt to create a paper trail and improve compliance for this important part of the economy.

3.3 Data Sources and Descriptives

The data used in this paper is a combination of various sources of detailed administrative data from Ecuador. The core data consists of the universe of personal income tax declarations by self-employed individuals in the years 2006-2015. These self-employment tax records are merged to extensive firm-registry data (*Catastro de RUC*), where all self-employed are obliged to provide basic information on their business activities, including among other things the location within administrative districts. The relevant administrative region I focus on in this paper is the *cantón*, of which there are 221 in Ecuador.[9]

A central source of data used in this analysis is on registered professions. In Ecuador, the civil registry holds data on the registered professions of all citizens. This information

[9] In this paper, I focus solely on *cantón* as the relevant administrative district. The smaller administrative division of *parroquia* is also available in the firm-registry data, however, population and socio-demographic indicators are not differentiated by parroquia in urban areas. Moreover, throughout the analysis I use data on 219 *cantones* due to data availability issues for the remaining two.

is for example displayed on the official id card. Whenever an individual graduates from an educational institution or changes occupation, they are asked to provide this information to the civil registry. In case of professions demanding a degree from a tertiary educational institution (e.g. doctors, accountants, lawyers), this is only possible after providing documentary evidence of graduating and of the institution's accreditation. Obviously, while this process provides security against false statements of the professions, it does not guarantee that this information is continuously updated for all citizens. At birth, every individual is registered as a student (*estudiante*) which stays unchanged until they change their status (thereby this is the most frequent category). In many cases, the classification of the profession is quite fuzzy and includes terms like "employee" (*empleado*), "private-sector employee" (*empleado privado*) or "day laborer" (*jornalero*). I have coded these profession into overarching categories which are self-contained and which permit a clear mapping into effects of the deductions policy, making up about 35% of all self-employed individuals.[10] The sample of individual self-employment records used in this paper focuses on this subset of individuals. Table C.1 in the Appendix compares these individuals with clear profession categories to all other self-employed individuals. In general, the two groups are quite balanced, especially in terms of self-employment profits (8,193 versus 8,045 USD) and demographic characteristics. There are, however, significant mechanical differences in terms of educational achievement: Individuals belonging to these clearly specified professions are much more likely to have finished tertiary-level education (75% versus 27%). Since most of these clearly specified professions need an official degree, obviously the share of tertiary-educated individuals in this subset is larger. The resulting dataset includes a strongly increasing number of self-employed individuals over time, ranging from 142,190 in 2007 to 211,184 in 2015.

There are additional sources of data used for gaining information about the administrative regions (*cantón*) used in this paper. A key element stems from the universe of personal income tax declarations of wage earners. This data is used to aggregate information on key individual income and deduction indicators at the regional level. Specifically, I calculate the number of individuals with income above the income tax exemption threshold and the dollar amount of claimed deductions by taxpayers. Additionally, I take aggregate data from the Ecuadorian Census[11] to gain population figures by cantón. Throughout this analysis, all dollar values are expressed in 2013 real USD. This ensures that the income tax exemption

[10] In total, the Ecuadorian civil registry lists almost 2600 different professional categories. Many of them are, however, extremely specific ("pediatrician") or very unspecific, as mentioned above. The profession I focus on make up about 35% of the overall sample and include the categories of doctors, dentists, other health professionals (see details in Table C.4), lawyers, architects, engineers, drivers, economist, accountants, teaching industry professionals, tailors/shoemakers, business administrators, journalists, and food industry professionals. While there are numerous very infrequent categories, another 32% of the sample is made up of the three extremely unspecific professions categories "student", "private sector employee" and "employee".

[11] Full population-wide Census from 2010, administered by INEC (*Instituto Nacional de Estadística y Censos*).

Table 3.1 – Descriptive Statistics Aggregates

	(1) Above Mean	(2) Below Mean	(3) Above p90	(4) Below p90
Population	151,643	34,114	312,459	39,614
	(416,957)	(41,359)	(6710,595)	(52,525)
2007				
Total Deductions per capita	0.0005	0.0005	0.0014	0.0004
	(0.0028)	(0.0062)	(0.0046)	(0.0056)
Share Population w/ High Income	0.0129	0.0019	0.0235	0.0029
	(0.0172)	(0.0017)	(0.0250)	(0.0033)
2009				
Total Deductions per capita	66.85	5.99	129.00	10.78
	(84.00)	(5.28)	(117.85)	(11.28)
Share Population w/ High Income	0.0175	0.0027	0.0318	0.0040
	(0.0197)	(0.0020)	(0.0272)	(0.0040)
2012				
Total Deductions per capita	82.50	14.43	149.56	20.38
	(89.76)	(11.88)	(124.22)	(18.03)
Share Population w/ High Income	0.0217	0.0044	0.0388	0.0059
	(0.0234)	(0.0032)	(0.0325)	(0.0048)
2015				
Total Deductions per capita	117.02	28.74	211.73	35.64
	(130.02)	(29.98)	(175.78)	(35.82)
Share Population w/ High Income	0.0273	0.0064	0.0489	0.0081
	(0.0335)	(0.0056)	(0.0476)	(0.0078)
Observations	59	160	21	198

Notes: This table presents descriptive statistics at the regional (*cantón*) level. The 119 regions studied (two cantones drop out due to data restrcitions) are divided by treatment status according to their position relative to mean receipt demand (columns (1) and (2)) and relative to the 90th percentile of receipt demand (columns (3) and (4)). Reported are mean values in the subgroup with standard deviations in parantheses.

threshold is always at the same level (refer to Section 3.2 for details) and that estimates are comparable over the years. Moreover, all outcome variables are winsorized at the 99th percentile to reduce the effect outliers have on the estimation.

Table 3.1 provides descriptive statistics at the aggregate regional level. The 119 regions (*cantones*)[12] used are separated according to the level of demand for receipts they face. Column (1) depicts regions with above mean demand for receipts, column (2) regions with below mean demand receipts (for details on how this is calculated, please refer to the following Section 3.4). In columns (3) and (4), the same regions are divided according to the stronger treatment indicator of having demand for receipts above or below the 90th percentile of the distribution of claimed deductions. It is apparent that regions with larger

[12] In total, there are 221 cantones in Ecuador, however, two drop out of the sample due to insufficient observations.

demand for receipts are more populous. Moreover, by construction, the total deductions per capita and the share of population with high income (defined as above the exemption threshold) are significantly higher in regions with a large demand for receipts. This table also provides evidence for the fact that over time, the usage of deductions and income have greatly increased over time. The usage of deductions was basically zero in 2007 (before their introduction) and increased up to 117 USD per capita (inhabitant, not taxpayer) in the 59 regions with above mean usage. Likewise, the small share of individuals with income above the exemption threshold (about 10,000 USD) goes up from 1.29% in 2007 to 2.73% in 2015.

The corresponding descriptive statistics for the sample of individuals with data on reported professions presented above are presented in Table 3.2. Columns (1) to (3) present self-employed in regions with above mean demand for receipts, columns (4) to (6) regions with below mean demand for receipts. The individuals among these regions are separated into doctors, further health professionals, and all other professions. In terms of demographics, the two groups of individuals are relatively balanced, with the notable exception of a higher share of women in the further health professionals. The income variables are depicted for various years and it becomes apparent that reported self-employment income increases strongly over time. By far the largest jump, however, occurs between 2007 (before the deductions policy) and the following years. This jump is stronger in regions with high demand for receipts (columns (1)-(3)) than in regions with low demand for receipts (columns (4) - (6)). Lastly, the overall number of individuals submitting self-employment tax declarations also increases over time. This table is replicated in Table 3.3 with treatment indicated by regions above the 90th percentile of the deduction usage distribution. The qualitative findings are extremely similar.

3.4 Research Design and Results

In this section, I estimate the effect of the deductions policy on economic activity, in particular the behavior of self-employed individuals. The analysis begins by showing the importance of the reform for aggregate levels of reported self-employment profits and continues to use individual tax return data to refine the estimation strategy and rule out a number of alternative explanations.

The core idea behind the identification strategies presented in this paper relies on the fact that the deductions in personal income tax are only useful to individuals with sufficiently high income. As described in Section 3.2, personal income tax (PIT) is levied at the relatively high level of about 10,000 USD annual income.[13] Moreover, there is strong regional variation in the number of high-income individuals and hence in the usage of the deduction

[13] This exemption threshold changes yearly according to inflation, its value ranges from 7850 USD in 2008 to

Table 3.2 – Descriptive Statistics Individual Tax Returns — Relative to Mean

	Above mean canton			Below mean canton		
	(1) Doc	(2) Health	(3) Other	(4) Doc	(5) Health	(6) Other
Age	41.52	39.68	43.21	41.79	37.94	43.05
	(12.15)	(11.00)	(12.53)	(11.72)	(10.38)	(11.95)
Female	0.44	0.78	0.33	0.38	0.79	0.28
	(0.50)	(0.41)	(0.47)	(0.48)	(0.41)	(0.45)
Tertiary Education	1.00	0.95	0.70	1.00	0.93	0.54
	(0.06)	(0.23)	(0.46)	(0.06)	(0.25)	(0.50)
Married	0.65	0.56	0.68	0.65	0.53	0.65
	(0.48)	(0.50)	(0.47)	(0.48)	(0.50)	(0.48)
Observations	32,343	19,476	239,424	4,802	3,434	61,708
2007						
Self-Employment Inc	720.86	860.87	712.66	978.26	887.51	816.11
	(2,627.05)	(2,572.49)	(2,820.88)	(2,962.18)	(2,608.54)	(3,046.54)
Log Self-Employment Inc	1.01	1.49	0.87	1.33	1.39	0.96
	(2.71)	(3.13)	(2.57)	(3.06)	(3.09)	(2.69)
Observations	14,930	6,036	96,896	1,899	905	21,518
2009						
Self-Employment Inc	10,708.88	4,999.19	9,365.57	6,054.70	4,039.01	8,158.89
	(16,193.38)	(9,036.62)	(15,132.16)	(8,986.24)	(6,669.90)	(12,890.89)
Log Self-Employment Inc	6.80	5.27	6.09	6.01	4.88	6.11
	(3.86)	(4.11)	(4.20)	(3.95)	(4.15)	(4.12)
Observations	18,368	8,305	115,877	2,503	1,256	26,456
2012						
Self-Employment Inc	10,076.95	4,091.08	8,394.56	5,167.81	2,899.50	7,209.15
	(17,673.12)	(8,847.57)	(15,225.44)	(9,067.55)	(6,304.89)	(12,317.52)
Log Self-Employment Inc	5.99	4.45	5.58	5.18	3.85	5.64
	(4.20)	(4.16)	(4.30)	(4.13)	(4.07)	(4.20)
Observations	24,516	12,618	147,983	3,636	2,197	35,757
2015						
Self-Employment Inc	11,215.56	4,021.44	8,514.55	5,779.97	2,995.12	7,368.90
	(18,437.22)	(8,933.05)	(15,439.61)	(10,075.26)	(7,094.74)	(12,216.59)
Log Self-Employment Inc	6.25	4.28	5.64	5.45	3.89	5.86
	(4.16)	(4.13)	(4.27)	(4.10)	(4.01)	(4.14)
Observations	23,299	11,879	136,405	3,427	2,163	33,998

Notes: This table presents descriptive statistics for the sample of individuals with self-employment tax returns. Columns (1)-(3) refer to self-employed working in regions with above mean usage of per capita tax deductions, columns (4)-(6) to those in regions with per capita usage of tax deductions below mean. The self employed are further disaggregated by their reported professions: columns (1) and (4) report self employed registered as doctors or dentists, columns (2) and (5) those registered as other health professionals (including nurses, medical technicians, psychologists, pharmacists, midwifes, physical therapists, optometrists, and nutritionists), and columns (3) and (6) refer to all remaining self-employed. Reported are means, standard deviations given in parentheses.

Table 3.3 – Descriptive Statistics Individual Tax Returns — Relative to p90

	Above p90 canton			Below p90 canton		
	(1) Doc	(2) Health	(3) Other	(4) Doc	(5) Health	(6) Other
Age	41.90	40.15	43.38	40.60	37.76	42.87
	(12.25)	(11.07)	(12.67)	(11.61)	(10.39)	(12.02)
Female	0.44	0.78	0.33	0.40	0.79	0.30
	(0.50)	(0.41)	(0.47)	(0.49)	(0.41)	(0.46)
Tertiary Education	1.00	0.95	0.73	1.00	0.94	0.56
	(0.06)	(0.22)	(0.44)	(0.06)	(0.24)	(0.50)
Married	0.65	0.56	0.68	0.65	0.55	0.67
	(0.48)	(0.50)	(0.47)	(0.48)	(0.50)	(0.47)
Observations	27,186	15,902	181,005	9,959	7,008	120,127
2007						
Self-Employment Inc	710.63	895.43	705.68	874.82	787.99	774.02
	(2,630.94)	(2,655.04)	(2,800.88)	(2,779.75)	(2,373.50)	(2,963.44)
Log Self-Employment Inc	0.98	1.51	0.87	1.26	1.39	0.92
	(2.68)	(3.16)	(2.56)	(2.97)	(3.05)	(2.64)
Observations	12,804	4,933	73,738	4,025	2,008	44,676
2009						
Self-Employment Inc	11,603.64	5,422.03	9,871.91	5,795.57	3,550.82	7,922.14
	(17,011.75)	(9,576.43)	(15,777.07)	(8,810.24)	(6,221.05)	(12,758.44)
Log Self-Employment Inc	6.98	5.44	6.14	5.90	4.70	6.02
	(3.81)	(4.11)	(4.22)	(3.96)	(4.10)	(4.13)
Observations	15,650	6,756	88,997	5,221	2,805	53,336
2012						
Self-Employment Inc	11,027.41	4,443.68	8,986.19	5,071.08	2,740.83	6,868.89
	(18,667.72)	(9,368.33)	(16,083.73)	(9,221.78)	(6,119.31)	(12,132.14)
Log Self-Employment Inc	6.17	4.61	5.63	5.10	3.83	5.53
	(4.19)	(4.18)	(4.33)	(4.13)	(4.04)	(4.21)
Observations	20,663	10,210	112,379	7,489	4,605	71,361
2015						
Self-Employment Inc	12,136.59	4,306.81	9,090.65	6,102.09	2,924.24	7,049.12
	(19,370.97)	(9,384.97)	(16,315.52)	(10,734.16)	(6,873.84)	(12,186.71)
Log Self-Employment Inc	6.39	4.34	5.68	5.50	3.96	5.69
	(4.16)	(4.16)	(4.30)	(4.09)	(4.00)	(4.16)
Observations	19,560	9,538	103,238	7,166	4,504	67,165

Notes: This table presents descriptive statistics for the sample of individuals with self-employment tax returns. Columns (1)-(3) refer to self-employed working in regions with the value of per capita tax deductions above the 90th percentile, columns (4)-(6) to those in regions with per capita usage of tax deductions below the 90th percentile. The self employed are further disaggregated by their reported professions: columns (1) and (4) report self employed registered as doctors or dentists, columns (2) and (5) those registered as other health professionals (including nurses, medical technicians, psychologists, pharmacists, midwifes, physical therapists, optometrists, and nutritionists), and columns (3) and (6) refer to all remaining self-employed. Reported are means, standard deviations given in parentheses.

possibilities and the accompanying demand for receipts. This section estimates a series of double and triple difference estimators where the main treatment indicator at the regional level is determined by the actual usage of deductions within a given administrative region. I calculate the total value of claimed deductions of all taxpayers in a region relative to the size of the population. Treatment is determined by a region's position in the distribution of the per-capita usage of deductions in 2009. This allows some time for the reform to become effective and is the most direct measure of the actual demand for receipts possible.[14] I propose two alternative treatment indicators: Being in a region with above average per capita deductions in 2009 and being in a region with per capita deductions above the 90th percentile of the regional deductions distribution.

3.4.1 Aggregate Regional Data

This section examines the effect the deduction opportunities have on reported economic activity aggregated at the regional level. While taking an aggregate view rules out some more robust identification strategies presented in the next subsection, it allows for a better analysis of the economic relevance of the reform. The measure of economic activity I focus on is reported profit of self-employed individuals. Using the universe of self-employed tax declarations, I aggregate various measures of reported self-employment profits at the regional level. To relate these aggregates to the size of the economy, I calculate the per-capita versions based on regional population figures from the Ecuadorian national census.

My analysis focuses on three measures of reported regional per-capita self-employment business activity. The first is reported professional profits, defined as profits claimed by self-employed exercising their profession. Examples of professional income that fall in this category includes freelance doctors, dentists, lawyers, journalists and accountants. Profits in terms of the tax declaration is income minus deductions of expenses necessary for that particular income generation — an example are costs of maintaining a doctors practice. The second measure I use in the analysis is per-capita reported total self-employment profits. As detailed in Section 3.2, this includes further income sources such as certain capital gains and dividends, rental income, and agricultural income. The third measure I focus on is reported business profits of sole proprietorships. Sole proprietorships are businesses owned by individuals subject to personal income taxation, but which due to their size[15] are obliged to report not just revenue and deductible costs, but comprehensive accounting information.

10800 USD in 2015. As all monetary values are deflated to 2013 USD, I use the exemption threshold of 2013 (10180 USD) throughout this analysis.

[14] As shown by Bohne and Nimczik (2018), the usage of the deduction possibilities depends on the availability of information, which spreads sluggishly. The results remain unchanged if I use the deductions distribution of other years or if I use the alternative concept of number of high income individuals living in a given region (results available upon request).

[15] These self-employed business owners have capital exceeding 60,000 USD or annual revenue in excess of 100,000 USD. See Section 3.2 for details.

These are effectively medium to large formal sector businesses mostly already formalized before the reform and owned by individuals subject to personal income taxation.

I measure the effect of the reform using a difference-in-differences estimation framework. As discussed above, the identifying variation stems from regional heterogeneity in the demand for receipts. Specifically, I estimate the following regression equation:

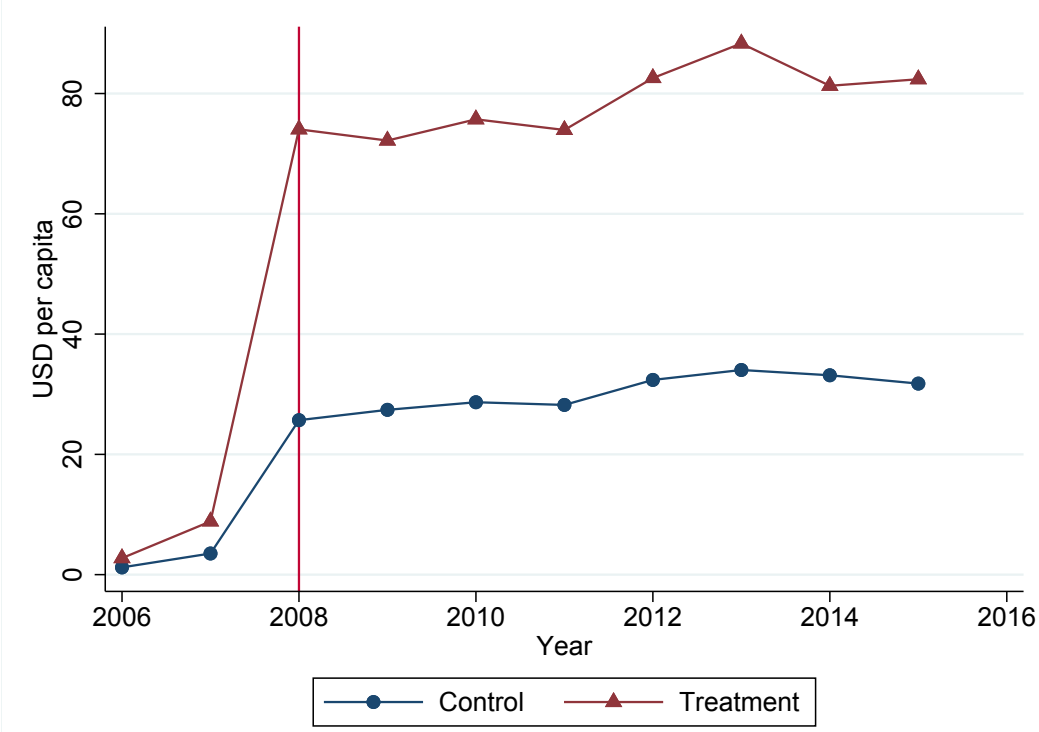
$$Y_{jt} = \gamma_j + \lambda_t + \beta Treat_j \times Post_t + \epsilon_{jt} \tag{3.1}$$

with Y_{jt} being one of the outcome variables described above for region j at time t. Region (γ_j) and year (λ_t) fixed effects are included throughout and control for any unobserved level differences between regions and over time. $Treat_j$ is an indicator function equal to one in case a region is subject to a high demand for receipts — either above the mean or above the 90th percentile of the distribution of claimed deductions. $Post_t$ is equal to one in all periods after the introduction of the reform in 2008. The coefficient of interest β is the difference-in-differences estimator measuring the additional change in outcome occurring in regions with a high demand for receipts compared to the change in outcomes in regions with a low demand for receipts. Under the assumption of parallel trends in outcomes in absence of the reform, this parameter measures the treatment effect of the deductions policy.

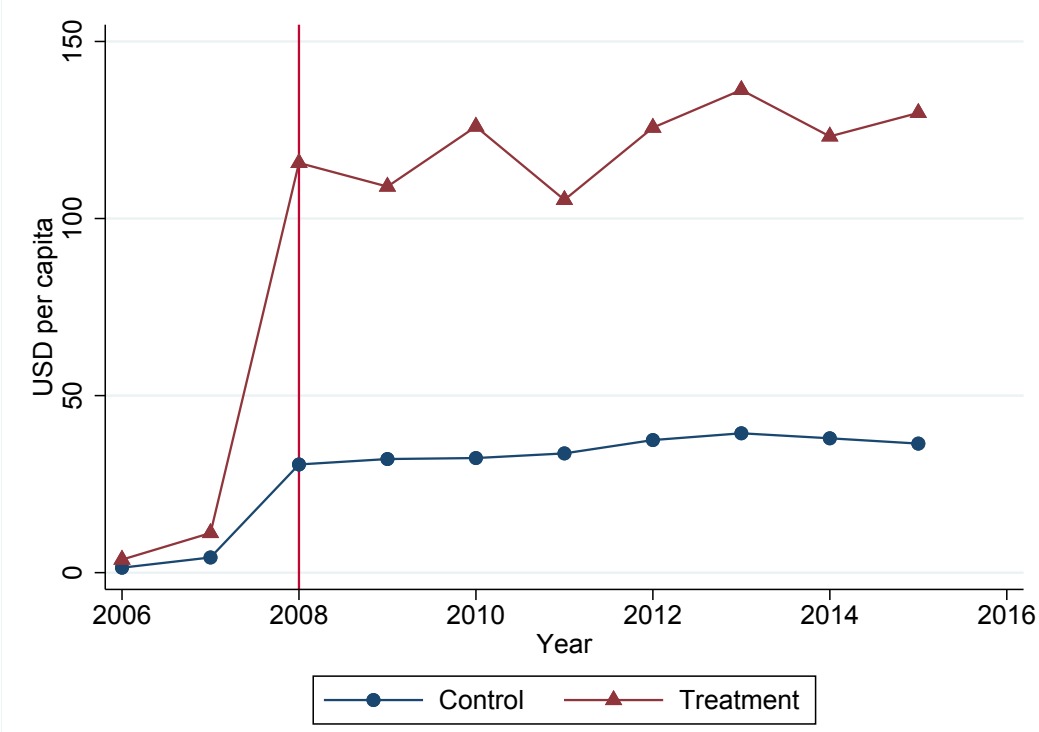
Figure 3.1 depicts the trends in reported professional profit, differentiating between regions with a high demand for receipts (“treatment") and regions with a low demand for receipts (“control"). Among both groups of regions, reported professional profits increased significantly after the introduction of the reform. The pre-reform levels of below 10 USD reported professional profits per inhabitant living in the region lead to believe that self-employed professionals were basically reporting none of their income before the introduction of the reform. The increase in reported profits, however, differs greatly depending on treatment status: regions with a high demand for receipts had a much higher per-capita increase in reported professional profits than regions with a low demand for receipts. The difference is even larger in Panel (b) of Figure 3.1, in which I take treatment as the stronger version with claimed deductions above the 90th percentile. Most importantly for my identification strategy, pre-trends seem to be very parallel (and close to zero) between the two groups. Moreover, with the introduction of the reform, reported profits seem to jump to sizeable levels of up to more than 100 USD per inhabitant (for regions with the largest demand for receipts) and stay relatively stable over time. This apparent stable treatment effect further motivates my use of the difference-in-differences estimator with just one coefficient measuring the overall before/after effect.

The two following Figures 3.2 and 3.3 show trends for the two remaining outcome variables: Total per-capita self-employment income and profits of sole proprietorships. The pre-trends of total self-employment profits are just as parallel as in the previous figure.

Figure 3.1 – Pre-Trends Professional Profits



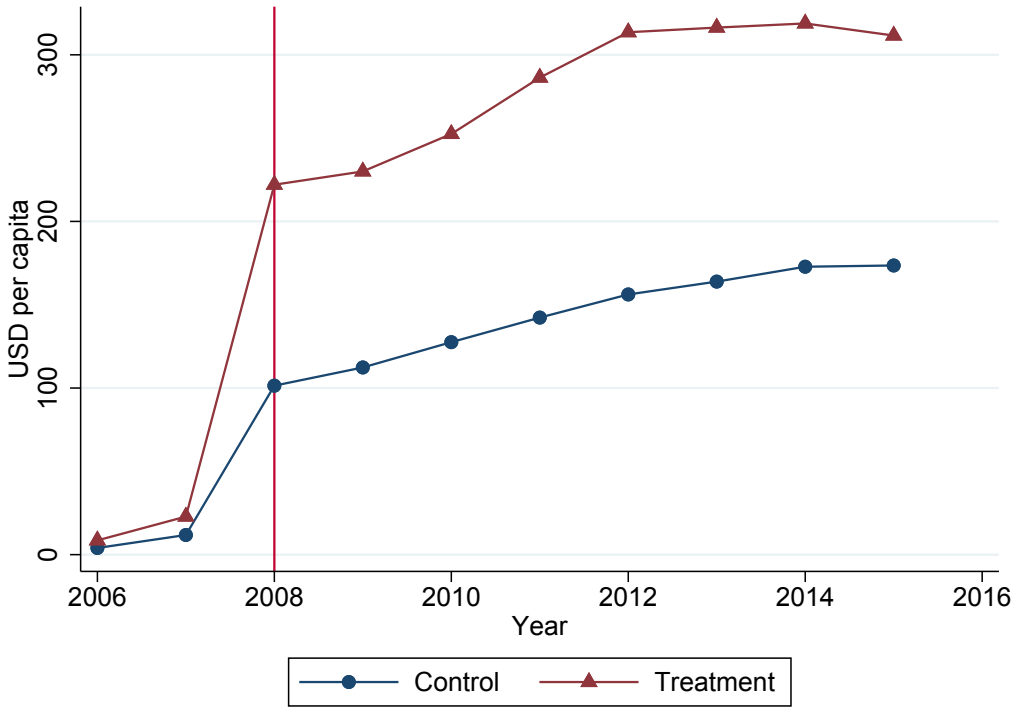
(a) Treatment: Above Mean



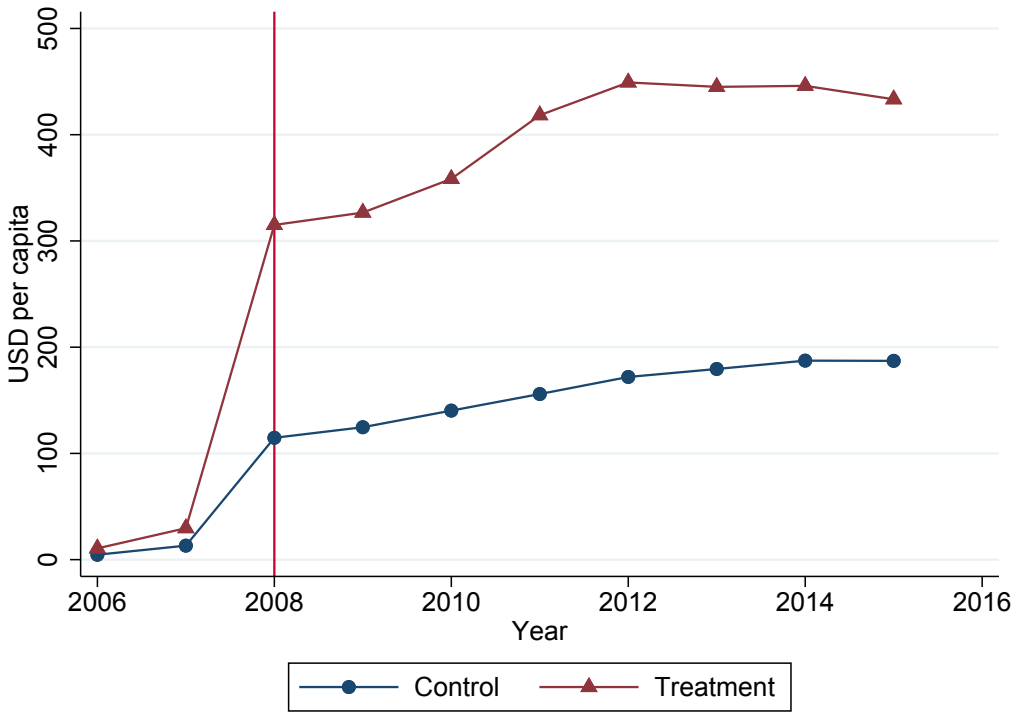
(b) Treatment: Above p90

This figure depicts the trends in outcomes depending on the treatment status of a given region, with treatment being defined as demand for receipts above mean (Panel (a)) or above the 90th percentile (Panel (b)). Observation is at the region (*cantón*) level. The outcome variable in this figure is aggregated per-capita professional profits, defined as the sum of all professional profits reported in a given region, divided by the number of inhabitants.

Figure 3.2 – Pre-Trends Total Self-Employment Profits



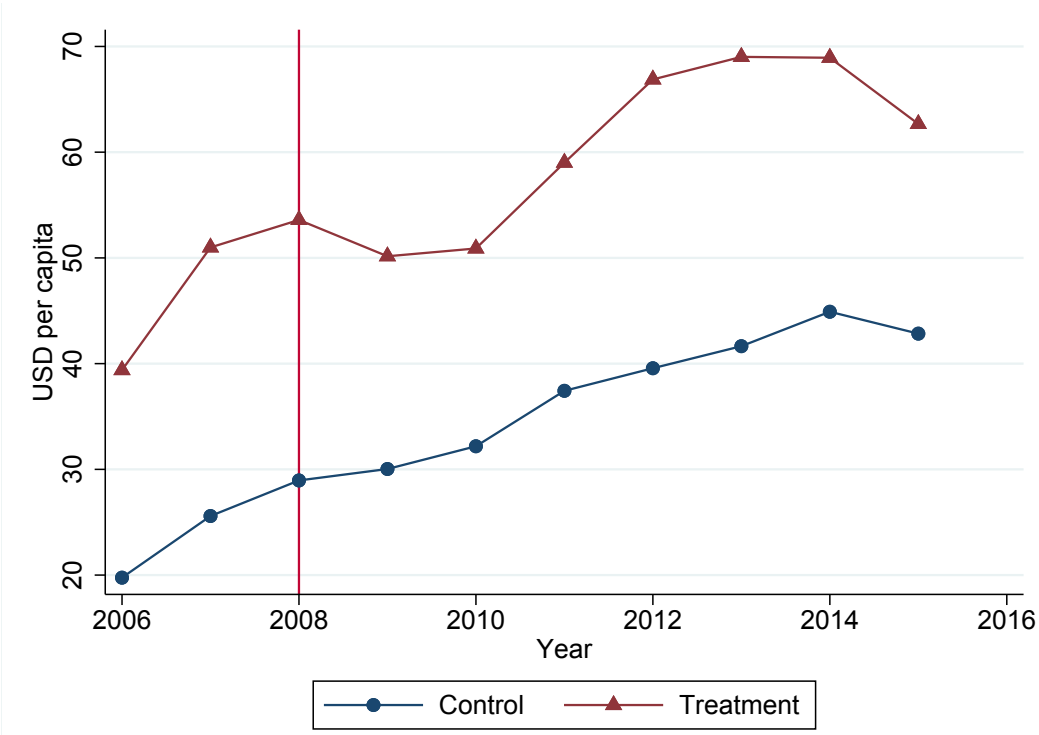
(a) Treatment: Above Mean



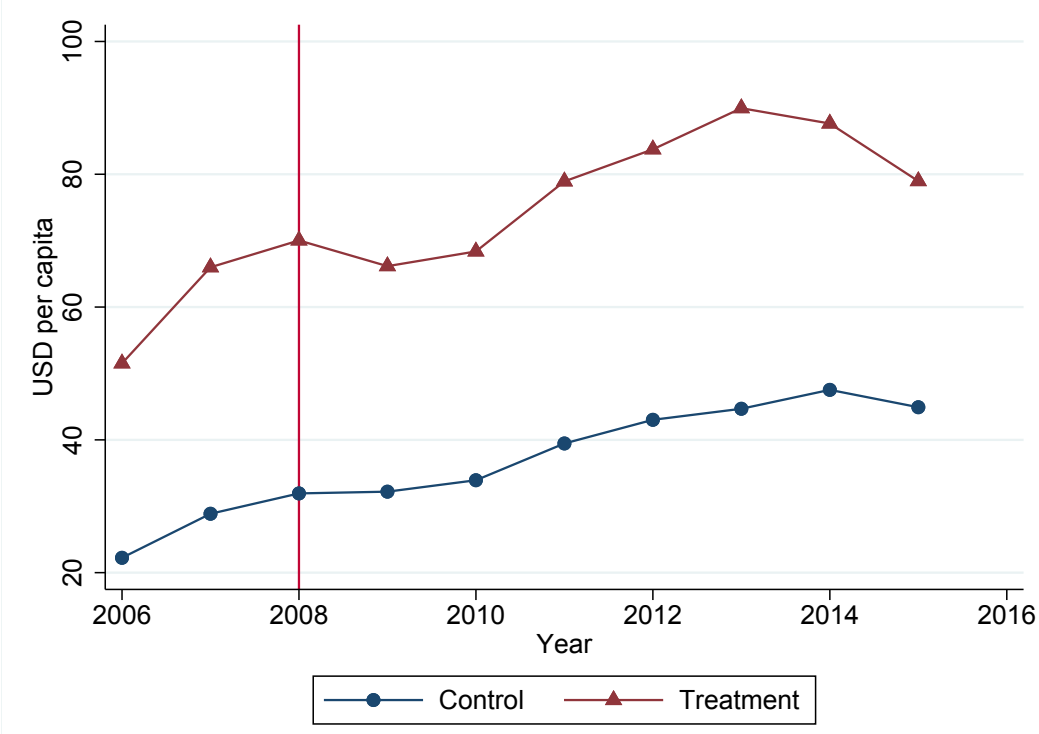
(b) Treatment: Above p90

This figure depicts the trends in outcomes depending on the treatment status of a given region, with treatment being defined as demand for receipts above mean (Panel (a)) or above the 90th percentile (Panel (b)). Observation is at the region (*cantón*) level. The outcome variable here is aggregated per-capita total self-employment profits, defined as the sum of all self-employment profits reported in a given region, divided by the number of inhabitants.

Figure 3.3 – Pre-Trends Profits Sole Proprietorships



(a) Treatment: Above Mean



(b) Treatment: Above p90

This figure depicts the trends in outcomes depending on the treatment status of a given region, with treatment being defined as demand for receipts above mean (Panel (a)) or above the 90th percentile (Panel (b)). Observation is at the region (*cantón*) level. The outcome variable of interest is aggregated per-capita profits of sole proprietorships, defined as the sum of all profits of sole proprietorships reported in a given region, divided by the number of inhabitants.

However, since this measure includes other sources of income, the overall per-capita values are substantially higher in Figure 3.2 than in the previous one. For this aggregated variable, the time trend after the jump in 2008 is clearly positive, evidence for the fact that, in contrast to the professional profits, the other variables making up overall self-employment profits are actually increasing over time. In the last set of trends in Figure 3.3, the overall picture is quite different. First, pre-trends do not seem very parallel between the two sets of regions. Likewise, even at this descriptive level there is not a clear effect of the reform. This points to the conclusion that for sole proprietorships, which are medium to large formal sector businesses, the introduction of the deduction opportunities did not seem to have a large impact on reported profits.

These descriptive effects are confirmed in the regression results in Table 3.4. Following the estimation framework laid out in regression equation 3.1, Panel A estimates the effect of the reform on professional profits. The reform led to a positive and robust increase of reported professional profits at the regional level. The effects are sizeable and statistically significant in absolute (45 to 82 USD per capita) as well as relative terms (17% to 33%). In Panel B, I present results for the effect on total overall per-capita self-employment profit. These effects are large and significant in levels, but become very small an insignificant in logs. Lastly, Panel C presents results for the per-capita profits of sole proprietorships and finds, in line with the descriptive trends, no or even a negative effect of the reform on this aggregated measure.

To summarize, this section provides evidence for a strong impact of the reform on reported profits of self-employed individuals aggregated at the regional level. The effect is very strong and robust for professional profits, becomes smaller for overall self-employment profits and vanishes completely for profits by sole proprietorships. The pre-trends are remarkably parallel in the first two cases. The very strong increases in reported profits for professionals from about zero to substantial levels show that this reform was instrumental in getting a segment of the economy to participate and report their revenue and profits in the formal economy. It is noteworthy that in this setting, the self-employed increased their actual *profits* - in this case revenue minus costs. In the setting of corporate taxation in Ecuador, Carrillo et al. (2017) recently showed that when faced with increased enforcement, firms increase their reported revenue but offset this by increasing reported costs, resulting in a null effect for the tax authority. One possible explanation for this is that while these self-employed increased their reported profits, this does not necessarily imply paying (higher) taxes. After their profits are calculated, self-employed, just like salaried workers, have the same possibility to themselves make use of the extensive deduction possibilities and reduce their tax liability. However, this setting can still be beneficial from a policy perspective as it gets self-employed to report their profits in the first place and thereby increases paper trails and information the tax authority can use to collect other revenue sources, also in the future.

Table 3.4 – DD Results Aggregate Effets

	(1)	(2)	(3)	(4)
Panel A: Professional Profits				
	Per-capita Professional Profits		Log Per-capita Professional Profits	
DD (mean × post)	45.21***		0.175*	
	(9.884)		(0.0932)	
DD (p90 × post)		81.79***		0.334**
		(22.20)		(0.151)
Observations	2,190	2,190	2,157	2,157
R^2	0.3887	0.4222	0.8215	0.8221
Panel B: Total Self-Employment Profits				
	Per-capita Total Self-employment Profits		Log Per-capita Total Self-employment Profits	
DD (mean × post)	129.8***		-0.0882	
	(27.44)		(0.0703)	
DD (p90 × post)		230.1***		0.00152
		(57.11)		(0.0935)
Observations	2,190	2,190	2,190	2,190
R^2	0.6309	0.6527	0.9444	0.9443
Panel C: Profits Sole Proprietorships				
	Per-capita Profits Sole Proprietorships		Log Per-capita Profits Sole Proprietorships	
DD (mean × post)	0.429		-0.137**	
	(2.439)		(0.0557)	
DD (p90 × post)		5.069		-0.139*
		(3.786)		(0.0785)
Observations	2,190	2,190	2,190	2,190
R^2	0.2636	0.2650	0.2642	0.2627

Notes: This table presents results from the difference-in-differences regression presented in equation 3.1. The level of observation is at the region (*cantón*) level. The outcome in Panel A, column (1) and (2) is the sum of all professional profits reported in a given region, divided by the number of inhabitants. Columns (3) and (4) of the same Panel report the log of this variable, that is log aggregated professional profits, divided by the number of inhabitants. Panel B reports results for total self-employment profits, including additional sources such as dividends and rental income. Panel C depicts results with the outcome variable profits for sole proprietorships, i.e. businesses already formalized and above certain size thresholds. In each panel, the interaction term from equation 3.1 is reported with treatment given either by the above mean or above p90 indicator function. Standard errors in parentheses are clustered at the cantón level, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.4.2 Individual Self-Employed Data

This section studies the responses of individual taxpayers to the increased demand for deductions. Using the sub-sample of self-employed tax records with data on registered professions (please refer to Section 3.3 for a detailed sample description), I replicate the findings at the aggregate level with micro-level data while controlling for unobserved individual heterogeneity. The additional data on registered professions allow me to exploit the fact that the increased demand for receipts affects varying professions to different degrees. In triple difference regressions, I show that doctors react significantly stronger to the reform than other professionals.

The previous section documented how the demand for receipts led to increases in regional aggregates of reported professional profits. However, as described in Section 3.2, total self-employment profits also include other sources of income such as dividends, rental income, and profits from sole proprietorships, which may be subject to considerable income shifting among themselves. At the regional level, the results for these broader measures of self-employment business activity were not as clear-cut. In the following analysis, I estimate the effect of an increase in demand for receipts on individual self-employed individuals. To circumvent any issues regarding the shifting of income between categories, I only consider individual total self-employment profits aggregated over all income categories as the outcome variable of interest.[16]

I start by replicating, at the individual level, the same difference-in-differences estimation strategy as used in Section 3.4.1. In particular, I run the follow regression:

$$Y_{it} = \alpha_i + \lambda_t + \beta Treat_i \times Post_t + \epsilon_{it} \tag{3.2}$$

where Y_{it} are total self-employment profits of individual i at time t. $Treat_i$ is an indicator for whether the self-employed individual i is active in a region (canton) with above average (above 90th percentile) level of deductions, as measured by total claimed deductions per capita. $Post_t$ is an indicator for an observation in the years after the introduction of the reform, and β provides the difference-in-differences estimator. In this specification I control for general time trends through λ_t and individual unobserved heterogeneity through the inclusion of individual fixed effects (α_i).

The results of these double difference regressions are reported in Panel A of Table 3.5. Individual self-employed in regions with above-average demand for receipts increased their reported self-employment profits by about 2400 USD more in the years after the reform than individuals in areas with below average demand for receipts. This effect is statistically highly significant and becomes even stronger (about 3700 USD) for self-employed residing

[16] In robustness analyses, I find the same or stronger effects when restricting the outcome variable to purely professional profits.

in regions in the 90th percentile of the distribution of claimed deductions. The magnitude of these effects becomes clear when looking at log reported business profits[17] as the outcome: The reform led to a hike in reported profits by about 30% for self-employed exposed to above average demand for receipt and almost 50% for those in the 90th percentile of the deductions distribution.

Table 3.5 – Effects Individual Tax Returns

	(1)	(2)	(3)	(4)
Panel A: Double Differences				
	Self-employment Profits		Log Self-employment Profits	
DD (mean × post)	2402.8***		0.295***	
	(90.84)		(0.0254)	
DD (p90 × post)		3716.0***		0.481***
		(77.91)		(0.0201)
Individual & Year FE	Yes	Yes	Yes	Yes
Observations	1,880,549	1,880,549	1,880,549	1,880,549
R^2	0.1178	0.1202	0.2840	0.2843
Panel B: Triple Differences				
	Self-employment Profits		Log Self-employment Profits	
DDD (mean × post × doc)	5268.6***		0.978***	
	(293.9)		(0.0935)	
DDD (p90 × post × doc)		5337.9***		0.987***
		(251.9)		(0.0689)
Individual & Year FE	Yes	Yes	Yes	Yes
Observations	1,880,549	1,880,549	1,880,549	1,880,549
R^2	0.1190	0.1215	0.2846	0.2850

Notes: This table depicts results using the sample of individual self-employed tax return data. Outcome variables are total self-employment profits and log(self-employment profits + 1). Panel A depicts the results from the interaction term in the double difference regression equation (3.1). Panel B depicts the results for the triple interaction term in regression equation (3.3). The second treatment layer is given by an indicator for a self-employed individual with registered profession of either medical doctor or dentist. Standard errors in parentheses, clustered at the individual level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The identifying assumption for this difference-in-differences analysis is that self-employed in areas with high demand for receipts face the same underlying trends as self-employed in areas with low demand for receipts. One major reason why this assumption may not

[17] Defined as log(self-employment profits +1) in order to account for some individuals claiming zero profits.

hold is that the regions differ in their underlying economic trends. Positive regional economic growth could lead to both an increase in the demand for receipts and higher true profits of self-employed individuals. The remarkably parallel pre-trends in aggregated self-employment profits between the different regions presented in Section 3.4.1 speak a different language. However, self-employed in treated regions could still be affected by underlying factors increasing their true earnings in the absence of the deductions policy. To circumvent these difficulties, I apply a triple difference methodology exploiting additional data on the reported professions of individuals and the fact that only certain goods and services can be deducted (see Section 3.2 for details). This identification strategy additionally compares self-employed individuals particularly affected by the reform to self-employed not so affected within a given region. Intuitively, this controls for changes in overall economic trends between the regions. One subset of professions particularly affected by the reform are doctors and dentists, since health expenses are one of the most salient deduction categories and health services have previously typically faced low levels of paper trails since they are exempt from VAT payments (SRI, 2018).

The regression setup for the triple difference regression takes on the following form:

$$Y_{it} = \alpha_i + \lambda_t + \beta_1 Treat_i \times Post_t + \beta_2 Doc_i \times Post_t + \delta Treat_i \times Post_t \times Doc_i + \epsilon_{it} \quad (3.3)$$

with Y_{it}, α_i and λ_t defined as before. The two differing layers of treatment are represented by the indicator functions $Treat_i$ for regions with a high demand for receipts (above mean or above the 90th percentile) and the indicator function Doc_i, which takes on the value one if an individual is either a doctor or a dentist. This setup requires the full set of interactions between the treatment indicators and the dummy $Post_t$ and the coefficient δ measures the triple difference estimate we are interested in. Panel B of Table 3.5 reports the results from this regression. Doctors in regions with high demand for receipts increase their reported profits significantly more than other professionals in the same regions, already taking into account the general increase in reported profits due to being in a region with a high demand for receipts. This effect of more than 5000 USD higher profits is not only large in absolute terms, but also particularly large in relative terms: Reported profits increase by almost one log point.

In placebo tests, I replicate the triple difference equation (3.3) for various groups of self-employed whose services are unaffected by the deductions policy. Table 3.6 presents the results of triple difference regressions with the second treatment layer being veterinary physicians (Panel A) and a pooled group of further unaffected professions consisting of economists, business administrators, and journalists (Panel B). The services of all of these professionals do not fit into any of the five deduction categories provided by the government.

Moreover, the latter group of professionals typically do not provide services to consumers subject to personal income taxation. The results in Table 3.6 are clear: The triple difference estimator for these groups of unaffected professions is generally close to zero and not statistically significant. In case of (marginal) statistical significance, the estimator is even negative, implying that the additional effect of the reform by being a member of one of these professions even reduces reported profits. These placebo differences, along with stable pre-trends, lend credibility to the double and triple differences estimation strategy applied in this paper.

Table 3.6 – Placebo Differences

	(1)	(2)	(3)	(4)
Panel A: Vet				
	Self-employment Profits		Log Self-employment Profits	
DDD (mean × post × vet)	-1227.8		0.141	
	(781.7)		(0.205)	
DDD (p90 × post × vet)		-1522.0**		0.0125
		(759.4)		(0.187)
Individual & Year FE	Yes	Yes	Yes	Yes
Observations	1,880,549	1,880,549	1,880,549	1,880,549
R^2	0.1178	0.1202	0.2840	0.2843
Panel B: Unaffected Professionals				
	Self-employment Profits		Log Self-employment Profits	
DDD (mean × post × unaffected)	447.4		0.170	
	(425.2)		(0.119)	
DDD (p90 × post × unaffected)		-627.7*		0.0143
		(351.1)		(0.0882)
Individual & Year FE	Yes	Yes	Yes	Yes
Observations	1,880,549	1,880,549	1,880,549	1,880,549
R^2	0.1178	0.1202	0.2840	0.2844

Notes: This Table presents the results of triple difference regressions as in equation (3.3) for two sets of placebo professional groups: Veterinary physicians (Panel A) and a further group of unaffected professionals (Panel B) consisting of economists, business administrators and journalists. Reported coefficients correspond to the triple interaction between being in an above mean (p90) region, belonging to the specific group of professionals and the observations being in a time period after the introduction of the reform. Outcome variable is individual total overall self-employment profit in columns (1)-(2) and log(self-employment profit + 1) in columns (3)-(4). Standard errors in parentheses, clustered at the individual level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

One further threat to identification in the triple difference setting could be shocks to the relevant subgroup of professionals that differ according to the demand for receipts in a given region. While it is unclear why this might be the case from economic fundamentals,

this could happen due to a side effect of the reform: The deductions policy could increase true demand for certain products and services, for example doctors consultations. However, there are a number of factors speaking against this observation. First, as the deduction categories include expenses in health, housing, nutrition, clothing and education, they likely affect almost all expenses a typical household faces. Therefore, relative prices of expenditures should stay unchanged. However, there might still be an income effect. To this end, Table C.4 in the Appendix replicates the triple difference estimator for a broader group of health professionals including nurses, medical technicians, psychologists, pharmacists, midwifes, physical therapists, optometrists and nutritionists (but excluding doctors and dentists). These professionals should also be affected from a general increase in the demand for health services, however, their incentives in the tax system are not as clear-cut as those faced by doctors or dentists. Many of these professionals might work in corporations, have customers who are not final consumers paying income taxes, or be completely active in the informal economy. The results of this triple difference estimator show that for this subgroup of professionals in the health industry, the reform had no or even a negative effect on their reported profits.

Having established the robustness of the presented estimates, it is still unclear whether this overall effect is driven by self-employed switching from the informal to the formal sector or by self-employed increasing the intensity of their (reported) formal sector business activities. One way to look at this is to focus on the subset of firms that were economically active before the introduction of the reform. In the Appendix, Table C.2 replicates the results for the subset of self-employed already filing tax declarations before the reform and throughout the whole sample period. The results are very similar and if anything, the magnitude of the effects is larger. This shows that increases in the intensive margin are an important driver of the observed overall effects of the reform.[18]

3.5 Conclusion

This paper analyzes an innovative reform drawn up to improve tax compliance and increase participation in the formal economy. The target population is mostly self-employed business owners operating outside of the VAT system. Formalization is achieved by incentivizing individual taxpayers to ask for receipts, which they can use to reduce their personal income tax liability. Due to a relatively high threshold for personal income tax liability in Ecuador, only a small portion of the population is affected by these incentives and these individuals are distributed unevenly between regions. Exploiting this regional heterogeneity in the density of high income individuals allows to estimate the effect of the reform using

[18] The placebo tests also look very similar for this subset of balanced observations (see Table C.3 in the Appendix).

a difference-in-differences framework comparing regions with a high demand for receipts to regions with a low demand for receipts. Merging the individual tax return records of self-employed business owners with their reported professions from the civil registry provides an additional layer of variation arising from the fact that the deductions are limited to certain categories such as health expenditures. This additional heterogeneity allows to estimate a triple-difference regression which, on top of controlling for level differences between regions, additionally controls for potential time-changing shocks affecting regions differentially according to their demand for receipts. In both double and triple difference regressions, I find very strong effects on reported business profits by self-employed individuals, especially in a subcategory measuring profits by professional service providers particularly affected by the reform.

The policy implications of the findings are very clear: This reform proved to be an effective measure in driving formalization of a sector of the economy previously mostly eluding government control due to lacking third party reports or paper trails of transactions. In fact, the reform led a sector of the economy previously almost completely operating in the informal sector to become formalized. It needs to be pointed out that while these findings are from a developing country context, reported business activities of self-employed individuals is an area also subject to high amounts of tax avoidance and evasion in many developed countries.

Yet while this reform was very successful in formalizing self-employed business activities, I cannot make statements about the cost-effectiveness of this reform. The reform additionally has strong spillover effects for the VAT system, mostly due to the fact that affected businesses are also likely to remit more VAT. However, this paper focuses on the effects on reported self-employed business activities and a comprehensive cost-benefit analysis (foregone PIT vs revenue increases due to formalization) would additionally need to take into account the spillover effects on VAT revenue. Moreover, the effects in this paper are overall effects driven by both the intensive margin (already formalized businesses reporting more revenue) and the extensive margin (new businesses joining the set of formal businesses). While this paper provides evidence that the intensive margin is an important component of this overall effect,[19] future work could go into the direction of better disentangling these two effects.

[19] Tables C.2 and C.3 provide evidence for this effect among a subsample of firms with completely balanced observations, that is firms which were already formalized before the reform and continue to operate after the reform.

Appendix A

Appendix to Chapter 1

Table A.1 – Effect of Moneybox on Savings

	(1) Savings	(2) Log(Savings)	(3) Savings Wide	(4) Log(Savings Wide)
Person received a moneybox	117.643*	1.755***	226.708	0.599***
	(61.370)	(0.191)	(224.239)	(0.153)
N	881	881	881	881
Mean Control Group	527.596	3.587	2861.914	6.508

Note: This table reports results from an OLS regression comparing individuals who received a moneybox to those who did not receive a moneybox. *Savings* measures cash plus money held in the moneybox (0 for respondents without moneybox), *Savings Wide* gives the sum of our previous *savings narrow* indicator plus savings in bank, *Iddir* (funeral society) and *Equb* (informal savings club). All variables are winsorized at the 95% percentile. Control variables include all baseline variables used during randomization (savings in cash, bank and informal arrangements, socio-demographics, land and consumption variables) as well as enumerator indicators and our measure of financial literacy. The last row gives the mean value of the outcome variable in the control group (no moneybox). Robust standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2 – Alternative Mechanisms

Panel A: Recommendations and Expected Problems

	(1) Mean Control Group	(2) Recommendation	(3) Financial Literacy	(4) Recommendation × Problems	(5) Problems
Savings	496.009	211.147*	-12.491	-39.020	-40.156
		(124.178)	(76.610)	(160.883)	(141.636)

Panel B: Recommendations and Financial Literacy

	(1) Mean Control Group	(2) Recommendation	(3) Financial Literacy	(4) Recommendation × Financial Literacy
Savings	496.009	150.271	-52.161	56.946
		(107.137)	(146.091)	(158.793)

Panel C: Recommendations and Present-Biasedness

	(1) Mean Control Group	(2) Recommendation	(3) Financial Literacy	(4) Recommendation × Present-Biasedness	(5) Present-Biased
Savings	496.009	245.207**	-5.671	-216.725	147.703
		(97.593)	(76.619)	(176.381)	(157.771)

Panel D: Recommendations and Risk-Lovingness

	(1) Mean Control Group	(2) Recommendation	(3) Financial Literacy	(4) Recommendation × Risk-loving	(5) Risk-loving
Savings	496.009	168.343	-6.832	27.819	1.683
		(112.288)	(76.997)	(163.991)	(141.168)

Note: N=599. The dependent variable *Savings* is calculated as the value of cash savings the household head carries at the time of the interview plus savings in the moneybox. Column (1) refers to average savings in the control group (no recommendation). Columns (2) through (5) give the coefficients for receiving a recommendation, being financially literate and a range of alternative mechanisms as well as their interactions with recommendation. Further control variables include all variables used for stratification at baseline plus enumerator controls. Robust standard errors in parentheses, * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A.3 – Alternative Savings Measures

	(1) Bank Savings	(2) *Equb*	(3) *Iddir*	(4) Savings Wide
Recommendation	50.238	-40.957	13.746	180.118
	(157.002)	(221.486)	(26.793)	(319.324)
Bank savings baseline	0.793***			
	(0.128)			
Equb baseline		0.619***		
		(0.170)		
Iddir baseline			0.008	
			(0.042)	
Savings wide baseline				0.530***
				(0.069)
Cash savings baseline	0.004	0.040**	0.001	-0.017
	(0.014)	(0.020)	(0.002)	(0.033)
Bank savings baseline	0.012	-0.049**	0.001	0.006
	(0.019)	(0.023)	(0.003)	(0.037)
Iddir member baseline	-131.006	-521.967***	31.356	-790.720**
	(148.064)	(201.869)	(32.617)	(312.881)
Eqqub member baseline	112.983	452.568	-0.492	373.047
	(156.867)	(306.414)	(27.122)	(387.682)
Female household head	193.417	308.512	-6.078	537.178
	(226.224)	(321.876)	(35.926)	(442.044)
Married household head	2.734	504.476*	-16.094	647.517*
	(205.511)	(272.963)	(34.768)	(391.651)
Education household head	40.809**	23.758	4.106	43.424
	(19.945)	(26.692)	(3.358)	(39.499)
Land size baseline	-18.385**	-1.105	0.360	-33.688
	(9.315)	(23.012)	(2.128)	(27.495)
Revenue baseline	0.010*	0.015***	0.000	0.026***
	(0.005)	(0.006)	(0.001)	(0.009)
Debt baseline	0.017	-0.013	-0.002	0.016
	(0.019)	(0.023)	(0.003)	(0.034)
Consumption expenditure baseline	0.016	0.129	0.024*	0.177
	(0.072)	(0.147)	(0.014)	(0.182)
Financial literacy	61.484	7.299	-19.288	-55.359
	(158.480)	(207.573)	(24.363)	(296.418)
N	599	599	599	599
Mean	1188.416	1172.650	136.778	2993.852

Note: *Bank savings* is the self-reported current balance of an individual's savings accounts (column (1)). Equb gives the amount of money people currently hold with *Equb*, an informal savings arrangement (column (2)). Iddir is the amount of money that a person would receive from the funeral society *Iddir* in case of death (column (3)). *Savings wide* gives is the sum of our previous *savings narrow* indicator + bank savings + Equb + Iddir (column (4)). All variables are winsorized at the 95% percentile. The respective *baseline* variables are the average value between baseline and treatment implementation. The last row gives the mean value of the outcome variable in the control group (no recommendation). Robust standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B

Appendix to Chapter 2

B.1 Institutional Details

Table B.1 – Tax Brackets (in US $)

Marginal Rate	06	07	08	09	10	11	12	13	14	15
5%	7,680	7,850	7,850	8,570	8,910	9,210	9,720	10,180	10,410	10,800
10%	15,360	15,700	10,000	10,910	11,350	11,730	12,380	12,970	13,270	13,770
12%	–	–	12,500	13,640	14,190	14,670	15,480	16,220	16,590	17,210
15%	30,720	31,400	15,000	16,370	17,030	17,610	18,580	19,470	19,920	20,670
20%	46,080	47,100	30,000	32,740	34,060	35,210	37,160	38,930	39,830	41,330
25%	61,440	62,800	45,000	49,110	51,080	52,810	55,730	58,390	59,730	61,980
30%	–	–	60,000	65,480	68,110	70,420	74,320	77,870	79,660	82,660
35%	–	–	80,000	87,300	90,810	93,890	99,080	103,810	106,200	110,190

Note: Columns denote the years to which the tax brackets apply. The numbers indicate the value of the lower bound above which income is taxed at the relevant marginal rate. For example: In 2014, all income between 10,410 USD and 13,270 USD is taxed at the marginal rate of 5%.

Figure B.1 – Tax Declaration Form F107 for Wage Earners

SRI
SERVICIO DE RENTAS INTERNAS

FORMULARIO 107
RESOLUCIÓN No. NAC-DGERCGC12-00829

COMPROBANTE DE RETENCIONES EN LA FUENTE DEL IMPUESTO A LA RENTA POR INGRESOS DEL TRABAJO EN RELACIÓN DE DEPENDENCIA No.

EJERCICIO FISCAL	102		FECHA DE ENTREGA	103	AÑO	MES	DIA

100 Identificación del Empleador (Agente de Retención)

105	RUC 0 0 1	106	RAZÓN SOCIAL O APELLIDOS Y NOMBRES COMPLETOS

200 Identificación del Trabajador (Contribuyente)

201	CÉDULA O PASAPORTE	202	APELLIDOS Y NOMBRES COMPLETOS

Liquidación del Impuesto

SUELDOS Y SALARIOS	301	+	
SOBRESUELDOS, COMISIONES, BONOS Y OTROS INGRESOS GRAVADOS	303	+	
PARTICIPACIÓN UTILIDADES	305	+	
INGRESOS GRAVADOS GENERADOS CON OTROS EMPLEADORES	307	+	
DÉCIMO TERCER SUELDO	311		
DÉCIMO CUARTO SUELDO	313		
FONDO DE RESERVA	315		
OTROS INGRESOS EN RELACIÓN DE DEPENDENCIA QUE NO CONSTITUYEN RENTA GRAVADA	317		
(-) APORTE PERSONAL IESS CON ESTE EMPLEADOR (únicamente pagado por el trabajador)	351	-	
(-) APORTE PERSONAL IESS CON OTROS EMPLEADORES (únicamente pagado por el trabajador)	353	-	
(-) DEDUCCIÓN GASTOS PERSONALES - VIVIENDA	361	-	
(-) DEDUCCIÓN GASTOS PERSONALES - SALUD	363	-	
(-) DEDUCCIÓN GASTOS PERSONALES - EDUCACIÓN	365	-	
(-) DEDUCCIÓN GASTOS PERSONALES - ALIMENTACIÓN	367	-	
(-) DEDUCCIÓN GASTOS PERSONALES - VESTIMENTA	369	-	
(-) EXONERACIÓN POR DISCAPACIDAD	371	-	
(-) EXONERACIÓN POR TERCERA EDAD	373	-	
IMPUESTO A LA RENTA ASUMIDO POR ESTE EMPLEADOR	381	+	
BASE IMPONIBLE GRAVADA 301+303+305+307-351-353-361-363-365-367-369-371-373+381 ≥ 0	399	=	
IMPUESTO A LA RENTA CAUSADO	401	=	
VALOR DEL IMPUESTO RETENIDO Y ASUMIDO POR OTROS EMPLEADORES DURANTE EL PERÍODO DECLARADO	403		
VALOR DEL IMPUESTO ASUMIDO POR ESTE EMPLEADOR	405		
VALOR DEL IMPUESTO RETENIDO AL TRABAJADOR POR ESTE EMPLEADOR	407		
INGRESOS GRAVADOS CON ESTE EMPLEADOR (informativo) 301+303+305+381	349	=	

IMPORTANTE: Sírvase leer cada una de las siguientes instrucciones.

1.- El trabajador que, en el mismo período fiscal haya reiniciado su actividad con otro empleador, estará en la obligación de entregar el formulario 107 entregado por su anterior empleador a su nuevo empleador, para que aquel, efectúe el cálculo de las retenciones a realizarse en lo que resta del año.

2.- El campo 307 deberá ser llenado con la información registrada en el campo 349 del Formulario 107 entregado por el anterior empleador, y/o con la proyección de ingresos de otros empleadores actuales, en caso de que el empleador que registra y entrega el presente formulario haya efectuado la retención por los ingresos percibidos con éstos últimos.

3.- La deducción total por gastos personales no deberá superar el 50% del total de ingresos gravados, y en ningún caso será mayor al equivalente a 1.3 veces la fracción básica exenta de Impuesto a la Renta de personas naturales.

4.- A partir del año 2011 debe considerarse como cuantía máxima para cada tipo de gasto, el monto equivalente a la fracción básica exenta de Impuesto a la Renta en:
vivienda 0.325 veces, educación 0.325 veces, alimentación 0.325 veces, vestimenta 0.325, salud 1.3 veces.

5.- El trabajador deberá presentar el Anexo de Gastos Personales que deduzca, de cumplir las condiciones establecidas por el Servicio de Rentas Internas.

6.- De conformidad con la Resolución No. NAC-DGER2008-0566 publicada en el Registro Oficial No. 342 el 21 de mayo del 2008, el beneficio de la exoneración por tercera edad se configura a partir del ejercicio en el cual el beneficiario cumpla los 65 años de edad. El monto de la exoneración será el equivalente al doble de la fracción básica exenta de Impuesto a la Renta.

7.- A partir del año 2013, conforme lo dispuesto en la Ley Orgánica de Discapacidades el monto de la exoneración por discapacidad será el equivalente al doble de la fracción básica exenta de Impuesto a la Renta.

8.- El presente formulario constituye la declaración de Impuesto a la Renta del trabajador, siempre que durante el período declarado la persona únicamente haya prestado sus servicios en relación de dependencia con el empleador que entrega este formulario, y no existan valores de gastos personales que deban ser reliquidados. En caso de pérdida de este documento el trabajador deberá solicitar una copia a su empleador.
Por el contrario, el trabajador deberá presentar obligatoriamente su declaración de Impuesto a la Renta cuando haya obtenido rentas en relación de dependencia con dos o más empleadores o haya recibido además de su remuneración ingresos de otras fuentes como por ejemplo: rendimientos financieros, arrendamientos, ingresos por el libre ejercicio profesional, u otros ingresos, los cuales en conjunto superen la fracción básica exenta de Impuesto a la Renta de personas naturales, o cuando tenga que reliquidar gastos personales con aquellos efectivamente incurridos, teniendo presente los límites referidos en las notas 3 y 4 de este documento.

DECLARO QUE LOS DATOS PROPORCIONADOS EN ESTE DOCUMENTO SON EXACTOS Y VERDADEROS, POR LO QUE ASUMO LA RESPONSABILIDAD LEGAL QUE DE ELLA SE DERIVEN (Art. 101 de la L.R.T.I.)

FIRMA DEL AGENTE DE RETENCIÓN	FIRMA DEL TRABAJADOR CONTRIBUYENTE	FIRMA DEL CONTADOR
		199 RUC CONTADOR 0 0 1

Figure B.2 – Tax Declaration Form for Projecting Decuctions

FORMULARIO SRI-GP

DECLARACIÓN DE GASTOS PERSONALES A SER UTILIZADOS POR EL EMPLEADOR EN EL CASO DE INGRESOS EN RELACION DE DEPENDENCIA

EJERCICIO FISCAL	2015	CIUDAD Y FECHA DE ENTREGA/RECEPCION	CIUDAD	AÑO	MES	DIA
			QUITO			

Información / Identificación del empleado contribuyente (a ser llenado por el empleado)

101	CEDULA O PASAPORTE	102	APELLIDOS Y NOMBRES COMPLETOS

INGRESOS GRAVADOS PROYECTADOS (sin decimotercera y decimocuarta remuneración) (ver Nota 1)

(+) TOTAL INGRESOS GRAVADOS CON ESTE EMPLEADOR (con el empleador que más ingresos perciba)	103	USD$	
(+) TOTAL INGRESOS CON OTROS EMPLEADORES (en caso de haberlos)	104	USD$	
(=) TOTAL INGRESOS PROYECTADOS	105	**USD$**	

GASTOS PROYECTADOS

(+) GASTOS DE VIVIENDA	106	USD$	
(+) GASTOS DE EDUCACION	107	USD$	
(+) GASTOS DE SALUD	108	USD$	
(+) GASTOS DE VESTIMENTA	109	USD$	
(+) GASTOS DE ALIMENTACION	110	USD$	
(=) TOTAL GASTOS PROYECTADOS (ver Nota 2)	111	**USD$**	

NOTAS:
1.- Cuando un contribuyente trabaje con DOS O MÁS empleadores, presentará este informe al empleador con el que perciba mayores in gresos, el que efectuará la retención considerando los ingresos gravados y deducciones (aportes personales al IESS) con todos los empleadores. Una copia certificada, con la respectiva firma y sello del empleador, será presentada a los demás empleadores para que se abstengan de efectuar retenciones sobre los pagos efectuados por concepto de remuneración del trabajo en relación de dependencia.
2. La deducción total por gastos personales no podrá superar el 50% del total de sus ingresos gravados (casillero 105), y en ningún caso será mayor al equivalente a 1.3 veces la fracción básica exenta de Impuesto a la Renta de personas naturales. A partir del año 2011 debe considerarse como cuantía máxima para cada tipo de gasto, el monto equivalente a la fracción básica exenta de Impuesto a la Renta en: vivienda 0.325 veces, educación 0.325 veces, alimentación 0.325 veces, vestimenta 0.325, salud 1.3 veces.

Identificación del Agente de Retención (a ser llenado por el empleador)

112	RUC	113	RAZON SOCIAL, DENOMINACION O APELLIDOS Y NOMBRES COMPLETOS
	1760013210001		SERVICIO DE RENTAS INTERNAS

Firmas

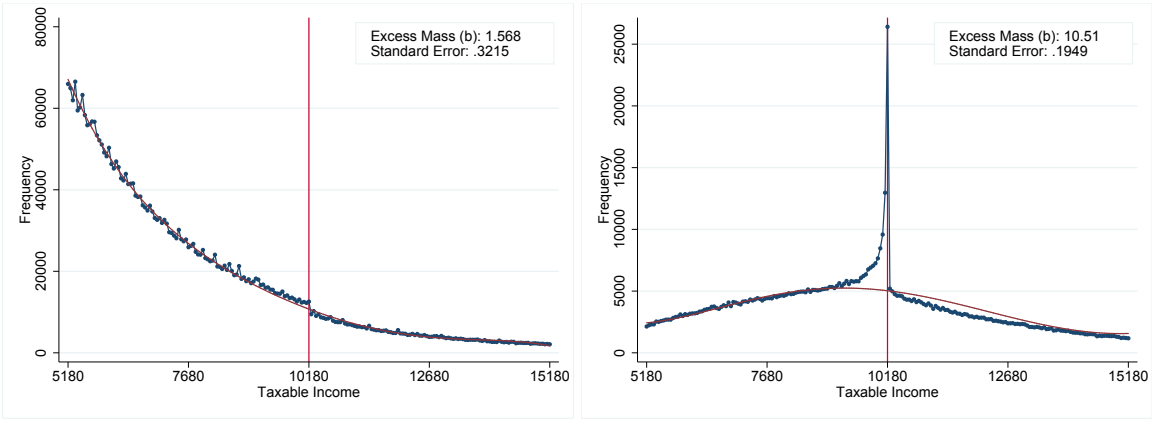
EMPLEADOR / AGENTE DE RETENCION	EMPLEADO CONTRIBUYENTE
	FIRMA DEL SERVIDOR

B.2 Subgroup Analyses

Further evidence for the fact that bunching is driven by reporting behavior can be found in Figure B.3. Individuals who do not file deductions for personal expenses do not display high levels of bunching (Figure B.3a). In contrast, individuals who file deductions (Figure B.3b) form a substantial excess mass to the left of the exemption threshold. The estimate here is extremely high (ten times as many individuals) and significant. Moreover, when only looking at gross income pooled in our sample period, our estimate of the bunching estimator is extremely small and insignificant (Figure B.5). Summing up, we find that in line with large parts of the literature, the reactions to tax incentives are mostly driven by reporting behavior rather than real labor supply responses. Furthermore, deductions for personal expenses are the primary tool used to avoid taxes.

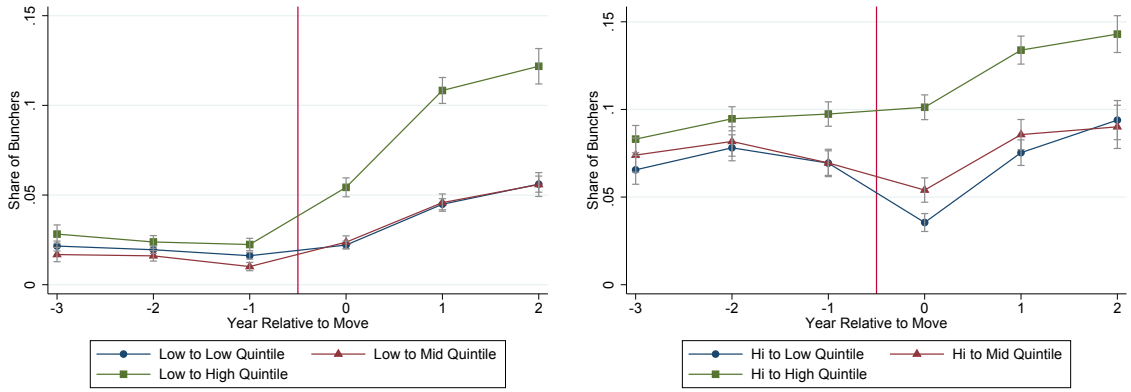
In the job switcher analysis in Section 2.4.1, the asymmetry of the response is further emphasized by the evidence in Figure B.4. The left panel shows bunching shares among workers who start from a firm in the lower quintile of the bunching distribution while the right panel refers to movers who start in the upper quintile. Among workers starting in the lower bunching quintile we see very similar patterns as before: individuals who move to the high quintile experience strong and sustained increases in bunching, whereas individuals moving to the low or mid quintile exhibit much smaller increases. Considering workers starting in the high bunching quintile we see some small additional increases among those going back to the high quintile, whereas taxpayers moving to the mid or low quintile have a temporary decrease in their probability to adjust their taxable income.

Figure B.3 – The impact of filing deductions



(a) No Deductions **(b)** With Deductions

Figure B.4 – Event Study Job Switchers



(a) from low bunching **(b)** from high bunching

Figure B.5 – Bunching Estimates Gross Income

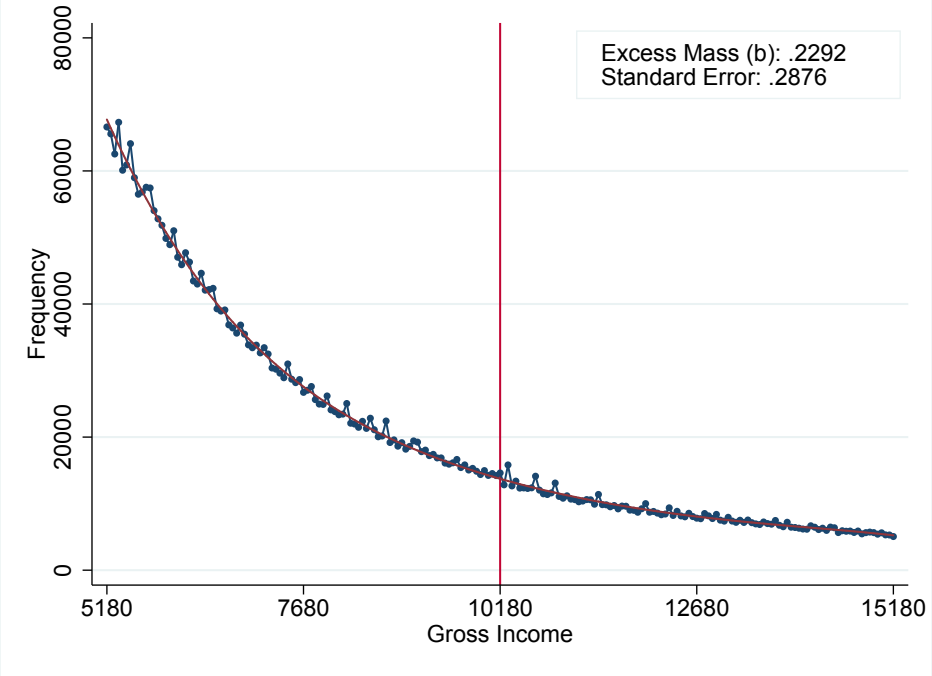
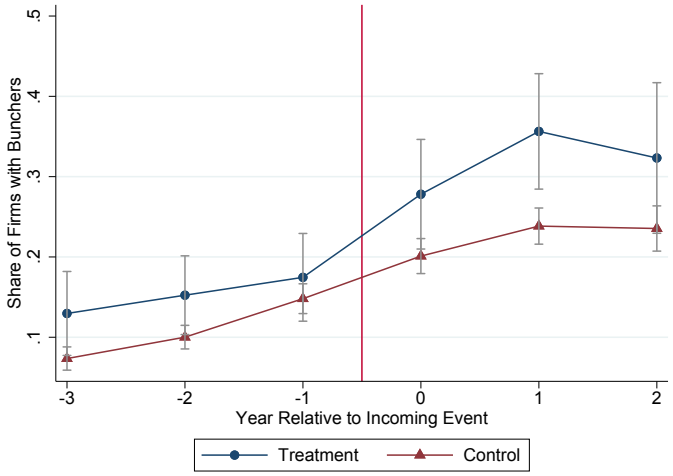
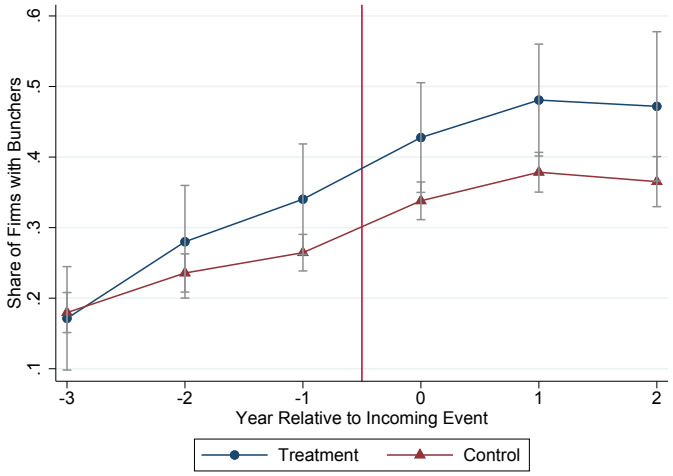


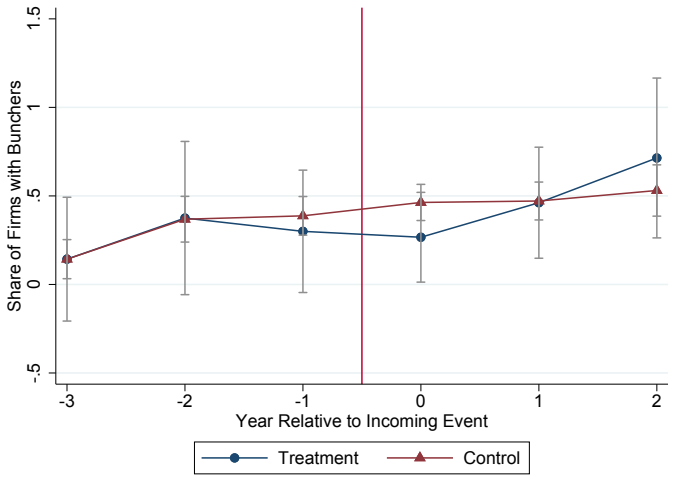
Figure B.6 – Peer Learning Event Study - Firm Size



(a) Small Firms



(b) Medium Sized Firms



(c) Large Firms

B.3 Robustness Checks: Outcomes

In this section, we perform the entire analyses of the main part for two different measures of tax avoidance. Our base measure of tax avoidance, bunching at the first kink in the tax schedule, is subject to an ad hoc choice of the bunching window around the kink (in our choice $1000 to the left of the kink). To check robustness with respect to this measure, we perform the different analyses using a more general indicator for avoiding tax payments that turns on if taxable income is below the first kink while gross income is above the first kink.

Our second robustness check employs an even more general measure of tax avoidance, the filing of deductions. Hence, we perform the analysis using an indicator whether an individual files any deduction.

B.3.1 Taxable Income below Kink

Table B.2 – Bunching Individuals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Experience exper	0.0086	0.055	0.082	0.089	0.031	0.029	0.028
	(.)	(0.0037)	(0.0042)	(0.0042)	(0.0091)	(0.0088)	(0.011)
Experience^2		-0.0053	-0.012	-0.014	-0.0042	-0.0037	0.0014
		(0.00027)	(0.0013)	(0.0012)	(0.0020)	(0.0020)	(0.0025)
Experience^3			0.00050	0.00061	0.00010	0.000070	-0.00023
			(0.000097)	(0.000091)	(0.00013)	(0.00013)	(0.00016)
Married				0.015	0.012	0.010	
				(0.0038)	(0.0042)	(0.0034)	
Age				0.0073	0.0030	0.0039	
				(0.0014)	(0.0022)	(0.0015)	
Age^2				-0.000084	-0.000040	-0.000050	
				(0.000014)	(0.000022)	(0.000015)	
Female				0.020	0.024	0.016	
				(0.0056)	(0.0066)	(0.0052)	
Secondary Education				0.087	0.071	0.047	
				(0.016)	(0.018)	(0.018)	
Tertiary Education				0.11	0.080	0.053	
				(0.016)	(0.018)	(0.019)	
Foreign				0.0028	-0.011	-0.016	
				(0.0065)	(0.0081)	(0.0074)	
Number of Jobs				-0.091	-0.094	-0.092	-0.074
				(0.0046)	(0.0048)	(0.0039)	(0.0023)
Log Gross Income					0.084	0.085	0.11
					(0.0069)	(0.0057)	(0.0072)
Gross Income Growth					-0.0094	-0.0096	-0.019
					(0.0034)	(0.0033)	(0.0050)
Corporate Firm						0.0016	0.017
						(0.0065)	(0.0046)
Firm Age						0.000022	-0.00015
						(0.00041)	(0.00012)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	No	Yes	Yes
Industry FE	No	No	No	No	No	Yes	Yes
Worker FE	No	No	No	No	No	No	Yes
R^2	0.067	0.071	0.071	0.093	0.089	0.095	0.423
Observations	618,356	618,356	618,356	618,356	508,417	508,417	508,417

The table shows results from linear regressions with a binary indicator for taxable income below the first kink (while gross income above the first kink) as dependent variable. The sample is restricted to potential bunchers in 2008 to 2015. Further (unreported) control variables include firmsize, firm age. Standard errors (in parentheses) are clustered at the firm level.

Table B.3 – Job Switchers - Descriptives

	Descriptive Statistics						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	From Mid	Mid to Low	Diff	Matched Diff	Mid to High	Diff	Matched Diff
Demographics							
Age	31.63	32.90	1.87	-0.20	30.58	-0.45	0.25
	(8.76)	(9.40)	(0.15)	(0.15)	(7.99)	(0.15)	(0.16)
Married	0.45	0.43	-0.01	0.01	0.48	0.04	0.02
	(0.50)	(0.50)	(0.01)	(0.01)	(0.50)	(0.01)	(0.01)
Female	0.24	0.22	-0.01	-0.00	0.29	0.06	-0.03
	(0.43)	(0.42)	(0.01)	(0.01)	(0.45)	(0.01)	(0.01)
Tertiary Education	0.19	0.17	0.01	0.00	0.26	0.11	-0.04
	(0.39)	(0.37)	(0.01)	(0.01)	(0.44)	(0.01)	(0.01)
Pre-Switch							
Gross Income	5964.57	5350.51	-467.12	-33.99	7101.76	1284.14	109.97
	(5734.17)	(5431.36)	(86.60)	(84.85)	(6745.05)	(106.88)	(131.24)
Taxable Income	5665.17	5109.49	-439.23	-2.23	6670.02	1121.30	171.70
	(4917.99)	(4636.13)	(73.78)	(72.12)	(5845.90)	(91.82)	(111.60)
Share Deduction Filers	0.05	0.04	-0.01	0.00	0.09	0.04	0.01
	(0.23)	(0.20)	(0.00)	(0.00)	(0.28)	(0.00)	(0.01)
Buncher	0.03	0.02	-0.01	0.00	0.04	0.01	0.01
	(0.17)	(0.14)	(0.00)	(0.00)	(0.20)	(0.00)	(0.00)
Post-Switch							
Gross Income	6288.45	5053.08	-1206.54	-103.47	8185.55	1925.92	247.28
	(5587.02)	(4607.13)	(82.79)	(72.34)	(6482.91)	(108.03)	(133.81)
Taxable Income	5822.17	4893.36	-994.87	-76.67	7120.04	1231.81	-97.42
	(4560.98)	(4053.18)	(69.86)	(62.76)	(5048.67)	(86.25)	(104.94)
Share Deduction Filers	0.09	0.03	-0.03	-0.00	0.23	0.16	0.12
	(0.29)	(0.18)	(0.00)	(0.00)	(0.42)	(0.01)	(0.01)
Buncher	0.04	0.02	-0.01	0.00	0.07	0.04	0.02
	(0.19)	(0.15)	(0.00)	(0.00)	(0.25)	(0.00)	(0.00)
Observations	19,968	7,560			5,035		

Notes: This table reports summary statistics for the job switcher sample, consisting of all individuals who switch their job between 2010 and 2014 (regarding only their first move) and for whom it is possible to observe at least two consecutive years before and after the move. Pre-move gives mean values in the two years before the move, post-move the respective values in the first two years at the destination firm. Individuals are grouped into quintiles depending on the share of co-workers with taxable income below the first kink (and gross income above the kink) for any given year. Columns (2) to (4) represent individuals starting in the mid (third) quintile of the distribution of shares in the year before the move and moving to a firm in the low (first), mid (third) or high (fifth) quintile.

Table B.4 – Job Switchers

	Mid to Low				Mid to High			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full Sample		Matching	Displaced	Full Sample		Matching	Displaced
A. Overall Effect								
After event year	-0.007	-0.004	-0.004	-0.023	0.110	0.104	0.094	0.101
	(0.003)	(0.003)	(0.004)	(0.012)	(0.009)	(0.007)	(0.009)	(0.018)
B. Effects by Relative Year								
Anticipatory Effects								
Event year - 3	0.006	0.004	0.003	0.005	-0.012	-0.012	-0.021	-0.008
	(0.004)	(0.004)	(0.005)	(0.016)	(0.006)	(0.005)	(0.012)	(0.015)
Event year - 2	0.000	-0.001	0.001	0.013	-0.007	-0.006	-0.014	-0.002
	(0.003)	(0.003)	(0.004)	(0.010)	(0.005)	(0.005)	(0.009)	(0.010)
Post Treatment Effects								
Event year	-0.007	-0.005	-0.004	-0.008	0.047	0.044	0.034	0.042
	(0.003)	(0.003)	(0.004)	(0.013)	(0.008)	(0.007)	(0.009)	(0.014)
Event year + 1	-0.008	-0.005	-0.003	-0.026	0.119	0.113	0.096	0.131
	(0.004)	(0.004)	(0.005)	(0.013)	(0.010)	(0.009)	(0.011)	(0.023)
Event year + 2	-0.001	0.001	-0.002	-0.020	0.187	0.179	0.161	0.143
	(0.006)	(0.006)	(0.008)	(0.018)	(0.016)	(0.015)	(0.018)	(0.034)
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	70,292	70,292	68,988	8,084	59,294	59,294	47,089	8,278
C. Timing								
Event year - 2	0.005	-0.002		-0.005	0.006	0.001		-0.091
	(0.004)	(0.004)		(0.015)	(0.004)	(0.004)		(0.028)
Event year	-0.000	0.005		0.001	0.026	0.022		0.120
	(0.003)	(0.005)		(0.017)	(0.005)	(0.007)		(0.025)
Event year + 1	0.015	0.016		0.017	0.071	0.059		0.281
	(0.004)	(0.008)		(0.027)	(0.006)	(0.011)		(0.047)
Controls	No	Yes		Yes	No	Yes		Yes
Observations	25,048	25,048		1,499	23,947	23,947		1,559

The panels of this table denote the results from regression equations (2.1), (2.2) and (2.3) respectively with an indicator for taxable income below the first kink (and gross income above the first kink). Standard errors (in parentheses) are clustered at the destination firm by year level.

Table B.5 – Extensive Margin of Firms with Taxable Income below Kink over time by cohort

	2008	2009	2010	2011	2012	2013	2014	2015	Obs
Cohort									
2008	0.33	0.52	0.61	0.62	0.74	0.79	0.81	0.81	489
2008	(0.47)	(0.50)	(0.49)	(0.49)	(0.44)	(0.41)	(0.39)	(0.39)	
2009		0.43	0.54	0.59	0.70	0.77	0.80	0.83	528
2009		(0.50)	(0.50)	(0.49)	(0.46)	(0.42)	(0.40)	(0.38)	
2010			0.38	0.51	0.66	0.73	0.76	0.79	555
2010			(0.48)	(0.50)	(0.48)	(0.44)	(0.43)	(0.41)	
2011				0.44	0.61	0.69	0.73	0.76	1100
2011				(0.50)	(0.49)	(0.46)	(0.44)	(0.43)	
2012					0.50	0.65	0.71	0.74	1657
2012					(0.50)	(0.48)	(0.45)	(0.44)	
2013						0.58	0.68	0.72	2203
2013						(0.49)	(0.47)	(0.45)	
2014							0.56	0.66	3280
2014							(0.50)	(0.47)	
2015								0.55	4847
2015								(0.50)	

Note: Share of firms in given cohort with at least one employee with taxable income below and gross income above the kink. Cohorts conditioned on the firm's year of entry into the formal sector. Further conditioned on employing potential bunchers in all subsequent years. Standard deviations given in parentheses.

Table B.6 – Intensive Margin of Firms with Taxable Income below Kink over time by firm cohort

		2008	2009	2010	2011	2012	2013	2014	2015
Cohort									
2008	Share	0.39	0.52	0.51	0.47	0.51	0.56	0.58	0.62
2008	SD	(0.29)	(0.28)	(0.29)	(0.26)	(0.27)	(0.26)	(0.26)	(0.26)
2008	Obs	36	83	104	130	173	201	219	208
2009	Share		0.49	0.57	0.48	0.57	0.56	0.57	0.57
2009	SD		(0.27)	(0.27)	(0.28)	(0.28)	(0.25)	(0.27)	(0.26)
2009	Obs		41	79	113	134	159	181	179
2010	Share			0.53	0.51	0.54	0.56	0.60	0.64
2010	SD			(0.31)	(0.27)	(0.27)	)0.27)	(0.26)	(0.25)
2010	Obs			30	77	101	140	159	160
2011	Share				0.47	0.53	0.56	0.59	0.62
2011	SD				(0.31)	(0.28)	(0.27)	(0.27)	(0.26)
2011	Obs				55	122	189	237	242
2012	Share					0.53	0.54	0.58	0.59
2012	SD					(0.28)	(0.26)	(0.26)	(0.28)
2012	Obs					77	158	247	266
2013	Share						0.57	0.57	0.62
2013	SD						(0.28)	(0.29)	(0.26)
2013	Obs						94	207	240
2014	Share							0.54	0.62
2014	SD							(0.29)	(0.27)
2014	Obs							133	200
2015	Share								0.61
2015	SD								(0.28)
2015	Obs								96

Note: Average share of employees with taxable income below and gross income above the kink among those with gross income in a range where it is possible to reduce taxable income below the kink. Average values for the subset of firms with at least one employee with taxable income below but gross income above the kink. Cohorts conditioned on year of entry into formal sector and having potential bunchers in all subsequent years. Further conditioned on firms employing at least 5 potential bunchers in given year. The number of observations varies between year of observation since the conditioning on having at least 5 potential bunchers leads to a yearly changing composition of the cohort. Standard deviations given in parentheses.

Table B.7 – Information Transmission: Extensive Margin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Incoming Below Kink	0.064	0.042	0.024		0.042	0.040	0.019	0.025		0.019
	(0.0075)	(0.0100)	(0.013)		(0.0099)	(0.017)	(0.020)	(0.022)		(0.020)
Incoming Below Kink above p90		0.047	0.060		0.045		0.052	0.047		0.052
		(0.012)	(0.013)		(0.012)		(0.028)	(0.029)		(0.028)
Incoming Below Kink between p50 and p90			0.034					-0.014		
			(0.014)					(0.028)		
Knowledgeable Accountant				0.040	0.039				0.057	0.057
				(0.0057)	(0.0057)				(0.0082)	(0.0082)
Avg. Age	-0.0028	-0.0028	-0.0028	-0.0029	-0.0028	-0.0014	-0.0014	-0.0014	-0.0015	-0.0014
	(0.00056)	(0.00056)	(0.00056)	(0.00056)	(0.00056)	(0.0015)	(0.0015)	(0.0015)	(0.0015)	(0.0015)
Share Female	0.037	0.037	0.037	0.037	0.038	0.024	0.025	0.024	0.024	0.026
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
Between 25 and 250 Employees	0.037	0.037	0.037	0.047	0.038	0.061	0.062	0.062	0.063	0.062
	(0.0065)	(0.0065)	(0.0065)	(0.0063)	(0.0065)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
More than 250 Employees	0.10	0.096	0.096	0.13	0.097	0.16	0.16	0.16	0.17	0.16
	(0.013)	(0.013)	(0.013)	(0.012)	(0.013)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Manufacturing	0.051	0.051	0.051	0.055	0.051					
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)					
Construction	0.041	0.041	0.041	0.042	0.041					
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)					
Trade; Repairing	0.060	0.061	0.060	0.063	0.062					
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)					
Hotel and Restaurant	0.013	0.013	0.013	0.011	0.011					
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)					
Transport, Storage, Communication	0.043	0.043	0.043	0.046	0.044					
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)					
Financial Sector	0.078	0.078	0.078	0.086	0.082					
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)					
Real Estate, Business and Renting	0.051	0.051	0.050	0.052	0.051					
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)					
Education	0.0015	0.0018	0.0013	0.0029	0.0041					
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)					
Health and Social Services	0.052	0.051	0.051	0.054	0.053					
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)					
Other	0.043	0.043	0.043	0.048	0.045					
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)					
Firm FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	35025	35025	35025	35025	35025	35025	35025	35025	35025	35025

The outcome variable is a binary indicator for a firm having at least one avoider with taxable income below but gross income above kink. Further time-varying controls: lagged bunching behavior, share married, share with tertiary education, average gross income at firm, as well as year and province fixed effects.

Table B.8 – Information Transmission: Intensive Margin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Incoming Below Kink	-0.023	-0.013	-0.022		-0.013	-0.012	-0.011	-0.014		-0.011
	(0.010)	(0.013)	(0.016)		(0.013)	(0.016)	(0.019)	(0.020)		(0.019)
Incoming Below Kink above p90		-0.018	-0.012		-0.019		-0.0015	0.00071		-0.0015
		(0.016)	(0.017)		(0.016)		(0.024)	(0.024)		(0.024)
Incoming Below Kink between p50 and p90			0.015					0.0070		
			(0.016)					(0.024)		
Knowledgeable Accountant				0.0080	0.0090				-0.0031	-0.0030
				(0.0089)	(0.0089)				(0.011)	(0.011)
Avg. Age	-0.00033	-0.00036	-0.00036	-0.00013	-0.00030	0.0038	0.0038	0.0039	0.0039	0.0039
	(0.0017)	(0.0016)	(0.0016)	(0.0016)	(0.0016)	(0.0041)	(0.0041)	(0.0041)	(0.0041)	(0.0041)
Share Female	0.11	0.11	0.11	0.11	0.11	0.041	0.041	0.041	0.043	0.041
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.092)	(0.092)	(0.092)	(0.092)	(0.092)
Between 25 and 250 Employees	-0.051	-0.051	-0.052	-0.056	-0.051	-0.030	-0.030	-0.031	-0.031	-0.031
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
More than 250 Employees	-0.17	-0.17	-0.17	-0.18	-0.17	-0.063	-0.063	-0.064	-0.065	-0.063
	(0.018)	(0.018)	(0.018)	(0.017)	(0.018)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
Manufacturing	0.028	0.028	0.029	0.027	0.029					
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)					
Construction	0.0078	0.0079	0.0074	0.0078	0.0079					
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)					
Trade; Repairing	0.050	0.050	0.049	0.050	0.050					
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)					
Hotel and Restaurant	0.027	0.027	0.027	0.027	0.026					
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)					
Transport, Storage, Communication	0.036	0.035	0.035	0.035	0.035					
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)					
Financial Sector	0.040	0.040	0.040	0.039	0.041					
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)					
Real Estate, Business and Renting	0.028	0.028	0.028	0.028	0.028					
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)					
Education	-0.068	-0.069	-0.070	-0.067	-0.069					
	(0.035)	(0.035)	(0.035)	(0.034)	(0.035)					
Health and Social Services	0.0020	0.0034	0.0033	0.0041	0.0036					
	(0.033)	(0.033)	(0.032)	(0.033)	(0.033)					
Other	0.019	0.019	0.018	0.018	0.019					
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)					
Firm FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	7264	7264	7264	7264	7264	7264	7264	7264	7264	7264

The outcome variable is the share of avoiders among potential bunchers. Further controls: lagged bunching behavior, share married, share with tertiary education, average gross income at firm, as well as year and province fixed effects. Standard errors clustered at firm level.

Table B.9 – Peer Learning Event Study - Descriptives

	Descriptive Statistics			
	(1) Full Sample	(2) Treated	(3) Diff	(4) Matched Diff
Demographics				
Avg Age	36.02	36.36	0.48	-0.31
	(6.03)	(6.14)	(0.24)	(0.30)
Share Married	0.52	0.52	0.01	0.00
	(0.24)	(0.23)	(0.01)	(0.01)
Share Female	0.37	0.39	0.03	-0.03
	(0.27)	(0.26)	(0.01)	(0.01)
Share Tertiary Education	0.32	0.34	0.02	0.01
	(0.26)	(0.27)	(0.01)	(0.01)
Pre-Event				
Firmsize	50.74	50.82	0.10	-2.73
	(120.13)	(96.22)	(4.84)	(6.30)
Corporate Firm	0.85	0.87	0.03	-0.00
	(0.36)	(0.34)	(0.01)	(0.02)
Avg Gross Income	6903.01	7528.15	888.66	252.44
	(4052.51)	(4359.96)	(162.46)	(209.82)
Avg Taxable Income	6231.00	6658.35	607.49	148.79
	(3177.46)	(3310.22)	(127.53)	(160.47)
Share with Avoiders	0.33	0.39	0.09	0.05
	(0.47)	(0.49)	(0.02)	(0.02)
Post-Event				
Avg Gross Income	7761.76	8193.82	614.19	178.71
	(3949.91)	(4171.53)	(158.74)	(198.53)
Avg Taxable Income	6925.06	7135.82	299.60	45.20
	(3073.59)	(3062.69)	(123.71)	(149.11)
Share with Avoiders	0.41	0.49	0.12	0.07
	(0.49)	(0.50)	(0.02)	(0.02)
Observations	2,954	876		

Notes: This table shows descriptive statistics for the sample of firms used in the event study quantifying the peer learning channel. The sample consists of all firms receiving one incoming employee between 2010 and 2014 and for which it is possible to observe at least two consecutive years before and after the event. Treated refers to firms receiving incoming avoiders with taxable income below but gross income above the kink prior to joining their new firm. Column (3) displays the difference between treated and control and column (4) this same difference for the matched sample. Matching was done on average age, share married, female and tertiary educated, firmsize, corporate status of firm and average gross income pre and post event. Pre-event refers to the year before the arrival of new co-workers and post-event to the first year after the arrival of the new coworkers.

Table B.10 – Peer Learning - Regression Results

	(1)	(2)	(3)
	Full Sample		Matching
A. Overall Effect			
DiD estimate	0.036	0.033	0.035
	(0.016)	(0.016)	(0.021)
B. Effects by Relative Year			
Anticipatory Effects			
Event year - 2	0.023	0.025	0.039
	(0.017)	(0.017)	(0.022)
Post Treatment Effects			
Event year	0.052	0.049	0.051
	(0.019)	(0.019)	(0.025)
Event year + 1	0.040	0.038	0.040
	(0.021)	(0.021)	(0.027)
Event year + 2	0.038	0.038	0.060
	(0.028)	(0.027)	(0.037)
Controls	No	Yes	Yes
Observations	15,913	15,913	9,418

Notes: The table reports results from the event-study regression equation (2.5) at the firm level. Outcome variable is the leave-out firm avoidance decision and event year refers to the year of incoming employees. Event year - 1 is excluded and serves as the base category. Firm and year fixed effects are included throughout. Columns (1) and (2) refer to the full sample, and column (3) uses matching on observables. We control for average gross income, average age, share married, share female, share tertiary educated, firmsize, corporate status of firm, as well as industry and province dummies and dummies for the year of the incoming event. Standard errors (in parentheses) are clustered at the firm level.

Table B.11 – Experts Event Study - Descriptives

	Descriptive Statistics			
	(1) Full Sample	(2) Treated	(3) Diff	(4) Matched Diff
Demographics				
Avg Age	36.00	36.07	-0.03	0.06
	(7.40)	(7.29)	(0.15)	(0.18)
Share Married	0.47	0.47	-0.00	0.01
	(0.28)	(0.27)	(0.01)	(0.01)
Share Female	0.41	0.41	0.00	-0.00
	(0.30)	(0.30)	(0.01)	(0.01)
Share Tertiary Education	0.28	0.28	-0.00	0.01
	(0.28)	(0.27)	(0.01)	(0.01)
Pre-Event				
Firmsize	43.99	49.11	7.87	-2.07
	(178.40)	(224.39)	(3.63)	(5.25)
Avg Gross Income	5217.38	5801.08	783.54	29.86
	(4529.22)	(4891.55)	(92.75)	(122.52)
Avg Taxable Income	4766.13	5215.96	597.02	-12.16
	(3671.01)	(3929.51)	(75.36)	(98.88)
Share with below	0.21	0.26	0.07	0.02
	(0.41)	(0.44)	(0.01)	(0.01)
Post-Event				
Avg Gross Income	5227.81	5882.51	907.73	176.91
	(4668.04)	(5166.17)	(84.52)	(120.25)
Avg Taxable Income	4770.90	5267.86	691.88	127.25
	(3837.20)	(4128.78)	(69.57)	(95.53)
Share with below	0.21	0.27	0.09	0.03
	(0.41)	(0.45)	(0.01)	(0.01)
Observations	16,389	4,201		

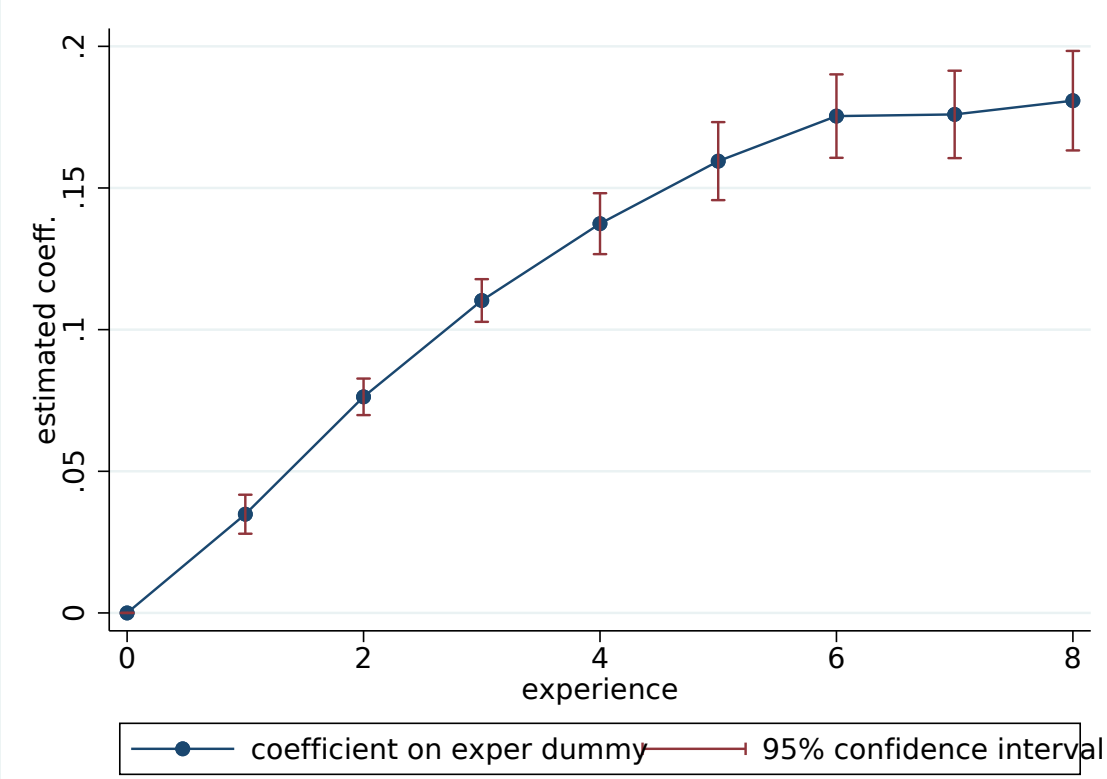
Notes: This table shows descriptive statistics for the sample of firms used in the event study quantifying the experts channel. The sample is based on the universe of accountant switches between 2010 and 2014 for which it is possible to observe at least two consecutive years before and after the event and the firms employ potential bunchers throughout. Cases in which firms simultaneously received knowledgeable and non-knowledgeable accountants were excluded. Treated refers to firms receiving new accountants previously working at a firm in which employees were avoiding paying taxes by having taxable income below but gross income above the kink. Columns (3) displays the difference to a control group consisting of firms receiving an accountant previously working at a firm with potential bunchers but with zero tax avoiders. Column (4) displays the difference to the matched sample. Matching was done on average age, share married, female and tertiary educated, firmsize, and average gross income pre and post event. Pre-event refers to the year before the arrival of the new accountants and post-event to the first year after the arrival of the new accountants.

Table B.12 – Experts Event Study - Regression Results

	(1)	(2)	(3)
	Full Sample		Matching
A. Overall Effect			
DiD estimate	0.119	0.015	0.014
	(0.007)	(0.008)	(0.013)
B. Effects by Relative Year			
Anticipatory Effects			
Event year - 2	-0.013	-0.000	-0.003
	(0.007)	(0.009)	(0.013)
Post Treatment Effects			
Event year	0.078	0.014	0.008
	(0.007)	(0.009)	(0.016)
Event year + 1	0.133	0.015	0.012
	(0.009)	(0.010)	(0.016)
Event year + 2	0.172	0.017	0.023
	(0.011)	(0.013)	(0.020)
Controls	No	Yes	Yes
Observations	60,483	60,483	28,243

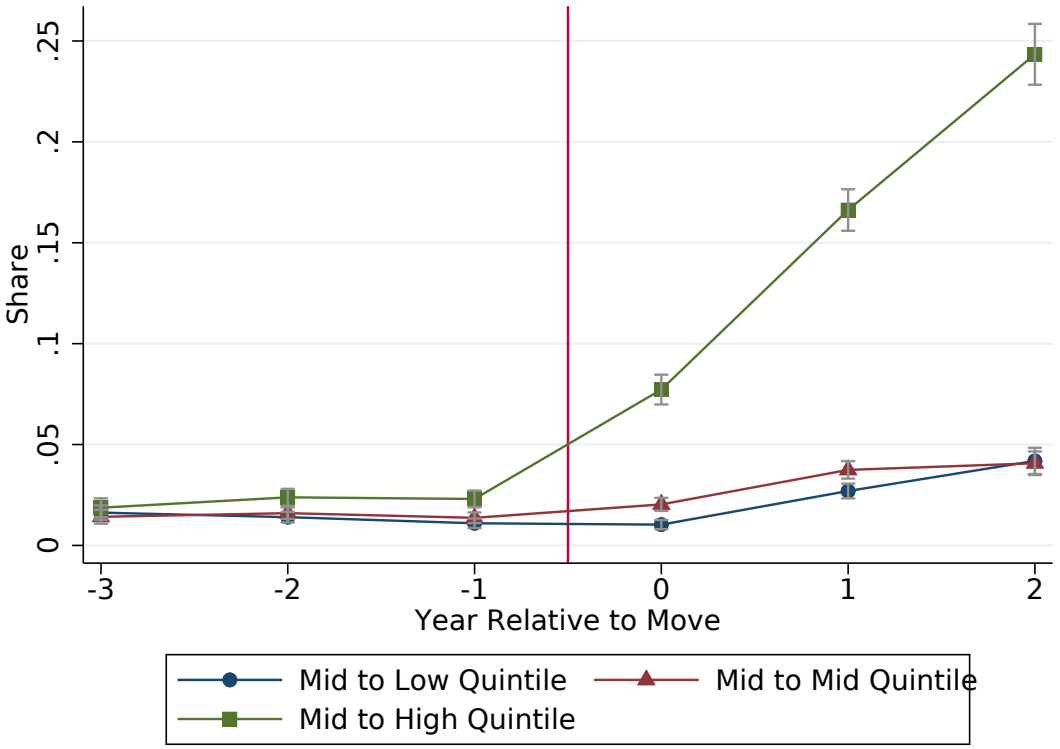
Notes: The table reports results from the event study regressions quantifying the experts channel detailed in Section (2.4.2). Outcome variable is the firm avoiding decision and event year refers to the year of the incoming accountant. Event year - 1 is excluded and serves as the base category. Firm and year fixed effects are included throughout. Columns (1) and (2) refer to the full sample, and column (3) uses matching on observables. We control for average gross income, average age, share married, share female, share tertiary educated, firmsize, as well as industry and province dummies and dummies for the year of the accountant switch. Standard errors (in parentheses) are clustered at the firm level.

Figure B.7 – Coefficients on experience dummies



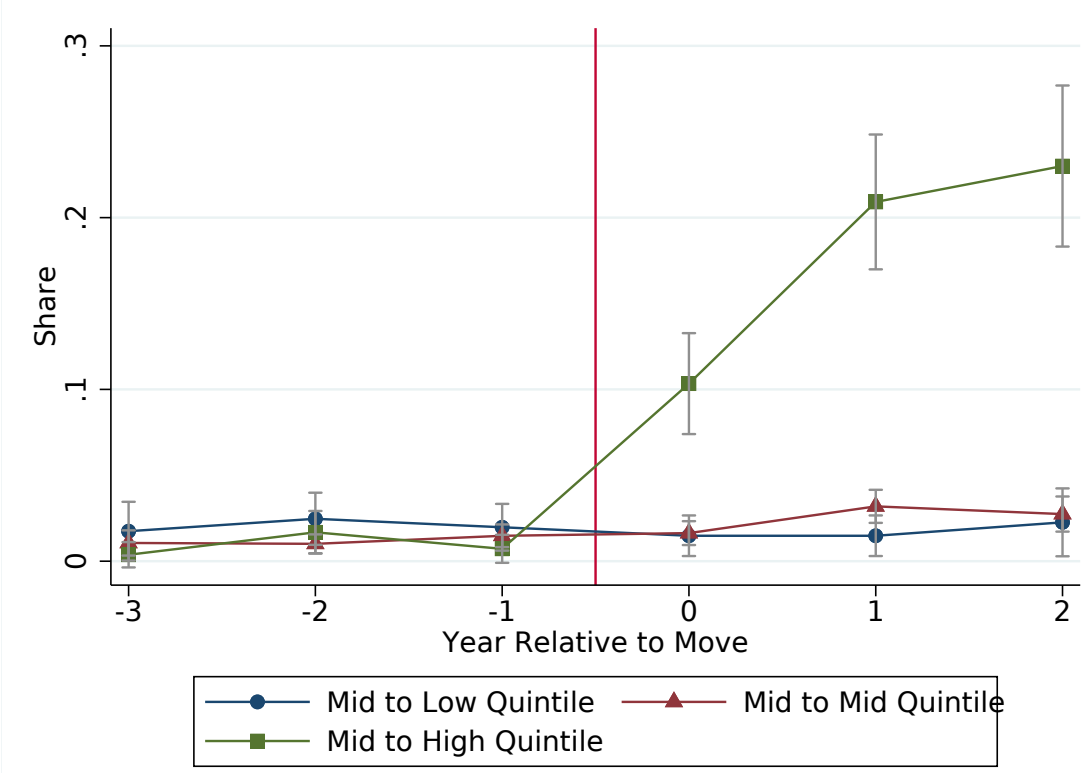
This figure depicts estimated coefficients from a linear regression of an avoiding indicator on dummy variables for each year of experience in the formal sector. We control for individual fixed effects, income dynamics and a broad range of firm characteristics.

Figure B.8 – Event Study Job Switchers



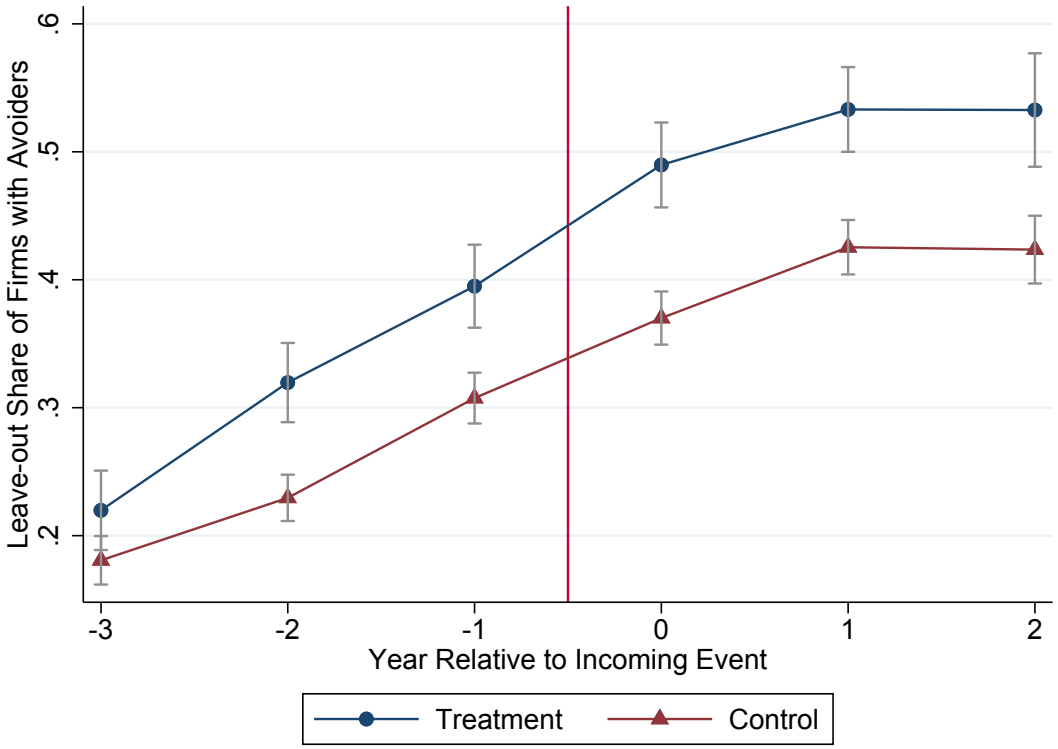
This figure shows an event study with avoiding shares of job switchers around the time of the job transition. The vertical line indicates the time of the transition. We observe avoiding among individuals who come from a firm in the medium quintile of the distribution of co-worker bunching shares and differentiate between those who switch to a firm in the bottom, medium, and top quintile.

Figure B.9 – Event Study Job Switchers – Sample of Displaced Workers



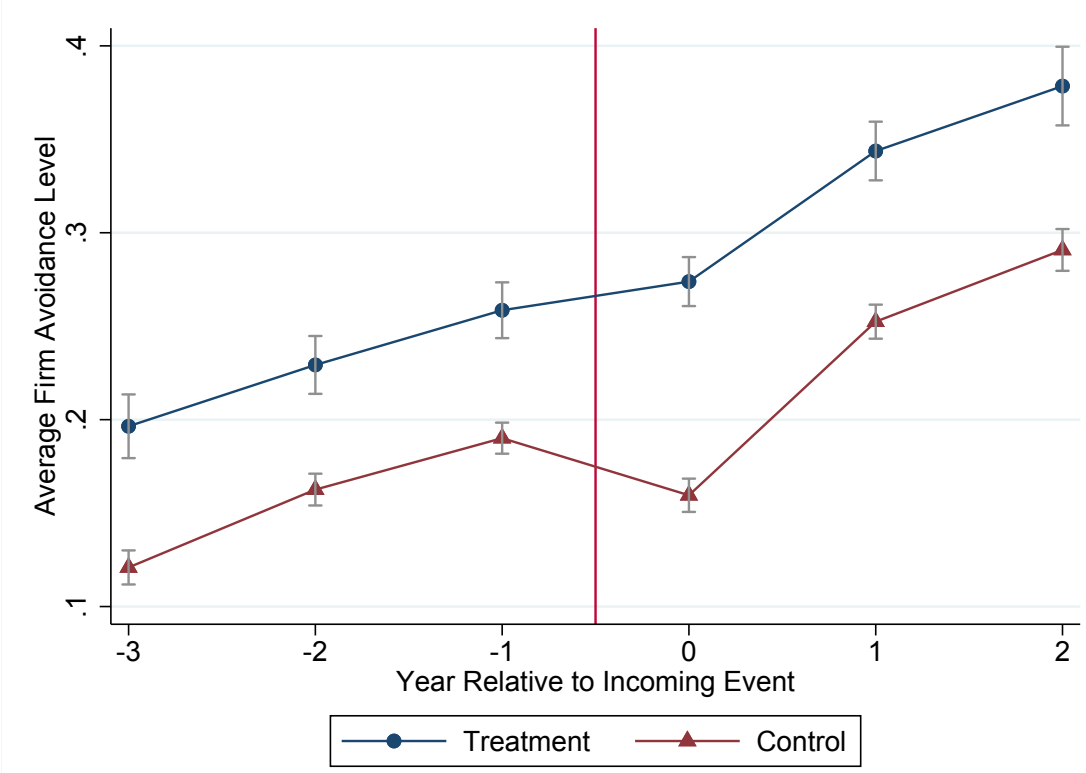
This figure shows an event study with avoiding shares of job switchers around the time of the job transition in the subsample of workers who exogenously loose their job due to a firm closure.

Figure B.10 – Peer Learning Event Study



This figure shows an event study on the firm level with the share of firms employing avoiders around the hiring of a new co-worker (leaving out the new worker from the calculation). The vertical line denotes the arrival of the new worker. The treatment group is formed by firms that receive a new co-worker who was avoiding in her previous firm while the control group is formed by firms with a new co-worker who was not avoiding (despite being a potential avoider with gross income in the range above the kink).

Figure B.11 – Experts Event Study



This figure shows an event study on the firm level with the average share of avoiders around the entry of a new accountant into the firm. The vertical line denotes the arrival of the new accountant. The treatment group is formed by firms that receive an accountant who was previously working for a firm with avoiders while the control group is formed by firms with a new accountant who was working for a firm without any avoiders.

B.3.2 Filing Deductions

Table B.13 – Bunching Individuals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Experience	0.029	0.10	0.15	0.16	-0.099	-0.10	-0.022
	(0.0023)	(0.0050)	(0.0075)	(0.0083)	(0.011)	(0.0097)	(0.0089)
Experience^2		-0.0084	-0.021	-0.023	0.020	0.020	0.014
		(0.00036)	(0.0014)	(0.0015)	(0.0021)	(0.0020)	(0.0019)
Experience^3			0.00092	0.0010	-0.0012	-0.0012	-0.00093
			(0.000093)	(0.000094)	(0.00013)	(0.00012)	(0.00012)
Married				0.025	0.012	0.010	
				(0.0046)	(0.0045)	(0.0036)	
Age				0.016	-0.0028	-0.0012	
				(0.0022)	(0.0030)	(0.0020)	
Age^2				-0.00019	0.0000039	-0.000010	
				(0.000022)	(0.000030)	(0.000021)	
Female				0.0077	0.018	0.014	
				(0.0080)	(0.0087)	(0.0067)	
Secondary Education				0.18	0.10	0.075	
				(0.021)	(0.024)	(0.023)	
Tertiary Education				0.26	0.15	0.12	
				(0.021)	(0.024)	(0.024)	
Foreign				0.061	0.021	0.018	
				(0.0090)	(0.0087)	(0.0080)	
Number of Jobs				-0.11	-0.093	-0.092	-0.080
				(0.0061)	(0.0063)	(0.0049)	(0.0026)
Log Gross Income					0.33	0.33	0.25
					(0.0050)	(0.0050)	(0.0050)
Gross Income Growth					-0.075	-0.075	-0.053
					(0.0021)	(0.0020)	(0.0026)
Corporate Firm						0.014	0.016
						(0.0081)	(0.0057)
Firm Age						-0.000046	-0.000013
						(0.00061)	(0.00017)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	No	Yes	Yes
Industry FE	No	No	No	No	No	Yes	Yes
Worker FE	No	No	No	No	No	No	Yes
R^2	0.138	0.145	0.145	0.180	0.305	0.312	0.606
Observations	618,356	618,356	618,356	618,356	508,417	508,417	508,417

The table shows results from linear regressions with a binary indicator for filing any deductions as dependent variable. The sample is restricted to potential bunchers in 2008 to 2015. Further (unreported) control variables include firmsize, firm age. Standard errors (in parentheses) are clustered at the firm level.

Table B.14 – Job Switchers - Descriptives

	Descriptive Statistics						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	From Mid	Mid to Low	Diff	Matched Diff	Mid to High	Diff	Matched Diff
Demographics							
Age	31.66	33.43	2.37	-0.12	30.48	-0.58	0.03
	(8.92)	(9.74)	(0.16)	(0.18)	(7.94)	(0.15)	(0.16)
Married	0.46	0.45	-0.00	-0.01	0.48	0.03	0.01
	(0.50)	(0.50)	(0.01)	(0.01)	(0.50)	(0.01)	(0.01)
Female	0.27	0.25	-0.03	0.01	0.27	-0.01	-0.00
	(0.44)	(0.43)	(0.01)	(0.01)	(0.44)	(0.01)	(0.01)
Tertiary Education	0.20	0.18	0.02	-0.02	0.27	0.11	-0.03
	(0.40)	(0.39)	(0.01)	(0.01)	(0.44)	(0.01)	(0.01)
Pre-Switch							
Gross Income	5622.45	5089.50	-272.77	119.02	6581.08	1218.81	-152.99
	(4591.53)	(4464.57)	(73.19)	(74.25)	(5158.91)	(81.10)	(104.01)
Taxable Income	5375.94	4882.35	-271.50	52.78	6237.73	1083.89	-13.54
	(3925.14)	(3827.79)	(62.57)	(64.05)	(4398.04)	(69.13)	(86.26)
Share Deduction Filers	0.05	0.04	-0.00	0.01	0.07	0.03	-0.01
	(0.21)	(0.19)	(0.00)	(0.00)	(0.26)	(0.00)	(0.01)
Buncher	0.03	0.03	0.00	-0.00	0.05	0.02	0.01
	(0.18)	(0.16)	(0.00)	(0.00)	(0.21)	(0.00)	(0.00)
Post-Switch							
Gross Income	6259.49	4719.54	-1141.97	2.72	8541.21	2679.70	372.29
	(5122.91)	(3736.68)	(71.82)	(65.95)	(6297.83)	(94.57)	(124.52)
Taxable Income	5813.79	4627.89	-957.02	44.46	7463.84	1878.93	71.73
	(4216.59)	(3376.41)	(61.73)	(58.38)	(5025.47)	(76.93)	(99.78)
Share Deduction Filers	0.10	0.02	-0.03	-0.01	0.25	0.20	0.14
	(0.30)	(0.14)	(0.00)	(0.00)	(0.43)	(0.01)	(0.01)
Buncher	0.04	0.02	-0.01	0.00	0.07	0.04	0.02
	(0.19)	(0.15)	(0.00)	(0.00)	(0.25)	(0.00)	(0.00)
Observations	19,365	6,216			5,525		

Notes: This table reports summary statistics for the job switcher sample, consisting of all individuals who switch their job between 2010 and 2014 (regarding only their first move) and for whom it is possible to observe at least two consecutive years before and after the move. Pre-move gives mean values in the two years before the move, post-move the respective values in the first two years at the destination firm. Individuals are grouped into quintiles depending on the share of co-workers filing any deductions for any given year. Columns (2) to (4) represent individuals starting in the mid (third) quintile of the distribution of shares in the year before the move and moving to a firm in the low (first), mid (third) or high (fifth) quintile.

Table B.15 – Job Switchers

	Mid to Low				Mid to High			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full Sample		Matching	Displaced	Full Sample		Matching	Displaced
A. Overall Effect								
After event year	-0.036	-0.026	-0.016	-0.025	0.264	0.240	0.197	0.223
	(0.005)	(0.005)	(0.006)	(0.016)	(0.015)	(0.011)	(0.016)	(0.025)
B. Effects by Relative Year								
Anticipatory Effects								
Event year - 3	0.005	0.000	-0.005	-0.037	-0.035	-0.025	-0.009	-0.057
	(0.006)	(0.006)	(0.006)	(0.022)	(0.009)	(0.008)	(0.016)	(0.025)
Event year - 2	0.005	0.000	-0.002	0.002	-0.021	-0.012	-0.017	0.005
	(0.004)	(0.004)	(0.006)	(0.015)	(0.007)	(0.007)	(0.015)	(0.020)
Post Treatment Effects								
Event year	-0.030	-0.024	-0.017	-0.025	0.169	0.159	0.150	0.222
	(0.005)	(0.005)	(0.006)	(0.016)	(0.021)	(0.016)	(0.020)	(0.060)
Event year + 1	-0.032	-0.025	-0.015	-0.019	0.283	0.261	0.215	0.223
	(0.006)	(0.006)	(0.007)	(0.019)	(0.016)	(0.014)	(0.019)	(0.033)
Event year + 2	-0.041	-0.036	-0.025	-0.062	0.335	0.308	0.219	0.186
	(0.009)	(0.008)	(0.011)	(0.021)	(0.018)	(0.019)	(0.025)	(0.077)
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	65,694	65,694	56,447	7,740	63,225	63,225	51,604	7,570
C. Timing								
Event year - 2	0.005	-0.002		-0.029	0.006	0.001		-0.064
	(0.004)	(0.004)		(0.024)	(0.004)	(0.004)		(0.034)
Event year	-0.000	0.005		-0.005	0.026	0.022		0.173
	(0.003)	(0.005)		(0.022)	(0.005)	(0.007)		(0.045)
Event year + 1	0.015	0.016		0.025	0.071	0.059		0.221
	(0.004)	(0.008)		(0.037)	(0.006)	(0.011)		(0.064)
Controls	No	Yes		Yes	No	Yes		Yes
Observations	25,048	25,048		2,028	23,947	23,947		1,855

The panels of this table denote the results from regression equations (2.1), (2.2) and (2.3) respectively with an indicator for filing any deductions. Standard errors (in parentheses) are clustered at the destination firm by year level.

Table B.16 – Extensive Margin of Firms with Deduction Filers over time by cohort

	2008	2009	2010	2011	2012	2013	2014	2015	Obs
Cohort									
2008	0.45	0.63	0.72	0.75	0.86	0.89	0.90	0.92	645
2008	(0.50)	(0.48)	(0.45)	(0.43)	(0.35)	(0.32)	(0.30)	(0.27)	
2009		0.57	0.70	0.75	0.82	0.88	0.89	0.91	699
2009		(0.50)	(0.46)	(0.43)	(0.38)	(0.33)	(0.31)	(0.28)	
2010			0.54	0.67	0.78	0.86	0.89	0.90	775
2010			(0.50)	(0.47)	(0.41)	(0.35)	(0.32)	(0.29)	
2011				0.55	0.75	0.84	0.87	0.88	1425
2011				(0.50)	(0.44)	(0.37)	(0.34)	(0.32)	
2012					0.65	0.81	0.86	0.86	2105
2012					(0.48)	(0.39)	(0.34)	(0.35)	
2013						0.71	0.82	0.83	2724
2013						(0.45)	(0.38)	(0.37)	
2014							0.72	0.80	3802
2014							(0.45)	(0.40)	
2015								0.68	4996
2015								(0.47)	

Note: Share of firms in given cohort with at least one employee filing deductions. Cohorts conditioned on the firm's year of entry into the formal sector and having employees with gross income above the kink in all subsequent years. Standard deviations given in parentheses.

Table B.17 – Intensive Margin of Firms with Deduction Filers over time by firm cohort

		2008	2009	2010	2011	2012	2013	2014	2015
Cohort									
2008	Share	0.60	0.71	0.72	0.70	0.78	0.78	0.81	0.81
2008	SD	(0.31)	(0.27)	(0.27)	(0.25)	(0.24)	(0.24)	(0.22)	(0.22)
2008	Obs	64	129	174	220	258	301	324	327
2009	Share		0.71	0.78	0.75	0.79	0.81	0.81	0.82
2009	SD		(0.28)	(0.24)	(0.25)	(0.24)	(0.24)	(0.23)	(0.22)
2009	Obs		56	110	167	186	227	259	252
2010	Share			0.71	0.74	0.77	0.80	0.81	0.84
2010	SD			(0.28)	(0.25)	(0.23)	(0.23)	(0.23)	(0.21)
2010	Obs			48	126	172	221	271	266
2011	Share				0.68	0.77	0.78	0.80	0.82
2011	SD				(0.30)	(0.24)	(0.23)	(0.23)	(0.23)
2011	Obs				96	198	291	369	391
2012	Share					0.72	0.77	0.80	0.82
2012	SD					(0.28)	(0.24)	(0.23)	(0.23)
2012	Obs					141	288	394	410
2013	Share						0.75	0.78	0.82
2013	SD						(0.27)	(0.25)	(0.22)
2013	Obs						165	343	376
2014	Share							0.71	0.81
2014	SD							(0.28)	(0.25)
2014	Obs							206	332
2015	Share								0.79
2015	SD								(0.23)
2015	Obs								158

Note: Average share of employees filing deductions among those with gross income above the kink. Values for given cohort, conditional on firm having at least one employee filing deductions. Cohorts conditioned on year of entry into formal sector and having employees with gross income above the kink in all subsequent years. Further conditioned on firms employing at least 5 workers with gross income above the kink in given year. The number of observations varies between year of observation since the conditioning on having at least 5 workers with gross income above the kink leads to a yearly changing composition of the cohort. Standard deviations given in parentheses.

Table B.18 – Information Transmission: Extensive Margin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Incoming Deducter	0.020	0.00066	0.012		0.00083	-0.016	-0.024	0.011		-0.023
	(0.0056)	(0.0074)	(0.0095)		(0.0073)	(0.013)	(0.015)	(0.017)		(0.015)
Incoming Deducter above p90		0.040	0.032		0.039		0.021	-0.0024		0.020
		(0.0083)	(0.0092)		(0.0083)		(0.021)	(0.021)		(0.021)
Incoming Deducter between p50 and p90			-0.022					-0.074		
			(0.0093)					(0.019)		
Knowledgeable Accountant				0.031	0.030				0.045	0.045
				(0.0050)	(0.0050)				(0.0068)	(0.0068)
Avg. Age	-0.0042	-0.0042	-0.0042	-0.0042	-0.0041	-0.0034	-0.0034	-0.0035	-0.0035	-0.0035
	(0.00052)	(0.00052)	(0.00052)	(0.00052)	(0.00052)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)
Share Female	0.030	0.030	0.031	0.031	0.031	-0.0046	-0.0043	-0.0046	-0.0018	-0.0018
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
Between 25 and 250 Employees	0.029	0.028	0.029	0.033	0.029	0.054	0.054	0.055	0.053	0.054
	(0.0057)	(0.0057)	(0.0057)	(0.0055)	(0.0057)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
More than 250 Employees	0.078	0.072	0.073	0.088	0.073	0.13	0.13	0.13	0.13	0.13
	(0.010)	(0.011)	(0.010)	(0.0095)	(0.010)	(0.025)	(0.025)	(0.024)	(0.025)	(0.025)
Manufacturing	0.035	0.035	0.034	0.036	0.035					
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)					
Construction	0.021	0.021	0.021	0.021	0.021					
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)					
Trade; Repairing	0.027	0.027	0.028	0.028	0.028					
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)					
Hotel and Restaurant	0.018	0.018	0.018	0.016	0.017					
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)					
Transport, Storage, Communication	0.011	0.011	0.011	0.012	0.012					
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)					
Financial Sector	0.042	0.042	0.042	0.046	0.045					
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)					
Real Estate, Business and Renting	0.023	0.024	0.024	0.024	0.024					
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)					
Education	-0.032	-0.032	-0.031	-0.031	-0.030					
	(0.022)	(0.021)	(0.021)	(0.021)	(0.021)					
Health and Social Services	0.022	0.021	0.021	0.023	0.023					
	(0.017)	(0.017)	(0.017)	(0.016)	(0.016)					
Other	0.0085	0.0087	0.0088	0.011	0.011					
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)					
Firm FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	35025	35025	35025	35025	35025	35025	35025	35025	35025	35025

The outcome variable is a binary indicator for a firm having at least one deducter. Further time-varying controls: lagged bunching behavior, share married, share with tertiary education, average gross income at firm, as well as year and province fixed effects. Standard errors clustered at firm level.

Table B.19 – Information Transmission: Intensive Margin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Incoming Deducter	0.018	0.0075	-0.024		0.0075	-0.035	-0.020	-0.050		-0.020
	(0.026)	(0.030)	(0.037)		(0.030)	(0.040)	(0.042)	(0.050)		(0.042)
Incoming Deducter above p90		0.020	0.039		0.020		-0.033	-0.012		-0.033
		(0.038)	(0.039)		(0.038)		(0.067)	(0.070)		(0.067)
Incoming Deducter between p50 and p90			0.056					0.070		
			(0.039)					(0.060)		
Knowledgeable Accountant				-0.0052	-0.0060				0.0085	0.0084
				(0.023)	(0.023)				(0.027)	(0.027)
Avg. Age	-0.015	-0.015	-0.015	-0.015	-0.015	-0.014	-0.014	-0.014	-0.014	-0.014
	(0.0034)	(0.0034)	(0.0034)	(0.0035)	(0.0034)	(0.0089)	(0.0090)	(0.0090)	(0.0089)	(0.0090)
Share Female	0.19	0.19	0.19	0.19	0.19	0.022	0.019	0.023	0.028	0.018
	(0.072)	(0.072)	(0.072)	(0.072)	(0.072)	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)
Between 25 and 250 Employees	0.29	0.29	0.29	0.30	0.29	0.16	0.16	0.16	0.16	0.16
	(0.027)	(0.027)	(0.027)	(0.026)	(0.027)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
More than 250 Employees	0.30	0.30	0.29	0.31	0.30	0.26	0.25	0.25	0.25	0.26
	(0.050)	(0.050)	(0.050)	(0.049)	(0.050)	(0.073)	(0.072)	(0.072)	(0.071)	(0.072)
Manufacturing	0.012	0.011	0.012	0.012	0.011					
	(0.072)	(0.072)	(0.072)	(0.072)	(0.072)					
Construction	-0.077	-0.076	-0.078	-0.077	-0.076					
	(0.071)	(0.071)	(0.071)	(0.071)	(0.071)					
Trade; Repairing	-0.032	-0.032	-0.035	-0.033	-0.032					
	(0.067)	(0.066)	(0.067)	(0.067)	(0.066)					
Hotel and Restaurant	-0.073	-0.073	-0.072	-0.075	-0.072					
	(0.080)	(0.080)	(0.080)	(0.080)	(0.080)					
Transport, Storage, Communication	-0.17	-0.17	-0.17	-0.17	-0.17					
	(0.069)	(0.069)	(0.069)	(0.069)	(0.069)					
Financial Sector	-0.14	-0.14	-0.14	-0.14	-0.14					
	(0.097)	(0.097)	(0.097)	(0.098)	(0.097)					
Real Estate, Business and Renting	-0.091	-0.090	-0.092	-0.091	-0.090					
	(0.068)	(0.068)	(0.068)	(0.068)	(0.068)					
Education	-0.27	-0.27	-0.27	-0.27	-0.27					
	(0.083)	(0.083)	(0.083)	(0.084)	(0.083)					
Health and Social Services	-0.14	-0.14	-0.14	-0.14	-0.14					
	(0.087)	(0.088)	(0.087)	(0.087)	(0.088)					
Other	-0.076	-0.075	-0.077	-0.076	-0.075					
	(0.077)	(0.077)	(0.077)	(0.077)	(0.077)					
Firm FE	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	7264	7264	7264	7264	7264	7264	7264	7264	7264	7264

The outcome variable is the share of deducters at a firm. Further controls: lagged bunching behavior, share married, share with tertiary education, average gross income at firm, as well as year and province fixed effects. Standard errors clustered at firm level.

Table B.20 – Peer Learning Event Study - Descriptives

	Descriptive Statistics			
	(1) Full Sample	(2) Treated	(3) Diff	(4) Matched Diff
Demographics				
Avg Age	36.02	36.11	0.15	0.15
	(6.03)	(6.04)	(0.22)	(0.24)
Share Married	0.52	0.52	0.01	0.00
	(0.24)	(0.23)	(0.01)	(0.01)
Share Female	0.37	0.38	0.03	-0.02
	(0.27)	(0.27)	(0.01)	(0.01)
Share Tertiary Education	0.32	0.34	0.03	0.01
	(0.26)	(0.27)	(0.01)	(0.01)
Pre-Event				
Firmsize	50.74	49.65	-1.93	-2.10
	(120.13)	(103.84)	(4.46)	(4.45)
Corporate Firm	0.85	0.87	0.03	0.01
	(0.36)	(0.34)	(0.01)	(0.01)
Avg Gross Income	6903.01	7489.99	1038.90	81.77
	(4052.51)	(4499.99)	(149.21)	(172.22)
Avg Taxable Income	6231.00	6626.24	699.54	-56.25
	(3177.46)	(3482.79)	(117.24)	(136.95)
Share with Deducters	0.39	0.45	0.10	0.03
	(0.49)	(0.50)	(0.02)	(0.02)
Post-Event				
Avg Gross Income	7761.76	8205.73	785.80	61.69
	(3949.91)	(4199.75)	(145.90)	(164.29)
Avg Taxable Income	6925.06	7167.07	428.33	-99.44
	(3073.59)	(3155.16)	(113.82)	(126.60)
Share with Deducters	0.47	0.54	0.11	0.06
	(0.50)	(0.50)	(0.02)	(0.02)
Observations	2,954	1,285		

Notes: This table shows descriptive statistics for the sample of firms used in the event study quantifying the peer learning channel. The sample consists of all firms receiving one incoming employee between 2010 and 2014 and for which it is possible to observe at least two consecutive years before and after the event. Treated refers to firms receiving incoming co-workers using deductions prior to joining their new firm. Column (3) displays the difference between treated and control and column (4) this same difference for the matched sample. Matching was done on average age, share married, female and tertiary educated, firmsize, corporate status of firm and average gross income pre and post event. Pre-event refers to the year before the arrival of new co-workers and post-event to the first year after the arrival of the new coworkers.

Table B.21 – Peer Learning - Regression Results

	(1)	(2)	(3)
	Full Sample		Matching
A. Overall Effect			
DiD estimate	0.022	0.023	0.049
	(0.015)	(0.015)	(0.020)
B. Effects by Relative Year			
Anticipatory Effects			
Event year - 2	0.022	0.028	0.038
	(0.015)	(0.015)	(0.019)
Post Treatment Effects			
Event year	0.031	0.035	0.045
	(0.017)	(0.017)	(0.023)
Event year + 1	0.037	0.040	0.079
	(0.019)	(0.019)	(0.025)
Event year + 2	0.014	0.018	0.069
	(0.024)	(0.024)	(0.032)
Controls	No	Yes	Yes
Observations	15,913	15,913	13,847

Notes: The table reports results from the event-study regression equation (2.5) at the firm level. Outcome variable is the leave-out firm deduction decision and event year refers to the year of incoming employees. Event year - 1 is excluded and serves as the base category. Firm and year fixed effects are included throughout. Columns (1) and (2) refer to the full sample, and column (3) uses matching on observables. We control for average gross income, average age, share married, share female, share tertiary educated, firmsize, corporate status of firm, as well as industry and province dummies and dummies for the year of the incoming event. Standard errors (in parentheses) are clustered at the firm level.

Table B.22 – Experts Event Study - Descriptives

	Descriptive Statistics			
	(1) Full Sample	(2) Treated	(3) Diff	(4) Matched Diff
Demographics				
Avg Age	36.00	36.10	0.02	-0.15
	(7.40(	(7.28)	(0.15)	(0.17)
Share Married	0.47	0.47	0.00	0.00
	(0.28)	(0.27)	(0.0)1	(0.01)
Share Female	0.41	0.41	0.00	-0.00
	(0.30)	(0.30)	(0.01)	(0.01)
Share Tertiary Education	0.28	0.28	-0.01	0.01
	(0.28)	(0.27)	(0.01)	(0.01)
Pre-Event				
Firmsize	43.99	47.12	5.42	-2.61
	(178.40)	(211.66)	(3.49)	(4.80)
Avg Gross Income	5217.38	5813.76	848.54	55.52
	(4529.22)	(5080.19)	(89.15)	(115.40)
Avg Taxable Income	4766.13	5232.50	656.64	11.70
	(3671.01)	(4117.10)	(72.43)	(93.26)
Share using Deductions	0.27	0.32	0.09	0.02
	(0.44)	(0.47)	(0.01)	(0.01)
Post-Event				
Avg Gross Income	5227.81	5887.27	968.05	100.96
	(4668.04)	(5437.23)	(81.09)	(119.54)
Avg Taxable Income	4770.90	5285.93	758.88	96.06
	(3837.20)	(4415.80)	(66.74)	(96.23)
Share using Deductions	0.27	0.33	0.10	0.04
	(0.44)	(0.47)	(0.01)	(0.01)
Observations	16,389	4,824		

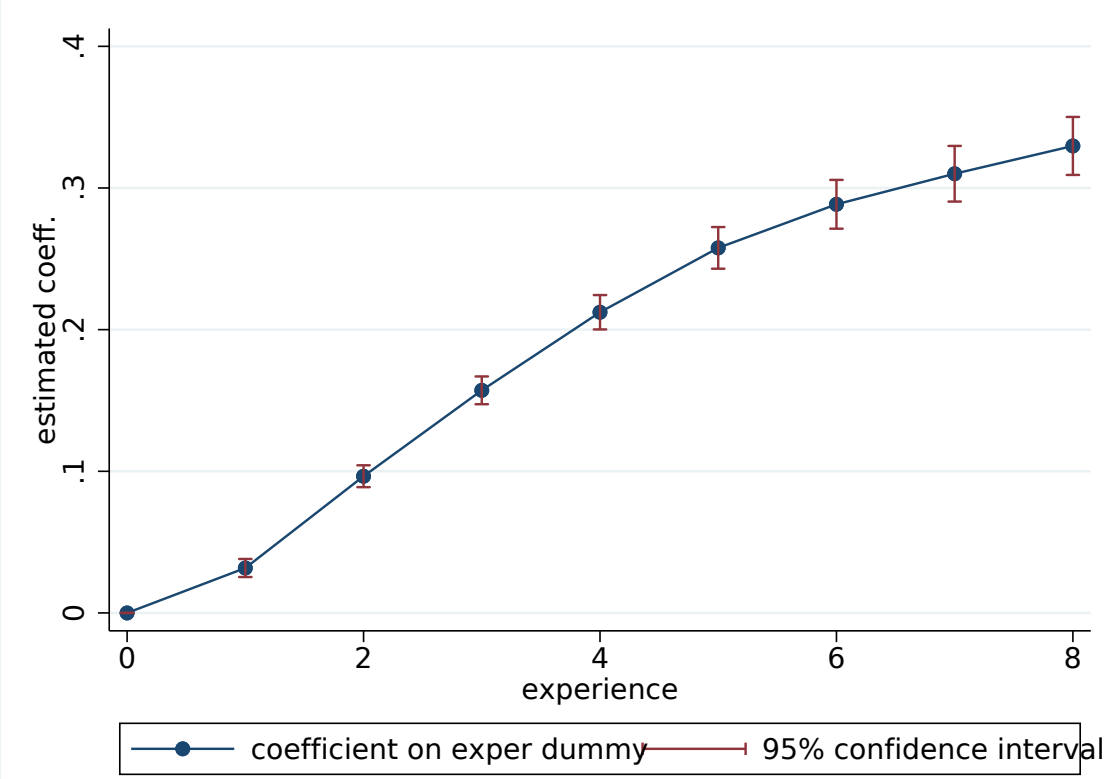
Notes: This table shows descriptive statistics for the sample of firms used in the event study quantifying the experts channel. The sample is based on the universe of accountant switches between 2010 and 2014 for which it is possible to observe at least two consecutive years before and after the event and the firms employ potential bunchers throughout. Cases in which firms simultaneously received knowledgeable and non-knowledgeable accountants were excluded. Treated refers to firms receiving new accountants previously working at a firm in which employees were using deductions. Columns (3) displays the difference to a control group consisting of firms receiving an accountant previously working at a firm with potential bunchers but zero employees using deductions. Column (4) displays the difference to the matched sample. Matching was done on average age, share married, female and tertiary educated, firmsize, and average gross income pre and post event. Pre-event refers to the year before the arrival of the new accountants and post-event to the first year after the arrival of the new accountants.

Table B.23 – Experts Event Study - Regression Results

	(1)	(2)	(3)
	Full Sample		Matching
A. Overall Effect			
DiD estimate	0.130	0.015	0.031
	(0.007)	(0.008)	(0.012)
B. Effects by Relative Year			
Anticipatory Effects			
Event year - 2	-0.010	0.005	-0.011
	(0.007)	(0.008)	(0.012)
Post Treatment Effects			
Event year	0.083	0.005	0.017
	(0.007)	(0.009)	(0.014)
Event year + 1	0.147	0.018	0.022
	(0.008)	(0.010)	(0.015)
Event year + 2	0.192	0.034	0.054
	(0.010)	(0.012)	(0.018)
Controls	No	Yes	Yes
Observations	60,483	60,483	32,075

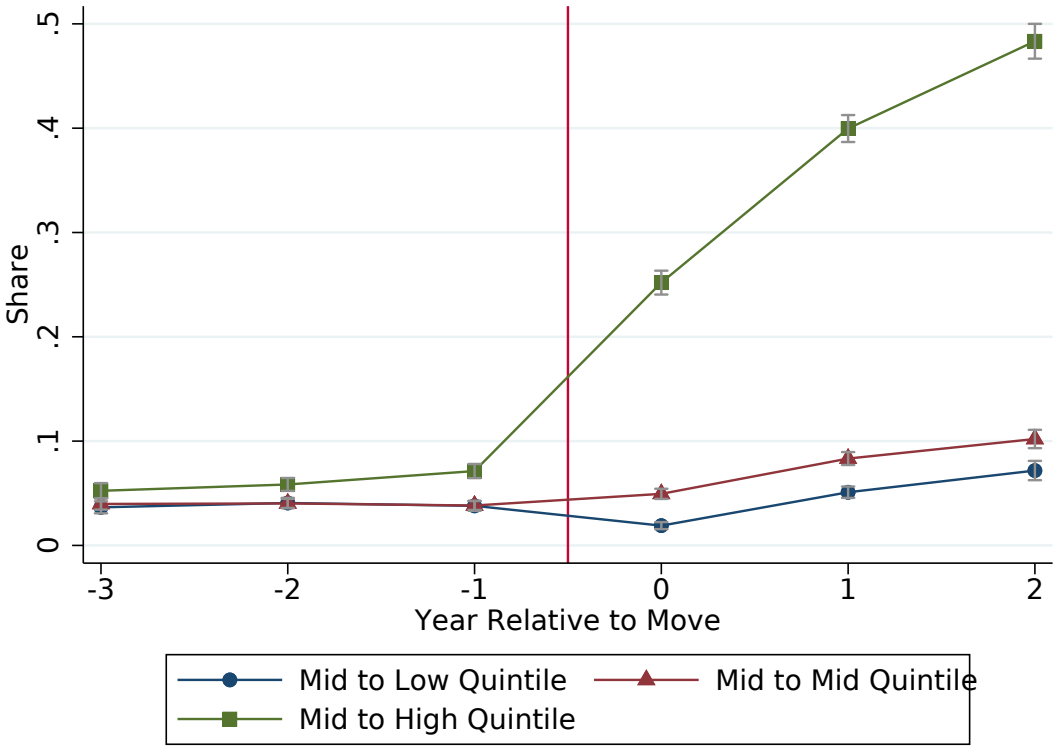
Notes: The table reports results from the event study regressions quantifying the experts channel detailed in Section (2.4.2). Outcome variable is the firm deduction decision and event year refers to the year of the incoming accountant. Event year - 1 is excluded and serves as the base category. Firm and year fixed effects are included throughout. Columns (1) and (2) refer to the full sample, and column (3) uses matching on observables. We control for average gross income, average age, share married, share female, share tertiary educated, firmsize, as well as industry and province dummies and dummies for the year of the accountant switch. Standard errors (in parentheses) are clustered at the firm level.b

Figure B.12 – Coefficients on experience dummies



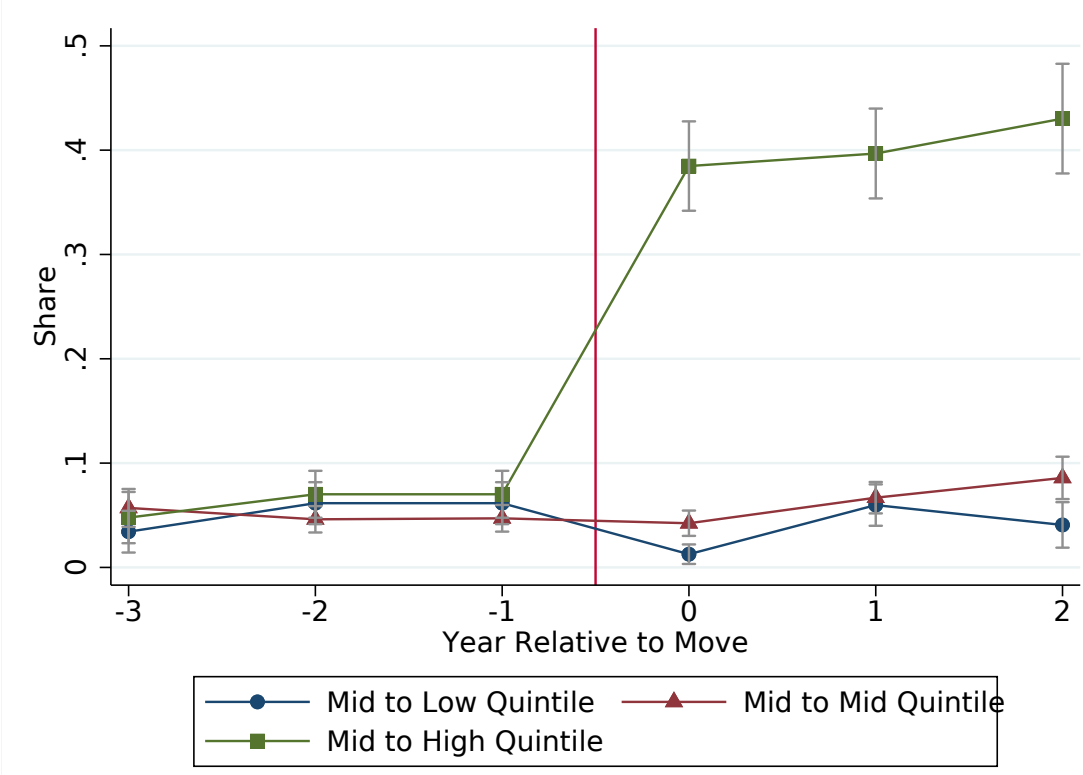
This figure depicts estimated coefficients from a linear regression of a deducting indicator on dummy variables for each year of experience in the formal sector. We control for individual fixed effects, income dynamics and a broad range of firm characteristics.

Figure B.13 – Event Study Job Switchers



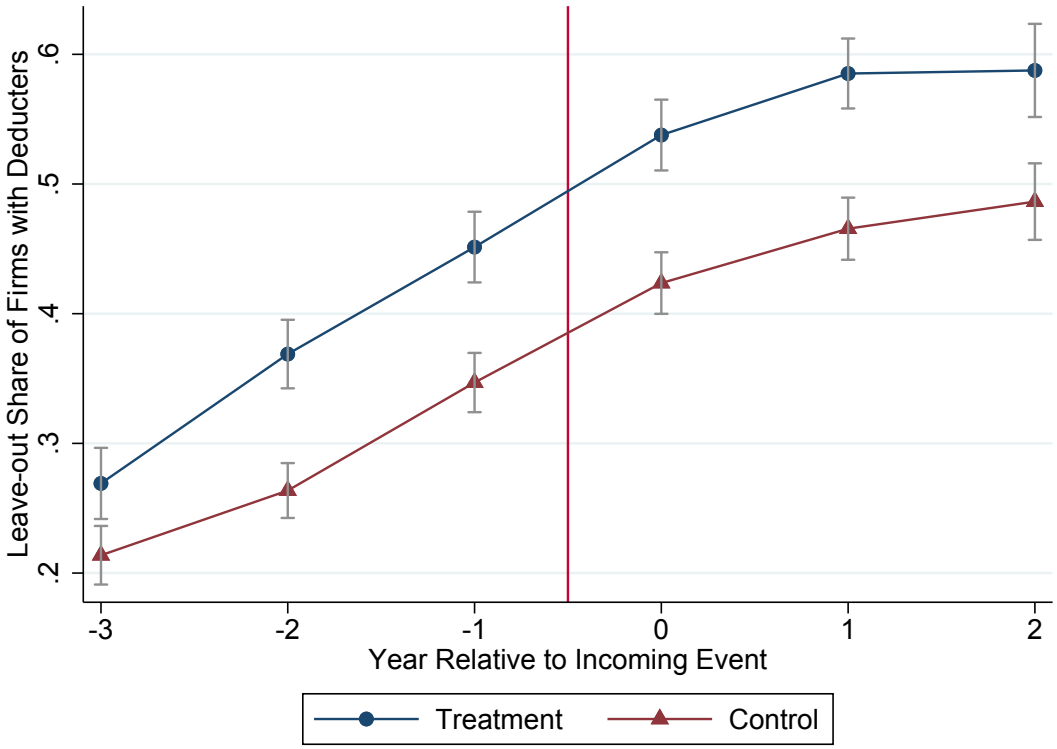
This figure shows an event study with deducting shares of job switchers around the time of the job transition. The vertical line indicates the time of the transition. We observe deducting among individuals who come from a firm in the medium quintile of the distribution of co-worker deducting shares and differentiate between those who switch to a firm in the bottom, medium, and top quintile.

Figure B.14 – Event Study Job Switchers – Sample of Displaced Workers



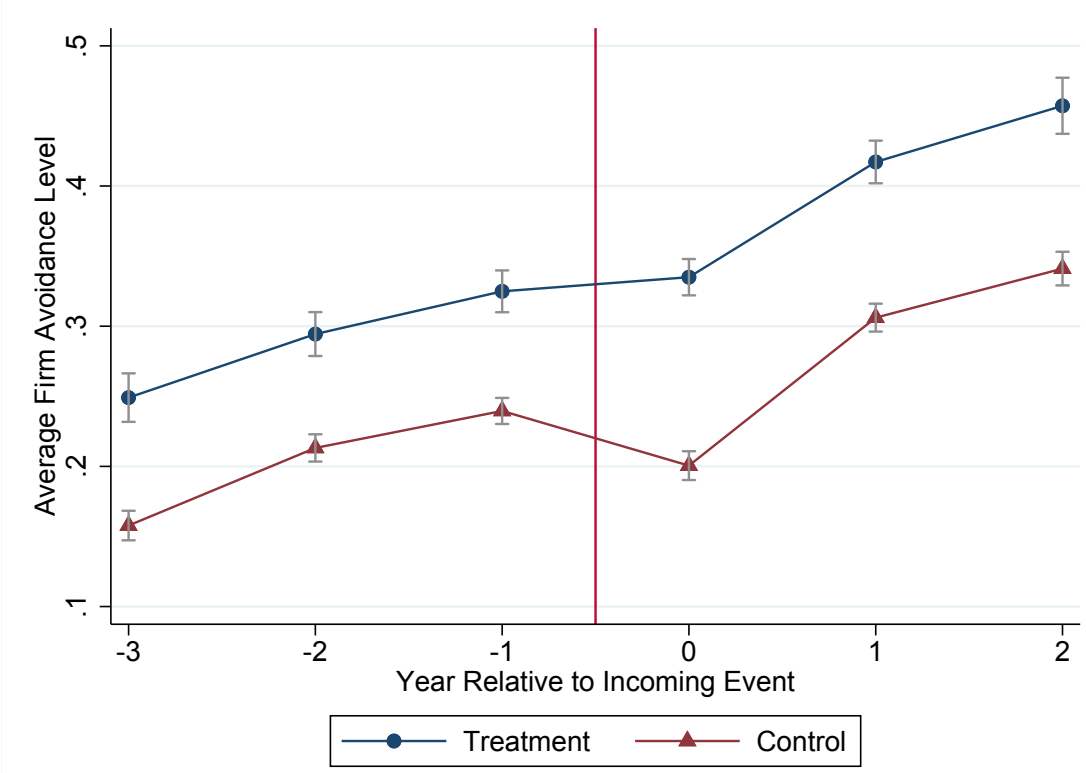
This figure shows an event study with deducting shares of job switchers around the time of the job transition in the subsample of workers who exogenously loose their job due to a firm closure.

Figure B.15 – Peer Learning Event Study



This figure shows an event study on the firm level with the share of firms employing deducters around the hiring of a new co-worker (leaving out the new worker from the calculation). The vertical line denotes the arrival of the new worker. The treatment group is formed by firms that receive a new co-worker who was deducting in her previous firm while the control group is formed by firms with a new co-worker who was not deducting (despite being a potential deducter with gross income in the range above the kink).

Figure B.16 – Experts Event Study



This figure shows an event study on the firm level with the average share of deducters around the entry of a new accountant into the firm. The vertical line denotes the arrival of the new accountant. The treatment group is formed by firms that receive an accountant who was previously working for a firm with deducters while the control group is formed by firms with a new accountant who was working for a firm without any deducters.

Appendix C

Appendix to Chapter 3

Table C.1 – Comparison Availability of Professions Data

	With Data on Clearly Defined Profession	Without Data on Clearly Defined Profession
Self-employment Profit	8,193.69	8,045.50
	(34,956.80)	(34,471.54)
Age	46.27	43.18
	(12.27)	(12.97)
Female	0.33	0.38
	(0.47)	(0.49)
Married	0.70	0.61
	(0.46)	(0.49)
Tertiary	0.75	0.27
	(0.43)	(0.44)
Observations	1,880,649	3,479,112

Notes: This table provides descriptive statistics on the universe of individuals with self-employed tax declarations. Individuals are divided into subgroups as to whether or not they belong to one of the clearly defined professions presented in Section 3.3. The table reports means for each subgroup and standard deviations in parentheses.

Table C.2 – Effects Individual Tax Returns - Balanced Panel

	(1)	(2)	(3)	(4)
Panel A: Double Differences				
	Self-employment Profits		Log Self-employment Profits	
DD (mean × post)	3890.7***		0.496***	
	(166.7)		(0.0356)	
DD (p90 × post)		5885.8***		0.781***
		(141.0)		(0.0277)
Individual & Year FE	Yes	Yes	Yes	Yes
Observations	519,730	519,730	519,730	519,730
R^2	0.1914	0.1960	0.4951	0.4962
Panel B: Triple Differences				
	Self-employment Profits		Log Self-employment Profits	
DDD (mean × post × doc)	6343.2***		0.799***	
	(468.4)		(0.112)	
DDD (p90 × post × doc)		6611.3***		0.820***
		(382.9)		(0.0807)
Individual & Year FE	Yes	Yes	Yes	Yes
Observations	519,730	519,730	519,730	519,730
R^2	0.1927	0.1975	0.4961	0.4971

Notes: This table depicts results using the individual self-employed tax return data for the subsample of individuals observed without interruption throughout the whole sample period 2006-2015. Outcome variables are overall self-employment profits and log(self-employment profits + 1). Panel A depicts the results from the interaction term in the double difference regression equation (3.1). Panel B depicts the results for the triple interaction term in regression equation (3.3). The second treatment layer is given by an indicator for a self-employed individual with registered profession of either medical doctor or dentist. Standard errors in parentheses, clustered at the individual level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.3 – Placebo Differences - Balanced Panel

	(1)	(2)	(3)	(4)
Panel A: Vet				
	Self-employment Profits		Log Self-employment Profits	
DDD (mean × post × vet)	-2688.2**		-0.0468	
	(1248.8)		(0.269)	
DDD (p90 × post × vet)		-3825.7***		-0.405*
		(1209.3)		(0.246)
Individual & Year FE	Yes	Yes	Yes	Yes
Observations	519,730	519,730	519,730	519,730
R^2	0.1915	0.1961	0.4951	0.4962
Panel B: Unaffected Professionals				
	Self-employment Profits		Log Self-employment Profits	
DDD (mean × post × unaffected)	141.7		0.0129	
	(763.9)		(0.166)	
DDD (p90 × post × unaffected)		-1281.8**		-0.0820
		(628.7)		(0.123)
Individual & Year FE	Yes	Yes	Yes	Yes
Observations	519,730	519,730	519,730	519,730
R^2	0.1914	0.1961	0.4952	0.4963

Notes: This table depicts results using the individual self-employed tax return data for the subsample of individuals observed without interruption throughout the whole sample period 2006-2015. It presents the results of triple difference regressions as in equation (3.3) for two sets of placebo professional groups: Veterinary physicians (Panel A) and a further group of unaffected professionals (Panel B) consisting of economists, business administrators and journalists. Reported coefficients correspond to the triple interaction between being in an above mean (p90) region, belonging to the specific group of professionals and the observations being in a time period after the introduction of the reform. Outcome variable is individual total self-employment profits in columns (1)-(2) and log(self-employment profits + 1) in columns (3)-(4). Standard errors in parentheses, clustered at the individual level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.4 – Triple Differences Health Workers

	(1)	(2)	(3)	(4)
	Self-employment Profits		Log Self-employment Profits	
DDD (mean × post × health)	-1031.9***		0.139	
	(315.6)		(0.147)	
DDD (p90 × post × health)		-1499.1***		0.164
		(243.7)		(0.108)
Individual & Year FE	Yes	Yes	Yes	Yes
Observations	1,880,549	1,880,549	1,880,549	1,880,549
R^2	0.1192	0.1216	0.2848	0.2852

Notes: This table presents results of a triple difference regression as in equation (3.3). The second treatment layer is given by a broader group of health professionals excluding doctors and dentists but including nurses, medical technicians, psychologists, pharmacists, midwifes, physical therapists, optometrists, and nutritionists. Reported coefficients correspond to the triple interaction between being in an above mean (p90) region, belonging to the specific group of health professionals and the observations being in a time period after the introduction of the reform. Outcome variable is individual total overall self-employment profits in columns (1)-(2) and log(self-employment profits + 1) in columns (3)-(4). Standard errors in parentheses, clustered at the individual level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure C.1 – Tax Declaration Form F102A for Self-Employed

FORMULARIO 102A
RESOLUCIÓN Nº NAC-DGERCGC13-00881

DECLARACIÓN DEL IMPUESTO A LA RENTA PERSONAS NATURALES Y SUCESIONES INDIVISAS NO OBLIGADAS A LLEVAR CONTABILIDAD

No.

100 IDENTIFICACIÓN DE LA DECLARACIÓN

102 AÑO

IMPORTANTE: POSICIONE EL CURSOR SOBRE EL CASILLERO PARA OBTENER AYUDA SOBRE SU LLENADO

104 Nº. DE FORMULARIO QUE SUSTITUYE

105 Nº. DE EMPLEADOS EN RELACIÓN DE DEPENDENCIA

200 IDENTIFICACIÓN DEL SUJETO PASIVO

201 RUC (0 0 1)

202 APELLIDOS Y NOMBRES COMPLETOS / RAZÓN O DENOMINACIÓN SOCIAL DE LA SUCESIÓN INDIVISA

RENTAS GRAVADAS DE TRABAJO Y CAPITAL	AVALÚO		INGRESOS			GASTOS DEDUCIBLES			RENTA IMPONIBLE (INGRESOS - GASTOS DED.)		
ACTIVIDADES EMPRESARIALES CON REGISTRO DE INGRESOS Y EGRESOS			481	+		491	(-)				
INGRESOS SUJETOS A IMPUESTO A LA RENTA ÚNICO			510								
LIBRE EJERCICIO PROFESIONAL			511	+		521	(-)				
OCUPACIÓN LIBERAL (INCLUYE COMISIONISTAS, ARTESANOS, AGENTES, REPRESENTANTES Y DEMÁS TRABAJADORES AUTÓNOMOS)			512	+		522	(-)				
ARRIENDO DE BIENES INMUEBLES	503		513	+		523	(-)				
ARRIENDO DE OTROS ACTIVOS	504		514	+		524	(-)				
RENTAS AGRÍCOLAS	505		515	+		525	(-)				
INGRESO POR REGALÍAS			516	+							
INGRESOS PROVENIENTES DEL EXTERIOR			517	+							
RENDIMIENTOS FINANCIEROS			518	+							
DIVIDENDOS			519	+							
OTRAS RENTAS GRAVADAS			520	+		530	(-)				
SUBTOTAL			529	=		539	=				
RENTA IMPONIBLE ANTES DE INGRESOS POR TRABAJO EN RELACIÓN DE DEPENDENCIA								529-539	549	=	
SUELDOS, SALARIOS, INDEMNIZACIONES Y OTROS INGRESOS LÍQUIDOS DEL TRABAJO EN RELACIÓN DE DEPENDENCIA			541	+		551	(-)		559	+	
SUBTOTAL BASE GRAVADA								**549+559**	**569**	=	

OTRAS DEDUCCIONES Y EXONERACIONES			APLICABLE AL PERÍODO		
GASTOS PERSONALES - EDUCACIÓN			571	(-)	
GASTOS PERSONALES - SALUD			572	(-)	
GASTOS PERSONALES - ALIMENTACIÓN			573	(-)	
GASTOS PERSONALES - VIVIENDA			574	(-)	
GASTOS PERSONALES - VESTIMENTA			575	(-)	
EXONERACIÓN POR TERCERA EDAD			576	(-)	
EXONERACIÓN POR DISCAPACIDAD	560	PORCENTAJE DE DISCAPACIDAD	577	(-)	
50% UTILIDAD ATRIBUIBLE A LA SOCIEDAD CONYUGAL POR LAS RENTAS QUE LE CORRESPONDA	570	IDENTIFICACIÓN DEL CÓNYUGE (C.I. O PASAPORTE)	578	(-)	
SUBTOTAL OTRAS DEDUCCIONES Y EXONERACIONES		**SUMAR DEL 571 AL 578**	**579**	=	

TOTAL GASTOS PERSONALES
SUMAR DEL 571 AL 575
580 (=)

OTRAS RENTAS EXENTAS	VALOR IMPUESTO PAGADO		INGRESOS		
INGRESOS POR LOTERÍAS, RIFAS Y APUESTAS	581		583	+	
HERENCIAS, LEGADOS Y DONACIONES	582		584	+	
PENSIONES JUBILARES			586	+	
OTROS INGRESOS EXENTOS			587	+	
SUBTOTAL OTRAS RENTAS EXENTAS			**589**	=	

RESUMEN IMPOSITIVO			
BASE IMPONIBLE GRAVADA	**569-579**	**832**	=
TOTAL IMPUESTO CAUSADO		**839**	=
(-) ANTICIPO PAGADO		840	(-)
(=) IMPUESTO A LA RENTA CAUSADO MAYOR AL ANTICIPO DETERMINADO	839-840>0	842	=
(=) CRÉDITO TRIBUTARIO GENERADO POR ANTICIPO	839-840<0	843	=
(-) RETENCIONES EN LA FUENTE QUE LE REALIZARON EN EL EJERCICIO FISCAL		845	(-)
(-) RETENCIONES EN LA FUENTE QUE LE REALIZARON EN EL EJERCICIO FISCAL EN RELACIÓN DE DEPENDENCIA		846	(-)
(-) CRÉDITO TRIBUTARIO POR DIVIDENDOS		847	(-)
(-) RETENCIONES POR INGRESOS PROVENIENTES DEL EXTERIOR CON DERECHO A CRÉDITO TRIBUTARIO		848	(-)
(-) ANTICIPO DE IMPUESTO A LA RENTA PAGADO POR ESPECTÁCULOS PÚBLICOS		849	(-)
(-) CRÉDITO TRIBUTARIO DE AÑOS ANTERIORES		850	(-)
(-) CRÉDITO TRIBUTARIO GENERADO POR IMPUESTO A LA SALIDA DE DIVISAS		851	(-)
(-) EXONERACIÓN Y CRÉDITO TRIBUTARIO POR LEYES ESPECIALES		852	(-)
SUBTOTAL IMPUESTO A PAGAR	**842-843-845-846-847-848-849-850-851-852>0**	**855**	=
SUBTOTAL SALDO A FAVOR	**842-843-845-846-847-848-849-850-851-852<0**	**856**	=
(+) IMPUESTO A LA RENTA ÚNICO		857	(+)
(-) CRÉDITO TRIBUTARIO PARA LA LIQUIDACIÓN DEL IMPUESTO A LA RENTA ÚNICO		858	(-)
IMPUESTO A LA RENTA A PAGAR		**859**	
SALDO A FAVOR CONTRIBUYENTE		**869**	
ANTICIPO DETERMINADO PRÓXIMO AÑO	**50% Impuesto a la Renta Causado Menos Retenciones 871+872**	**879**	=
ANTICIPO A PAGAR PRIMERA CUOTA		871	(+)
SEGUNDA CUOTA		872	(+)
PAGO PREVIO (Informativo)		890	

DETALLE DE IMPUTACIÓN AL PAGO (Para declaraciones sustitutivas)

INTERÉS	897	USD	IMPUESTO	898	USD	MULTA	899	USD

VALORES A PAGAR Y FORMA DE PAGO (luego de imputación al pago en declaraciones sustitutivas)				
TOTAL IMPUESTO A PAGAR	859-898	902	+	
INTERÉS POR MORA		903	+	
MULTA		904	+	
TOTAL PAGADO		**999**	=	
MEDIANTE CHEQUE, DÉBITO BANCARIO, EFECTIVO U OTRAS FORMAS DE PAGO		905	USD	
MEDIANTE COMPENSACIONES		906	USD	
MEDIANTE NOTAS DE CRÉDITO		907	USD	

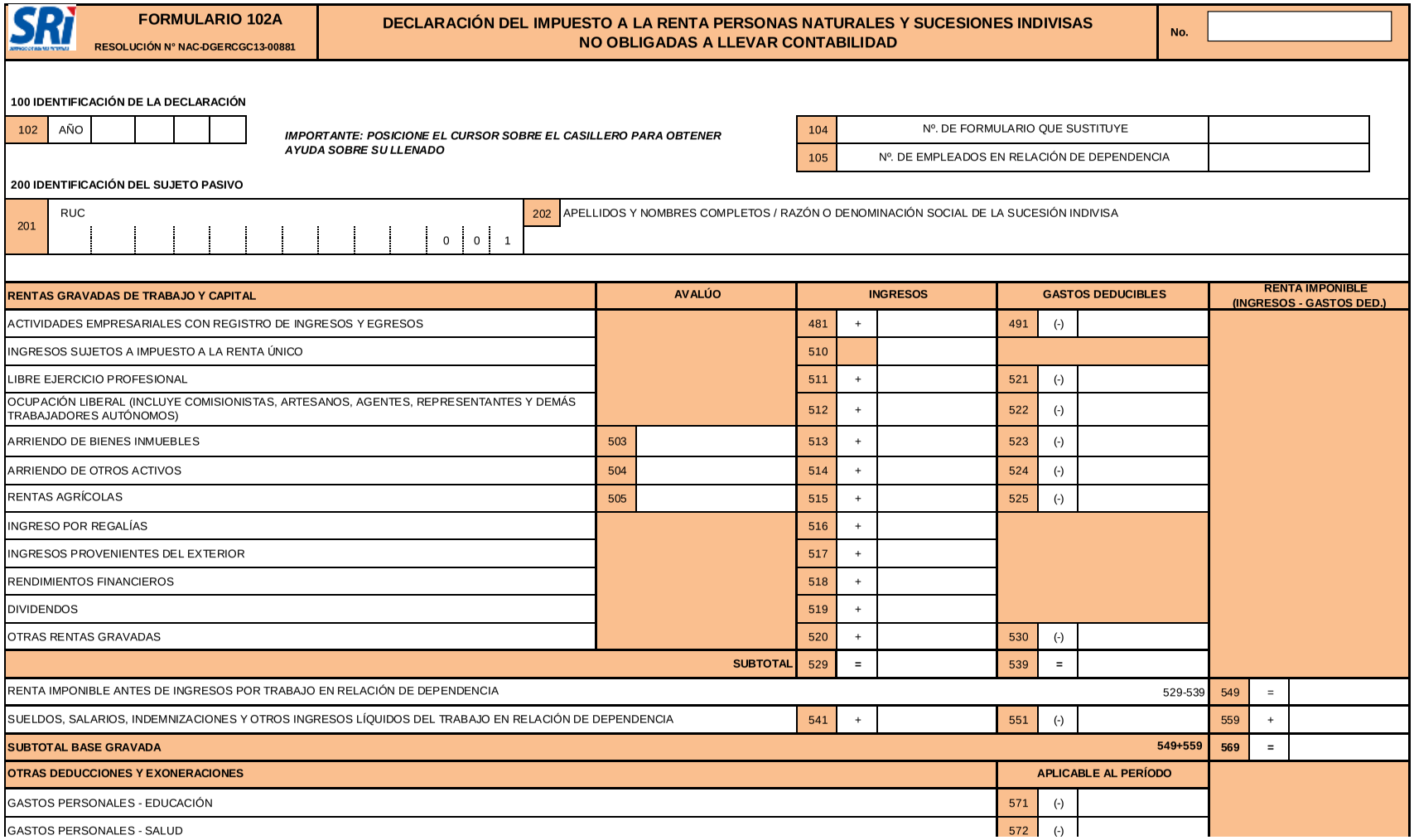
DETALLE DE NOTAS DE CRÉDITO CARTULARES						DETALLE DE NOTAS DE CRÉDITO DESMATERIALIZADAS		DETALLE DE COMPENSACIONES			
908	N/C No	910	N/C No	912	N/C No			916	Resol No.	918	Resol No.
909	USD	911	USD	913	USD	915	USD	917	USD	919	USD

DECLARO QUE LOS DATOS PROPORCIONADOS EN ESTE DOCUMENTO SON EXACTOS Y VERDADEROS, POR LO QUE ASUMO LA RESPONSABILIDAD LEGAL QUE DE ELLA SE DERIVEN (Art. 101 de la L.R.T.I.)

FIRMA SUJETO PASIVO

NOMBRE : | 198 Cédula de Identidad o No. de Pasaporte

Figure C.2 – Tax Declaration Form F102A for Self-Employed



FORMULARIO 102A
RESOLUCIÓN N° NAC-DGERCGC13-00881

DECLARACIÓN DEL IMPUESTO A LA RENTA PERSONAS NATURALES Y SUCESIONES INDIVISAS NO OBLIGADAS A LLEVAR CONTABILIDAD

No.

100 IDENTIFICACIÓN DE LA DECLARACIÓN

102 AÑO

IMPORTANTE: POSICIONE EL CURSOR SOBRE EL CASILLERO PARA OBTENER AYUDA SOBRE SU LLENADO

104	Nº. DE FORMULARIO QUE SUSTITUYE	
105	Nº. DE EMPLEADOS EN RELACIÓN DE DEPENDENCIA	

200 IDENTIFICACIÓN DEL SUJETO PASIVO

201 RUC ... 0 0 1

202 APELLIDOS Y NOMBRES COMPLETOS / RAZÓN O DENOMINACIÓN SOCIAL DE LA SUCESIÓN INDIVISA

RENTAS GRAVADAS DE TRABAJO Y CAPITAL	AVALÚO		INGRESOS			GASTOS DEDUCIBLES			RENTA IMPONIBLE (INGRESOS - GASTOS DED.)		
ACTIVIDADES EMPRESARIALES CON REGISTRO DE INGRESOS Y EGRESOS			481	+		491	(-)				
INGRESOS SUJETOS A IMPUESTO A LA RENTA ÚNICO			510								
LIBRE EJERCICIO PROFESIONAL			511	+		521	(-)				
OCUPACIÓN LIBERAL (INCLUYE COMISIONISTAS, ARTESANOS, AGENTES, REPRESENTANTES Y DEMÁS TRABAJADORES AUTÓNOMOS)			512	+		522	(-)				
ARRIENDO DE BIENES INMUEBLES	503		513	+		523	(-)				
ARRIENDO DE OTROS ACTIVOS	504		514	+		524	(-)				
RENTAS AGRÍCOLAS	505		515	+		525	(-)				
INGRESO POR REGALÍAS			516	+							
INGRESOS PROVENIENTES DEL EXTERIOR			517	+							
RENDIMIENTOS FINANCIEROS			518	+							
DIVIDENDOS			519	+							
OTRAS RENTAS GRAVADAS			520	+		530	(-)				
SUBTOTAL			529	=		539	=				
RENTA IMPONIBLE ANTES DE INGRESOS POR TRABAJO EN RELACIÓN DE DEPENDENCIA					529-539				549	=	
SUELDOS, SALARIOS, INDEMNIZACIONES Y OTROS INGRESOS LÍQUIDOS DEL TRABAJO EN RELACIÓN DE DEPENDENCIA			541	+		551	(-)		559	+	
SUBTOTAL BASE GRAVADA					**549+559**				**569**	=	

OTRAS DEDUCCIONES Y EXONERACIONES	APLICABLE AL PERÍODO		
GASTOS PERSONALES - EDUCACIÓN	571	(-)	
GASTOS PERSONALES - SALUD	572	(-)	

This figure depicts an excerpt of the tax declaration form F102A for self-employed individuals not obliged to file an extensive tax declaration. The excerpt focuses on the boxes relevant for documenting the various sources of self-employment income.

Bibliography

AGHION, P., U. AKCIGIT, M. LEQUIEN, AND S. STANTCHEVA (2017): "Tax Simplicity and Heterogeneous Learning," NBER Working Paper No. 24049.

AJZEN, I. (2002): "Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior," *Journal of Applied Social Psychology*, 32, 665–683.

ALLINGHAM, M. G. AND A. SANDMO (1972): "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1, 323–338.

ALSTADSAETER, A., N. JOHANNESEN, AND G. ZUCMAN (2017): "Tax Evasion and Inequality," NBER Working Paper No. 23772.

AMERIKS, J., A. CAPLIN, J. LEAHY, AND T. TYLER (2007): "Measuring self-control problems," *The American Economic Review*, 97, 966–972.

ASATRYAN, Z. AND A. PEICHL (2017): "Responses of Firms to Tax, Administrative and Accounting Rules: Evidence from Armenia," CESifo Working Paper No. 6754.

ASHRAF, N., D. KARLAN, AND W. YIN (2006): "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," *Quarterly Journal of Economics*, 121, 635–672.

BACHAS, P. AND M. SOTO (2017): "Not(ch) Your Average Tax System: Corporate Taxation Under Weak Enforcement," Working paper, UC Berkeley.

BALSVIK, R. (2011): "Is labor mobility a channel for spillovers from multinationals? Evidence from Norwegian manufacturing," *The Review of Economics and Statistics*, 93, 285–297.

BANERJEE, A. AND E. DUFLO (2006): "The Economic Lives of the Poor," *Journal of Economic Perspectives*, 21, 141–167.

BARBER, B. M. AND T. ODEAN (2001): "Boys will be boys: Gender, Overconfidence, and Common Stock Investment," *Quarterly Journal of Economics*, 261–292.

BARBERIS, N. AND R. THALER (2003): "A survey of Behavioral Finance," *Handbook of the Economics of Finance*, 1, 1053–1128.

BASTANI, S. AND H. SELIN (2014): "Bunching and Non-Bunching at Kink Points of the Swedish Tax Schedule," *Journal of Public Economics*, 109, 36–49.

BAYER, P. J., B. D. BERNHEIM, AND J. K. SCHOLZ (2009): "The Effects of Financial Education in the Workplace: Evidence from a Survey of Employers," *Economic Inquiry*, 47, 605–624.

BECCHETTI, L., N. SOLFERINO, M. E. TESSITORE, ET AL. (2015): "Some Insights on Procrastination: a Curse or a Productive Art?" *Review of Behavioral Economics*, 2, 331–351.

BENDER, S., N. BLOOM, D. CARD, J. VAN REENEN, AND S. WOLTER (2016): "Management Practices, Workforce Selection and Productivity," NBER Working Paper No. 22101.

BENZARTI, Y. (2017): "How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs," NBER Working Paper No. 23903.

BERG, G. AND B. ZIA (2013): "Harnessing Emotional Connections to Improve Financial Decisions: Evaluating the Impact of Financial Education in Mainstream Media," *World Bank Policy Research Working Paper*.

BESLEY, T. AND T. PERSSON (2013): "Taxation and Development," *Handbook of Public Economics*, 5, 51–110.

BEST, M. (2014): "The Role of Firms in Workers' Earnings Responses to Taxes: Evidence From Pakistan," Working paper, London School of Economics.

BEST, M., A. BROCKMEYER, H. J. KLEVEN, J. SPINNEWIJN, AND M. WASEEM (2015): "Production vs Revenue Efficiency With Limited Tax Capacity: Theory and Evidence From Pakistan," *Journal of Political Economy*, 123, 1311–1355.

BHARGAVA, S. AND D. MANOLI (2015): "Psychological Frictions and Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment," *American Economic Review*, 105, 3489–3529.

BISIN, A. AND K. HYNDMAN (2014): "Present-bias, Procrastination and Deadlines in a Field Experiment," NBER Working Paper No. 19874.

BLOMQUIST, S. AND W. NEWEY (2017): "The Bunching Estimator Cannot Identify the Taxable Income Elasticity," Working Paper 24136, National Bureau of Economic Research.

BOHNE, A. AND J. S. NIMCZIK (2018): "Information Frictions and Learning Dynamics: Evidence from Tax Avoidance in Ecuador," IZA Discussion Paper No. 11536.

BROCKMEYER, A., M. HERNANDEZ, S. KETTLE, AND S. D. SMITH (2018): "Casting a Wider Tax Net: Experimental Evidence from Costa Rica," Working paper, The World Bank.

BROWN, J. R. AND A. PREVITERO (2014): "Procrastination, Present-biased Preferences, and Financial Behaviors," Working paper, University of Illinois at Urbana-Champaign.

BRUNE, L., X. GINÉ, J. GOLDBERG, AND D. YANG (2016): "Facilitating Savings for Agriculture: Field Experimental Evidence from Malawi," *Economic Development and Cultural Change*, 64, 187–220.

CALIENDO, F. AND K. X. HUANG (2007): "Overconfidence in Financial Markets and Consumption over the Life Cycle," *FRB of Philadelphia Working Paper No. 07-3*.

CANELAS, C. (2014): "Minimum Wage and Informality in Ecuador," WIDER Working Paper 2014/006, UNU-WIDER.

——— (2015): "Poverty and Informality in Ecuador," WIDER Working Paper 2015/112, UNU-WIDER.

CARRILLO, P., M. S. EMRAN, AND A. RIVADENEIRA (2012): "Do Cheaters Bunch Together? Profit Taxes, Withholding Rates and Tax Evasion," Working paper, George Washington University.

CARRILLO, P., D. POMERANZ, AND M. SINGHAL (2017): "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement," *American Economic Journal: Applied Economics*, 9, 144–164.

CARVALHO, L. S., S. PRINA, AND J. SYDNOR (2016): "The effect of saving on risk attitudes and intertemporal choices," *Journal of Development Economics*, 120, 41–52.

CHATTERJEE, S., M. FINKE, AND N. HARNESS (2011): "The Impact of Self-efficacy on Wealth Accumulation and Portfolio Choice," *Applied Economics Letters*, 18, 627–631.

CHETTY, R., J. FRIEDMAN, T. OLSEN, AND L. PISTAFERRI (2011): "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *Quarterly Journal of Economics*, 126, 749–804.

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, 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, 1–31.

COBB-CLARK, D. A., S. C. KASSENBOEHMER, AND M. SINNING (2013): "Locus of Control and Savings," *Ruhr Economic Paper No. 455*.

DE PAULA, A. AND J. A. SCHEINKMAN (2010): "Value-Added Taxes, Chain Effects, and Informality," *American Economic Journal: Macroeconomics*, 2, 195–221.

DOERR, U., O. M. TOMAN, AND U. SCHMIDT (2011): "Overconfidence and Risk management of Ethiopian Farmers," Working paper, Kiel Institute for the World Economy.

DOERRENBERG, P., A. PEICHL, AND S. SIEGLOCH (2017): "The Elasticity of Taxable Income in the Presence of Deduction Possibilities," *Journal of Public Economics*, 151, 41–55.

DUFLO, E., M. KREMER, AND J. ROBINSON (2009): "Nudging Farmers to use Fertilizer: Theory and Experimental Evidence from Kenya," NBER Working Paper No. 15131.

DUPAS, P., A. KEATS, AND J. ROBINSON (forthcoming): "The Effect of Savings Accounts on Interpersonal Financial Relationships: Evidence from a Field Experiment in Rural Kenya," *Economic Journal*.

DUPAS, P. AND J. ROBINSON (2013): "Why Don't the Poor Save More? Evidence from Health Savings Experiments," *American Economic Review*, 103, 1138–1171.

DWENGER, N., H. KLEVEN, I. RASUL, AND J. RINCKE (2016): "Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany," *American Economic Journal: Economic Policy*, 8, 1–31.

FABBRI, M. (2015): "Shaping Tax Norms through Lotteries," *International Review of Law and Economics*, 44, 8–15.

FOUARGE, D., N. SALAMANCA, R. MONTIZAAN, ET AL. (2013): "Locus of Control and Investment in Risky Assets," Working paper, Maastricht University, Research Centre for Education and the Labour Market (ROA).

GARRARD, P. AND G. ROBINSON (2015): *The Intoxication of Power: Interdisciplinary Insights*, Palgrave Macmillan.

GELBER, A. M., D. JONES, AND D. W. SACKS (2017): "Estimating Earnings Adjustment Frictions: Evidence from the Social Security Earnings Test," NBER Working Paper No. 19491.

GERARD, F. AND G. GONZAGA (2016): "Informal Labor and the Efficiency Cost of Social Programs: Evidence from the Brazilian Unemployment Insurance Program," NBER Working Paper No. 22608.

GERVAIS, S. AND T. ODEAN (2001): "Learning to be overconfident," *Review of Financial Studies*, 14, 1–27.

GORDON, R. AND W. LI (2009): "Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation," *Journal of Public Economics*, 93, 855–866.

GRUBB, M. D. (2009): "Selling to Overconfident Consumers," *The American Economic Review*, 99, 1770–1807.

——— (2015): "Overconfident Consumers in the Marketplace," *The Journal of Economic Perspectives*, 29, 9–35.

HAMILTON, S. (2018): "Optimal Deductibility: Theory, and Evidence from a Bunching Decomposition," Working paper, University of Michigan.

HOLT, C. AND S. LAURY (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

ILO (2014): "Recent Experiences of Formalization in Latin America and the Caribbean," Report, International Labor Organization, Regional Office for Latin America and the Caribbean.

JACOBSON, L. S., R. J. LALONDE, AND D. G. SULLIVAN (1993): "Earnings Losses of Displaced Workers," *The American Economic Review*, 83, 685–709.

JENSEN, A. (2016): "Employment Structure and the Rise of the Modern Tax System," Working paper, London School of Economics.

JOHN, A. (2017): "When Commitment Fails - Evidence from a Field Experiment," Working paper, CREST.

KARLAN, D. AND L. L. LINDEN (2017): "Loose Knots: Strong versus Weak Commitments to Save for Education in Uganda," Working paper, Yale University.

KARLAN, D., M. MCCONNELL, S. MULLAINATHAN, AND J. ZINMAN (2016): "Getting to the top of mind: How Reminders Increase Saving," *Management Science*, 62, 3393–3411.

KARLAN, D., A. L. RATAN, AND J. ZINMAN (2014): "Savings by and for the Poor: A Research Review and Agenda," *Review of Income and Wealth*, 60, 36–78.

KAST, F., S. MEIER, AND D. POMERANZ (2018): "Saving more in Groups: Field experimental evidence from Chile," *Journals of Development Economics*, 133, 275–294.

KEEN, M. AND J. SLEMROD (2017): "Optimal Tax Administration," *Journal of Public Economics*, 152, 133–142.

KIRCHLER, E. AND B. MACIEJOVSKY (2002): "Simultaneous over-and underconfidence: Evidence from Experimental Asset Markets," *Journal of Risk and Uncertainty*, 25, 65–85.

KLEVEN, H. J. (2016): "Bunching," *Annual Review of Economics*, 8, 435–464.

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, 651–692.

KLEVEN, H. J., C. T. KREINER, AND E. SAEZ (2016): "Why can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries," *Economica*, 83, 219–246.

KLEVEN, H. J. AND M. WASEEM (2013): "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *The Quarterly Journal of Economics*, 128, 669–723.

KOPCZUK, W. AND C. POP-ELECHES (2007): "Electronic Filing, Tax Preparers, and Participation in the Earned Income Tax Credit," *Journal of Public Economics*, 91(7-8), 1351–1367.

KOPCZUK, W. AND J. SLEMROD (2006): "Putting Firms into Optimal Tax Theory," *American Economic Review*, 96, 130–134.

KUMLER, T., E. VERHOOGEN, AND J. FRIAS (2015): "Enlisting Employees in Improving Payroll-Tax Compliance: Evidence from Mexico," NBER Working Paper No. 19385.

LI, Y., M. SHAN, AND M. Z. LI (2016): "Advance Selling Decisions with Overconfident Consumers," *Journal of Industrial and Management Optimization*, 12, 891–905.

LINARDI, S. AND T. TANAKA (2013): "Competition as a savings incentive: A field experiment at a homeless shelter," *Journal of Economic Behavior & Organization*, 95, 240–251.

MAHON, J. AND E. ZWICK (2017): "The Costs of Corporate Tax Complexity," Working paper, Harvard University.

MALMENDIER, U. AND G. TATE (2005): "CEO Overconfidence and Corporate Investment," *The Journal of Finance*, 60, 2661–2700.

MARCHESE, C. (2009): "Rewarding the Consumers for Curbing the Evasion of Commodity Taxes?" *FinanzArchiv: Public Finance Analysis*, 65, 383–402.

MATIKKA, T. (forthcoming): "The Elasticity of Taxable Income: Evidence from Changes in Municipal Income Tax Rates in Finland," *Scandinavian Journal of Economics*.

MION, G. AND L. D. OPROMOLLA (2014): "Managers' mobility, trade performance, and wage," *Journal of International Economics*, 94, 85–101.

MODICA, E., S. LAUDAGE, AND M. HARDING (2018): "Domestic Revenue Mobilisation," *OECD Taxation Working Papers*, No. 36.

NARITOMI, J. (2016): "Consumers as Tax Auditors," Working paper, London School of Economics.

PAETZOLD, J. AND H. WINNER (2016): "Taking the High Road? Compliance with Commuter Tax Allowances and the Role of Evasion Spillovers," *Journal of Public Economics*, 143, 1–14.

PARROTTA, P. AND D. POZZOLI (2012): "The effect of Learning by Hiring on Productivity," *RAND Journal of Economics*, 43, 167–185.

PERRY, G., W. MALONEY, O. ARIAS, P. FAJNYZLBER, A. MASON, AND J. SAAVEDRA (2007): *Informality: Exit and Exclusion*, World Bank.

PERRY, V. G. AND M. D. MORRIS (2005): "Who is in control? The Role of Self-perception, Knowledge, and Income in Explaining Consumer Financial Behavior," *Journal of Consumer Affairs*, 39, 299–313.

POMERANZ, D. (2015): "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax," *American Economic Review*, 105, 2539–69.

POOLE, J. P. (2013): "Knowledge Transfers from Multinational to Domestic Firms: Evidence from Worker Mobility," *The Review of Economics and Statistics*, 95, 393–406.

ROCHA, R., G. ULYSSEA, AND L. RACHTER (2018): "Do lower Taxes reduce Informality? Evidence from Brazil," *Journal of Development Economics*, 134, 28–49.

ROTTER, J. B. (1966): "Generalized Expectancies for Internal versus External Control of Reinforcement." *Psychological Monographs: General and Applied*, 80, 1.

SAEZ, E. (2010): "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 180–212.

SAEZ, E., B. SCHOEFER, AND D. SEIM (2017): "Payroll Taxes, Firm Behavior, and Rent Sharing: Evidence from a Young Workers' Tax Cut in Sweden," NBER Working Paper No. 23976.

SEIBOLD, A. (2017): "Reference Dependence in Retirement Behavior: Evidence from German Pension Discontinuities," Working paper, London School of Economics.

SERAFINELLI, M. (forthcoming): "'Good' Firms, Worker Flows and Local Productivity," *Journal of Labor Economics*.

SHIPTON, P. ET AL. (1990): "How Gambians Save and what their Strategies Imply for International Aid," Working paper, The World Bank.

SLEMROD, J. (2007): "Cheating Ourselves: The Economies of Tax Evasion," *Journal of Economic Perspectives*, 21, 25–48.

SONG, J., P. ALMEIDA, AND G. WU (2003): "Learning–by–Hiring: When Is Mobility More Likely to Facilitate Interfirm Knowledge Transfer?" *Management Science*, 49, 251–365.

STOYANOV, A. AND N. ZUBANOV (2012): "Productivity Spillovers Across Firms through Worker Mobility," *American Economic Journal: Applied Economics*, 4, 168–198.

TANG, N. AND A. BAKER (2016): "Self-esteem, financial knowledge and financial behavior," *Journal of Economic Psychology*, 54, 164–176.

TAUBINSKY, D. AND A. REES-JONES (forthcoming): "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment," *Review of Economic Studies*.

THALER, R. H. AND S. BENARTZI (2004): "Save more tomorrow: Using Behavioral Economics to Increase Employee Saving," *Journal of Political Economy*, 112, S164–S187.

VILLACRESES, N. (2014): "Gastos Personales: cual es su objetivo?" Tech. rep., Notas de Reflexion, Centro de Estudios Fiscales.

WASEEM, M. (2018): "Information, Asymmetric Incentives, Or Withholding? Understanding the Self-Enforcement of Value-Added Tax," Working paper, University of Manchester.

WORLD BANK (1998): "Financial Literacy Questionnaire," Tech. rep., The World Bank.

Curriculum Vitae, Albrecht Bohne

2012 – 2018	PhD Studies in Economics, Center for Doctoral Studies in Economics, University of Mannheim.
2011 – 2012	MSc in Economics, London School of Economics and Political Sciences
2008 – 2011	BSc in Economics, Ludwig-Maximilians Universität München (LMU Munich)

Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Dissertation selbstständig angefertigt und mich keiner anderen als der in ihr angegebenen Hilfsmittel bedient zu haben. Insbesondere sind sämtliche Zitate aus anderen Quellen als solche gekennzeichnet und mit Quellenangaben versehen.

Mannheim, 08.08.2018: _____

Albrecht Bohne