WORKING paper

Tax Simplicity and Heterogeneous Learning

Philippe Aghion[1], Ufuk Akcigit[2],
Matthieu Lequien[3] & Stefanie Stantcheva[4]

ABSTRACT

We study how strongly individuals respond to tax simplicity and how they learn about the complexities of the tax system. We use new French tax returns data on the self-employed from 1994 to 2012. France has three fiscal regimes for the self-employed, which differ in their monetary tax incentives and in their tax simplicity. These regimes are subject to eligibility thresholds: we find large excess masses (bunching) right below the latter. The regimes impact different agents heterogeneously and have changed extensively over time. We estimate a large value for tax simplicity of up to 650 euros per year per individual. Tax complexity has sizable costs: agents are not immediately able to understand what the right regime choice is, leave significant money on the table, and learn over time. These costs are "regressive", impacting more the uneducated, low income, and low skill agents.

[1] Collège de France, PSE and LSE, paghion@fas.harvard.edu
[2] Department of Economics, University of Chicago, uakcigit@uchicago.edu
[3] Banque de France, matthieu.lequien@banque-france.fr
[4] Department of Economics, Harvard University, sstantcheva@fas.harvard.edu

NON-TECHNICAL SUMMARY

We study how strongly individuals respond to tax simplicity and how they learn about the complexities of the tax system. We focus on the self-employed, who can more easily adjust to tax incentives and whose responses directly stem from their own understanding of the tax system. We use new French tax returns data from 1994 to 2012. France serves as a good quasi-laboratory: it has three fiscal regimes[5] for the self-employed, which differ in their monetary tax incentives and in their tax simplicity. Our methodology builds on the rich literature on the effect of taxes on income. In particular Saez (2010) found evidence of bunching around the first kink of the US Earned Income Tax Credit, but concentrated solely among the self-employed. Chetty et al (2011) show bunching at kinks of the Danish tax system and Kleven and Waseem (2013) do the same on Pakistanese tax data.

Bunching at the eligibility thresholds
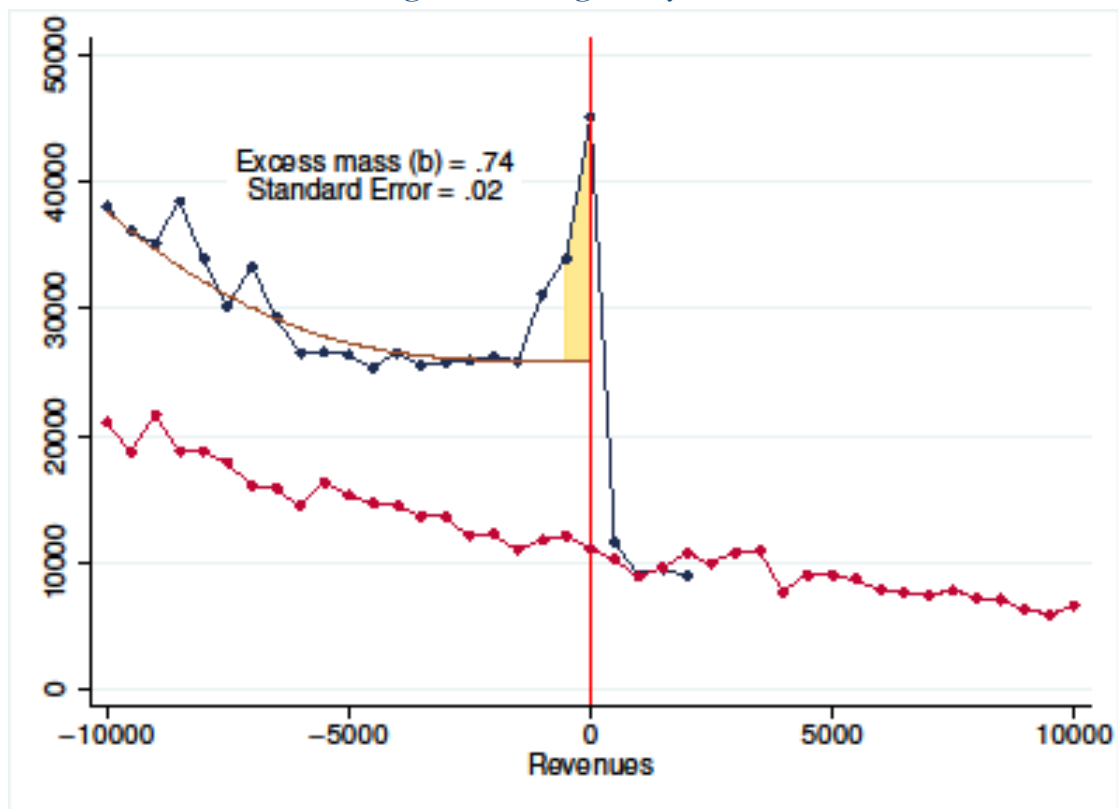
Notes: The figure represents the frequency of revenues, by bins of revenues centered around the eligibility threshold (the red vertical line). We pool data for 1999-2012 and for all agents either *micro-* or *auto-entrepreneurs*. The estimated excess mass is in yellow. There is significant bunching, equal to 74% of the average counterfactual frequency within 1,500 euros of the notch. The thick red curve serves as a placebo test: it shows the frequency distribution for the Industrial & Commercial Retail activities,

---

[5] The *régime réel* treats an individual's net business income (revenues minus costs) as taxable income, it requires a detailed tax accounting. The simplified regime (*micro-entreprise* and *auto-entrepreneur sans option libératoire*) cuts down on tax hassle and allows agents to claim a flat-rate rebate as a fraction of revenues instead of reporting their true business costs, The *régime auto-entrepreneur avec option libératoire* further increases tax simplicity by replacing all income taxes and social insurance contributions by a unique - and relatively low – flat rate payment proportional to gross revenues.

centered around the eligibility threshold for the I&C Services and Non Commercial activities (the actual threshold for the I&C Retail activities is higher).

Two key features are that, first, these regimes are subject to eligibility thresholds; we find large excess masses (bunching) right below the latter (Figure). Second, the regimes impact different agents heterogeneously and have changed extensively over time. Taken together, these two key elements give us measures of tax responses (the bunching) as well as the variation needed to jointly estimate a value of tax simplicity and taxable income elasticities. They also give us an opportunity to study how individuals learn about and respond over time to changing policy parameters. We estimate a large value for tax simplicity of up to 650 euros per year per individual depending on the regime and activity. We also find sizable costs of tax complexity; agents are not immediately able to understand what the right regime choice is, leave significant money on the table, and learn over time. The cost of complexity is "regressive" in that it affects mostly the uneducated, low income, and low skill agents. Agents who can be viewed as more informed and knowledgeable (e.g., the more educated or high-skilled) are more likely to make the correct regime choice and to learn faster.

Simplicité de la fiscalité et apprentissage hétérogène

RÉSUMÉ

Nous étudions comment les individus réagissent à la simplicité de la fiscalité et comment ils apprennent les complexités du régime fiscal. Nous utilisons de nouvelles données fiscales françaises sur les indépendants de 1994 à 2012. La France dispose de trois régimes fiscaux pour les indépendants, qui diffèrent par leurs incitations monétaires et leur simplicité administrative. Ces régimes sont soumis à des seuils d'éligibilité: nous trouvons de grandes masses en excès (*bunching*) juste en dessous de ces dernières. Les régimes ont un impact hétérogène selon les agents et ont connu de nombreux changements au fil du temps. Nous estimons une grande valeur pour la simplicité fiscale, pouvant atteindre 650 euros par an et par personne. La complexité fiscale entraîne des coûts importants: les individus ne comprennent pas immédiatement quel est le bon choix de régime, ils laissent beaucoup d'argent sur la table et apprennent au fil du temps. Ces coûts sont "régressifs" puisqu'ils touchent davantage les individus peu instruits, à faible revenu et peu qualifiés.

1 Introduction

"Simplicity is the ultimate sophistication," wrote Leonardo da Vinci. Many policy makers would probably agree with his statement. Designing a policy that fulfills its stated goals, provides clear and correct incentives, minimizes administrative hassle for individuals, and at the same time remains sufficiently simple for people to understand is an enormous challenge. Tax policy is a case in point: The best tax incentives may turn out to be ineffective if people do not understand them. Even worse, complexity may make the system more regressive if it is mostly the least sophisticated agents or those who cannot afford professional tax advice who cannot understand it and benefit from it. Many tax and transfer policies are targeted towards the bottom of the income distribution, where simplicity may be even more important, and where complexity may prevent the very same people targeted by these policies from taking advantage of them.

In this paper, we attempt to answer the following questions: Do people respond exclusively to monetary tax incentives or does tax simplicity come into consideration as well? We define tax simplicity as the combination of conceptual simplicity and practical simplicity: a system is simple if it is both easy to understand and logistically easy to handle. How much do people value tax simplicity? Do they understand and adapt rapidly to changes in the fiscal landscape or is there a costly process involved in learning about a complex tax system? Are certain agents quicker to learn and to understand the complexities of the tax code?

To answer these questions, we study the self-employed. This is for two related reasons: First, they are typically shown to be much less constrained than wage earners and can more easily adjust their incomes to tax incentives (Saez, 2010; Kleven and Waseem, 2013). This is important if we want to measure how people respond to simpler or more complex tax policies. Second, since the self-employed are their own decision makers, there is a more direct map between their own understanding of the tax system and their response to it. This link is weakened for wage earners, since it may be their company determining their pay structure and responses to taxes, based on its own (presumably, better) knowledge of the tax system. In addition to these key advantages, the self-employed are a very interesting group to study *per se*. They have become more numerous and important in recent years, through the rise of platforms such as Uber, Air BnB, or Task Rabbit, and the outsourcing of jobs previously done in-house. In recent work, Katz and Krueger (2016) and Katz and Krueger (2017) cast light on the rise of alternative work arrangements – those differing from conventional self-employment and regular employment – and on the ensuing fragmentation of the labor market.

Our study focuses on France, which serves as a particularly well-suited quasi-laboratory for studying the effects of tax simplicity and complexity. It has a very unique variety of fiscal "regimes"– or modes of taxation of self-employment– which differ not only in their monetary incentives, but also in their degree of tax simplicity. These fiscal regimes have changed significantly over time, offering the opportunity to study learning and dynamic adjustments. They also impact different groups of agents heterogeneously, thus providing valuable policy variation that helps our estimation.

Our first main contribution is to introduce and use new individual tax returns data from the French internal revenue service over the period 1994-2012. The tax returns data is combined with additional administrative and large-scale survey data, to yield information on employment, demographics, education and government benefits received. This highly valuable combination of administrative tax data and census-style survey data allows us to study the characteristics of agents who respond differently to tax incentives.

In Section 2, we start by describing the landscape of French policies related to self-employment. There are three regimes under which the self-employed may choose to operate, which differ along two main dimensions: monetary tax incentives and tax simplicity. In brief, the "standard regime" treats an individual's net business income (revenues minus costs) as taxable income, which is advantageous for businesses with employees, significant investments, or high operating costs. It does, however, come with the most involved and costly tax accounting requirements, which also limit the scope for misreporting. The "simplified regime" cuts down on tax hassle and allows agents to claim a flat-rate rebate as a fraction of revenues instead of reporting their true business costs, which can be very advantageous for agents with low operating costs. The "super simplified" regime further increases tax simplicity by replacing all income taxes and social insurance contributions by a unique – and relatively low– flat rate payment proportional to gross revenues. The simplified and super simplified regimes require that revenues are below an eligibility threshold. This threshold depends on the type of business activity, and has changed over time. Thus, broadly speaking, the simplified and super simplified regimes are well suited to agents with small and slow-growing activities, with relatively low operating costs and investments, and with strong preferences for tax simplicity.

In Section 3, we provide key new summary statistics on the self-employed for the period 1994-2012.[1] Section 4 formally models the three self-employed regimes, their financial (net-of-tax) payoffs, and the costs imposed by tax requirements, which we call "tax hassle costs." The eligibility thresholds create a special type of discontinuity, not only in monetary payoffs, but also in tax simplicity. We express this discontinuity – or "notch"– in monetary terms as a function of underlying parameters, such as an agent's tax bracket, activity type, tax hassle costs and operating costs. We find very significant behavioral responses (in terms of regime choice and income) to the notches created by the eligibility thresholds. We also highlight heterogeneity in responses: Agents who have other sizable sources of income, such as salaried income or pension income, exhibit much stronger bunching. The same holds true for agents who stand to gain more from fiscal optimization, namely those in higher tax brackets. Importantly, only agents with a at least a high school degree respond to the eligibility thresholds; those without one do not.

Our second main contribution is to use these regime notches and the ensuing excess masses to

---

[1] The self-employed are on average older than wage earners, more likely to be retired, more educated, more likely to be in high skill occupations, and have higher labor, capital, and total income. They are less likely to receive unemployment or social insurance benefits. The fraction of agents with self-employed income remained stable at around 5% of all tax filers aged 18-65 until 2009 and has risen since then. The fraction of agents who earn only self-employed income remained at 4% until 2009 and has increased sharply since then.

structurally estimate a value of tax simplicity. While notches are studied in a rich empirical literature on the effects of taxation on taxable income using "bunching" methods– as surveyed in Kleven (2016) and with key papers such as Saez (2010), Chetty et al. (2011), and Kleven and Waseem (2013)– our setting is quite special. In our case, it is not only the average tax that changes at the notch, but also the tax simplicity. Our analysis also benefits from some peculiar features in the French setting, where agents can be in very different tax brackets even conditional on the same self-employed revenue. Thus, there exists significant heterogeneity in the monetary incentives – and in the effective payoff from bunching across agents– around the eligibility threshold. The parameters of the tax system also have changed a lot over time, providing further heterogeneity and variation for our estimation. In Section 5, we put this heterogeneity and variation to use in order to jointly estimate the value of tax simplicity and the income elasticities based on the bunching at the eligibility thresholds. We estimate preferences for tax simplicity to be quite large and important, ranging from 160 euros to 650 euros depending on the regime and activity. Agents thus seem to strongly value tax simplicity.

In Section 6, we document the costs of tax complexity that manifest themselves in agents' suboptimal choices and sluggish learning. This is our third main contribution. First, we quantify the cost of making the wrong regime choice and show that it is regressive: it is the lower skilled, non-educated, and lower income agents who make the most mistakes and leave the most money on the table. Second, we document more patterns consistent with learning over time and slow adjustment due to the cost of understanding the (complexity of the) tax system: The excess masses and estimated behavioral elasticities grow over time. In addition, when the thresholds change, there often remains some bunching mass at the old, no longer relevant thresholds. Responses to the expansion of an existing and familiar regime are much faster than those to the introduction of a completely new regime.

Finally, in Section 7, we ask how the observed sharp excess masses and their movements over time actually come about, and offer two pieces of evidence in favor of misreporting of self-employed income. First, there is significant round-number bunching in the simplified regimes, but not in the standard regime, which is subject to more rigorous monitoring by certified accounting centers. We also find that those who live in households with two self-employed earners in a simplified regime exhibit peculiar patterns: the distribution of their individual self-employed revenues has stronger excess mass than that of agents in households with a single self-employed earner; the distribution of the sum of their self-employed revenues features sharp bunching at twice the eligibility threshold, and there appears to be income shifting from the higher to the lower earner as the former approaches the tolerance threshold.

Our paper is related to several studies on the effects of taxation on entrepreneurship and self-employment. Cullen and Gordon (2007) use U.S. tax returns data to show that different components of the tax system, such as the progressivity and the marginal tax rates, have had distinct and significant impacts on entrepreneurial risk-taking (see also Cullen and Gordon (2006)).[2] We contribute to the

---

[2]Gentry and Hubbard (2000) find that a progressive tax system discourages entry into entrepreneurship. Using the Panel Study of Income Dynamics, Bruce (2000) finds that reducing marginal tax rates on self-employed income reduces the probability of entry into self-employment, while reducing the average tax rate slightly increases entry.

literature on taxable income elasticities (Gruber and Saez, 2002; Saez, Slemrod, and Giertz, 2012), but focus on the self-employed.

Our analysis of how members of the same household jointly optimize (and misreport) their self-employed earnings echoes the analysis of the joint income decisions among wage earners in Eissa and Hoynes (2004), Eissa and Hoynes (2006), and Gelber (2014).

Our work also relates to how tax payers respond to costly information with inattention as in Hoopes, Slemrod, and Reck (2017) or with behavorial biases as in Lockwood and Taubinsky (2016) and Lockwood (2016).

More generally, our work also speaks to the literature on the determinants of entrepreneurship (see, among others, Hamilton (2000), Schoar (2010), Adelino, Schoar and Severino (2015), and Schmalz, Sraer, and Thesmar (2016)), but we focus specifically on the role of fiscal incentives, taxation, and administrative simplicity. Most closely related are papers on entrepreneurship in France, using other sources of administrative data. Lelarge, Sraer, and Thesmar (2008) look at the effects of credit constraints on entrepreneurship using variation from a French loan guarantee program. Hombert, Schoar, Sraer, and Thesmar (2017) show that unemployment insurance can stimulate self-employed activity in France.

A series of recent studies makes use of the new French administrative data. Fack and Landais (2010) and Fack and Landais (2016) study charitable contributions. French tax data also is used in two important contemporaneous papers that study income and wealth distributions in France, by Garbinti, Goupille-Lebret, and Piketty (2017) and Garbinti, Goupille-Lebret, and Piketty (2016).

A copious literature applies the bunching methodology to a wide range of topics such as intertemporal allocation in response to mortgage contracts changes (Best, Cloyne, Ilzetzki, and Kleven, 2015), transaction taxes in housing markets (Best and Kleven, 2016), corporate taxation (Best, Brockmeyer, Kleven, Spinnewijn, and Waseem, 2015), responses to the EITC (Chetty, Friedman, and Saez, 2013), the social insurance earnings test (Gelber, Jones, and Sacks, 2017; Gelber, Jones, Sacks, and Song, 2017), and fuel efficiency requirements (Slemrod and Salleel, 2012).

Finally, our work is related to the many empirical studies of misreporting in response to taxation, especially recent examples of which are Carillo, Pomeranz, and Singhal (2017), Feldman and Slemrod (2007), Pomeranz (2015), and Gordon and Slemrod (2000).

2   The Landscape of Self-Employment in France 1994-2012

In this section, we describe the (complex) landscape of self-employment in France over the period 1994 to 2012, by providing details on the institutional background in France, the different fiscal incentives in place, and their evolution over time.

2.1 A Primer on the French Personal Income Tax and Social Insurance System

We start with a brief note on the French tax and social insurance system with regards to the features that will be relevant for the self-employed.

Taxable income of a household is the sum of all the sources of income – including income from self-employed activities– minus exemptions and deductions (itemized and standard). Each household has a scaling factor called the number of parts, which is determined by the household composition. For a single adult, that scaling factor is one, for a married couple, it is 2. Each child adds 0.5, up to the third child which adds 1. A disabled child adds 1. For example, a married couple with a child has a number of parts equals to 2.5. A married couple with 3 children has a number of parts equals to 4, and a married couple with one disabled child has a number of parts equals to 3.

The tax bracket cutoffs are expressed in terms of the so-called family coefficient, defined as:

$$\text{Family coefficient} := \text{FC} = \frac{\text{household taxable income}}{\text{number of parts}}$$

Appendix A.2 shows how the tax liability of a household is determined. In brief, the family coefficient serves the same role as the taxable income in the U.S. for determining the tax bracket and total tax paid "per-part." To get the total tax liability of the household, the "per part" tax is inflated by the number of parts.[3]

An important feature of the tax system is that there is no unique map from taxable income to tax bracket. In fact, at a given taxable income, there can be a wide range of tax brackets based on family structure, which will be helpful in our analysis and for the estimation.

Employed and self-employed also have to make a sizable contribution (around 30%) of their earnings to the system of social insurance. These payments are collected and managed by entirely different government bodies than the income tax.[4] They go towards government-provided health insurance, workers' compensation, disability insurance, social insurance and public pensions, as well family-related and means-tested transfers. For the self-employed these social insurance contributions are levied on the same base as the income tax, but with a different timing and some adjustments. Social insurance contributions depend on the type of activity. Additionally, there are many different rates and contribution schemes for agents with different professions. This contributes to the significant heterogeneity in the total tax rate (income tax plus social insurance contribution rate) faced by agents, even conditional on the same total income.

---

[3]Figure A1 shows the income tax schedule for fiscal years 1994, 2006, and 2012. The tax schedule changes almost every year as part of the yearly budget voted by the French Parliament.

[4]These contributions do not appear on the tax returns.

2.2 Self-employed Regimes: A Complex System

Activity types:

For tax purposes, the self-employed are classified into three types of activities. These are important because they affect the policy parameters facing an agent, which we describe below. The three types are: (i) the "Industrial and Commercial Services" category, referred to as "I&C Services" below,[5] (ii) the "Industrial and Commercial Retail" category, referred to as "I&C Retail,"[6] and the (iii) the "Non Commercial" category.[7].

These activity types, defined for fiscal purpose, do not necessarily align well with the underlying economic characteristics of businesses. For instance, developing and selling software pertains to the Non Commercial type, while purchasing and selling equipment goods pertains to the I&C Retail category. Similarly, bakery, butchery, or restaurant businesses are counted as I&C Retail activities, while construction work, plumbery, carpenters, and auto or other repair shops and dry cleaning count as I&C Services. Moreover, all professional activities, such as consulting, private coaching, translation services, sales agents services, expert services, empty property subleasing, as well as all liberal professions (doctors, notaries, or lawyers in private practices) belong to the Non Commercial category.

Three self-employed regimes

In this paper, we focus on self-employed businesses that operate under the personal income tax code.[8] As of 2012, these self-employed could choose one of three regimes:

 (1) *The standard regime*[9]

 (2) *The simplified regime*[10]

 (3) *The super simplified regime*[11]

These three regimes can be characterized along seven main dimensions summarized for convenience in Table 1:

 1. *Eligibility requirements:* The super simplified and simplified regimes can only be chosen by agents with revenues below a threshold y_{kt}^*, which depends on the type of activity k (with k = I&C

---

[5]These are the so-called *Bénéfices Industriels et Commerciaux Services.*

[6]*Bénéfices Industriels et Commerciaux Vente.*

[7]*Bénéfices Non Commerciaux.*

[8]A self-employed individual who owns his business can also chose to incorporate and be subject to the corporate tax code. We do not study those individuals, who typically operate on a larger scale than the businesses studied here.

[9]*Régime Réel.*

[10]We lump together under this heading two regimes, which are indistinguishable in the tax data: the (1) *Régime Micro-entreprise* and the (2) *Régime Auto-Entrepreneur sans Option Libératoire.*

[11]*Régime Auto-Entrepreneur avec Option Libératoire.*

Retail, I&C Services or Non Commercial) and on the fiscal year t. Figure 3 shows the thresholds' evolution. The thresholds are not very high for the Services and Non Commercial activities (equal to 32,600 euros in 2012), but much higher for the Retail activities (81,500 euros). In the case of the super simplified regime, there is an additional condition on the family coefficient as of year $t-2$, which has to be below a year-specific threshold f_t^* that corresponds to the third tax bracket cutoff.[12] Figure 2 schematically represents the regime options. An agent with revenues below the threshold (y_{kt}^*) for his activity type in a given year can choose between the simplified, super simplified, and standard regimes. Above the threshold, the only possible option is the standard regime.[13] In addition, certain types of professions cannot operate under the simplified or super simplified regimes, most notably agricultural activities, leasing of durables and equipment, leasing of professional or non furnished buildings, and real estate businesses. Additional activities excluded from the super simplified regime include liberal professions such as lawyers, doctors, insurance agents, or accounting experts, and formally registered artists rewarded through copyright.

2. *Definition of the income tax and social security base:* In the standard regime, taxable income is net business income, i.e., the difference between gross revenues and costs, including depreciation of assets and investments according to standard accounting rules. In the simplified regime, taxable income is equal to revenues times a scaling factor $1-\mu$, where the rebate μ is determined by the tax administration. It depends on the activity type and has changed over time (see Figure 3). In the super simplified regime, taxable income is simply revenues (i.e., the rebate $\mu = 0$).[14] In the simplified and super simplified regimes, an agent can not claim any deficits.

3. *Income tax and social insurance contribution rates paid on the base:* In the standard and simplified regimes, the regular tax and social insurance contribution rates apply (both of which differ across households depending on several factors as explained above). In the super simplified regime, the agent pays a flat rate that simultaneously takes care of both income tax and the social insurance contributions. The flat rate differs by activity and has changed over time. It is completely unrelated to an agent's actual income tax bracket or tax rate that applies to the rest of his (non super simplified) income. Thus, even an agent in the zero income tax bracket still has to pay the flat rate times revenues for all his activities that fall under this regime. In the simplified regime, a minimal social security contribution is due even at zero revenues.

4. *Business registration procedure:* In the super simplified regime, the registration procedure is

---

[12]For instance, that cutoff was 26,420 euros for year 2010, so that for households to be eligible for the super simplified regime in 2012, their family coefficient in 2010 had to be lower than 26,420 euros.

[13]In theory, there is a limit of 750,000 euros for self-employed in the standard regime. We will not study that threshold, as it makes an agent shift between the personal and corporate income tax realms.

[14]A subtlety to note is that, to determine the overall tax bracket of the household, it is the revenues times $1-\mu$ where μ is the same rebate as in the simplified regime above that is added to the rest of a household's income. It is not the full amount of revenues that is added, which would make the super simplified regime very unattractive.

highly streamlined, straightforward, and quick, lowering the cost of starting a business. However, in all regimes, the requirements for professional qualifications and the quality and safety standards of each activity are identical.

5. *Accounting, and reporting requirements:* Self-employed in the standard regime have to keep detailed accounts to document their revenues and costs, following standard rigorous accounting practices. Businesses in this regime can join a certified accounting center (hereafter, CAC), which helps them keep and check their accounts and serves as a guarantee of sound fiscal conduct to the tax authority. In practice, almost all join a CAC because not doing so results in the business' taxable income being inflated by 25%. Self-employed in the simplified and super simplified regimes only need to report their revenues and are not required to comply with rigorous accounting practices. They are nevertheless required to keep private accounts for their activity, as well as receipts from purchases and sales in case of an audit (much like any tax payer who would, e.g., claim itemized deductions).

6. *VAT payments:* The standard regime is the only one subject to the VAT: self-employed in this regime charge VAT on their products sold and claim VAT on their inputs.

7. *Timing of payments:* In the standard and simplified regimes, tax payments occur annually at the normal tax filing date and social insurance payments occur separately through the regular social insurance procedure. In the super simplified regime, tax and social insurance payments are due monthly or quarterly, based on actual realized revenues (cash in hand), and are all taken care of at the same time, thus minimizing filing and hassle.

FIGURE 1: ELIGIBILITY THRESHOLDS AND REGIME CHOICE OPTIONS

Possible regime choice options

$y_{kt}^* = $ eligibility threshold
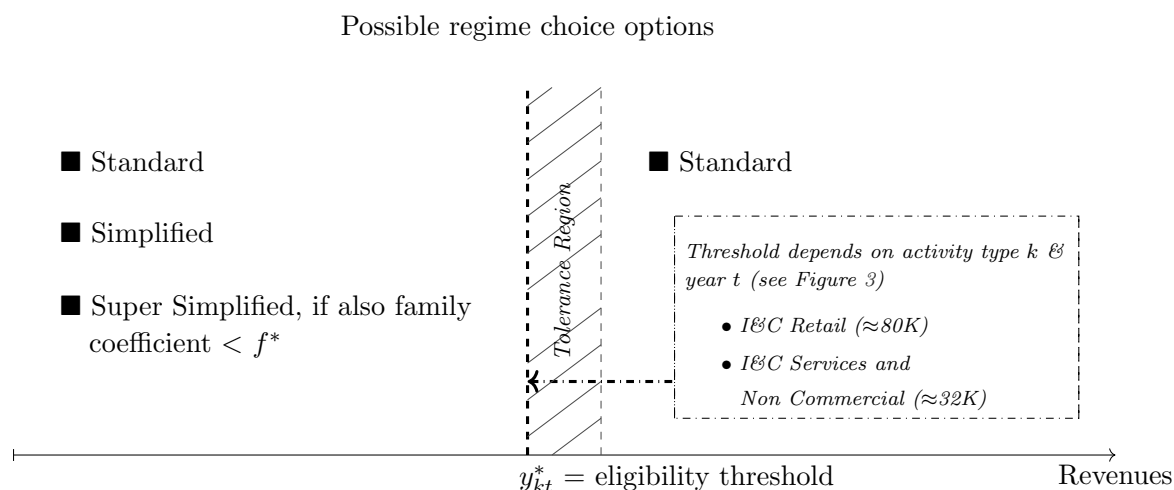
On balance, the key advantage of the standard regime is that it allows subtraction of input and operating costs from taxable income. This is advantageous for businesses with employees, significant

TABLE 1: Summary of the self-employed regimes

	(1) Standard	(2) Simplified	(3) Super simplified
Eligibility	None	Revenues $< y^*_{kt}$	Revenues $< y^*_{kt}$ $+ \mathrm{FC}_{t-2} < f^*$
Income tax & SI contribution base	Net business income	Gross revenues $\times$ (1- rebate)	Gross revenues
Income tax & SI contribution rate	Standard	Standard	Flat rate
Registration procedure	Standard	Standard	Simplified
Accounting requirements	Detailed	Only for audit	Only for audit
Subject to VAT	Yes	No	No
Timing of payments	Annual	Annual	Monthly or quarterly

investments, or high operating costs.[15] The main disadvantage is that it has higher tax "hassle" costs, i.e., a lower level of tax simplicity because it necessitates more stringent administrative, accounting, and reporting requirements. The key advantages of the simplified regime are that, first, the rebate μ may be very generous for agents with low operating costs as it de facto allows them to deduct more than they actually spent and, second, that the tax hassle costs, such as accounting requirements, are lighter. On the flip side, any large expenses or investments cannot be deducted. The key advantage of the super simplified regime is its maximal tax simplicity. In addition, the flat rate is low relative to the sum of the regular income tax and social insurance contribution rates.

Recall that we defined tax simplicity as the combination of conceptual and practical simplicity: a tax regime is simple if it is both easy to understand and easy to handle logistically, i.e., has low tax hassle and tax administrative burdens. Thus, the three regimes can be ranked by tax simplicity according to this definition as in Figure 2.

FIGURE 2: TAX SIMPLICITY BY REGIME CHOICE OPTIONS
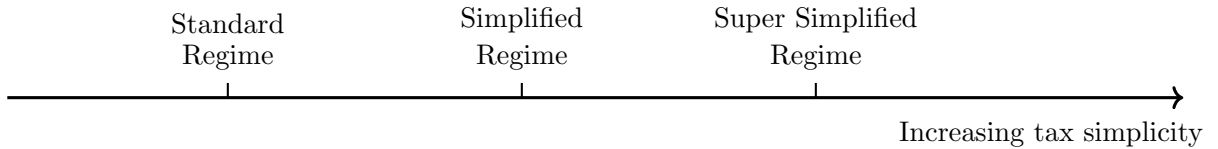
Based on these differences, we can imagine that, broadly speaking, agents who should chose the

---

[15]In addition, self-employed in this regime can benefit from tax credits, such as those for R&D spending, e.g., *Crédit d'impôt recherche* or *crédit d'impôt compétitivité et emploi*, and some government help in special zones, none of which are available when filing under one of the simplified regimes.

standard regime are those with high operating costs, with larger investments (since no investment can be deducted or depreciated in the other regimes), who want to hire employees, whose activity is expected to grow rapidly during the year, who may expect the need to claim a deficit, and who do not find tax simplicity as important.

Grace period and tolerance region:

Agents have to decide by February of fiscal year t which regime they want to be affiliated with for their income earned in fiscal year t. If the agent's revenues end up being higher than the threshold, there is a two-year grace period, as long as the revenues are still within a "tolerance region" (which is, e.g., 6.1% higher than the actual threshold in 2012 for the Services and Non Commercial Activities and 9,9% higher for the Retail Activities).[16] This is shown as the hatched area in Figure 1. If the eligibility threshold is crossed for more than two years or if the tolerance threshold is crossed, the special regime status is lost, and the agent has to file under the standard regime.

Ease of misreporting:

As noted above, agents in the standard regime who are not members of a certified accounting center (CAC) see their taxable income inflated by 25%. Figure A2 shows that at higher business income levels, almost 100% of all agents in the standard regime are CAC members.[17] A government report (Cour des Comptes, 2014) states that conditional on an audit, the size of the penalties among non-CAC members is larger than among CAC members of comparable size (around 26,000 euros versus 7,000 euros). In addition, the Cour des Comptes (2014) states that the discrepancies in taxes due and taxes actually paid among CAC members seem mostly due to genuine accounting mistakes and delays in payments, and almost never to outright tax evasion – as opposed to the discrepancies noticed among non-CAC members.

The large share of agents in the standard regime who are members of a CAC, especially around the threshold that we will focus on (where it is essentially 100%) lends support to the hypothesis that cheating is much easier in the simplified or super simplified regimes and much more difficult in the standard regime.

Two Key Reforms:

The thresholds and rates applicable to each regime have changed extensively over time as shown in Figure 3. These changes generate policy variation that is key for our analysis. Two major reforms stand out: The 1999 reform greatly extended the eligibility threshold for the simplified regime from 100,000 French Francs (15,244 euros) to 500,000 French Francs (76,220 euros) for I&C Retail activities and to

---

[16]This grace period and tolerance threshold do not apply in the first year after the business' creation.

[17]Why is the fraction not 100% at lower income levels? It is not explained by agents in the zero tax bracket (who do not care about the inflation of taxable income). It may be that agents at lower income levels under-report more than 25% of their taxable income and are not very averse to the risk of being audited.

175,000 French Francs (26,678 euros) for the I&C Services or Non Commercial activities. Before 1999, the thresholds were so low as to only be applicable for very small activities and the simplified regime was not a reasonable option for many self-employed. The 2008 reform created the super simplified regime. It stemmed from the political will to further increase tax simplicity by replacing the social insurance contributions and income taxes by a unique tax transfer proportional to self-employed revenue.

3 Data and Descriptive Statistics

We now describe our datasets and provide new summary statistics on the demographic and economic characteristics of the self-employed across activities, years, and regime types.

3.1 Data

Our main data consists of the entire tax returns of a representative sample of 500,000 households (out of around 33 million)[18] from the French Internal Revenue Service.[19] For fiscal year 2011, we have in addition the full population data of around 36 million households.[20] The income tax returns contain comprehensive income data at the individual and household levels, as well as key demographic information such as household composition, individual age, and gender.

We also make use of a quite unique data source, the *Enquete Revenus Fiscaux et Sociaux*, which consists of tax returns for a subsample of the population that are matched to large-scale employment survey data and benefits receipt data. This combined dataset covers the period 1996 to 2012 and has a sample size of around 100,000 respondents per year. It contains some highly useful variables such as education, type of profession and occupation, social insurance benefits and government transfers received, standard of living, and tax free capital income.

Finally, we also use the register of businesses when we study the effects of the reforms on the creation of new self-employed businesses and switches between regimes.[21]

3.2 Descriptive Statistics on Income and Demographics of the Self-Employed

These new datasets provide some key original summary statistics on the demographic characteristics and incomes of the self-employed in France.

TABLE 2: AVERAGE TOTAL TAX RATES BY REGIME, ACTIVITY AND TAX BRACKET

Panel A: Total Average Tax Rates in the Simplified and Super Simplified Regimes

| | Simplified 1999-2008 | | Super Simplified 2009-2012 | |
Bracket	I&C Services	Non Commercial	I&C Services	Non Commercial
1 (low)	48.0%	45.0%	23%	20.5%
2 (medium)	52.6%	49.7%	23%	20.5%
3+ (high)	63.2%	60.2%		20.5%

Panel B: Total Average Tax Rates in the Standard Regime

| | 1999-2008 | | 2009-2012 | |
Bracket	I&C Services	Non Commercial	I&C Services	Non Commercial
1 (low)	32.9%	31.5%	32.5%	31.1%
2 (medium)	36.0%	34.8%	35.1%	33.5%
3+ (high)	43.3%	42.1%		37.9%

Tax brackets and tax rates of the self-employed

We will often consider heterogeneous effects by tax brackets, so it helps to define them here. The tax system has changed over time, but we can in each year assign every agent to a tax bracket and compute his average income tax rate and his total tax rate (including social security contributions) in each regime. We group people into bracket 0 ("zero" tax rate), bracket 1 (which we also call "low" in some graphs and tables), bracket 2 ("medium") and brackets 3 and above ("high"). Table 2 gives a concrete picture of the levels of taxes faced by self-employed in these tax groups, depending on their self-employed regime and on the time period. The mean total average tax rates including social insurance contributions are summarized in panel A for the simplified and super simplified regimes, and in Panel B for the standard regime. Total average tax rates, can be very high, up to 63% in the "high" tax bracket (3+). In the Super simplified regime, total tax rates are much lower, highlighting the financial advantages of being in that regime. The "zero" tax bracket group faces, by definition, a zero income tax, but still pays social security contributions, so that its total tax is not zero.

---

[18]There are around 1200 variables.

[19]*Direction Générale des Finances Publiques (DGFiP).*

[20]Currently, we are working on incorporating the full population data for all years 2011-2015. Using the year 2011 for which we have both the subsample of 500,000 and the full population data, we can indeed check that all results look very similar when we use either the full sample or the subsample.

[21]The datasets are *Fichier de comptabilité unifié dans SUSE* (FICUS) and *Fichier approché des résultats d'Esane* (FARE).

Key characteristics of the self-employed

Figure 4 shows the evolution over time of the fraction of self-employed among the total population aged 18 to 65. The two vertical red lines represent the 1999 and the 2008 reforms respectively, which we will study in detail in Section 6. We distinguish two groups: those who have some self-employed income (in red) and those who have only self-employed income (in blue). The fraction of self-employed has remained stable over time until 2009, at around 5%, while the fraction with only self-employed income has been slightly decreasing from 4.2% in 1994 to 3.7% in 2009. But self-employment has seen a rise since 2009, especially in the number of agents who earn self-employed income in addition to salary income. In 2012 the fraction with any self-employed income had risen to 6%.

Table 6 shows the demographic characteristics (in Panel A), income (in Panel B) and income tax information (in Panel C) of three subgroups: (i) wage earners, (ii) self-employed, (iii) agents who have both self-employed and wage income. Averages are taken over the full sample period 1994-2012. The self-employed tend to be on average 8-9 years older than wage earners and are almost three times as likely to be retired. This goes against an often encountered idea whereby the self-employed are young entrepreneurs. There is no significant difference in the presence or number of children, but the self-employed are more likely to be married or in a civil union. Despite being, by definition, employed or earning self-employed income for at least part of the year, a significant portion of agents have also claimed unemployment benefits at some point during the year; however, self-employed are less likely to have done so than wage earners.[22] The variable "Educated" is equal to one for agents who have at least a professional or training high-school level degree, but not an academic high school level degree (the *Baccalauréat* (BAC)).[23] There is no significant difference in the fraction of self-employed and wage earners with a high-school degree, but the self-employed are significantly more likely to have completed at least a bachelor-level university degree. The variable "High skill" identifies agents that are in higher skilled occupations, such as licensed professionals, teachers, engineers or executives of the public service or private sector. The self-employed are significantly more likely to be in high skill occupations.

Panel B shows that those who receive self-employed income only earn on average around 33,000 euros. Those who earn self-employed income in addition to salaried income have on average 30,000 euros of self-employed income and only around 6,000 euros of wage income. Self-employed agents have more than three times as much capital income as wage earners (around 6,000 as compared to 1,900). The same goes for tax free capital income. The variable "standard of living" measures the total disposable income per adult equivalent at the household level. On average, those with at least some self-employed income have a 25% higher standard of living than wage earners.

Panel C shows the distribution across tax brackets of each group. Self-employed individuals are

---

[22]The interplay between unemployment insurance and self-employment could inform the studies on the design of unemployment benefits and their insurance value as in Landais, Kolsrud, Nilsson, and Spinnewijn (2015) and Landais and Spinnewijn (2017).

[23]In French, the two professional high school level degrees are the *Certificat d'Aptitude Profesionnelle (CAP)* or the *brevet d'études professionnelles (BEP)*.

more than three times as likely to fall in the highest tax brackets relative to wage earners.

Appendix Tables A1, A2 and A3 repeat Table 6 over different periods. There are several noteworthy findings. First, wage income adjusted for inflation has been consistently rising, but average self-employed income experienced a fall post 2008, at the same time as the number of agents with self-employed income rose (Figure 4). We return to this in more detail below when we study the 2008 reform that created the super simplified regime. Second, capital income increased significantly for all groups. Third, the proportion of self-employed who perceive unemployment benefits at some point during the year doubled from the 1999-2008 to the 2009-2012 period. Such an increase did not occur for wage earners.

Differences between Industrial and Commercial and Non Commercial activities

Table 7 shows the demographic and socioeconomic variables for the self-employed split by activity type. There is a significant gender gap: women are significantly more represented in Non Commercial activities, although they are underrepresented among the self-employed in general. Retirees concentrate in the I&C Retail and I&C Services activities. The most educated and highly skilled self-employed are in the Non Commercial activities, which includes liberal professions such as doctors, lawyers, or notaries. Perhaps as a result, self-employed with Non Commercial activities have higher standard of living and capital and labor incomes.

Differences between regimes

Table 8 shows the same summary statistics, but by type of regime. The first two columns compare the standard and simplified regimes, which existed throughout the full sample period; the last three columns compare the simplified, standard and super simplified regimes over the period 2009-2012 after the introduction of the super simplified. Agents in the simplified are on average older, much more likely to be retired. Agents in the super simplified are on average younger, less likely to be married, and much more likely to be claiming unemployment and social insurance benefits. Agents in the simplified and super simplified are less educated and in lower skill occupations than those in the standard regime.

As expected, people in the standard regime have nearly four times as much self-employment income as those in the simplified or super simplified regimes and a significantly higher standard of living (i.e., disposable income per adult equivalent). Those in the simplified regime have the most wage, followed by the super simplified regime. This is to be expected since self-employed who have larger business that they want to focus on (i.e., who earn little or no additional wage income) would choose the standard regime. Those who want to combine a smaller self-employed activity with some additional salaried work would typically choose the simplified or super simplified regime. Agents in the super simplified and simplified regimes are similarly distributed across tax brackets, and both significantly poorer overall than those in the standard regime.

4 Bunching in the Simpler Regimes

In this section, we provide graphical evidence of behavioral responses at the eligibility thresholds. We first start by outlining an analytical framework that shows the changes in tax incentives and tax simplicity around the thresholds, and captures in a schematic way the main institutional features we outlined in Section 2.

4.1 Modeling the Tax Discontinuities

A given individual can operate in one of the three regimes described above: the simplified (regime "m") the super simplified (regime "f") and the standard one (regime "r"). Effective operating costs, taking into account input costs and VAT payments are modeled as a fraction c_i of revenues y_i in each of the regimes $i = m, f, r$. Each regime entails a "tax hassle cost" a_i, reflecting the tax reporting and compliance costs (e.g., registration costs, administrative accounting requirements, costs of keeping track and understanding, etc.).

Let μ be the rebate on gross revenues in the simplified regime: the taxable income of agents in this regime is $(1 - \mu) \cdot y$. The agent's effective average income tax rate (which is the same regardless of regime choice) is τ^y and the social insurance contributions rate in regime i is τ_i^{ss}. Denote by τ_i the effective rates in regime i, levied on the tax base applicable in that regime, denoted by z_i (which also differs across regimes). In practice, an agent's effective average income tax rate and his social insurance contribution rate depend on his total income (self-employed income, wages and salaries, ordinary capital income, etc.), household composition, activity type, and occupation, as explained in Section 2. We do take this heterogeneity into account in our numerical estimations when assigning a tax rate to each agent. For simplicity of the exposition in this section, we express the rates as if they were homogeneous across all agents in a given regime.

The effective rates and tax bases are as follows:

Standard regime: $\qquad \tau_r = \tau^y + \tau_r^{ss}(1 - \tau^y) \qquad$ is levied on net income $\qquad z_r = (1 - c_r)y_r$

Simplified regime: $\qquad \tau_m = \tau^y + \tau_m^{ss} \qquad$ is levied on taxable income $\quad z_m = (1 - \mu)y_m$

Super simplified regime: $\qquad \tau_f \qquad$ is levied on gross revenues $\qquad z_f = y_f$

It is more convenient for the sake of comparability between regimes to rewrite the agents' payoffs in terms of the effective tax rates t_i expressed as a fraction of the revenues, i.e., such that:

Standard regime: $$t_r = c_r + (\tau^y + \tau_r^{ss}(1 - \tau^y))(1 - c_r)$$

Simplified regime: $$t_m = c_m + (\tau^y + \tau_m^{ss})(1 - \mu)$$

Super simplified regime: $$t_f = c_f + \tau_f$$

The thresholds of eligibility y^* for the simplified and super simplified regimes create "notches", i.e., discontinuities in the incentives. What is peculiar in our setting is that part of the change comes from monetary incentives and part come from the tax simplicity. We can combine both types of incentives into a single monetary equivalent, i.e., we can express the change in tax simplicity in terms of a monetary tax hassle cost change, ΔT. Doing so, at each eligibility threshold, an agent faces a total tax liability that can be generically written as a function of the tax before the threshold t, the change in marginal tax Δt and the change in average tax ΔT as:

$$T(y) = ty + (\Delta T + \Delta ty)I(y > y^*) \tag{1}$$

Consider first an agent who would prefer being in the simplified regime rather than in the standard regime below the threshold. For him, the value of the base tax is:

$$t = c_m + (\tau^y + \tau_m^{ss})(1 - \mu),$$

and the monetary wedges are:

$$\Delta t = [c_r - c_m + \tau^y(\mu - c_r) + \tau_r^{ss}(1 - c_r)(1 - \tau^y) - (1 - \mu)\tau_m^{ss}] \quad \text{and} \quad \Delta T = a_r - a_m$$

For an agent who would like to be in the super simplified regime below the eligibility threshold, the value of the base tax is:

$$t = c_f + \tau_f$$

and the monetary wedges are:

$$\Delta t = c_r(1 - \tau^y) + \tau^y - \tau_r^{ss}(1 - \tau^y)(1 - c_r) - c_f - \tau_f \quad \text{and} \quad \Delta T = a_r - a_f$$

Remarks:

The change in the overall monetary wedge Δt and the ensuing agent's reaction to the notch depend not just on the pure tax liability τ^y or on social contribution rates, but also on the relative operating costs c_i, the change in the taxable base (e.g., on the gap between operating costs and the rebate), the difference in hassle costs, and the income tax rate, which magnifies any tax base difference.

Even if agents were to locate at the same self-employed revenue level (e.g., at the eligibility thresh-

old), they can nevertheless still face widely disparate financial incentives as embodied in t, Δt and ΔT. The parameters in the formulas for t, Δt and ΔT depend on the activity type (which affects operating costs and the rebate) and on an agent's total income and family situation (as explained in Appendix A.2, agents with the same self-employed and total income can be in different income tax brackets because of the peculiarity of the French tax system). Table 2 illustrated how different the tax rates can be, even conditional on the same self-employed revenues.

Only agents who would choose to be in the simplified regime and earn revenues above the eligibility threshold absent the notch face a discontinuity in their tax schedule. These are, to a first order, agents with low operating costs c and high hassle costs a. For agents who do not want to be in the simplified regime there is no discontinuity at the thresholds. Thus, when we consider the excess mass of revenues, we should not count agents who below the threshold choose to be in the standard regime.[24] In terms of the model, agents affected by the notch have parameters such that $\Delta t > 0$.

In the graphs below, we mostly focus on the I&C Services and Non Commercial Activities, because the threshold for the I&C Retail activities is placed so high that the income distribution at that level is very thin. We do use the I&C Retail as a placebo group around the (much lower) threshold for the I&C Services and Non Commercial activities (where the distribution of I&C Retail income is sufficiently dense as well).

In graphs where we pool all self-employed agents (in all regimes), we need to work with taxable income z (rather than revenues y). This is because the self-employed in the standard regime have to report only their net income (revenues minus costs); those in the simplified and super simplified regimes only have to report their revenues (and not their costs). We convert these reports into the common and comparable (reported) "taxable" income z as defined above. Similarly, we convert the thresholds for the simplified regime into a taxable income equivalent.[25] In graphs in which we only consider agents in the simplified or super simplified regimes, we can directly work with revenues y.

4.2 Regime Choice Responses: Graphical Evidence

The choice between one of the simpler regimes (simplified or super simplified) and the standard regime depends on the combination of all factors described in Section 2, and is thus complex. An agent would tend to choose one of the simpler regimes if his operating costs are low, if his investment in the business was minimal, if he has no employees, or finds tax hassle particularly burdensome.

We start by checking whether the proportion of agents who choose one of the simpler regimes (simplified or super simplified) exhibits uneven behavior around the eligibility thresholds. Figure 5a

---

[24]Since we do not actually see revenues for agents in the standard regime, when we plot graphs in terms of revenues, we are automatically taking into account only agents who are in the simplified regime and ignoring those who chose the standard regime.

[25]This yields two thresholds (one for I&C Services equal to $z^*(1 - \mu_{\text{I\&C Services}})$ and one for Non Commercial activities $(z^*(1 - \mu_{\text{I\&C Retail}}))$, due to their different rebates. The thresholds for the super simplified regime is the same whether expressed in revenue or taxable income since $z_f = y_f$ and $\mu_{\text{Super simplified}} = 0$ in that regime, regardless of the activity.

shows the fraction of agents in any simplified regime among all self-employed, as a function of their taxable income. The vertical lines represent the eligibility thresholds expressed in taxable income for, respectively, the simplified I&C Services, the simplified Non Commercial, and the super simplified (I&C Services plus Non Commercial). There are jumps in the fraction of agents in a simplified regime at all those thresholds – a first indication for behavioral responses to the thresholds.[26]

For any given business structure and preferences between regimes, agents in higher tax brackets – holding income fixed– have stronger financial incentives to optimize their regime, since a given difference in payoffs between the two regimes gets amplified by a larger tax saving. This can be seen in the expression for Δt above, for instance around the simplified regime's eligibility threshold. Conditional on $\Delta t > 0$ (i.e., the agent has an incentive to stay in the simplified or super simplified regime), the financial gain from optimizing one's income tax base (captured by the term $\tau^y(\mu - c_r - (1-c_r)\tau_r^{ss}) > 0$) is increasing in τ^y holding all else fixed. Thus, an agent who would chose to be in the simplified regime given his primitives, would be even more inclined to do so if he were in a higher tax bracket. Recall that because of the way tax brackets are determined in the French tax code, there are several possible tax brackets at a given taxable income.

Figure 5b shows the fraction of agents in any simplified regime by tax bracket. All groups exhibit a strong threshold effect in regime choice, with excess mass right before the eligibility thresholds and a dip right after. We can compute the percent increase in the proportion by taking the ratio of the proportion right at the threshold relative to the proportion at 2,000 euros before the threshold (i.e., around where the excess mass starts forming) for each of the three tax bracket groups. For the low tax bracket, the increase is of 35%, for the medium tax group it is 47%, and for the high tax group it is 59%. As predicted, the high tax group has the largest (proportional) excess mass.[27]

4.3 Income Responses: Graphical Evidence

We now study the income responses to the notch created by the eligibility thresholds.

Method

We can identify the excess mass in self-employed revenues as illustrated in Figure 6 to the left of the eligibility threshold. Recall that except for revenues that fall in the tolerance region during the grace period (as described in Section 2 and Figure 1), we do not observe revenues for agents above the eligibility threshold.

---

[26]That the response is especially strong at the super simplified threshold may be due to either or to both of the following facts: First, this threshold applies to both I&C Services and Non Commercial activities, and hence attracts a larger mass of people, while the other thresholds apply to one activity type only because of the different rebate rates. Second, it may just be that agents respond more to the incentives provided by the super simplified regime.

[27]Appendix Figure A3 shows the same, but zooms in exclusively on the fraction in the super simplified regime among all self-employed who are eligible for the super simplified regime. For the low tax bracket, the increase in the proportion is of 47%, for the medium tax group it is 63%, and for the high tax group it is 118%.

Figure 7 shows the distribution of self-employed revenues for agents in the simplified and super simplified regimes, pooled over all years 1999-2012, for those in the I&C Services and Non Commercial Activities.[28] We pool together these years to increase our sample size and precision, but we show the results year-by-year and for different periods below. The revenues on the horizontal axis are centered around the eligibility threshold (the red vertical line) which varies over time and across activities. I.e., in each year, we plot the distribution of the difference between revenues and the eligibility threshold in that year and for that activity. We group individuals into 500 euro bins by re-centered revenue and plot the count in each bin. There is a spike in the income distribution right before the threshold.

To estimate the counterfactual density to the left of the threshold that would apply absent the notch, we fit a flexible polynomial to the data, excluding a range R_1 of income before the threshold T. We divide individuals into income bins of size 500 euros, so that individual incomes in bin B_j are in the interval $]B_j - 500, B_j]$. Bins are always constructed so that $T = B_T$, i.e. such that the threshold T is a bin upper bound. Typically, when income is normalized by the threshold, $B_j = \{..., -5000, -4500, -4000, ..., -500, 0, 500...\}$ and $B_T = 0$. C_j stands for the count of individuals in income bin B_j.

We fit a smooth polynomial to proxy for the counterfactual distribution to the left of the threshold, by running the following regression:

$$C_j = \sum_{i \in A} \beta_i (B_j)^i + \varepsilon_j \ , \ \forall B_j \leq T - R_1 \tag{2}$$

where A is the set of polynomial exponents, which is allowed to be fractional (i.e. A is a finite set, $A \subset \mathbb{Q}$ and $A \cap (\mathbb{Q} \setminus \mathbb{N}) \neq \emptyset$). R_1 is an interval to the left of the threshold which is excluded from the regression. We then use estimates from (2) to obtain the predicted counterfactual

$$\hat{C}_i = \sum_{j \in A} \hat{\beta}_j (B_i)^j$$

in the excluded range. The excess mass B is calculated as:

$$B = \sum_{i \in S} (C_i - \hat{C}_i)$$

where $S = \{i \in \mathbb{N}/B_i \in]T - R_2, T]\}$ and $R_2 \leq R_1$. B is therefore the number of individuals in excess relative to the counterfactual distribution in a "bunching zone" that is smaller than the excluded zone. We take R_2 to be one bin (of size 500 euros). We then define the normalized excess mass b to be the excess mass in the bunching zone, divided by the counterfactual number of individuals in the bin with

---

[28]As already explained above, we do not consider the I&C Retail Activities whose threshold is very high up.

upper bound $B_T = 0$, which is $\hat{C}_T$.[29]

$$b = \frac{B}{\hat{C}_T} \qquad\qquad (3)$$

The counterfactual distribution is represented by the red curve in the figures. The bunching zone's area excess mass is colored in yellow. For the graphs, we exclude an interval $R_1 = 1,500$ euros before the eligibility threshold and compute the excess mass over a small interval $R_2 = 500$ euros. This is a conservative estimate of the excess mass, since the graphs indicate that the density starts increasing relative to the counterfactual one even further away. Our estimations are quite robust to the choice of length of the excluded interval and the bunching zone. They are also very robust to "round number bunching" (i.e., to including "round number" dummies). Identification relies on the fact that the distribution of self-employed revenues, absent the eligibility thresholds, would be smooth around the eligibility thresholds.[30]

To compute standard errors for the excess mass b (as well as for the elasticities which we compute based on B in Section 5), we generate earnings distributions and excess mass estimates by resampling the residuals in (2). The standard error is taken to be the standard deviation of the distribution of estimates generated by bootstrapping.

Results

Figure 7 shows that the excess mass before the notch estimated according to this method is 70% of the average height of the counterfactual distribution within 1,500 euros of the notch. The red curve represents the distribution for a placebo group, namely the I&C Retail, for which the eligibility threshold is higher. There is, reassuringly, no bunching for the placebo group.

Figure 8 splits the sample further into agents in the simplified and agents in the super simplified regimes. Panel (a) shows bunching at the eligibility threshold for the simplified regime (over this period, the super simplified regime did not exist) and panel (b) shows bunching at the super simplified eligibility threshold for the period 2009-2012. There is even starker bunching at the eligibility threshold of the super simplified regime, equal to more than 3.5 times the counterfactual density right before the threshold. Indeed, the financial benefits and tax simplicity advantages of this regime are quite sizable. In each panel, the red curve represents the income distribution of agents with I&C Retail activities, which serve as a placebo group. Reassuringly, they exhibit no excess mass.

Heterogeneity in Bunching

Bunching by tax brackets. Recall that any cost or benefit advantage from being in one regime over the other is amplified for agents in higher tax brackets. Figure 9 shows the bunching mass for agents

---

[29]Both the numerator and the denominator are effectively the counts within a bin, so that their sizes are directly comparable.

[30]Extensive margin responses may occur (we show below that here is entry into self-employment after 2008), but they should not necessarily concentrate exactly below the threshold. Since the estimations are all based on local responses (sharp bunching) before the threshold, our estimates should not be significantly affected by extensive margin responses.

TABLE 3: Excess mass by tax bracket

Tax bracket	Excess mass b	Standard error $se(b)$
0 (Zero)	0.37	0.11
1 (Low)	0.76	0.05
2 (Medium)	0.77	0.03
3+ (High)	1.24	0.05

Notes: The table shows the excess masses b as defined in (3) and as represented graphically in Figure 9, as well as their standard errors.

in different tax brackets.[31] We again exploit the peculiar feature of the French tax system that leads to vastly different tax brackets even given the same self-employed income (i.e., around the eligibility threshold), as reported in Table 2. Table 3 reports the excess mass and its standard error for each tax bracket. They are all significant and agents in higher tax brackets exhibit a stronger and more significant bunching mass.

Agents with additional income sources. We can also measure bunching for agents who have additional important sources of income. In Figure 10, we distinguish between self-employed agents who have also some salaried income (panel (a)) and those who do not (panel (b)), as well as self-employed agents who have pension income (panel (c)) and those who do not (panel (d)).

Self-employed agents who also have some salary income appear much more sensitive to the threshold and their excess mass is almost twice as large as that of self-employed who have no other income. Similarly, agents who are retired are more than three times as sensitive to the threshold as those who are not retired. It may be that the availability of additional income gives an agent more flexibility to optimize his self-employed income, since he does not have to fully depend on it for his living. Put differently, agents with additional income have better outside options and are thus more elastic to financial incentives and/or sensitive to hassle costs when engaging in self-employed work.

Bunching by education level. Figure 11 shows bunching for non-educated agents and educated agents (according to the definitions from Section 3.2). There is no bunching for non-educated agents; there is very significant bunching for educated agents.[32]

5   Estimating the Value of Tax Simplicity

In this section, we jointly estimate the value of tax simplicity and the income elasticities based on the bunching at the eligibility thresholds. Recall that tax simplicity means that a tax regime is both easy

---

[31]Appendix Figure A4 repeats this for the two periods 1999-2008 and 2009-2012, with similar results.

[32]In Table 5 we show that educated agents are also more likely to make the correct regime choice.

to understand and light on logistical and administrative hassle. The key to this estimation is that there is a large variability in the incentives faced by different agents at the same eligibility threshold because of their different tax brackets and activity types. Combined with the fact that these incentives have also changed extensively over time, this yields a lot of different data moments that inform us about the parameters of interest. Ultimately, the structural parameters that can best explain the observed bunching across different tax brackets, activities and years will turn out to be a large preference for tax simplicity and a small elasticity to monetary incentives.

5.1 Methodolgy: Estimating Income Elasticities and Hassle Costs

The notch created by the eligibility threshold generates a response in self-employed revenues, which we denote by Δy^*. This revenue response Δy^* can be estimated from the excess mass using the expression

$$B \approx f_0(y^*)\Delta y^* \tag{4}$$

where $f_0(y^*)$ is the counterfactual density at the threshold (i.e., the density that applies in the absence of the eligibility threshold).[33] Recall that we can write an agent's tax liability as:

$$T(y) = ty + (\Delta T + \Delta ty)I(y > y^*) \tag{5}$$

with, for instance at the eligibility threshold for the simplified regime, $t = c_m + (\tau^y + \tau_m^{ss})(1 - \mu)$, $\Delta t = [c_r - c_m + \tau^y(\mu - c_r) + \tau_r^{ss}(1 - c_r)(1 - \tau^y) - (1 - \mu)\tau_m^{ss}]$, and $\Delta T = a \equiv a_r - a_m$ as defined in Section 4.1.

An agent's reaction to the notch – as captured by the revenue response Δy^*– depends on the wedges that characterize the change in incentives t, Δt, and ΔT. These wedges in turn depend on the operating cost structure (c_r, c_m), the hassle cost difference a, the rebate μ, the income tax rate τ^y, the social contributions rate τ^{ss}, and the income elasticity. The parameters τ^y, τ^{ss}, and μ are known policy parameters. The operating and hassle cost parameters (c_r, c_m, a) are unknown primitives, which may be heterogeneous across agents. Symmetrically, for the eligibility threshold of the super simplified regime, the wedges depend on $(c_r, c_f, a, \tau^y, \tau^{ss}, a_r - a_f)$.

The revenue response Δy^* derived from the excess mass can in turn be mapped into a reduced form elasticity, which does not require any parametric assumption on the utility function, and into a structural elasticity that requires specifying a parametric utility function. To infer the elasticities, we need to assume that agents who bunch come from a continuous interval of revenues above the eligibility threshold and that there is one agent who is the marginal buncher. We consider a static setting without optimization frictions, but do estimate the elasticities over different time periods to

---

[33]As shown in Kleven and Waseem (2013) one can easily account for heterogeneous elasticities. For an agent with elasticity e, the earnings response is Δy_e^*. If $f_0(y, e)$ is the joint revenue-elasticity counterfactual distribution and $f_0(y) = \int_e f_0(y, e)$ is the unconditional counterfactual revenue distribution, the excess mass can be mapped to the average earnings response $\mathbb{E}(\Delta y_e^*) \approx f_0(y^*)E(\Delta y_e^*)$.

account for learning.[34]

Structural Model:

To estimate the structural elasticity and the tax hassle costs, we parameterize the model as follows. Each agent has a type θ that captures his productivity and that reduces his cost of earning a given level of revenues. The disutility of generating revenues y for an agent of type θ is denoted by $h(y, \theta)$, increasing in y and decreasing in θ. We assume that:

$$h(y) = \frac{\theta}{1 + \frac{1}{\varepsilon}} \left(\frac{y}{\theta} \right)^{1 + \frac{1}{\varepsilon}}$$

An agent's utility from earning revenue y in regime $i = m, f, r$ is thus:

$$u_i(y) = y - T_i(y) - h(y, \theta) - a_i$$

where $T_i(y)$ is the total tax liability as a function of revenues described above.

Absent the notch, the marginal agent in the simplified regime would have chosen revenue level $y^* + \Delta y^*$. This level is characterized by the tangency or optimality condition:

$$y^* + \Delta y^* = (\theta^* + \Delta \theta^*)[(1 - c_m) - \tau_m(1 - \mu)]^\varepsilon \tag{6}$$

After introduction of the threshold, this agent locates exactly at the notch y^* and his utility is:

$$u_m^* = y^*(1 - c_m) - \tau_m(1 - \mu)y^* - h(y^*, \theta^* + \Delta \theta^*) - a_m$$

Let's denote by y_r^I the indifference point in the standard regime, such that the agent is indifferent between earning revenues exactly equal to the threshold y^* or y_r^I. y_r^I yields utility:

$$u_r^I = y_r^I(1 - c_r)(1 - \tau_r) - h(y_r^I, \theta^* + \Delta \theta^*) - a_r$$

The indifference income is interior, and hence characterized by the tangency condition in the standard regime:

$$y_r^I = (\theta^* + \Delta \theta^*)[(1 - c_r)(1 - \tau_r)]^\varepsilon \tag{7}$$

The indifference condition is:

$$y_r^I(1 - c_r)(1 - \tau_r) - h(y_r^I, \theta^* + \Delta \theta^*) - (a_r - a_m) = y^*(1 - c_r) - \tau_m(1 - \mu)y^* - h(y^*, \theta^* + \Delta \theta^*) \tag{8}$$

Equation (8) relates the elasticity ε and the hassle cost $(a_r - a_m)$ to the excess mass through Δy^*

---

[34]Since we estimate elasticities assuming no optimization frictions, we are likely providing a lower bound of the elasticities.

(which can be obtained from $\Delta\theta^*$ through equation (6)).

The elasticity ε is thus the solution to:

$$\frac{1}{1+\Delta y^*/y^*}\left[1+\frac{\Delta T/y^*}{1-t}\right] - \frac{1}{1+1/\varepsilon}\left[\frac{1}{1+\Delta y^*/y^*}\right]^{1+1/\varepsilon} - \frac{1}{1+\varepsilon}\left[1-\frac{\Delta t}{1-t}\right]^{1+\varepsilon} = 0 \qquad (9)$$

Equation (9) defines the elasticity of self-employed revenues as a function of the earnings response $\Delta y^*/y^*$ and the changes in the average tax $\frac{\Delta T/y^*}{1-t}$ and $\frac{\Delta t}{1-t}$. It is solved numerically given an estimate of Δy^* (derived in turn from the excess mass B) and the policy parameters.[35] To derive the elasticity of agents in the super simplified regime from the bunching at their eligibility threshold, we can use the same equation with the corresponding, Δy^*, t, Δt, and ΔT from Section 4.1.

Intuitively, a given observed bunching can be generated by either a preference for low tax hassle or a sensitivity to monetary tax incentives. The hassle cost by itself creates a "pure notch" – i.e., an increase in the average tax without a change in the marginal tax; even in the absence of monetary incentives, tax accounting and reporting requirements act like an average tax increase and have distortionary effects. There are thus different combinations of tax hassle costs and income elasticities that can explain a given level of excess mass.[36] Figure 12 plots the combinations of income elasticities and tax hassle costs (in euros) that can jointly generate the observed excess mass for different tax brackets and activity types. As hassle costs increase, the elasticity naturally decreases. Sufficiently large hassle costs mean that even a very small structural elasticity can generate the observed bunching behavior.

Structural Estimation Method

We show three sets of results based on the structural estimation of the model. First, we cover two polar cases, which provide bounds on the elasticities and tax hassle costs. The first limit case (case 1) is the one in which people do not value tax simplicity *per se*, i.e., hassle costs are assumed to be zero.[37] This would be the standard case in the taxable income literature. The second limit case (case 2) is the one in which agents are not sensitive at all to monetary tax incentives (either because they are inelastic or because they do not understand them or pay attention to them) and only value tax simplicity. This yields an upper bound on the possible hassle costs.[38]

Second, we structurally estimate the model using a generalized method of moments. Different agents in different years face widely different incentives as embodied in t and Δt at the eligibility threshold because they are in different tax brackets (recall the peculiarity of the French tax system as described in Section 2), in different types of activities (which leads to different rebates μ and different operating costs c), and because income taxes, social security contribution rates and rebates

---

[35] In these estimations, agents make a series of static decisions period by period. This is realistic in so far that for the small self-employed considered here, self-employed income is a flexible yearly choice variable, not subject to career concerns or strategic dynamic considerations, such as signaling. We do estimate the elasticities for different periods in Section 6 to account for the presence of learning, which we documented in the previous section.

[36] I.e., there are different combinations of ε and a which are solutions to equation (9).

[37] This case corresponds to the elasticities at $a = 0$ in Figure 12.

[38] This case corresponds to the hassle cost as $\varepsilon \to 0$ in Figure 12.

have changed extensively over time. As a result, we have many data moments, i.e., observations of the empirical responses Δy^* across tax brackets, activity types, and years, which we can target in order to find the parameters that fit best. The estimation is performed separately for agents in the simplified regime and agents in the super simplified regime.

Formally, let k index activity, i index the tax bracket, n index the regime (super simplified or simplified) and t index the group of years for which the thresholds y^* and rebates μ are constant. For each regime, activity, tax bracket, and year, there is a model-predicted bunching interval Δy^*_{nkit}. Its empirical counterpart in the data is $\hat{\Delta} y^*_{nkit}$. Each $\hat{\Delta} y^*_{nkit}$ is thus a data moment (note that these moments are not collinear because t and Δt from nonlinear equation (9) are different for every (n, i, k, t)).

The parameters we seek to estimate are the operating costs c_r and c_m, (or c_f for agents in the super simplified regime), the hassle costs $a_r - a_m$ (or $a_r - a_f$), and the elasticity ε. In principle, there can be heterogeneous parameters for each regime, activity, year, and tax bracket. We limit ourselves to the period from 2006 to 2012, since, as we will show below, there is learning that may prevent agents from immediately adjusting to the new incentives. We make the following assumptions, which allow us to limit the number of parameters estimated: (i) the tax hassle costs and the operating costs are allowed to differ by activity and by regime, but are the same across tax brackets, conditional on the same activity type and regime; (ii) the elasticities are allowed to differ by tax brackets and regimes, but are the same across activities, conditional on tax bracket and regime. As a result, we have a vector χ_i of 7 parameters to be estimated for each regime n (where n = simplified or super simplified) with:

$$\chi_n := \left(\varepsilon_{1n}, \varepsilon_{2n}, \varepsilon_{3n}, a_{\text{I\&C Services,n}}, a_{\text{Non Commercial,n}}, c_{\text{I\&C Services,n}}, c_{\text{Non Commercial,n}} \right)$$

where ε_{in} is the elasticity in tax bracket i for regime n, $a_{\text{I\&C Services,n}}$ and $c_{\text{I\&C Services,n}}$ are the hassle and operating costs in the I&C Services activities and $a_{\text{Non Commercial,n}}$ and $c_{\text{Non Commercial,n}}$ are the hassle and operating costs in the Non Commercial activities in regime n.

For each regime, we have $M = 12$ moments (three tax brackets time two activity types and two sets of years during which the policy parameters were constant). The loss function we minimize is $L(\chi_n)$:

$$L(\chi_n) = \sum_{m=1}^{M} \frac{1}{M} \left(\hat{\Delta} y^*_{nkit} - \Delta y^*_{nkit} \right)^2 \tag{10}$$

Reduced Form Elasticity:

We can also compute a reduced form approximation to the structural elasticity, which does not require any structural assumptions. To do so, we relate the revenue response Δy^* to the change in the implicit marginal tax rate between the notch point y^* and the last bunching point $y^* + \Delta y^*$. For any

notch, define the implicit average tax between y^* and $y^* + \Delta y^*$ to be t^*:[39]

$$t^* = \frac{T(y^* + \Delta y^*) - T(y^*)}{\Delta y^*}$$

Which we can approximate to a first order by:

$$t^* \approx t + \frac{\Delta t}{\Delta y^*} y^* + \frac{\Delta T}{\Delta y^*}$$

The reduced form elasticity is then defined as the percent change in revenues divided by the percent change in the tax:

$$\varepsilon_R \equiv \frac{\Delta y^*}{y^*} \frac{1 - t^*}{\Delta t^*} = \frac{\Delta y^*}{y^*} \frac{1 - t - \frac{\Delta t}{\Delta y^*} y^* - \frac{\Delta T}{\Delta y^*}}{\frac{\Delta t}{\Delta y^*} y^* + \frac{\Delta T}{\Delta y^*}} \tag{11}$$

5.2 Structural Estimation Results

Case 1: Income elasticities if agents do not care about tax simplicity

Table 9 shows the elasticity estimates for the first limit case, which assumes that agents do not value tax simplicity at all (i.e., hassle costs $a = 0$).[40] We split respondents into groups defined by the policy parameters τ^y, τ^{ss}, and μ. Thus, the table shows the elasticities separately for the I&C Services (with a rebate of 50%) and the Non Commercial Activities (with a rebate of 34%) and by tax brackets. The elasticities averaged across all tax brackets are reported under "All." Panels A and B report, respectively, the results for the simplified and the super simplified regimes eligibility thresholds.

Column 2 reports the operating costs c, expressed as a fraction of the rebate μ. For comparability, we use the value of the costs that we estimate in the full structural estimation below (Table 10). Column 4 shows the size of the revenue response Δy^* in euros; column 5 the jump in the average tax rate Δt^*. The earnings responses range from 2.4% to 8.1% of the threshold revenues (from 730 euros to 2420 euros) for the simplified regime and from 10.8% to 11.5% (from 3,460 to 3,700 euros) for the super simplified regime. These are very sizable fractions of total self-employed revenues. The earnings response is precisely estimated and significant, as was already suggested by the graphical evidence showing large and sharp excess masses.

The structural elasticities in column 7 are, on the other hand, not very large in either regime. For the simplified regime, they range from 0.04 to 0.18 for the I&C Services and from 0.05 to 0.17 for the Non Commercial depending on the tax bracket. The elasticities for the super simplified regime are larger for the lower tax brackets, but on average extremely similar to the ones estimated for the simplified regime (0.12 for the I&C Services and 0.08 for the Non commercial activities). All

---

[39]This calculation nests the special case of proportional notches in Kleven and Waseem (2013) and the case of pure notches, i.e., changes in average tax with no change in the marginal tax rate.

[40]We assume symmetric operating costs for agents, regardless of their regime, so that $c_r = c_m$.

elasticities are very precisely estimated. The fact that the primitive structural elasticities estimated from two different eligibility thresholds and regimes (which differ in their monetary incentives and simplicity) are the same is very nice and reassuring. These results highlight the well-understood fact that notches can cause distortions even when the underlying elasticities are not very large. The reduced form elasticities in column 6 are consistently about two times larger than the structural ones for all regimes, tax brackets, and activities.

In Kleven and Waseem (2013), the structural elasticities estimated from notches in the income tax schedule for the self-employed in Pakistan range from 0.025 to 1.02 across the income distribution, depending on the estimation methods, and taking into account optimization frictions. In Saez (2010), the earnings elasticity estimates of the self-employed at the first EITC kink range from 0.7 to 1.6 for those with one child and from 0.38 to 0.95 for those with two children depending on the bandwidth and period under consideration. Thus, the elasticities we find in this first polar case in which people do not care about tax simplicity are on average within the (wide) ranges found in the previous literature, perhaps somewhat on the lower side. In fact, these structural elasticities could be much larger if there were optimization frictions. If a fraction f of agents cannot adjust their income due to optimization frictions, the structural elasticities in Table 9 would be inflated by a factor of $1/(1-f)$.

Case 2: Tax hassle costs when agents do not pay attention to monetary incentives

The second interesting case is the one in which agents do not understand or do not pay attention to monetary incentives, but rather exclusively focus on the tax simplicity of each regime. By assuming that the excess masses observed are generated exclusively by a taste for tax simplicity, we can compute an upper bound on the tax hassle costs. For the I&C Service Activities, the upper bounds on tax hassle costs are, respectively, 240 euros, 390 euros and 600 euros for tax brackets 1, 2, and 3, respectively. If we use an estimate of the net, post tax median hourly wage for the self-employed,[41] these hassle costs are equivalent to about, respectively, 24, 39, and 60 hours of work. For the Non Commercial Activities, the corresponding upper bounds are 42, 54, and 60 hours of work (420, 536, and 600 euros). This is a significant number of work hours. An increase in tax simplicity could thus entail the same behavioral response as a sizable increase in monetary incentives. Yet, the tax simplicity may be easier to understand and thus more salient to agents than the monetary incentives (we develop this point further in Section 6).

Full structural estimation results

Table 10 provides the results from the full structural estimation. The tax hassle costs turn out to be on the higher side, while the elasticities are on the lower side. For the I&C services, the estimated hassle costs are 315 euros for the simplified regime and 162 euros for the super simplified regime. For the

_____

[41]The median hourly net of tax wage number of around 10 euros per hour come from Insee Premiere: "Salaires dans le secteur prive et les entreprises publiques," issue 1565, September 2015.

Non Commercial activities, the tax hassle costs are much larger: 460 and 650 euros respectively. The elasticities range from 0.01 to 0.06 across tax brackets for the simplified regime eligibility threshold and, very similarly, from 0.08 to 0.01 for the super simplified regime. In Appendix Figure A5, we provide a sensitivity analysis by showing the elasticity estimates for alternative values of the operating costs in the feasible ranges.[42] The full structural estimation results confirm that individuals may mostly be attracted by the tax simplicity, rather than by the monetary tax incentives, which they may not understand that well.

Heterogeneous Hassle Costs

We saw in Section 4 that agents bunch to different extents depending on whether they have additional sources of income. Panel C of Table 10 shows how the excess masses for these different groups are mapped into group-specific tax hassle costs and elasticities. Agents who have additional income (salaried or pension income) have much higher preferences for tax simplicity, or, equivalently, higher disutility from dealing with tax hassle. Their tax elasticities are somewhat larger as well, but it is really the tax hassle costs that account for the bulk of the difference in bunching. It makes sense that since these agents have better outside options, they will only engage in a self-self-employed activity if it does not cost them much extra time and hassle.

6   Tax Complexity and Learning

The tax simplicity of a given regime is highly valued by agents, but the overall tax system is complex. In this section, we show that tax complexity has a cost: many agents make suboptimal choices. First, we quantify the cost of making the wrong regime choice and show that it is regressive: it is the lower skilled, non-educated and lower income agents who make the most mistakes and leave most money on the table. Second, we document a more general pattern of slow adjustment following policy changes and the introduction of new features, which we argue is due to the cost of understanding the (complexity of the) tax system.

6.1   Wrong Regime Choice and Learning

We start by checking whether agents make the correct regime choice given their tax bracket, activity and income. We focus on the choice between the simplified and super simplified regimes because it is straightforward and based purely on the comparison of the flat rate τ_f and the total tax rate (income tax plus social insurance contributions, accounting for the rebate μ) facing the agent. In fact, we can

---

[42]The figure plots the structural and reduced-form elasticities as functions of the operating costs in the feasible range. Higher operating costs relative to the rebate mean that the tax change Δt is smaller and thus a larger elasticity is needed to rationalize the observed bunching. This is especially true for high tax bracket agents, since a higher income tax rate τ^y magnifies any cost advantage relative to the rebate. Nevertheless, the elasticity is relatively stable in operating costs, except if the operating cost is at the very high end of the feasible interval.

TABLE 4: Financial Loss from Not Choosing the Super Simplified Regime

Tax bracket/ Activity	I&C Retail ($\mu = 0.71$) ($\tau_f = 13\%$)	I&C Services ($\mu = 0.5$) ($\tau_f = 23\%$)	Non Commercial ($\mu = 0.34$) ($\tau_f = 20.5\%$)
Tax bracket 1	2%	3%	3%
Tax bracket 2	3%	4%	4%
Tax bracket 3	4%	6%	7%
Tax bracket 4	6%	10%	12%
Tax bracket 5	9%	16%	19%

Notes: The table shows the implicit loss as a fraction of revenues from choosing the simplified regime or the super simplified one for eligible agents in the different activities and tax brackets.

show that the simplified regime is strictly dominated from a financial point of view for all eligible agents. Table 4 shows the implicit financial loss from not choosing the super simplified regime over the simplified regime for different tax brackets and different types of activities (thus, different rebates μ). The financial advantage is very large for agents in high tax brackets or in an activity with lower rebate μ. For instance, for agents in the third tax bracket in I&C Services it amounts to 16% of revenues. In addition, the super simplified regime has lower tax hassle costs than the simplified one, which should further reinforce its desirability.

However, Figure 13 highlights that, puzzlingly, not all eligible agents choose the super simplified regime over the simplified one. It shows that the fraction of agents who choose the super simplified regime, conditional on choosing *any* simplified regime and being eligible for the super simplified regime is far from 100% and is not constant across revenue levels. The only plausible explanation for the observed bunching at the threshold here is selection: agents who best understand that the correct choice of regime is the super simplified one may also be the ones with the highest revenues who locate right below the threshold and take full advantage of this regime. The figure also plots the fraction making the correct regime choice by tax brackets. The fraction in the super simplified regime among all eligible ones in a simplified regime increases at the threshold especially for agents in higher brackets. Agents in higher tax brackets have the biggest financial loss from not choosing the super simplified regime and those who understand the regime choice best seem to locate more sharply at the threshold.

Slow Responses to the New Regime and Dominated Behavior

The fact that not all eligible agents choose the super simplified regime is indicative of a lack of understanding of the tax incentives. We now look at how the choice of regime has evolved over time. Recall that the super simplified regime was introduced in 2008. Panels (a) and (b) of Figure 14 shows that the adjustment to the 2008 reform happened only very gradually, which is consistent with agents having trouble initially understanding the monetary incentives and the tax simplicity of that regime.

Panels (a) and (c) show the number of agents in the standard, simplified, and super simplified regime, as well as the total of all self-employed ("All"). The numbers are normalized, respectively, by their values in 2004 and 1994.

Even though the financial losses from remaining in the simplified regime once the super simplified is available are large and are recurring year by year (as was shown in Table 4), we still see in the data that the fraction of agents who are in the simplified regime despite it being dominated comes from both incumbents in the simplified regime who do not make the switch to the newly created regime and from continued new entry into the dominated simplified regime. Panel (c) shows that the entry into the simplified regime continues steadily even after 2008. If agents were aware of the financial payoff associated with the simplified and super simplified regimes, they should no longer enter into the simplified one.

We also see (Appendix Figure A6) that very few agents (less than 10%) switch from the simplified to the super simplified regime, which is again indicative of inertia in the regime choice presumably stemming from slow learning about new opportunities. This is despite the fact that the switch between the super simplified and the simplified regimes in 2009 was made as easy as possible, with a very light administrative procedure, which makes is unlikely that there are switching costs preventing incumbents from making the right regime choice.[43]

That agents value tax simplicity is once again reflected in the significant growth in the number of entrepreneurs in the super simplified regime, the shift away from the simplified regime, and in the overall entry into self-employed after the reform.[44] The reform increased the tax simplicity of becoming self-employed and thus affected the trade-off between becoming self-employed or remaining non self-employed. Panel (d) shows that the creation of the super simplified regime seems to have fostered the entry of much smaller businesses, and pushed down average revenues per business for businesses in the super simplified, and overall.

The time that new entrants or incumbents need to learn about the new regime, and the resulting slow responses, stand in contrast to the rapid effects of the 1999 reform, studied below, in which an *existing* regime that people were already familiar with was expanded, and in which the learning burden was smaller.

Which agents make the wrong regime choice?

Tax complexity is regressive if it is especially lower income and less educated agents who are unable to grasp the complexities and make the correct tax choices. Note that there are no tax preparation softwares in France and that paid tax preparers are rare and typically only used by relatively high

---

[43]The administrative procedure for moving from the simplified regime to the super simplified regime following the 2008 reform, is particularly light: all that is required is for the individual to fill a form by December 31st of year $t-1$ to qualify for entry into the super simplified regime in year t. To make the transition even easier, in 2009 (the year when the super simplified regime was introduced) individuals were granted extra time until March 31st of year t.

[44]The reason why we merge the simplified and the super simplified regimes in the figure, is simply that the super simplified regime did not exist prior to the 2008 reform.

TABLE 5: Fraction of agents choosing the correct (dominant) regime

Non-educated	22.1%	Educated	31.5%
Low skill	28.7%	High skill	34.3%
Old	27.2%	Young	37.3%
Does not claim social insurance benefits	25.7%	Claims social insurance benefits	33.8%
Does not claim UI benefits	29.3 %	Claims UI benefits	37.0%
Low standard of living	29.0 %	High standard of living	39.4%

Notes: The table shows the fraction of agents that choose the super simplified regime among all those eligible and who are in either the super simplified or the simplified regime. The proportions in each row are shown for mutually exclusive and exhaustive groups. "Educated" characterizes agents that have a least a professional high school level equivalent or training degree (*CAP et BEP*). "High skill" refers to agents that are licensed professionals, teachers, engineers or executives of the public service or private sector. "Young" refers to agents in the bottom 25% of the age distribution among the self-employed. "Claims social insurance benefits" (respectively, "claims UI benefits") means that an agent receives some kind of social insurance (respectively, unemployment insurance) benefits. "High standard of living" characterizes agents whose standard of living (disposable income) is in the top 5% of the self-employed distribution. For a more detailed definition and discussion of these variables, see Section 3.2.

income households. Table 5 digs deeper into the characteristics of agents who choose the correct, dominant super simplified regime over the simplified regime. The goal is to see whether agents who could be labeled "more informed" or "more knowledgeable" are more likely to make the correct regime choice. Each row shows the percentage of agents within a given socio-economic group who choose the super simplified regime among those who are eligible for the super simplified and currently in either the simplified or the super simplified. The socio-economic characteristics are defined as in Section 3.2.[45] Agents who are more educated and who are in high-skilled occupations are more likely to make the right regime choice, conditional on being eligible. The same holds true for agents in the top 5% of the standard of living (i.e., adult equivalent disposable income at the household level).

Younger agents are more likely to make the correct choice. It may be because they are better informed about the simpler regimes through their use of the internet, where a lot of information can be found e.g. on discussion forums about self-employment or blogs with tips on how to start one's own business. It may appear counterintuitive that those who claim social insurance (SI) benefits (as well as those who claim unemployment insurance (UI) benefits) are more prone to picking the correct regime. However, this is likely because they are in contact with social or unemployment insurance centers and workers, which routinely send targeted information about self-employment opportunities and the "best practices" explained.

Who learns faster?

Panel (a) of Figure 15 shows the fraction of self-employed who choose the super simplified regime (among those who are eligible for the super simplified regime and who are in either the simplified or

---

[45]All agents considered here are eligible for the super simplified. The fraction eligible itself varies with the socio-economic characteristics. We consider what fraction make the right regime choice, conditional on being eligible.

the super simplified) split into a "low" and a "high" tax bracket: Those in higher tax brackets are typically more likely to make the correct (dominant) regime choice and to do so faster. This may be because –as Table 4 showed– the loss from not being in the super simplified regime increases with an agent's income tax bracket. It may also be that agents in higher tax brackets, who have higher income are generally higher ability or can hire tax professionals.[46]

Recall also that we showed in Table 5 that some group of agents are more likely to be informed and knowledgeable about policies. In panels (b) through (d) of Figure 15, we present the dynamic counterpart to these findings, i.e., the evolution of the regime choices of these groups over time. More knowledgeable groups (the younger ones, the high skilled ones, and the educated ones) are not only more likely to make the correct (dominant) regime choice, but they also converge faster to it. It is likely that these groups have lower learning costs, which explains their faster adjustment.

6.2   Additional Evidence of Costly Learning

To complement the study of the (wrong) regime choice decisions and their evolution over time, we now provide additional evidence of costly learning about tax complexity by agents.

Increasing Bunching over Time

We start by studying what happens to bunching when thresholds change over time. To do so, we split the sample period into four sub-periods. We omit the years from 1994 to 1998, which corresponded to a period with very low thresholds (88,000 French francs or 17,000 constant 2012 euros for I&C Services and for I&C Retail) that were excessively constraining and only made the simplified regime appealing to very small businesses. The major expansion of the thresholds in 1999 made the simplified regime more widely applicable and more attractive to a range of businesses. During the subsequent years, there was a change in threshold only in 2002; the threshold remained constant thereafter until 2008. The rebate also changed in 2002 and 2006. In 2008, the super simplified regime was introduced. It is thus natural to split the 1999-2012 period into the following sub-periods within which thresholds and rebates are constant: (i) 1999-2001, (ii) 2002-2005, (iii) 2006-2008, (iv) 2009-2012.

Figure 16 shows the bunching mass in each of these periods at the eligibility threshold of the simplified regime. The response progressively grows over time until the introduction of the super simplified regime, with excess masses of, respectively, 101%, 140%, and 161% of the counterfactual densities before the threshold. In the final period, after the introduction of the super simplified regime, as explained above, the simplified regime becomes dominated – as a result, bunching does diminish (but remains high at 122% and significant).

Table 11 maps these increasing excess masses for the three periods into corresponding income elasticities (in case 1, i.e., assuming no tax hassle costs). The elasticities are low and insignificant in

---

[46]Although, importantly, there is no perfect correlation between income and tax bracket because of how the tax liability is computed as explained in Section 2.

the first period (point estimate of 0.14), but then grow monotonically over time and become significant. They reach 0.36 in 2006-2008 for the I&C Services and 0.40 for the Non Commercial activities.

The same increased bunching over time also occurs in the super simplified after its introduction. Figure 17 shows the year-by-year distribution of income in the super simplified regime in the vicinity of the eligibility threshold. In 2009, the year after the introduction, the distribution is very noisy, with only a very small visible extra mass right before the threshold. Over time, that excess mass increases. In 2012, it is large and visible, indicating that agents understand the eligibility threshold better.

Bunching at the Old Thresholds

We found that bunching increases over time as agents understand the regimes better. We can also show that when the thresholds change form year to year, some agents fail to adjust and remain at the old, no longer applicable threshold. In Figure 18, the solid vertical red line shows the actual threshold applicable in that fiscal year, the solid black vertical line refers to the threshold for year $t-1$, and the dotted vertical black line refers to the threshold of year $t-2$. It shows that, when the thresholds change by a relatively small amount, bunching still happens at the old, no longer applicable threshold.

Fast Responses to the Expansion of an Existing Regime

Contrary to the 2008 reform studied above, which introduced the new super simplified regime, the 1999 reform merely expanded the already existing simplified regime. It multiplied its eligibility threshold by five for I&C Retail activities and by almost two for I&C Services and Non Commercial activities. Figure 14 shows that the effects of the reform were immediate and almost all took place in the very first year: it appears that the incentives were sufficiently clear and large. Since this reform did not change the overall tax simplicity of being self-employed, there was only a very modest increase in the overall number of self-employed: the sharp increase of 60% in the number of agents in the simplified regime in 1999 was counterbalanced by an almost equivalent decline in the number of agents in the standard regime.[47]

The increase in mean income per business (panel (d)) were also immediate in all regimes. Businesses in the simplified regime grew the most. This in turn may reflect, first, an extensive margin effect, whereby it is on average larger businesses which made the switch from the standard to simplified regime, i.e., those who may previously have found the simplified more attractive but were constrained by the threshold limit. Second there could be an intensive margin effect whereby individuals already in the simplified regime are induced to work harder as performing well reduces less the probability of being subsequently excluded from that regime (i.e., of crossing the eligibility threshold). Regardless of the exact source, the effects were instantaneous, which could be due to the fact that agents did not have to learn from scratch about a new regime but rather just keep track of the (large and salient)

---

[47]Both series are normalized by their own 1994 value so the magnitudes on the graph do not allow to check that the increase in one is in fact almost equal to the decline in the other.

expansion in thresholds.

Summary: Costly Learning and Regressivity of Complexity

The evidence presented in this section is consistent with the notion that agents learn over time about complex fiscal incentives and that more informed and more sophisticated agents learn more quickly. Agents may be balancing their heterogeneous cognitive costs of learning against the benefits from it. Rational inattention would occur if the learning cost are not worth incurring given the size of the benefits.

Consistent with the costly learning hypothesis are the larger and faster responses to larger changes in the thresholds (as during the 1999 reform) and the bunching at the old threshold following smaller changes. Larger changes are more salient (i.e., have lower attention costs) and the financial losses from not adapting to them are larger. It is also consistent with our finding that agents who stand to gain more from the new regime (i.e., those in higher tax brackets) tend to catch up faster with their optimization, which indicates an intentional learning effort and a cost of learning that is balanced against its benefits. Finally, it is in line with the fact that more knowledgeable or sophisticated groups (the younger ones, the high skilled ones, and the educated ones) are not only more likely to make the correct, dominant regime choice, but also that they converge faster to that correct choice; indeed, it is likely that these groups have lower learning costs.

One possible competing explanation would be the presence of real economic frictions that prevent agents from quickly changing their business scale to target the thresholds. However, the 1999 reform led to an immediate adjustment, suggesting that there is no fundamental technological or feasibility constraint preventing such a sharp and fast response. There could be fixed costs that justify adjusting one's economic activity only if the policy change is sufficiently big, as it was in 1999. But the fact that more educated and sophisticated individuals catch up faster with policy changes cannot easily be rationalized with this explanation, unless they also systematically face lower adjustment frictions.[48] In addition, while there could in principle be switching costs, which would affect the response to the introduction of the new super simplified regime, they were by design made very low, with barely more than a box to be checked at registration and, as shown earlier, the financial gains from not switching to the new regime when eligible are large, presumably swamping realistic switching costs. The cognitive burden involved in understanding the complex and evolving tax incentives, on the other hand, may have been large, especially on less sophisticated agents.

---

[48]We will also show that agents have considerable flexibility to misreport their income, as evidenced by round number bunching and shifting within the household (as well as by the sharp bunching observed) and misreporting is easier to adjust (see Section 7).

7 Misreporting or Real Responses?

Are the observed bunching responses the result of real adjustments in economic activity or merely due to misreporting? While it is very difficult to test directly for misreporting, we can offer two different pieces of evidence that suggest agents may be misreporting their income at least to some extent.

7.1 Round number bunching

Figure 19 shows the fraction of agents bunching at round numbers each year. We distinguish between self-employed agents in the standard regime (panel 19a), the simplified regime (panel 19b), and the super simplified regime (panel 19c). Before 2001, numbers are expressed in French francs. Starting in 2001, numbers are in euros, the official newly adopted currency. One euro was worth around six francs so that the denominations became smaller after 2001. There are two key findings. First, there is barely any round number bunching in the standard regime. Recall that agents in the standard regime are typically members of Certified Accounting Centers, which play a monitoring and auditing role and limit the scope for misreporting. In contrast, there is a lot of round number bunching for agents in the simplified or super simplified regimes. If we count bunching at all round numbers (from 100 to 10,000), more than 25% or a fifth of agents in the simplified regime and around 20% of those in the super simplified regimes engage in it. Second, there is an interesting change that occurs at the time of the transition to the euro for agents in the simplified regime, but not for agents in the standard regime. There is a drop in bunching at larger denominations (above 1000) and an increase in bunching at small denominations such as 100. This makes sense if one considers the value of the euro relative to the franc. While "rounding down" to the closest 1000 francs might have seem negligible, rounding to the closest 1000 euros would entail a significant amount of misreporting. This change is not seen for agents in the standard regime, which suggests that it is not due to a fundamental change in the way revenues, costs, or prices are adjusted after the introduction of the euro.

7.2 Income Shifting Within the Household

The eligibility thresholds apply at the individual level; if there are two self-employed agents in the same household, each has to remain below the eligibility thresholds in order to be in the simplified or super simplified regime. This provides a potential incentive to shift income within the household, i.e., to attribute some of the self-employed income earned by one person to the other one. Of course, even at the individual level, the mere presence of the eligibility threshold provides some incentive to misreport income. However, it may be easier for a household with two self-employed earners to shift income from the taxable base of one to that of the other than it is for a single self-employed person to misreport. Agents may have a sense that reporting the income in *some* tax form field, even if not the correct one, is less bad than not reporting it at all (i.e., outright evading).[49]

---

[49]Furthermore, if household members actually help each other in their respective self-employed activities, they may even perceive this income shifting as somewhat fair, if not entirely legal.

Figure 20 show the excess mass for two groups of agents: individuals in households of at least two people, but who are the only one self-employed earner in a simplified regime of the household (panel (a)) and individuals in households with two self-employed members in as simplified regime in panel (b). We will call the two household members "partners" for simplicity, although they are not always spouses or civil partners. The latter exhibit significantly larger excess masses, as one may expect if it is easier to hit the threshold by being also able to shift income between the two partners (in addition to all other possible response margins such as real responses and misreporting). There are other possible interpretations, e.g., people in households with two self-employed earners may be better informed about the thresholds and more willing to misreport (individually).

Panels (c) and (d) offer more direct evidence for shifting. They focus on the subpopulation of households with two self-employed members in a simplified regime. Panel (c) plots the revenues of the partner with the smallest self-employed revenues among the two partners against the revenues of the partner with the largest revenues. The revenues of the lower earner increase as the higher earner's revenues start approaching the threshold. Even more striking is the spike in earnings of the lower earner that occurs when the higher earner starts being close to the upper bound of the tolerance region. One possible interpretation of this pattern is that, as the higher earner becomes threatened to no longer be eligible for the simplified regimes, he shifts some of his revenues to his partner.[50]

Panel (d) plots the distribution of the total (sum) of the revenues of the two self-employed household members relative to twice the eligibility threshold, represented by the red vertical line. There is very sharp bunching at exactly twice the eligibility threshold. We compare this to the bunching by "placebo partners." To build placebo partners, we separately take the earners from the households with two self-employed earners and randomly match them to another self-employed earner in a simplified regime. We then plot the sum of the revenues within the placebo partners. There is no bunching at twice the threshold.[51] Again, it may be that agents in households with two self-employed earners just bunch more individually. However, it is also likely that, although eligibility requirements are at the individual level, partners optimize their joint revenues by maybe shifting at least some income between themselves to each exactly hit the thresholds (and by maybe not reporting at all the excess joint revenue).

7.3 Learning and Income Shifting Within the Household

Another piece of evidence in favor of agents learning about and understanding tax complexity is found in the evolution of income shifting within the household. Figure 21 repeats the analysis in Figure 20, panel (b) that showed bunching for individuals in households with two self-employed earner, but split by period. In the early period, there is barely any bunching. In the later period, there is a strong and

---

[50]The argument here relies on the increase in the secondary earner's revenues being sharp enough around the eligibility or tolerance thresholds. Otherwise, it may just be the result of selection or assortative matching.

[51]To check that this pattern is not driven by the fact that households with two earners in a simplified regime are more prone to bunching at the individual threshold in general, we also check that if we form placebo partners by randomly matching only people from within two self-employed earner households, there is no bunching either. See Figure 20.

significant bunching. Panel (c) shows that bunching at twice the threshold for the sum of revenues in two self-employed earners households is much stronger in the later period as well.

8   Conclusion

In this paper, we study the effects of tax incentives and tax simplicity on self-employed agents. We use new French tax returns data from 1994 to 2012, combined with large-scale survey data. France is a good quasi-laboratory to study the simplicity and complexity of tax policies. This is because it has three different fiscal regimes under which the self-employed can operate, which differ in their monetary tax incentives and tax simplicity. The parameters governing these regimes have changed over time, generating valuable policy variation. In addition, thanks to the peculiarities of the French tax and social insurance contributions system, different people face very different tax rates (i.e., monetary incentives), even conditional on the exact same self-employed revenues. Taken together, these variations provide us with many data moments that allow us to structurally estimate taxable income elasticities and the value of tax simplicity.

Self-employed agents sharply respond to the notch created by the eligibility threshold: there are consistently strong and significant excess masses right before the notch. The thresholds have changed extensively over time – sometimes by a little, sometimes by a lot– and the bunching mass follows them. These excess masses and the ensuing revenue responses can be decomposed structurally into elasticities of taxable income and preferences for tax simplicity (or, equivalently, disutility from tax complexity). Our income elasticity estimates are within the ranges found in earlier studies, but on the lower side.[52] We estimate preferences for tax simplicity to be relatively large and important, ranging from 160 to 650 euros per year depending on the activity and regime. Combined with our findings that agents have a hard time understanding complex tax incentives, this suggests that financial incentives may be less salient and thus less important to agents than tax simplicity.

We uncover significant heterogeneity in the behavioral responses. Agents who have other sizable sources of income, such as salaried income or pension income exhibit much stronger bunching. Correspondingly, we also estimate that they have larger preferences for tax simplicity. Importantly, educated agents respond much more to the eligibility thresholds than do non educated agents, presumably because they understand the financial incentives better.

Agents highly value tax simplicity in a regime, but the overall tax system is complex and not easy to understand. This complexity has costs: First, agents are not choosing the correct self-employed regime (i.e., the regime that would yield them the highest payoff) and leave significant money on the table. We show that agents who can be viewed as more informed or knowledgable (e.g., the more educated and higher skilled ones) are more likely to make the correct, dominant regime choice. They also learn faster what the right regime choice is. In this sense, tax complexity is regressive. Second, people need time to learn about the complex tax system in general. We document several pieces of

---

[52]Formally taking into account the presence of optimization frictions would further inflate these elasticities.

evidence for costly learning over time about tax policy features and changes: The bunching responses and corresponding behavioral elasticities grow over time; whenever the thresholds changed by a little, there remains some bunching mass at the old, no longer applicable threshold; individuals' responses to the expansion of an existing regime are much faster than those to the introduction of a completely new regime.

We also find some direct evidence for misreporting of self-employed income, in addition to the suggestive evidence offered by the sharp and moving bunching. There is significant round-number bunching in the simplified regimes, which changed after the introduction of the denomination in euros. None of this holds true for the standard regime, which is subject to more rigorous monitoring by certified accounting centers. We also find that people in households with two self-employed earners bunch more strongly at the individual eligibility threshold, but also bunch jointly (i.e., in terms of the sum of their revenues) at twice the eligibility threshold. They also appear to be shifting income from the higher earner to the lower one, when the former is about to cross the tolerance threshold.

In future work, it would be interesting to study whether tax simplicity improves the chances of success of a self-employed activity. Do self-employed who understand tax incentives better end up doing better even in the long-run? Do they become true "entrepreneurs" and ultimately job creators? In addition, there is value in thinking of policies that can reduce the regressivity of complexity, such as the provision of tax workshops or tax accounting services to the lower educated or lower income agents who appear to be making worse financial choices.

References

Adelino, M., A. Schoar, and F. Severino (2015). House prices, collateral, and self-employment. *Journal of Financial Economics 117*(2), 288–306.

Best, M., J. Cloyne, E. Ilzetzki, and H. Kleven (2015). Interest rates, debt and intertemporal allocation: evidence from notched mortgage contracts in the united kingdom. *Working paper*.

Best, M. and H. Kleven (2016). Housing market responses to transaction taxes: Evidence from notches and stimulus in the uk. *Working paper*.

Best, M. C., A. Brockmeyer, H. J. Kleven, J. Spinnewijn, and M. Waseem (2015). Production versus revenue efficiency with limited tax capacity: theory and evidence from pakistan. *Journal of Political Economy 123*(6), 1311–1355.

Bruce, D. (2000). Effects of the united states tax system on transitions into self-employment. *Labour Economics 7*, 545–574.

Carrillo, P., D. Pomeranz, and M. Singhal (2017). Dodging the taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics 9*(2), 144–164.

Chetty, R., J. N. 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*(2), 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. *The American Economic Review 103*(7), 2683–2721.

Cour des Comptes (2014). Les organismes de gestion agréés, 40 ans après. Technical report.

Cullen, J. and R. Gordon (2006). Tax Reform and Entrepreneurial Activity. *Tax Policy and the Economy 20*, 41–72.

Cullen, J. and R. Gordon (2007). Taxes and Entrepreneurial Risk Taking: Theory and Evidence for the U.S. *Journal of Public Economics 91*(7-8), 1479–1505.

Eissa, N. and H. Hoynes (2004). Taxes and the labor market participation of married couples: the earned income tax credit. *Journal of Public Economics 88*, 1931–1958.

Eissa, N. and H. Hoynes (2006). The hours of work response of married couples: Taxes and the earned income tax credit. In J. Agell and P. B. Sorensen (Eds.), *Tax Policy and Labor Market Performance*. MIT Press.

Fack, G. and C. Landais (2010). Are tax incentives for charitable giving efficient? evidence from france. *American Economic Journal: Economic Policy, 2*(2), 117–141.

Fack, G. and C. Landais (2016). The effect of tax enforcement on tax elasticities: Evidence from charitable contributions in france. *Journal of Public Economics 1330*(1), 23–40.

Feldman, N. and J. Slemrod (2007). Estimating tax noncompliance with evidence from unaudited tax returns. *Economic Journal 17*(518), 327–352.

Garbinti, B., J. Goupille-Lebret, and T. Piketty (2016). Inequality dynamics in france, 1900-2014: Evidence from distributional national accounts (dina).

Garbinti, B., J. Goupille-Lebret, and T. Piketty (2017). Accounting for wealth inequality dynamics: Methods, estimates and simulations for france (1800-2014).

Gelber, A. (2014). Taxation and the Earnings of Husbands and Wives: Evidence from Sweden. *Review of Economics and Statistics 96*(2), 287–305.

Gelber, A. M., D. Jones, and D. W. Sacks (2017). Estimating earnings adjustment frictions: Method and evidence from the social security earnings test. Technical report.

Gelber, A. M., D. Jones, D. W. Sacks, and J. Song (2017). Using kinked budget sets to estimate extensive margin responses: Method and evidence from the social security earnings test. Technical report.

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

Gordon, R. H. and J. Slemrod (2000). Are 'real' responses to taxes simply income shifting between corporate and personal tax bases? In J. Slemrod (Ed.), *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, pp. 240–280. Russell Sage Foundation and Harvard University Press.

Gruber, J. and E. Saez (2002). The elasticity of taxable income: evidence and implications. *Journal of Public Economics, Elsevier 84*(1), 1–32.

Hamilton, B. H. (2000). Does entrepreneurship pay? an empirical analysis of the returns to self-employment. *Journal of Political Economy 108*(3), 604–631.

Hombert, J., A. Schoar, D. Sraer, and D. Thesmar (2017). Can unemployment insurance spur entrepreneurial activity? *National Bureau of Economic Research Working Paper* (20717).

Hoopes, J., J. Slemrod, and D. Reck (2017). Taxpayer search for information: Implications for rational attention. *American Economic Journal: Economic Policy, forthcoming*.

Katz, L. and A. Krueger (2017). The role of unemployment in the rise in alternative work arrangements. *American Economic Review Papers and Proceedings 107*(5), 388–392.

Katz, L. F. and A. B. Krueger (2016). The rise and nature of alternative work arrangements in the united states, 1995-2015. *NBER Working Paper 22667*.

Kleven, H. (2016). Bunching. *Annual Review of Economics 8*, 435–464.

Kleven, H. and M. Waseem (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan. *Quarterly Journal of Economics 128*(2), 669–723.

Landais, C., J. Kolsrud, P. Nilsson, and J. Spinnewijn (2015). The optimal timing of unemployment benefits: Theory and evidence from sweden. *IZA Discussion Paper* (9185).

Landais, C. and J. Spinnewijn (2017). Unemployment, consumption smoothing and the value of insurance. *London School of Economics Working paper*.

Lelarge, C., D. Sraer, and D. Thesmar (2008). Entrepreneurship and credit constraints: Evidence from a french loan guarantee program. In J. Lerner and A. Schoar (Eds.), *Differences in Entrepreneurship*.

Lockwood, B. B. (2016). Optimal income taxation with present bias. *University of Pennsylvania Working Paper*.

Lockwood, B. B. and D. Taubinsky (2016). Regressive sin taxes. *University of Pennsylvania Working Paper*.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review 105*(8), 2539–2569.

Saez, E. (2010). Do taxpayers bunch at kink points ? *American Economic Journal : Economic Policy 2*(3), 180–212.

Saez, E., J. Slemrod, and S. Giertz (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature 50*(1), 3–50.

Schmalz, M. C., D. A. Sraer, and D. Thesmar (2016). Housing collateral and entrepreneurship. *The Journal of Finance*.

Schoar, A. (2010). The divide between subsistence and transformational entrepreneurship. *Innovation Policy and the Economy 10*(1), 57–81.

Slemrod, J. and J. Salleel (2012). Car notches: Strategic automaker responses to fuel economy policy. *Journal of Public Economics 96*(11-12), 981–999.

FIGURE 3: Eligibility Thresholds and Rebate for the Simpler Regimes

(a) Eligibility Thresholds y^*
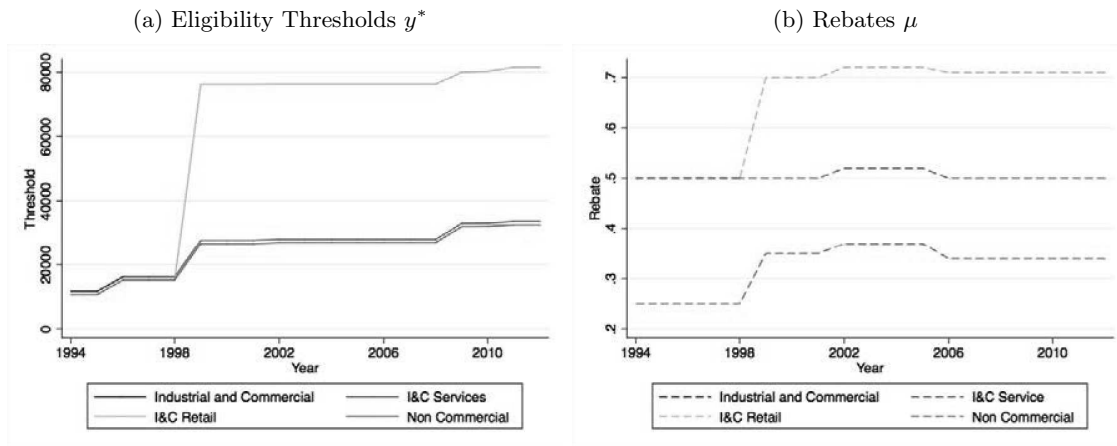
(b) Rebates μ

Notes: Panel 3a shows the thresholds and Panel 3b shows the rebates for the different activity types and over time. Before 1999, the I&C Retail and i&C Services cannot be distinguished in the tax data and are thus grouped together into the "Industrial and Commercial" category.

FIGURE 4: Evolution of Self-Employment 1994-2012
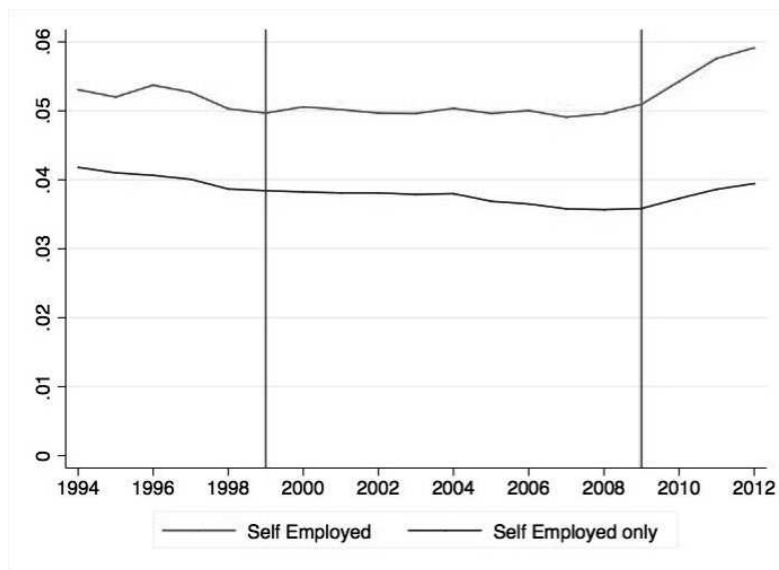
Notes: "Self-Employed" are individuals who earn any self-employment (and may or may not also receive additional wage income). "Self-Employed only" are individuals who only earn self-employed income. The numbers are the fraction of self-employed or self-employed only divided the number of all individuals aged between 18 and 65. The two red vertical lines represent the 1999 and the 2008 reforms presented in the text.

42

FIGURE 5: Regime Choice – Share Choosing the Simplified or Super Simplified Regime

(a) All

(b) By Tax Bracket
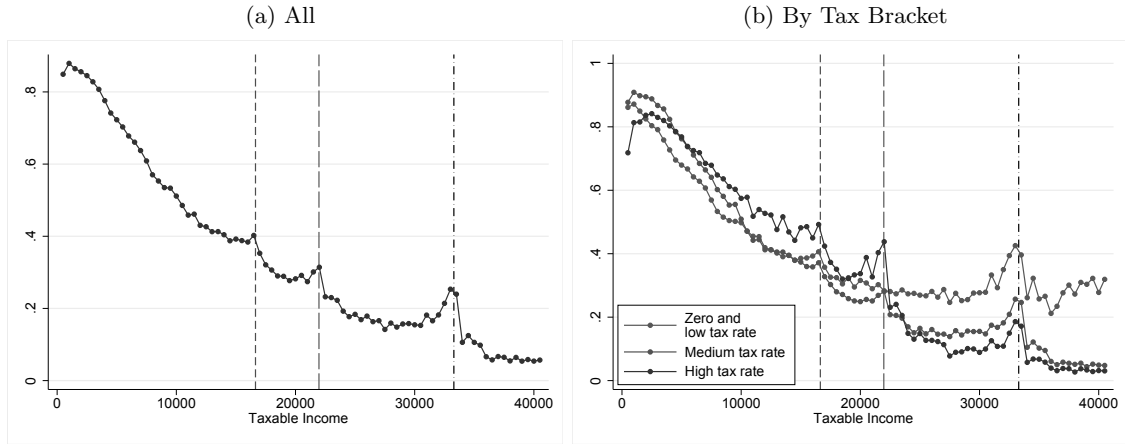
Notes: The figure represents the share of self-employed agents who choose either the simplified or the super simplified regime, based on the full population for 2011. The horizontal axis represents self-employed taxable income z, as defined in the text Section 4.1. The red dashed vertical line represents the eligibility threshold for the I&C Services (converted into a taxable income equivalent, i.e., multiplied by one minus the rebate μ applicable to I&C Services), the green dashed vertical line represents the threshold for the Non Commercial activities (also converted into a taxable income equivalent), and the blue dash-dot vertical line represents the threshold for the super simplified regime for both I&C Services and Non Commercial Activities. Panel (b) splits the sample into the different tax brackets as defined in Section 3.2 and Table 2.

FIGURE 6: Excess Mass Method
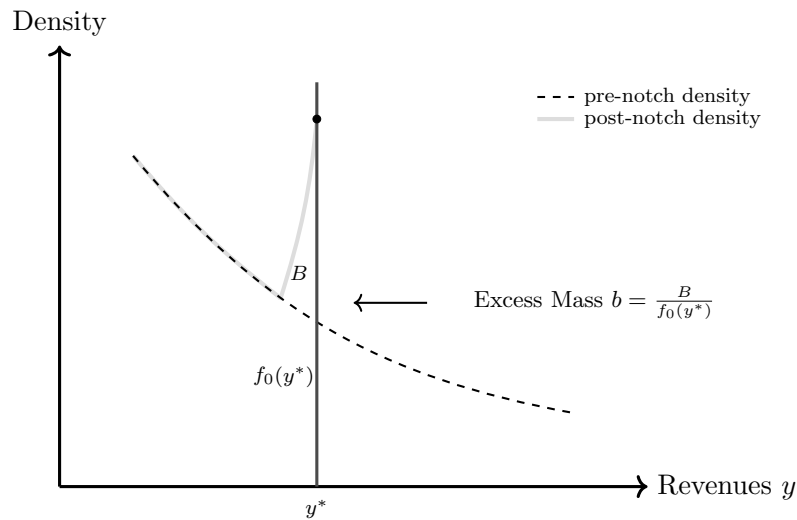
The figure illustrates how we measure the excess mass. The red vertical line represents the eligibility threshold y^*. The excess mass B is computed by taking the difference between the post-notch density (the light blue solid curve) relative to the pre-notch density represented by the dashed blue curve (and captured by a counterfactual smooth fitted polynomial). The excess mass measure is $b = B/f_0(y^*)$ where $f_0(y^*)$ is the counterfactual density right before the threshold.

FIGURE 7: Bunching at the Eligibility Thresholds
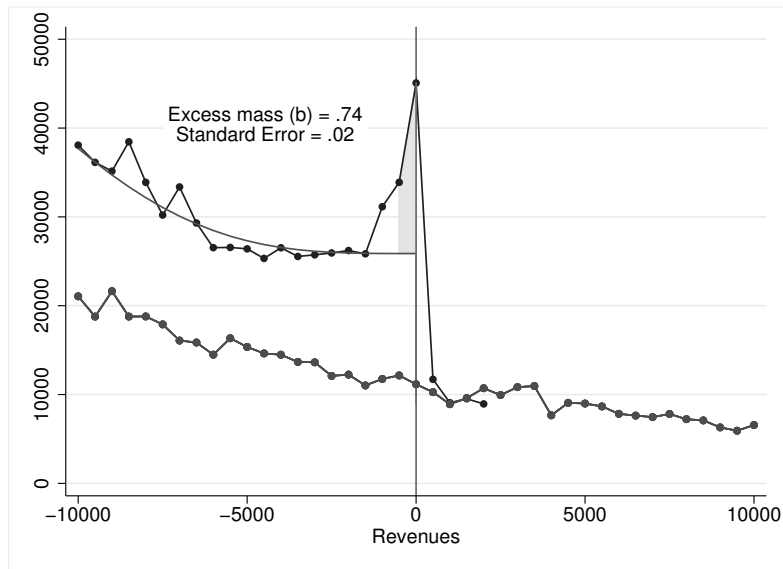
Notes: The figure represents the frequency of revenues, by bins of revenues centered around the eligibility threshold (the red vertical line). We pool data for 1999-2012 and for all agents in the simplified and super simplified regimes. The counterfactual distribution is represented by the red solid curve in the figures and is fitted using a smooth polynomial as explained in Section 4. The estimated excess mass is in yellow. There is significant bunching, equal to 74% of the average counterfactual frequency within 1,500 euros of the notch. The thick red curve serves as a placebo test: it shows the frequency distribution for the I&C Retail activities, centered around the eligibility threshold for the I&C Services and Non Commercial activities (the actual threshold for the I&C Retail activities is higher).

FIGURE 8: Bunching in the Simplified and Super Simplified Regimes

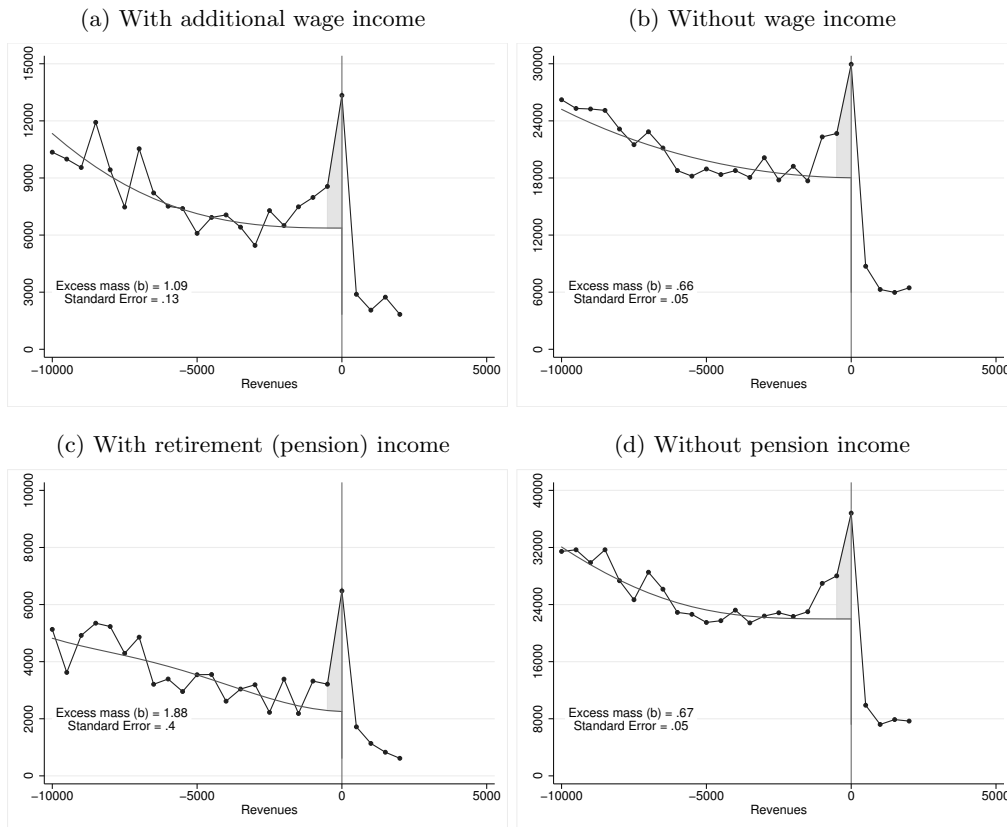(a) Simplified (1999-2008)   (b) Super Simplified (2009-2012)
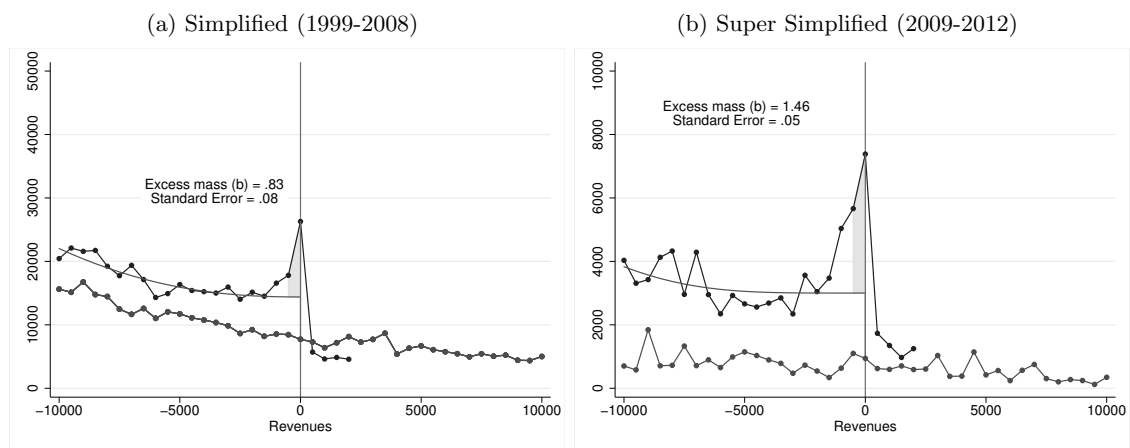
Notes: Panel (a) considers bunching among agents in the simplified regime only. Panel (b) considers agents in the super simplified regime only. See the notes to Figure 7.

44

FIGURE 9: Bunching by Tax Bracket
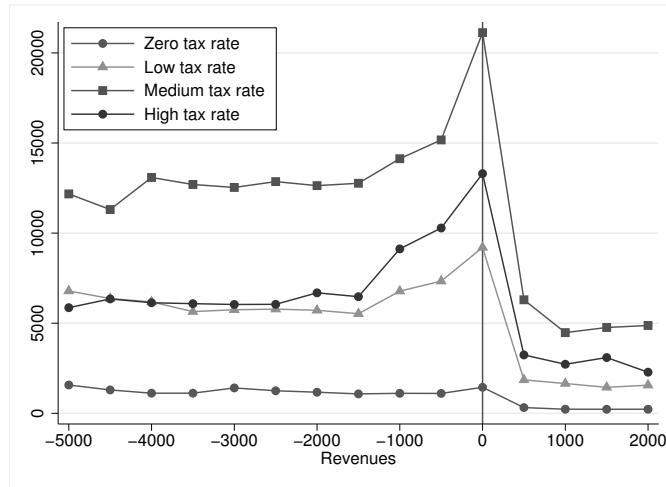
Notes: The figure shows bunching among agents in different tax brackets as defined in Section 3.2 and Table 2. Excess masses by tax bracket and their standard errors are reported in Table 3. See the notes to Figure 7.

FIGURE 10: Bunching for Agents with or without Additional Income Sources

(a) With additional wage income

(b) Without wage income

(c) With retirement (pension) income

(d) Without pension income

Notes: Bunching among different groups of self-employed agents: individuals who earn some wage income in addition to self-employed income (panel (a)) and those who do not (panel (b)), individuals with pension income (panel (c)) and those without (panel (d)). See the notes to Figure 7.

FIGURE 11: Bunching by Education Level
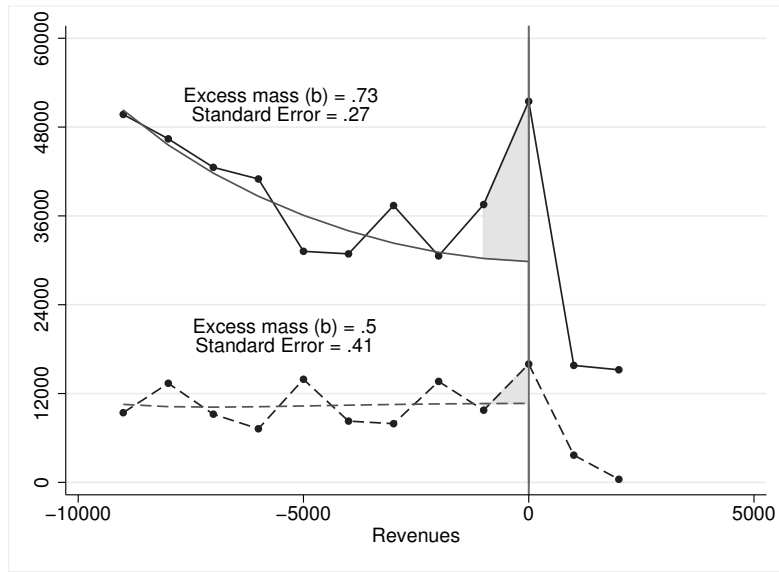
Notes: The figure shows bunching for educated agents (the solid, top curve) and for non educated agents (the dashed, bottom curve). Educated agents are those with at least a high-school training degree, as explained in detail in Section 3.2. See the notes to Figure 7.

FIGURE 12: Income Elasticities and Hassle Costs

(a) I&C Services

(b) Non Commercial
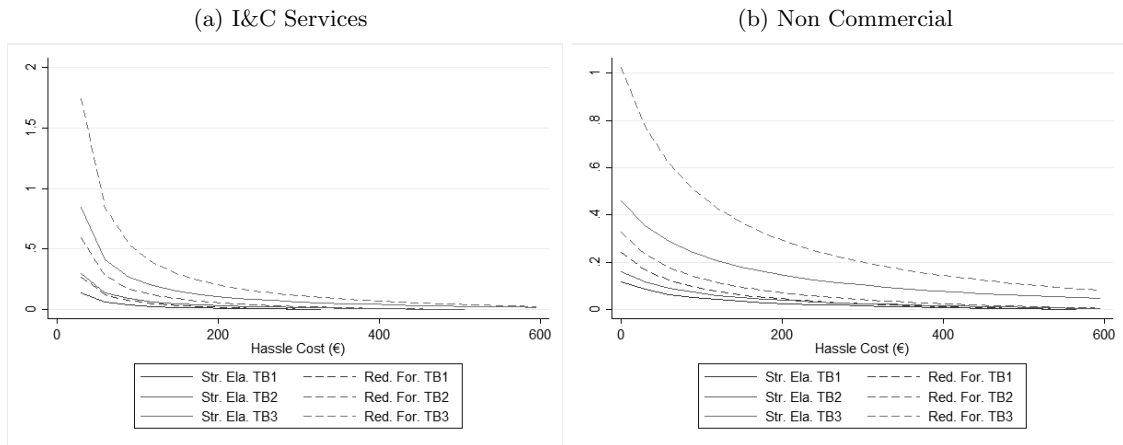
Notes: The figures show the combinations of income elasticities and tax hassle costs that can explain the observed bunching. The values of the structural elasticities (solid curves) and the reduced form elasticities (dashed curves) are shown for different tax groups, as a function of the tax hassle costs. The operating costs are set as in table 9.

46

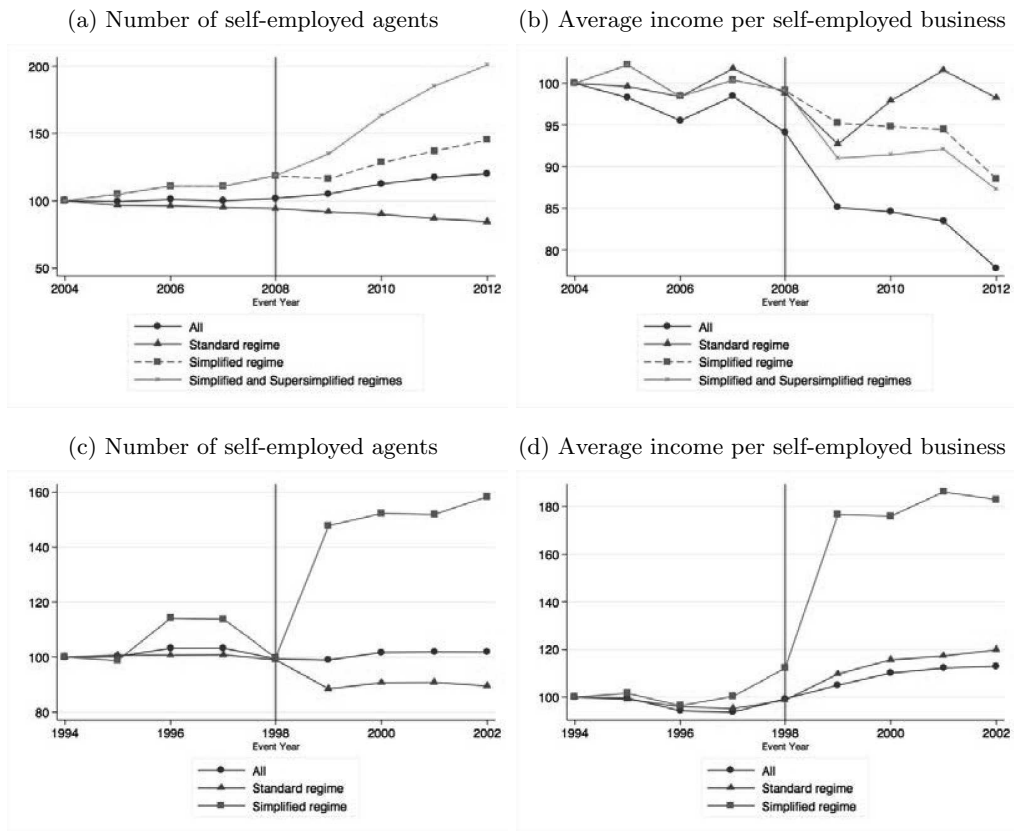FIGURE 13: Regime Choice – Share Choosing the Super Simplified Conditional on Choosing a Simpler Regime
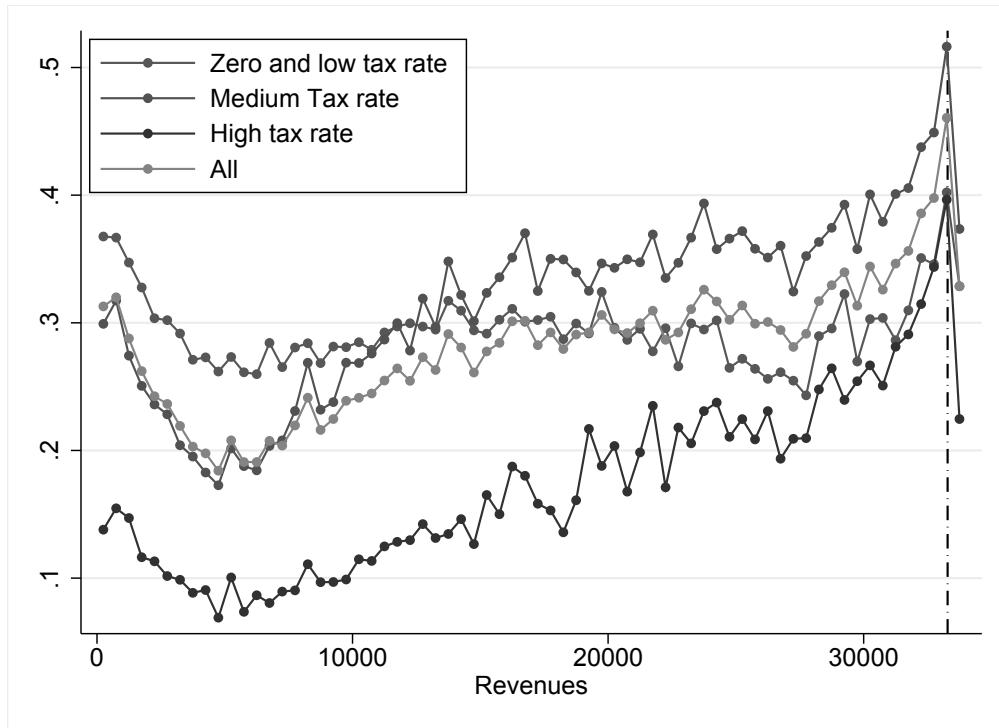
Notes: The figure shows the fraction of agents that choose the super simplified regime among all those eligible and who are in either the super simplified or the simplified regime, at each revenue level, based on the full population data for 2011. The horizontal axis shows self-employed gross revenues. The vertical line is the eligibility threshold for the super simplified regime. The results are shown for different tax brackets as defined in Section 3.2 and Table 2. "All" includes agents in all tax brackets.

FIGURE 14: Event Studies: the 1999 and 2008 Reforms

(a) Number of self-employed agents   (b) Average income per self-employed business

(c) Number of self-employed agents   (d) Average income per self-employed business

Notes: We normalize all variables by their starting level in 1994 for panels (a) and (b) and in 2004 for panels (c) and (d). Panels (a) and (c) plot the number of self-employed agents in each of the standard and simplified regimes, as well as the total number of self-employed. Panels (b) and (d) plot the average income per self-employed business for those in the standard regime (net business income) and those in the simplified regime (gross revenues).

FIGURE 15: Share of Agents Making the Correct Regime Choice

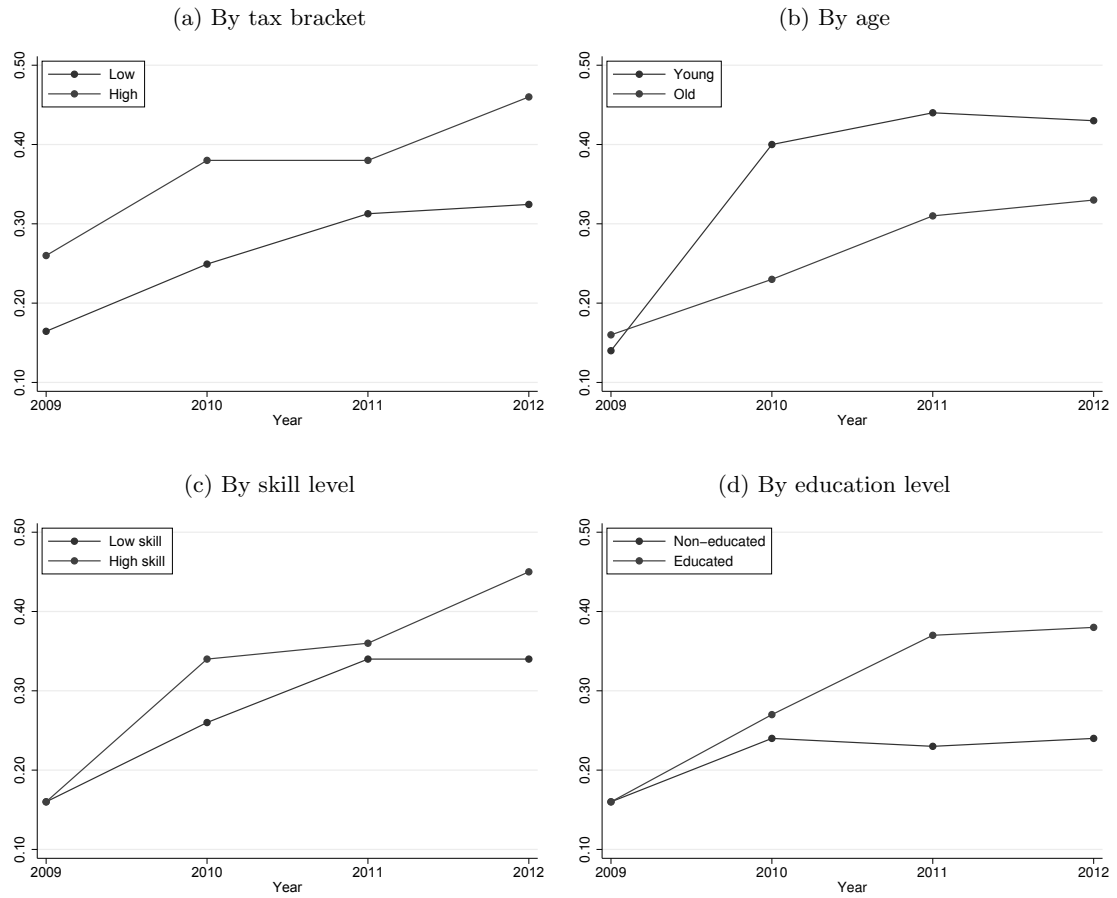(a) By tax bracket

(b) By age

(c) By skill level

(d) By education level

Notes: The figure shows the fraction of agents who choose the super simplified regime among all agents eligible for the super simplified regime and who choose one of the simpler regimes. The sample is split by tax bracket (panel (a)), age (panel (b)), skill level (panel (c)), and education level (panel (d)). The groups are as defined in Section 3.2 and Table 5. The "low" tax bracket encompasses tax brackets 0, 1, and 2 as defined in Table 2.

FIGURE 16: Bunching by Period

(a) 1999-2001

(b) 2002-2005

(c) 2006-2008
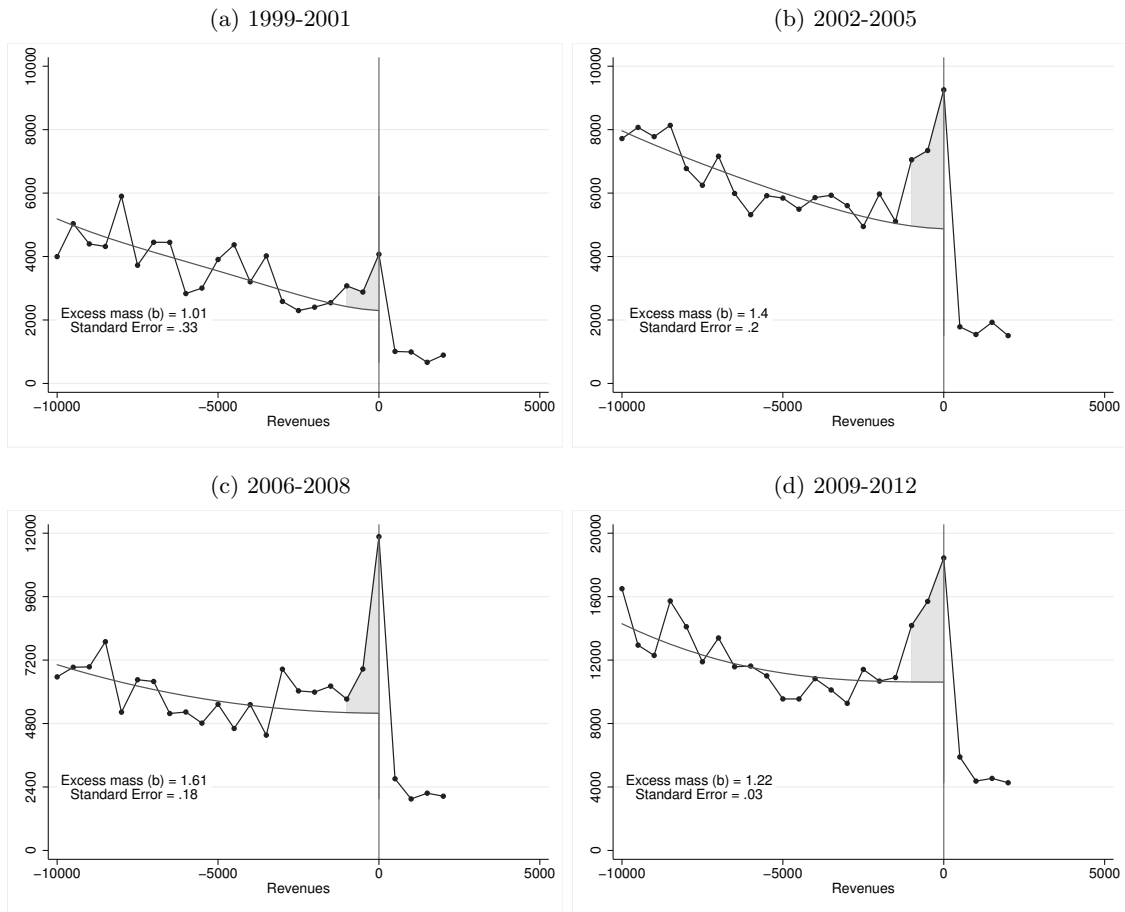
(d) 2009-2012

Notes: The figure shows bunching in four different periods, the choice of which is explained in the text: 1999-2001, 2002-2005, 2006-2008, and 2009-2012. Bunching increases over time until the introduction of the super-simplified regime in 2008. See the notes to Figure 7.

FIGURE 17: Bunching After the Introduction of the Super Simplified Regime
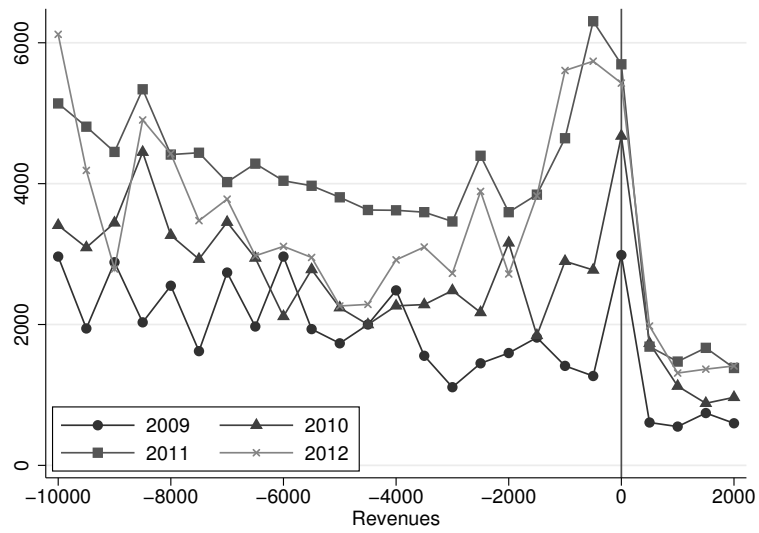
Notes: The figure shows bunching over time by agents in the super simplified regime. See the notes to Figure 7.

FIGURE 18: Bunching at the Old Threshold

(a) 2011

(b) 2012
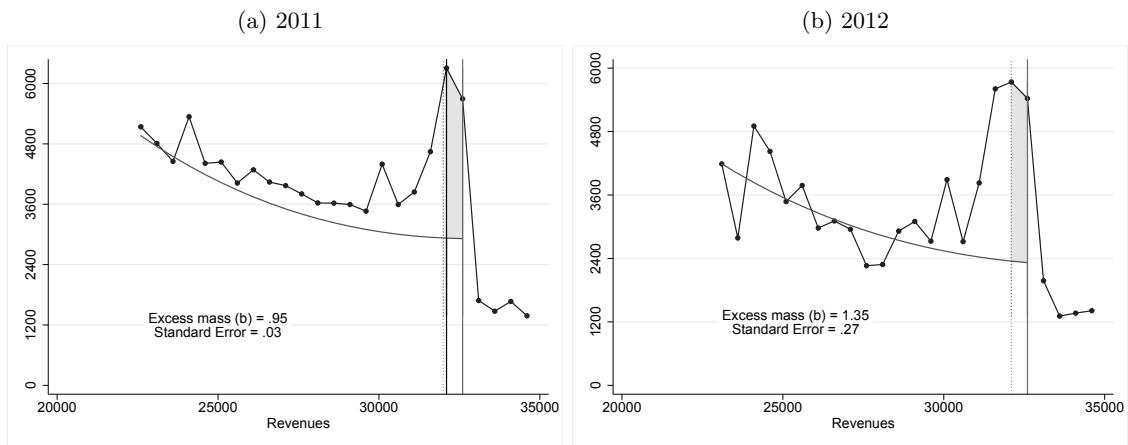
Notes: The red solid vertical line represents the eligibility threshold for year t. The black solid vertical line represents the threshold in year $t-1$. The dashed vertical line represent the threshold for year $t-2$. See the notes to Figure 7.

(a) Standard Regime

(b) Simplified Regime

(c) Super Simplified Regime
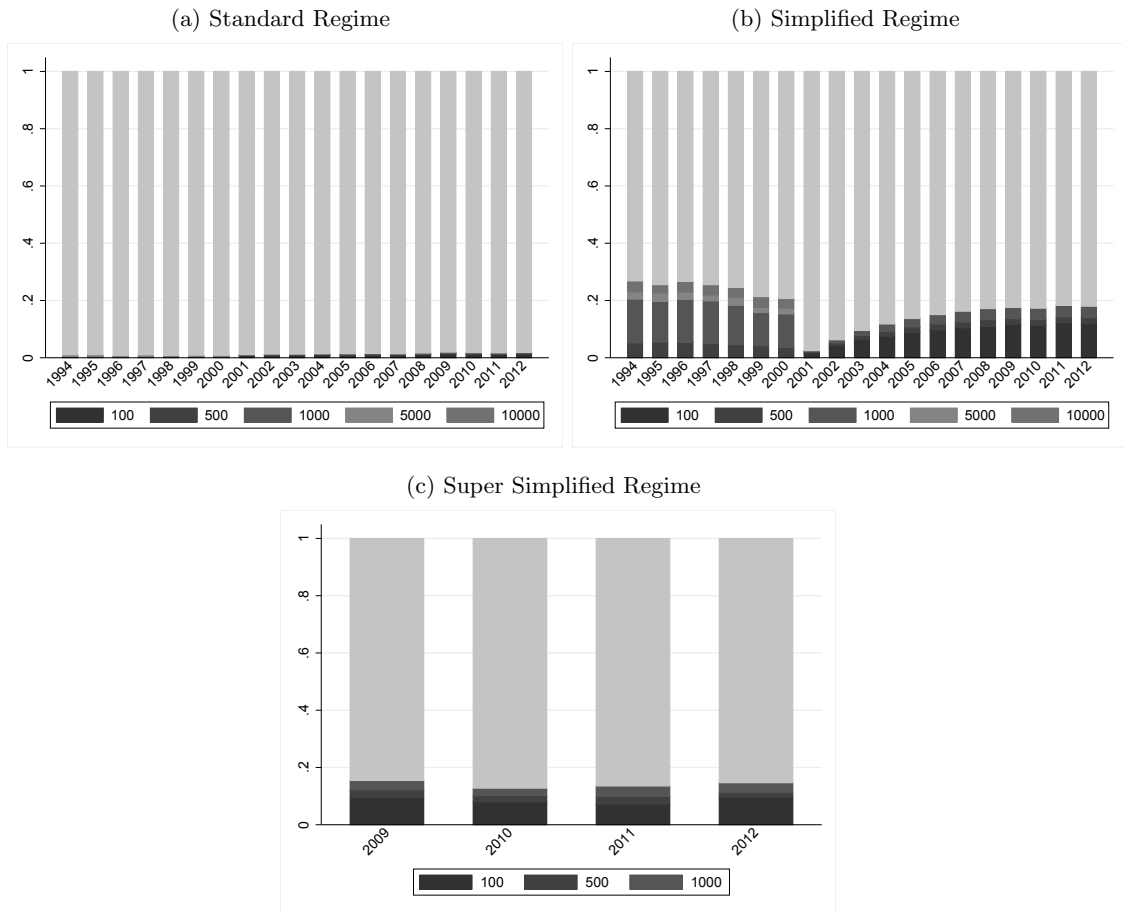
Notes: The figure shows the fraction of agents in each regime that bunches at round numbers. For the period 1994-2000, numbers are expressed in francs. Starting in 2001, numbers are in euros (the new currency applicable after 2001). The conversion rate from francs to euros was around 6.6 francs per euro.

(a) In households with one self-employed agent

(b) In households with two self-employed agents

Excess mass (b) = .78
Standard Error = .06

Excess mass (b) = 3.16
Standard Error = .67

(c) Lower earner's revenues

(d) Sum of revenues

Real couples
Placebo
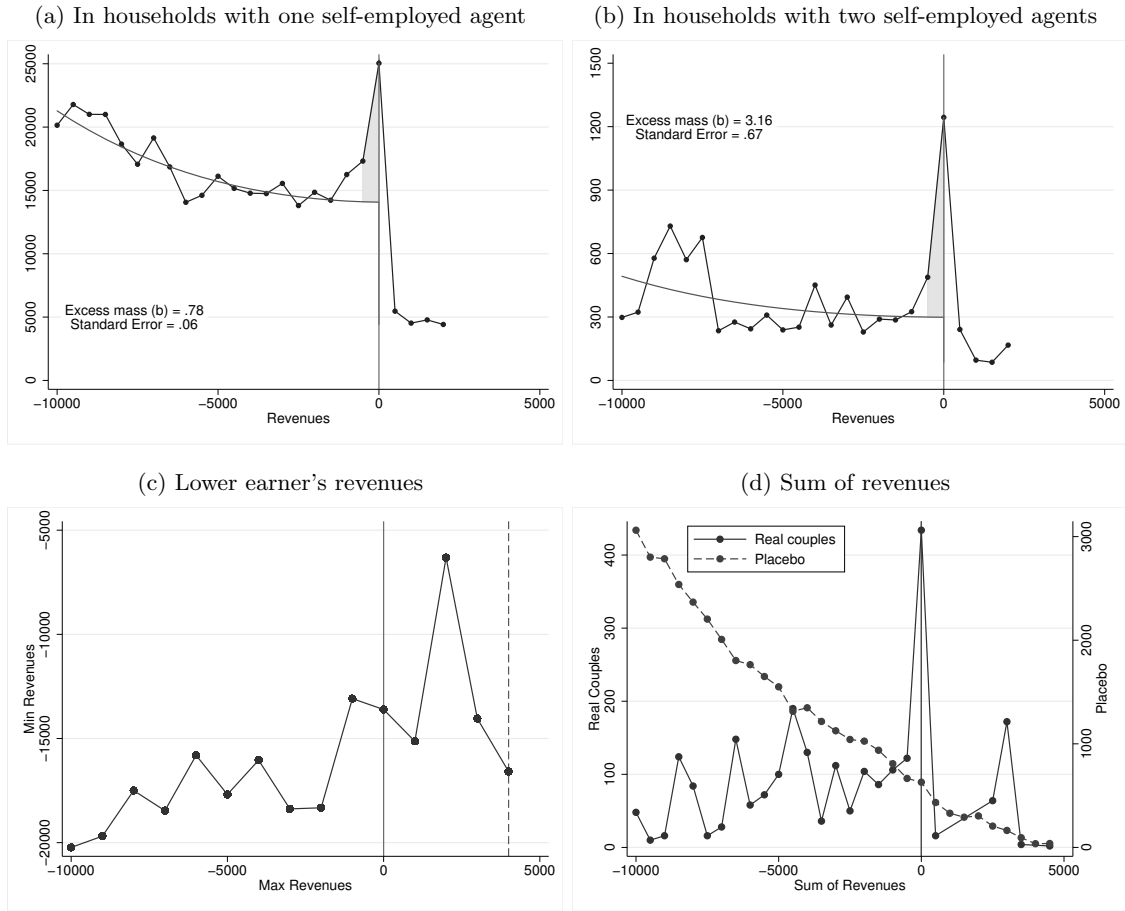
Notes: Panel (a) (respectively, Panel (b)) shows bunching among individuals in households with two (respectively, one) self-employed earner(s) in a simplified regime. Panel (c) shows the centered revenues (relative to the eligibility threshold) of the lower earner in a household with two self-employed earners in a simplified regime against the revenues of the higher earner. Panel (d) plots the sum of revenues of two partners in the same household who are both in a simplified regime; in this panel only, the red vertical line is at twice the individual eligibility threshold. The dotted red line shows the distribution of the sum of revenues of "placebo partners," which are formed by randomly matching each self-employed individual in a simplified regime to another self-employed individual in a simplified regime (from a different household). See also the notes to Figure 7.

FIGURE 21: Learning to Shift Income Within the Household

(a) Early Period 1999-2001

Excess mass (b) = 1.72
Standard Error = .99

(b) Later Period 2002-2008

Excess mass (b) = 3.7
Standard Error = .93

(c) Sum of Revenues in early and late period
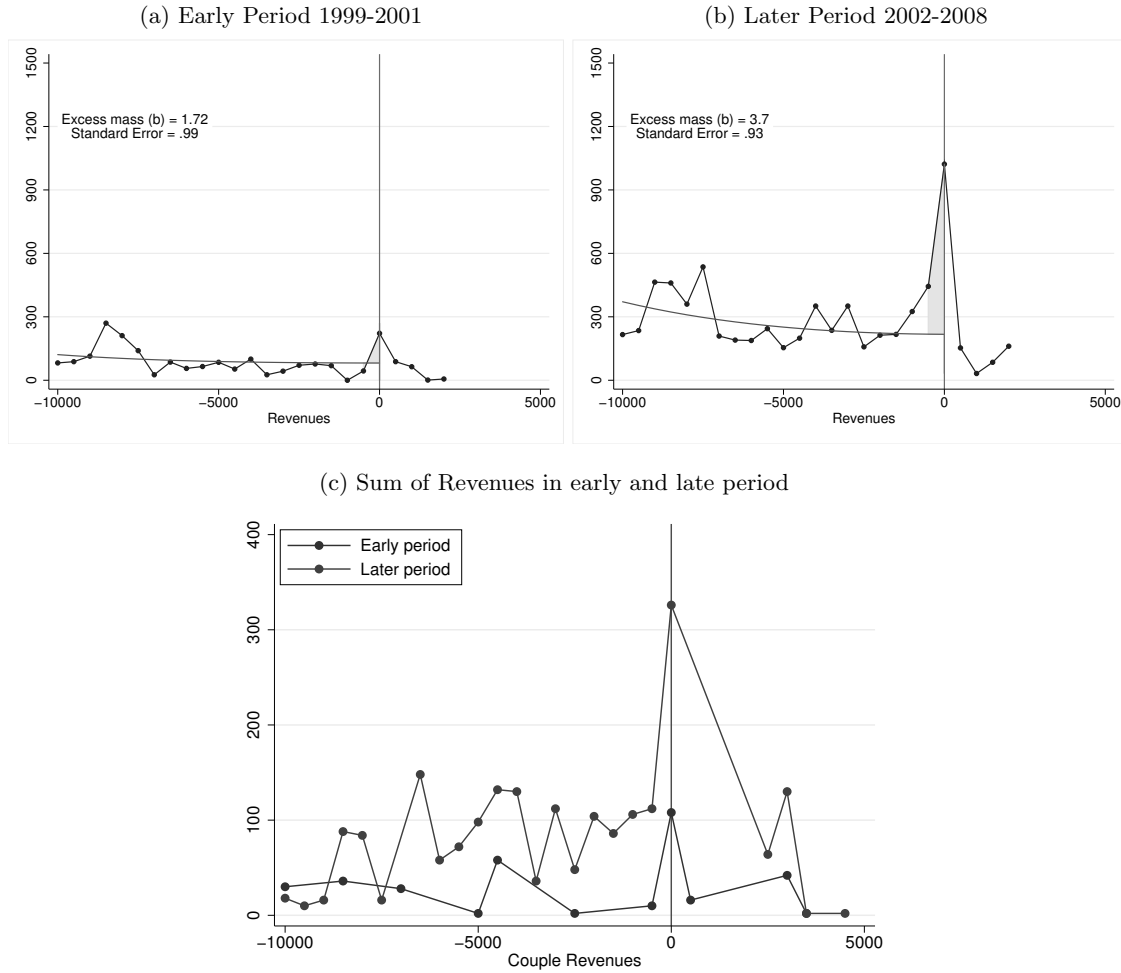
Notes: Panels (a) and (b) repeat panel (a) of Figure 20, splitting the period into two (an early period in panel (a) and a later period in panel (b)). Panel (c) repeats panel (d) from Figure 20, again split by period.

	All	With wage income only	With self-employed income only	With any self-employed income
Panel A				
Age	40	40	49	48
Female	0.47	0.48	0.32	0.33
Married or in Civil Union	0.50	0.49	0.63	0.62
Has any children	0.41	0.41	0.39	0.41
Number of Children	0.71	0.71	0.70	0.72
Retired	0.06	0.06	0.17	0.14
Claimed unemployment benefits	0.11	0.11	0.03	0.05
Claimed any social insurance benefits	0.48	0.48	0.38	0.39
Educated	0.72	0.72	0.73	0.76
Bachelor (at least)	0.15	0.15	0.21	0.24
High skill	0.12	0.11	0.19	0.20
Panel B				
Wage Income	19576	20549	0	6005
Self-employed Income	2004	0	32982	29934
Capital Income	2154	1875	5148	6047
Tax free capital income	1161	1072	2467	2351
Standard of living	42607	41845	50208	53312
Panel C				
Zero tax bracket	0.16	0.16	0.15	0.14
Low tax bracket	0.32	0.33	0.23	0.22
Medium tax bracket	0.38	0.39	0.31	0.32
High tax bracket	0.14	0.13	0.31	0.32
Panel D				
Population (in mill.)	532.7	497	26.3	35.6

Notes: "With wage income only" refers to individuals with only wage income. "With self-employed income only" refers to individuals with self-employed income only. "With any self-employed income" refers to people with any self-employment (who may also have wage income). "All" refers to any individual who earns either some wage income or some self-employed income (or both). All variables in panels A and C are expressed in percent of the full group represented in each column. "Age" is expressed in years. "Number of Children" is the average number of children. Panel B provides average income in each category in constant 2012 euros. Panel D gives the total population in each column for the whole period. See Section 3.2 for the definition of all variables.

	All	Industrial and Commercial (Retail and Service)	Non Commercial
Panel A			
Age	48	49	46
Female	0.33	0.28	0.41
Married or in Civil Union	0.63	0.65	0.59
Has any children	0.41	0.39	0.44
Number of Children	0.73	0.68	0.80
Retired	0.14	0.16	0.11
Claimed unemployement benefits	0.05	0.05	0.05
Claimed any social insurance benefits	0.40	0.39	0.41
Educated	0.76	0.67	0.90
Bachelor (at least)	0.24	0.10	0.49
High skill	0.20	0.08	0.43
Panel B			
Wage Income	6049	5265	7538
Self-employed Income	30505	22718	45376
Capital Income	6133	6040	6552
Tax free capital income	2303	1997	2790
Standard of living	53642	45317	69444
Panel C			
Zero tax bracket	0.13	0.16	0.08
Low tax bracket	0.22	0.26	0.14
Medium tax bracket	0.32	0.34	0.29
High tax bracket	0.33	0.24	0.49
Panel D			
Population (in mill.)	34,7	22.5	12.6

Notes: "Industrial and Commercial (Retail and Services)" refers to the group of agents who have some revenues from I&C Retail or I&C Services activities. "Non Commercial" refers to the group of agents who have some revenues in a Non Commercial Activity. "All" refers to agents who have some self-employed income in at least one of the two aforementioned categories. See the notes to Table 6.

TABLE 8: Self-Employed Earners by Regime

| | 1994-2008 | | 2009-2012 | | |
	Standard	Simplified	Standard	Simplified	Super-Simplified
Panel A					
Age	46	52	48	50	43
Female	0.30	0.34	0.34	0.38	0.37
Married or in Civil Union	0.68	0.58	0.61	0.52	0.44
Has any children	0.47	0.29	0.43	0.30	0.36
Number of Children	0.84	0.50	0.76	0.52	0.62
Retired	0.07	0.31	0.10	0.28	0.13
Claimed unemployment benefits	0.02	0.07	0.03	0.10	0.21
Claimed any social insurance benefits	0.40	0.33	0.41	0.38	0.54
Educated	0.77	0.68	0.83	0.76	0.81
High skill	0.22	0.15	0.26	0.19	0.15
Panel B					
Wage Income	3945	10439	4470	10868	7985
Self-employed Income	39446	11522	40925	11848	10307
Capital Income	5938	6174	7864	6713	2484
Tax free capital income	2294	2452	2464	2611	1032
Standard of living	56814	42434	66278	47553	39086
Panel C					
Zero tax bracket	0.09	0.19	0.08	0.18	0.23
Low tax bracket	0.20	0.28	0.16	0.24	0.27
Medium tax bracket	0.29	0.32	0.37	0.40	0.42
High tax bracket	0.42	0.22	0.39	0.18	0.08
Panel D					
Population (in mill.)	19.3	7.3	4.6	3.1	0.9

Notes: "Standard" refers to agents in the standard regime; "Simplified" to agents in the simplified regime; "Super simplified" to agents in the super simplified regime. See the notes to Table 6.

TABLE 9: Elasticity Estimates if no Preference for Tax Simplicity (Case 1)

Activity Type	Cost (% of rebate)	Tax bracket	Earnings Response Δy^*	ATR Jump Δt^*	Reduced-Form Elasticity e_R	Structural Elasticity e
Panel A – Simplified						
		1	730	0.33	0.07*** (0.018)	0.04*** (0.009)
I&C Services	0.5	2	1,090	0.36	0.14*** (0.021)	0.07*** (0.010)
		3	1,930	0.41	0.39*** (0.062)	0.18*** (0.027)
		All			0.18*** (0.031)	0.09*** (0.015)
		1	1,000	0.70	0.08** (0.038)	0.04** (0.018)
Non Commercial	0.1	2	1,240	0.76	0.10*** (0.017)	0.05*** (0.008)
		3	2,420	0.89	0.36*** (0.040)	0.17*** (0.017)
		All			0.22*** (0.029)	0.10*** (0.013)
Panel B – Super Simplified						
		1	3,460	0.60	0.56*** (0.099)	0.25*** (0.039)
I&C Services	0.3	2-3	3,660	2.30	0.11*** (0.034)	0.05*** (0.014)
		All			0.26*** (0.056)	0.12*** (0.022)
		1	3,000	0.36	1.02** (0.487)	0.45** (0.194)
Non Commercial	0.3	2-3	3,700	2.63	0.12*** (0.015)	0.06*** (0.006)
		All			0.17*** (0.042)	0.08*** (0.018)

Notes: The table shows the structural and reduced form elasticity estimates for agents in the simplified regime for the period 2006-2012. See Section 5. Standard errors are computed using 400 bootstrap iterations and are shown in parentheses. * p= 0.1, ** p=0.05, *** p=0.01.

TABLE 10: Full Structural Estimation Results– Tax Hassle Costs and Elasticities

Cost I&C Services (% of rebate)	Cost Non Commercial (% of rebate)	Hassle Cost I&C Services a_S	Hassle Cost Non Commercial a_{NC}	Tax bracket	Structural Elasticity e
Panel A – Simplified Regime					
				1	0.01
0.5	0.1	315	456	2	0.02
				3	0.06
Panel B – Super Simplified Regime					
				1	0.08
0.3	0.3	162	648	2-3	0.01
Panel C – By Additional Income Sources					
With salaried income					
				1	0.01
0.5	0.2	304	145	2	0.03
				3	0.07
Without salaried income					
				1	0.02
0.5	0.2	149	144	2	0.01
				3	0.04
With pension income					
0.5	0.2	305	580	1-2-3	0.02
Without pension income					
0.5	0.2	150	299	1-2-3	0.01

Notes: The table shows the results from the full structural estimation described in Section 5. For the simplified regime, the period is 2006-2012. For the super simplified, it is 2009-2012. See Section 5 for the computational details. Hassle costs are expressed in euros.

TABLE 11: Elasticity Estimates over Time

Activity Type	Cost (% of revenues)	Period	Earnings Response Δy^*	ATR Jump Δt^*	Reduced-Form Elasticity e_R
	0.52	1999-2001	1020	0.37	0.14 (0.619)
I&C Services	0.56	2002-2005	980	0.29	0.17** (0.085)
	0.52	2006-2008	1470	0.35	0.36** (0.172)
	0.15	1999-2001	1220	0.65	0.14 (0.154)
Non Commercial	0.22	2002-2005	1420	0.67	0.20*** (0.064)
	0.12	2006-2008	1690	0.59	0.40*** (0.126)

Notes: The table shows the structural and reduced form elasticity estimates for agents in the simplified regime, by period, as explained in Section 6. They correspond to case 1 described in Section 5, in which it is assumed that agents do not have a preference for tax simplicity ($a = 0$). Standard errors are computed using 400 bootstrap iterations and are shown in parentheses. * p= 0.1, ** p=0.05, *** p=0.01.

Online Appendix – Not for Publication

A.1 Additional Tables and Figures

FIGURE A1: Income Tax Rates in France 1994-2012
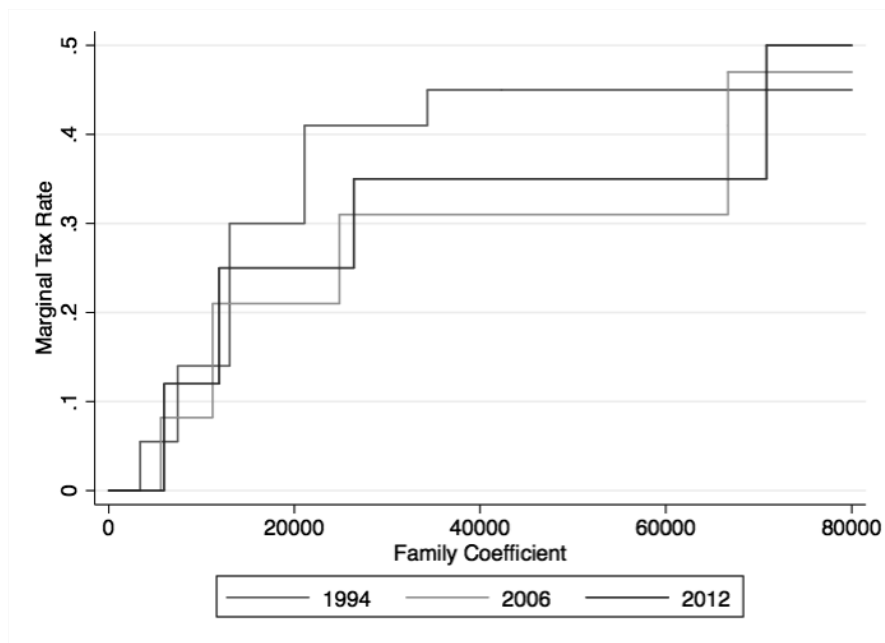
Notes: The figure shows the marginal tax rates in different tax brackets and for a selected set of years. The x axis shows the family coefficient, i.e., taxable income divided by the number of parts, as explained in Section 2.

FIGURE A2: Agents in the Standard Regime Affiliated with a Certified Accounting Center (CAC)
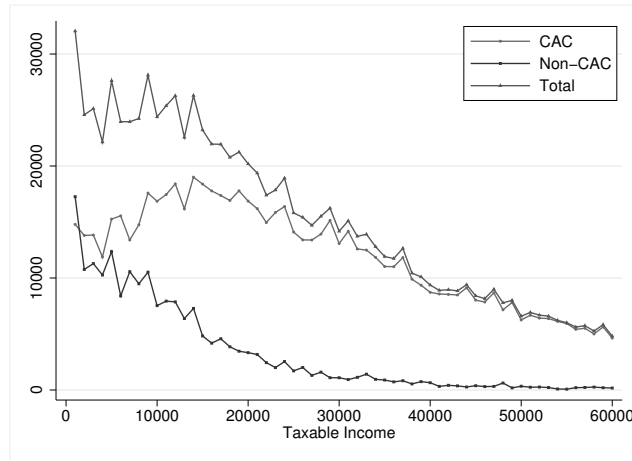

Notes: The figure shows the number of agents in the standard regime who are members of a CAC and the number of those who are not, as well as the total number of agents in the standard regime. The figure is based on the 2011 full population data. The x axis represents taxable income in the standard regime, i.e., net business income. At low income levels, there is a sizable fraction of agents who are not CAC members. That fraction declines rapidly and converges to zero at around 30,000 euros.

FIGURE A3: Share of Agents in the Super Simplified Regime
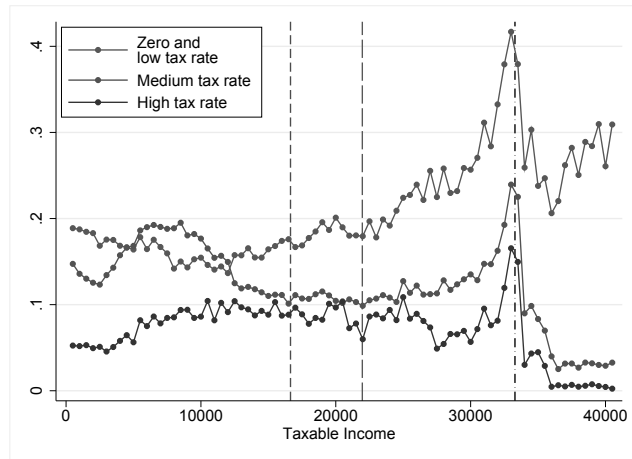
Notes: The figure shows the share of agents in the super simplified regime among all those in any self-employed regime who are eligible for the super simplified one. The results are shown for different tax brackets as defined in Section 2. See the notes to Figure 5a.

FIGURE A4: Bunching in the Simplified & Super Simplified regimes by Tax Brackets and Period

(a) 1999-2008   (b) 2009-2012
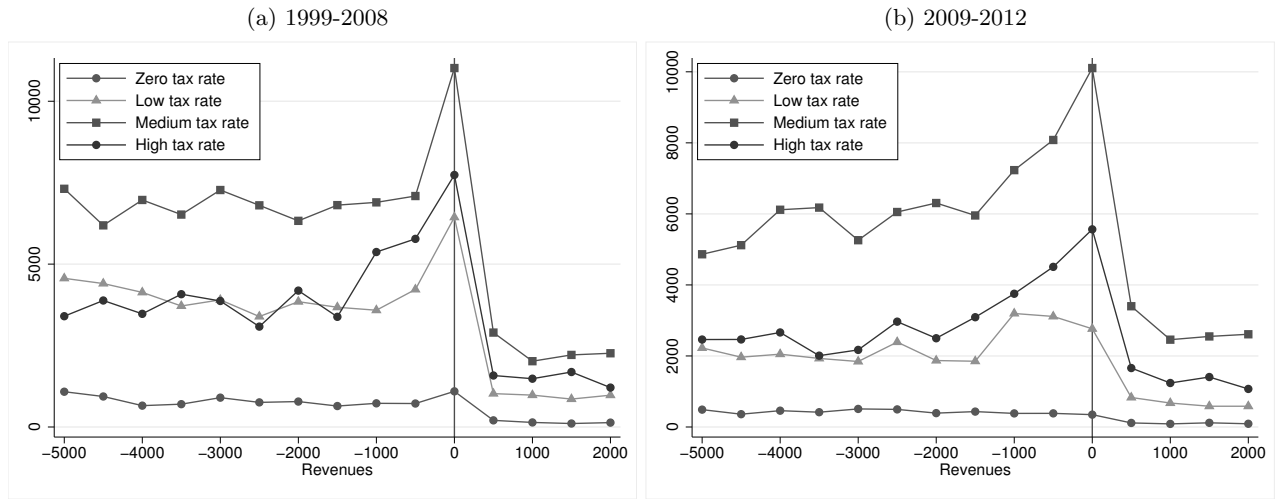
Notes: See the notes to Figure 9.

FIGURE A5: Sensitivity of Elasticity Estimates to Operating Costs c

(a) Simplified - I&C Services   (b) Simplified - Non commercial
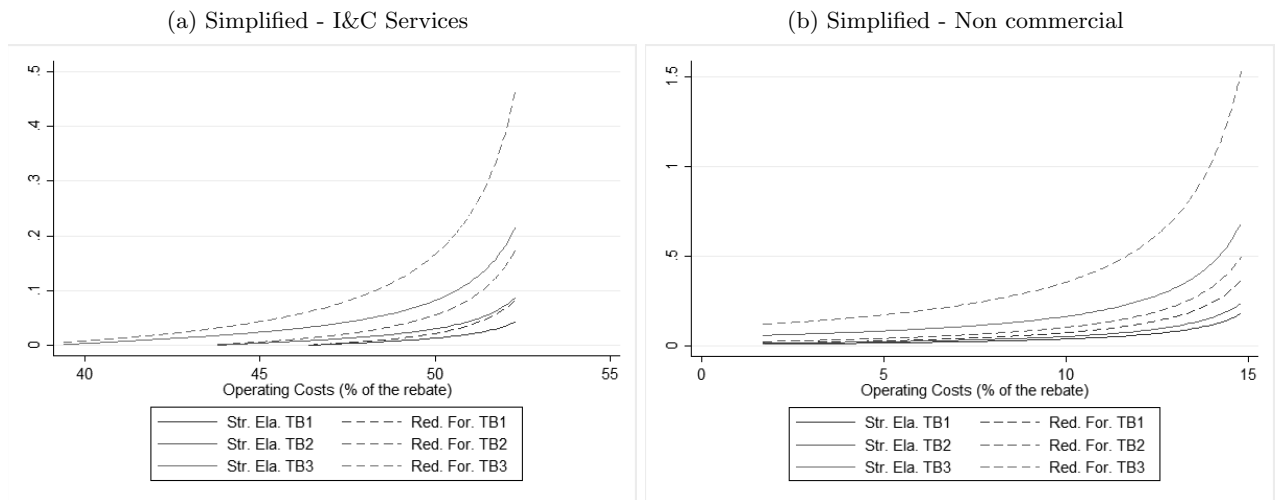
Notes: The figures show the value of the structural elasticities (solid curves) and the reduced form elasticities (dashed curves) for different tax groups, as a function of the operating costs. Numbers correspond to the benchmark case in which hassle costs are set to zero as in Table 9.

FIGURE A6: 2008 Reform: new entrants and switchers
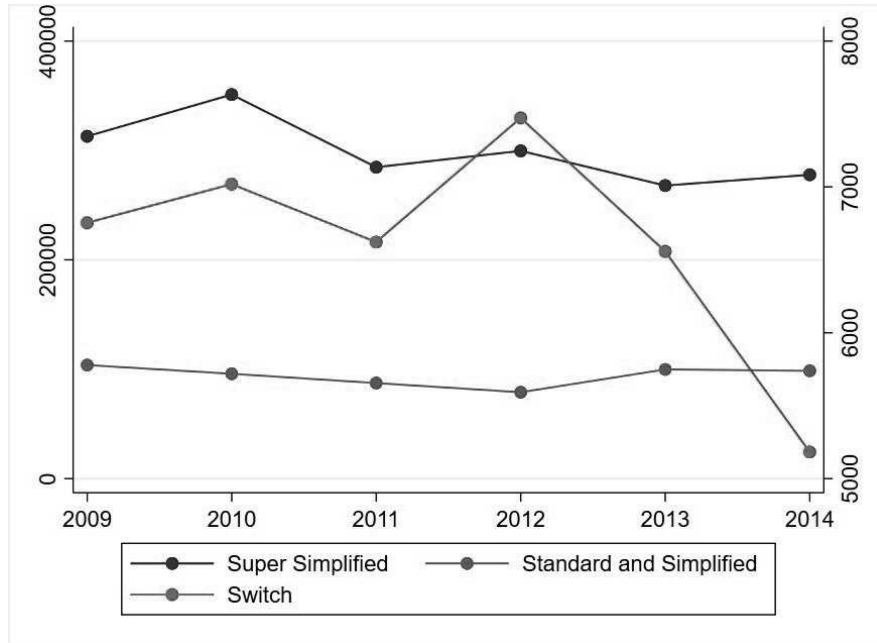
Notes: The figure shows the number of agents that have just created their businesses under the Super Simplified regime or under the Standard or Simplified regime (left y-axis) as well as agents who already operated a business under the Standard or Simplified regime and switched to the Super Simplified regime (right y-axis).

FIGURE A7: Income Shifting Within the Household: Sum of revenues
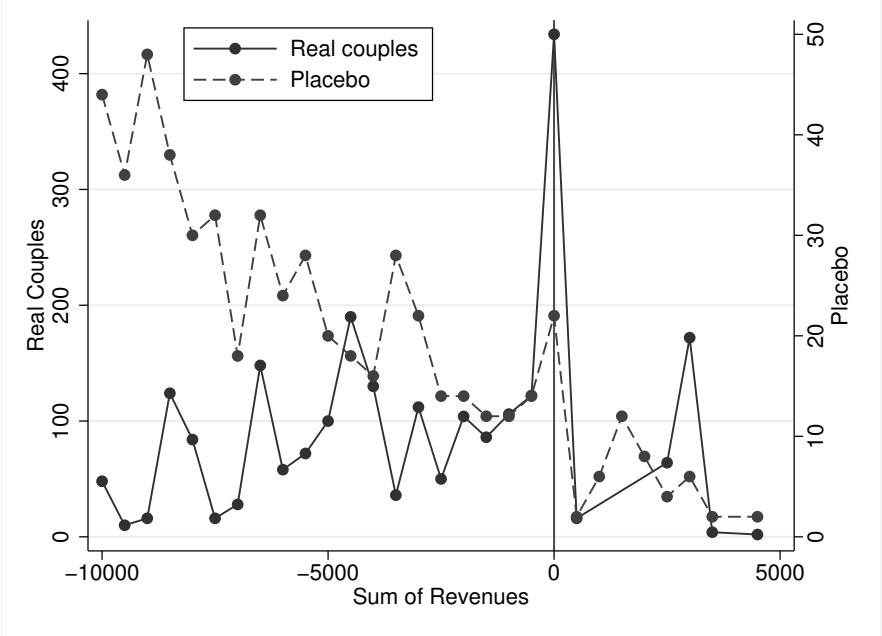
Notes: The figure shows bunching among individuals in households with two self-employed earners in a simplified regime at twice the eligibility threshold (the red vertical line). See the notes to Figure 20. Placebo partners are here formed by randomly matching each agent with another self-employed agent who is also in a household with two self-employed earners in a simplified regime.

FIGURE A8: Bunching Estimates by Year

(a) 1999   (b) 2000   (c) 2001

(d) 2002   (e) 2003   (f) 2004

(g) 2005   (h) 2006   (i) 2007

(j) 2008   (k) 2009   (l) 2010

(m) 2011   (n) 2012
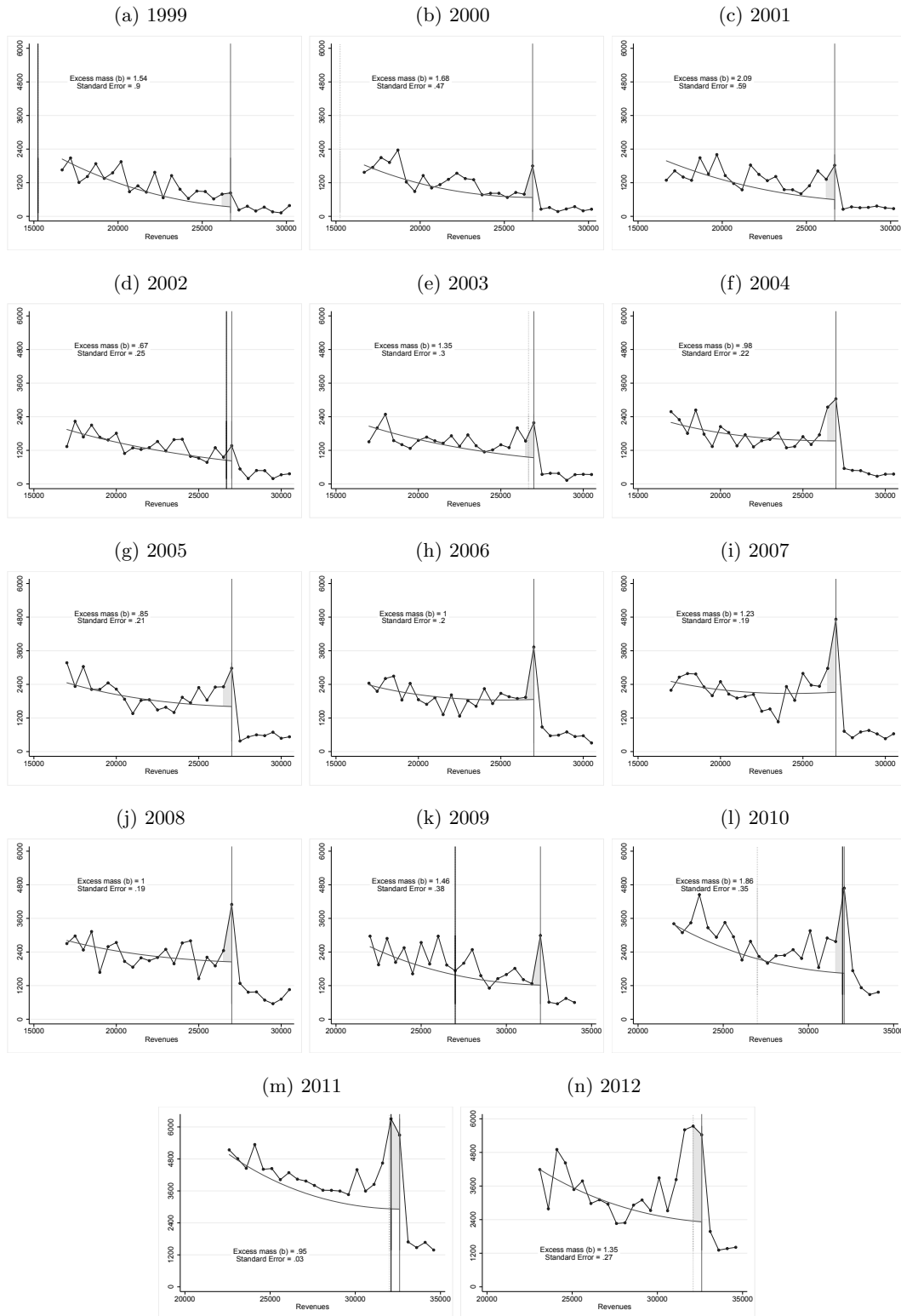
Notes: See the notes to Figure 7. The red solid (respectively, black solid and black dashed) vertical line represents the threshold in year t (respectively, $t-1$ and $t-2$).

66

TABLE A1: Summary Statistics: Self-Employed and Wage Earners 1994-1998

	All	With wage income only	With self-employed income only	With any self-employed income
Panel A				
Age	40	39	48	47
Married or in Civil Union	0.56	0.55	0.68	0.68
Has any Children	0.43	0.43	0.42	0.44
Number of children	0.77	0.77	0.76	0.79
Retired	0.08	0.07	0.16	0.14
Panel B				
Wage Income	18377	19329	0.00	5275
Self-employed Income	1927	0	30254	28438
Capital Income	1457	1250	3642	4306
Panel C				
Zero Tax bracket	0.19	0.20	0.16	0.14
Low Tax bracket	0.36	0.37	0.26	0.25
Medium Tax bracket	0.31	0.31	0.28	0.28
High Tax bracket	0.14	0.12	0.30	0.32
Panel D				
Population (in mill.)	130	121.2	6.8	8.8

Notes: See Table 6.

TABLE A2: Summary Statistics: Self-Employed and Wage Earners 1999-2008

	All	With wage income only	With self-employed income only	With any self-employed income
Panel A				
Age	40	40	49	48
Female	0.46	0.47	0.31	0.31
Married or in Civil Union	0.50	0.49	0.64	0.63
Has any children	0.40	0.40	0.39	0.40
Number of Children	0.70	0.70	0.70	0.72
Retired	0.05	0.04	0.16	0.14
Claimed unemployment benefits	0.09	0.10	0.02	0.03
Panel B				
Wage Income	19750	20711	0	5824
Self-employment Income	2066	0	34963	32032
Capital Income	2203	1904	5531	6533
Panel C				
Zero Tax bracket	0.15	0.15	0.14	0.13
Low Tax bracket	0.32	0.32	0.22	0.21
Medium Tax bracket	0.38	0.38	0.30	0.31
High Tax bracket	0.15	0.14	0.33	0.35
Panel D				
Population (in mill.)	283.8	265.5	13.7	18.3

Notes: See Table 6.

TABLE A3: Summary Statistics: Self-Employed and Wage Earners 2009-2012

	All	With wage income only	With self-employed income only	With any self-employed income
Panel A				
Age	41	41	50	48
Female	0.48	0.49	0.35	0.36
Married or in Civil Union	0.45	0.44	0.57	0.55
Has any Children	0.39	0.39	0.36	0.37
Number of children	0.67	0.67	0.63	0.66
Retired	0.07	0.06	0.21	0.17
Claimed unemployment benefits	0.13	0.14	0.04	0.07
Panel B				
Wage Income	20470	21503	0	7145
Self-employment Income	1939	0	31506	26980
Capital Income	2803	2493	6008	6801
Panel C				
Zero Tax bracket	0.13	0.13	0.18	0.16
Low Tax bracket	0.29	0.30	0.20	0.20
Medium Tax bracket	0.46	0.47	0.36	0.38
High Tax bracket	0.11	0.10	0.26	0.26
Panel D				
Population (in mill.)	118.8	110.3	5.8	8.5

Notes: See Table 6.

A.2 French Tax Calculation Primer

The French tax schedule typically looks as follows:

Bracket	Lower Bond	Upper bond	Marginal rate
1	$\underline{y_0} = 0$	$\underline{y_1}$	τ_1
2	$\underline{y_1}$	$\underline{y_2}$	τ_2
3	$\underline{y_2}$	$\underline{y_3}$	τ_3
4	$\underline{y_3}$	$\underline{y_4}$	τ_4
5	$\underline{y_4}$	∞	τ_5

In order to determine the tax amount to be paid by a household, the first thing to compute is the Family coefficient y which is defined as the ratio between taxable income Y and the number of parts

N of the household :

$$y = \frac{Y}{N} \tag{A1}$$

The household that has a family coefficient $y \in [\underline{y_{M-1}}; \underline{y_M}]$ belongs to the bracket M. Then, the amount of tax the household has to pay is :

$$T(y, N) = N \times [\sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times (y - \underline{y_{M-1}})] \tag{A2}$$

For instance, for a household with a family coefficient $y \in [\underline{y_2}; \underline{y_3}]$, we have :

$$T(y, N) = N \times (\tau_1 \times \underline{y_1} + \tau_2 \times (\underline{y_2} - \underline{y_1}) + \tau_3 \times (y - \underline{y_2})) \tag{A3}$$

Cap of the Family Coefficient

Let's assume that the number of parts is $N_b + N_a$ where N_b is the base number of parts, and N_a is the additional number of parts. To calculate the cap, one first calculates the tax that would apply without the additional parts: $y^b = Y/N_b$. We must then consider two possible situations : if the additional number of parts N_a (i) does place the household in a higher tax bracket, or (ii) does not place the household in a higher tax bracket.

Situation 1

If the additional number of parts N_a does not place the household in a higher tax bracket, then :

$$T(y^b, N_b) = N_b \times [\sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times (y^b - \underline{y_{M-1}})] \tag{A4}$$

The difference in taxes is :

$$T(y^b, N_b) - T(y, N) = (N_b - N) \times \sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times (N_b y^b - N_b \underline{y_{M-1}} - Ny + N \underline{y_{M-1}}) \tag{A5}$$

By definition, we have $Y = N_b y^b = Ny$, then :

$$T(y^b, N_b) - T(y, N) = (N_b - N) \times \sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times \underline{y_{M-1}}(N - N_b) \tag{A6}$$

We can re-arrange the expression to obtain :

$$T(y^b, N_b) - T(y, N) = (N_b - N) \times [\sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) - \tau_M \times \underline{y_{M-1}})] \tag{A7}$$

Situation 2

If the additional number of parts N_a places the household in a higher tax bracket, then :

$$T(y^b, N_b) = N_b \times [\sum_{m=1}^{M} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_{M+1} \times (y^b - \underline{y_M})] \tag{A8}$$

The difference in taxes is :

$$T(y^b, N_b) - T(y, N) = (N_b - N) \times \sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times (N_b \underline{y_M} - N_b \underline{y_{M-1}} - Ny + N\underline{y_{M-1}}) + \tau_{M+1} N_b \times (y^b - \underline{y_M}) \tag{A9}$$