Tax Simplicity or Simplicity of Evasion? Evidence from Self-Employment Taxes in France*

Philippe Aghion Ufuk Akcigit

Maxime Gravoueille Matthieu Lequien

Stefanie Stantcheva

May 3, 2023

Abstract

We exploit individual panel information from introducing of new and simpler tax regimes for the self-employed in France, in order to assess the extent to which individuals' shift towards the new regimes is driven by a quest for tax simplicity, and the extent to which the demand for tax simplicity is itself at least partly driven by tax evasion motives. We find evidence of a quest for tax simplicity from observing a significant amount of bunching at the eligibility thresholds for the simpler self-employment tax regimes and from the fact that bunching is increasing in the degree of simplicity of the self-employment regime. We also argue that tax evasion plays an important role in accounting for individuals' attraction towards simpler tax regimes. Finally, we quantitatively assess the importance of simplicity and evasion motives for choosing a simpler self-employment regime. More precisely, we combine bunching estimates and a structural model to jointly estimate the real income elasticities, the value of tax simplicity, and the evasion elasticity. We find that the parameters values which generate the best fit with the observed bunching across different tax brackets and years, imply noticeable preference for tax simplicity with a sizeable evasion elasticity behind it, and a negligible real income elasticity.

*Philippe Aghion: Collège de France, INSEAD, and London School of Economics (philippe.aghion@insead.edu); Ufuk Akcigit: University of Chicago, NBER, CEPR, Brookings, and CESifo (uakcigit@uchicago.edu); Maxime Gravoueille: Paris School of Economics and Collège de France (maximegravoueille@gmail.com); Matthieu Lequien: INSEE; Stefanie Stantcheva: Harvard, NBER, and CEPR (sstantcheva@fas.harvard.edu). This paper previously circulated under the title "Tax Simplicity and Heterogeneous Learning." We thank Antonin Bergeaud, Pierre Boyer, Larry Katz, Henrik Kleven, Etienne Lehmann and participants at the NBER Entrepreneurship Summer Institute, Paris seminar participants for useful comments and feedback. Stantcheva gratefully acknolwedges support from NSF CAREER Grant #1654517.

1 Introduction

Simplicity is the ultimate sophistication," wrote Leonardo da Vinci. Presumably policy makers would likely agree with this statement. Designing a policy that fulfils its stated goals, provides clear and correct incentives without unintended consequences, minimizes administrative hassle for individuals, and at the same time remains sufficiently simple for people to understand is an enormous challenge. This is particularly true for tax policy: the best designed tax incentives may turn out to be ineffective if people do not understand them. Even worse, complexity tends to make the tax system become 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. As it turns out, many tax and transfer policies have been targeted towards the bottom end of the income distribution, yet the lack of simplicity have prevented the targeted taxpayers from fully taking advantage of these policies.

While tax simplicity has undeniable advantages, in this paper we argue that it may also have a cost, namely to also sometimes favor tax evasion. More specifically, in this paper we exploit individual panel information from introducing of new and simpler tax regimes for the self-employed in France, in order to assess the extent to which individuals' shift towards the new regimes is driven by a quest for tax simplicity, and the extent to which the demand for tax simplicity is itself at least partly driven by tax evasion motives. 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. There are three tax regimes under which the self-employed in France may choose to operate – a standard regime, a simplified regime, and a super-simplified regime introduced more recently - which differ along two main dimensions: the monetary tax burden and the amount of red taping they impose upon individuals. Studying the observed choices of self-employed individuals between these three regimes and changes in these choices, we can assess individuals' demand for tax simplicity and the importance of tax evasion motives.

Our focus on the self-employed stems from three main considerations. First, those are typically shown to be less constrained than wage earners and can more easily adjust their incomes to tax incentives (Saez, 2010; Kleven and Waseem, 2013), which matters if we want to assess how people respond to simpler or more complex tax policies. Second, self-employment in France is a particularly well-suited quasi-laboratory for studying the effects of tax simplicity and complexity. Indeed, it displays a high variety of fiscal "regimes"which differ not only with regard to monetary incentives, but also in their degrees of tax simplicity. Institutional parameters of these fiscal regimes have changed significantly over time, providing valuable policy variation that helps our estimation.

There are at least three motives for choosing a simpler tax regime: (1) *monetary incentives*: by reporting less income and/or filing for an often less taxed regime, individuals end up paying fewer taxes in the simpler regime than if they had remained in the standard regime; (2) *tax simplicity*: by remaining in the simpler regimes, individuals save on hassle costs and reduce their administrative burdens; (3) *ease of misreporting*: it is much easier to misreport income in the simpler regimes than

in the standard regimes. In this paper we argue the quest for simplicity, plays an important role in explaining individuals' behavior around the eligibility thresholds for the simplified and super simplified regimes, and that this quest is itself at least partly driven by tax evasion motives.

In a nutshell, we find evidence of a quest for tax simplicity from observing a significant amount of bunching at the eligibility thresholds for the simpler self-employment tax regimes and from the fact that bunching is increasing in the degree of simplicity of the self-employment regime: it is higher at the eligibility threshold for the super simplified regime than at the threshold for the simplified regime. We also show that the observed bunching at the eligibility thresholds is to a large extent driven by tax evasion, itself made possible by and increasing the attraction of simpler tax regimes.

More precisely, we exploit new individual panel data from the French tax authority (the DG-Fip) to analyze individuals' choice of tax regime to then infer the extent to which this choice is driven by a quest for simplicity and/or by tax evasion motives. The self-employed can choose between three regimes, which we can rank by decreasing degree of complexity. The "standard regime" treats individuals' net business incomes (revenues minus costs) as taxable incomes, which is advantageous for corporate businesses with many employees, significant investments, or high operating costs. However, this regime entails involved tax accounting requirements, aimed at limiting the scope for misreporting. The "simplified regime" cuts down on tax complexity by allowing agents to claim a flat-rate rebate as a fraction of revenues instead of reporting their true business costs, this is particularly advantageous for agents with low operating costs. Finally, the "super simplified" regime enhances tax simplicity further by replacing all income taxes and social insurance contributions by a unique – and relatively low– flat rate payment proportional to gross revenues. However, to qualify for the simplified and super simplified regimes a self-employed individual must report revenues below some corresponding eligibility thresholds, the threshold being lower for the super-simplified regime than for the simplified regime. These thresholds in turn vary with the type of business activity, and they have also evolved over time. Overall, the eligibility thresholds for the simplified and super-simplified regimes induce discontinuities in monetary incentives, evasion opportunities, and in the degree of tax simplicity.

We first exploit individuals' bunching behavior around the eligibility thresholds to provide evidence of a quest for tax simplicity. Indeed, the eligibility thresholds create discontinuities in individuals' payoffs, which can be thought of as "notches", where not only the tax burden, but also the hassle costs and the ability to evade taxation, can potentially change. What complicates our assessment of individuals' response to the notches, is that we do not consistently observe revenues for agents above the eligibility thresholds for simpler regimes. Yet, in Section 3, we show that both the simplified and super simplified regimes exhibit sharp spikes in the density distribution of individuals right below the threshold. Most importantly, bunching is higher at the eligibility threshold for the super simplified regime than at the threshold for the simplified regime: this is true globally but also across activities. This in turn reflects a quest for simplicity on top of pure monetary incentives.

Second, we show tax evasion motives partly explain the quest for simplicity. The sharp bunching observed is in itself a smoking gun for evasion and avoidance responses. We further show dynamic bunching evidence, in addition to the static one: namely, individuals who bunch at the eligibility ceiling have sharply higher growth rates than other self-employed individuals but lower growth rates than those who cross the ceiling. Additional evidence for evasion comes from the fact that revenue statements are more often round numbers close to the thresholds than far from the thresholds, which in turn can be seen as evidence that the reported figure is more likely to have been forged. A second piece of evidence is that in households with two self-employed individuals, the highest earner appears to shift some of their income to their partner as their own income approaches the eligibility ceiling. Finally, we show that there is some "hidden employment" whereby employers contract out work previously done in-house, effectively circumventing costly labor contracts and relabeling self-employed work as employment.

This latter piece of evidence fits with the debates around the rise of platforms such as Uber or Task Rabbit, and the outsourcing of jobs previously done in-house. In recent work, Katz and Krueger (2019) 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.

Finally, we quantitatively assess the importance of simplicity and evasion motives for choosing a simpler self-employment regime. More precisely, we use our reduced-form bunching estimates as data moments to match in the estimation of a structural model of self-employed behavior. The model allows us to infer the value of tax simplicity and the evasion elasticity. We find that the parameters values which generate the best fit with the observed bunching across different agents and regimes imply a significant preference for tax simplicity and a sizeable evasion elasticity.

Our paper lies at the intersection of several strands of the literature. Most closely related to our analysis are the literature on tax simplicity, on tax evasion, and on bunching by small firms or self-employed individuals.

Tax simplicity: Farhi and Gabaix (2020) develop a theory of optimal taxation with behavioral agents which display misperceptions. Craig and Slemrod (2022) analyze the interplay between taxation and taxpayer education when individuals have an incomplete understanding of the tax system. Feldman et al. (2016) try to determine whether tax complexity causes misperceptions by looking the effects of tax liability changes. Relatedly, Abeler and Jäger (2015) and Bhargava and Manoli (2015) seek to understand how individuals react when facing complex tax systems, and suggest individuals underreact to change in tax incentives because of psychological frictions. Benzarti (2020) uses a quasi-experiment to estimate the cost of filing taxes. Blumenthal and Slemrod (1992), Slemrod (2005) and Zwick (2021) also investigate the effects of compliance costs in complex tax systems, while Warskett et al. (1998) and Grottke and Lorenz (2017) look at the role the institutional context (such as the interplay between public authorities, tax preparers and taxpayers) in shaping tax complexity. de Paula and Scheinkman (2010) and Tazhitdinova (2018) look at how tax enforcement can reduce tax evasion in the context of VAT and charitable givings. Finally,

Pirttilä and Selin (2011) show in the context of a dual income tax system in Finland, that a decrease in the marginal tax rate targeted to capital incomes increased income shifting for self-employed. We contribute to this literature by exploiting individual panel information on the choice between different tax regimes to provide evidence of a quest for tax simplicity, and by showing that this quest is at least partly driven by tax evasion motives.

Tax evasion: our work relates to multiple empirical studies of misreporting in response to taxation. Engström and Holmlund (2009) and Johns and Slemrod (2010) document significant income underreporting among the Swedish and US self-employed population. Harju and Matikka (2016) show that tax incentives are a motive for evasion, as income-shifting accounts for the majority of the overall taxable elasticity of income for business owners. Similarly, Saez (2010) and LaLumia et al. (2015) demonstrate that self-employed earners respond to tax incentives created by the EITC in the US. On the contrary, Parker (2003) finds no effect of tax incentives on the occupational choice to be self-employed and on tax evasion in Great-Britain. Kleven et al. (2011) use a tax enforcement field experiment in Denmark, to study the extent to which external tax auditing reduces the scope for tax evasion. Pomeranz (2015) uses randomized experiments on Chilean firms to assess the extent to which, once tax evasion is taken into account, seemingly equivalent taxation devices become markedly different. Carrillo et al. (2017) use information about a policy intervention on Ecuadorian firms to show the importance confronting taxpayer reports with third party information. Almunia and Lopez-Rodriguez (2018) use quasi-experimental information to analyze the effects of firm-size-dependent tax enforcement on tax compliance. Naritomi (2019), Brockmeyer et al. (2019) and Boning et al. (2020) also provides evidence of the role of tax enforcement in reducing tax evasion. We contribute to this literature by showing how tax evasion motives may hide behind the choice for self-employment and tax simplicity.

Bunching and structural estimation: A growing literature applies the bunching methodology to a wide range of topics such as inter-temporal allocation in response to mortgage contracts changes, transaction taxes in housing markets, or corporate taxation. Thus, Saez (2010) uses bunching information from US tax return data to estimate the elasticity of reported income with respect to the marginal tax rate. Gelber et al. (2020) use information on bunching in the earnings distribution at the budget set kinks to reassess the impact of changes in the effective marginal tax rate. Bergolo et al. (2021) study underreporting through tax deductions in Uruguay, and le Maire and Schjerning (2013) investigate the role of income shifting in explaining taxable income bunching in Danemark for the self-employed individuals. Kleven and Waseem (2013) exploit bunching information using administrative data from Pakistan to assess the impact of optimization frictions on individual responses to tax changes. Bastani and Selin (2014) and Alinaghi et al. (2021) also find that optimization frictions partly explain the observed bunching patterns in response to income taxes in Sweden and New-Zealand. Devereux et al. (2014) and Coles et al. (2022) focus on responses to corporate taxes. Mortenson and Whitten (2020) document behaviors that seek to maximize tax credit refunds in the US, and find that bunching is mainly driven by the self-employed. Chetty et al. (2013) use differences in manipulation of self-employed income across US areas as a proxy for

knowledge of the EITC program, in order to estimate wage earnings responses from this program. Chetty et al. (2011) use information on bunching at kinks using Danish tax records, to show that the labor supply response to tax changes, depends upon interaction between adjustment costs on the workers side and the working hours set by firms. Tazhitdinova (2020) also document the interplay between labor demand and labor supply in shaping earnings responses to tax incentives, using a salient discontinuity created by the "mini-jobs" program in Germany. Bíró et al. (2022) study the role of the minimum wage and tax enforcement in the Hungarian labor market where informality and imperfect enforcement are prevalent. We contribute to this literature by combining our computed reduced form bunching moments with a structural model to jointly estimate the value of tax simplicity and the tax evasion elasticity.

The remaining part of the paper is organized as follows. Section 2 presents the evolving landscape of self-employment in France over the period 1994-2015, describing the various self-employment regimes and the dynamic sequence of self-employment reforms. It also presents the data and provides some descriptive statistics. Section 3 provides evidence of a quest for tax simplicity by looking at individuals' bunching at the eligibility thresholds for the simpler self-employment regimes. Using the observed bunching characteristics, section 4 provides evidence to the effect that the quest for tax simplicity is partly driven by tax evasion motives. Section 5 uses the bunching moments to perform the structural estimation. Finally, section 6 concludes.

2 Institutional Background, Data and Descriptive Statistics

We begin with a brief description of the French self-employment regimes. We then present our data and provide descriptive statistics.

2.1 The Landscape of Self-Employment Regimes in France

Our study focuses on the period 2006-2015, during which there were interesting reforms of the taxation of self-employed income. We do not expand the analysis to after 2015, as the setting changed.

Activities. Self-employed individuals are classified into three types of activities, namely: 1) the "Industrial and Commercial Services" category, referred to as *I&C Services* below, 2) the "Industrial and Commercial Retail" category, referred to as *I&C Retail*, and 3) the *Non Commercial* category.[1] These are defined mainly for tax purposes and 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

[1]These are the so-called *Bénéfices Industriels et Commerciaux Services* for "Industrial and Commercial Services", *Bénéfices Industriels et Commerciaux Vente* for "Industrial and Commercial Retail" and *Bénéfices Non Commerciaux* for "Non Commercial".

activities, while construction work, plumbing, carpentry, 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.

Self-employed regimes. We focus on self-employed businesses that are taxed at the personal income tax schedule. Starting in 2009, the self-employed can choose between one of three regimes: the *super simplified* regime (created in 2009), the *simplified* regime (created in 1999) and the *standard* regime. The 2009 reform introducing the super simplified regime stemmed from the political will to further increase tax simplicity by reducing accounting requirements and tax hassle. The super simplified and simplified regimes have eligibility income ceilings (see below) above which individuals have to move to the standard regime.

Note that a self-employed individual who owns her business can also choose to incorporate and be subject to the corporate tax system. Self-employed with revenues above €750,000 have to incorporate. We do not study those individuals for two reasons. First, they typically operate on a larger scale than the businesses analyzed here. Second, individuals in either the super simplified or the simplified regimes that were to cross the eligibility threshold would face as the most immediate alternative the standard regime. This set of three regimes therefore captures well the choice sets of agents.

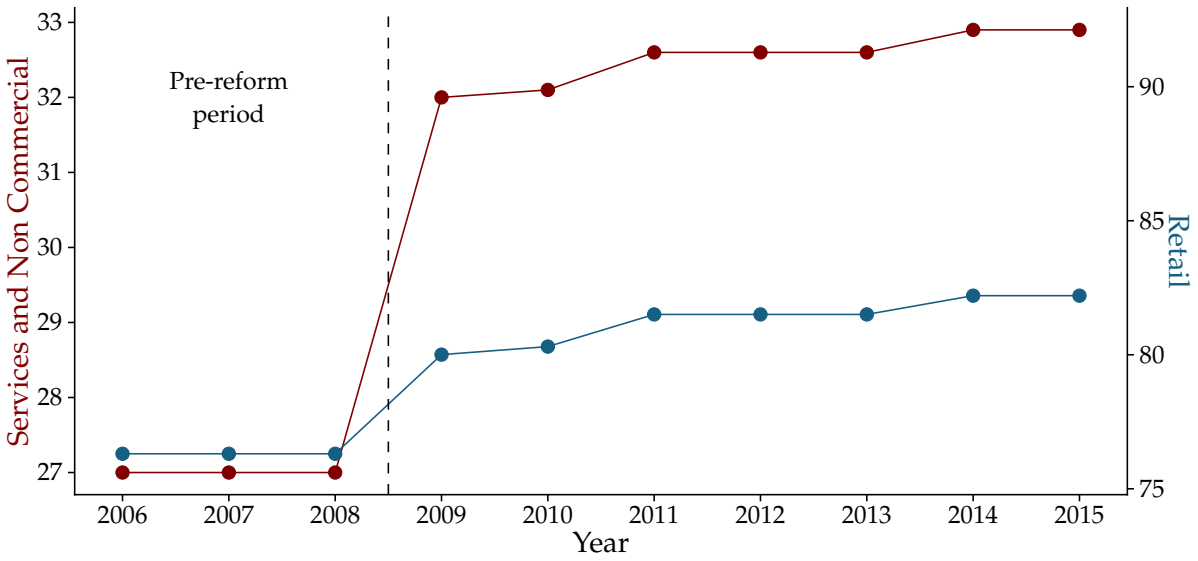
Eligibility requirements. The super simplified and simplified regimes can only be chosen by agents with revenues below a given threshold y^*_{kt}, which depends on the type of activity k, where $k \in \{\text{I\&C Retail, I\&C Services, Non Commercial}\}$, and on the fiscal year t. Figure 1 shows the thresholds' evolution. The thresholds for the Services and Non Commercial activities are lower than those for the Retail activities (32,600 euros in 2012 as contrasted with 81,500 euros). In the case of the super simplified regime, there is an additional requirement: family income in year $t-2$ has to be below a year-specific threshold f^*_t that corresponds to the third tax bracket cutoff.[2] An individual with income below the threshold y^*_{kt} for activity k in year t can choose between the simplified, the super simplified, and the standard regimes.[3]

Above the threshold, taxpayers are defaulted into the standard regime. To avoid a costly and abrupt change, there is a *tolerance region*. Thus, individuals with incomes with at most 6.1% of the threshold in 2012 for the Services and Non Commercial Activities and 9.9% of the threshold for the Retail Activities are in the tolerance region and can remain in the simpler regimes provided they do not cross the formal region more than two years in a row.

[2] For instance, that cutoff was 26,420 euros for year 2010.

[3] 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.

Figure 1: Eligibility threshold for simpler regimes, by activity and year



Notes: The figure plots the evolution of the eligibility threshold by activity over time, in thousand euros. The eligibility threshold correspond to the self-employed revenues, before the application of any potential rebate. The I&C Services and Non Commercial activities are on the left axis, while the I&C retail category is on the right axis. The vertical dashed line correspond to the introduction year of the super simplified regime in 2009.

Tax base and taxes. In the standard regime, the taxable income is the net business income, i.e. the difference between gross revenues and costs, including the depreciation of assets and investments according to standard accounting rules. In the simplified regime, the taxable income is equal to revenues times a scaling factor $1 - \mu$, where the rebate factor μ is determined by the tax administration. It depends on the activity type: 71% for Retail, 50% for Services, and 34% for Non Commercial activities.[4] In the super simplified regime, taxable income is simply equal to revenues (i.e.the rebate $\mu = 0$).[5] Under the simplified and super simplified regimes, an individual cannot claim any losses.

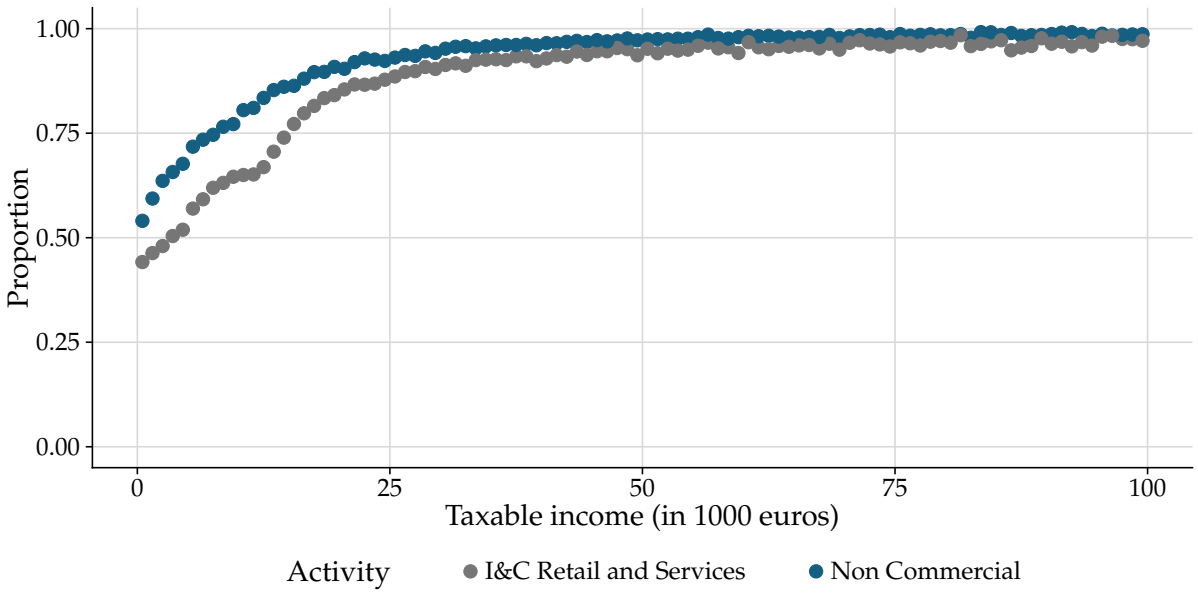
In the standard and simplified regimes, the regular tax and social insurance contribution rates apply, both of which differ across households depending on various factors as explained in the Appendix C.

In the super simplified regime, the individual instead pays a flat rate that covers both the income tax and the social insurance contributions. The flat rate differs by activity and it has changed over time, but it is unrelated to the individual's actual income tax bracket or to tax rate that applies to the remaining part of her income, not subject to the super simplified regime. Thus, even an individual in the zero income tax bracket is taxed at same flat rate on all her activities that fall

[4]The minimal rebate amount is capped at 305 euros.

[5]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 equivalent simplified regime 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.

Figure 2: Take up of CAC by activity



Notes: The figure plots the proportion of agents in the standard regime who are members of a CAC, by activity. The figure is based on the 2006-2015 data. Each dot represents the average number of agents, by year. 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.

under this regime.

Accounting and reporting simplicity. Each of those three regimes has different accounting and reporting requirements.

Self-employed individuals 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 call upon a "certified accounting center" (hereafter, CAC), which helps them keep their accounts and also serves as a guarantor of sound fiscal conduct vis-a-vis the tax authority. The financial incentives to join a CAC - namely the exemption from membership and accounting expenses and the avoidance of a 25% inflation of the tax base - have led a large share of agents in the standard regime to join such centers. Figure 2 shows that at the taxable income levels relevant for our analysis (between 15,000 and 35,000 euros), almost all agents in the standard regime are CAC members. A governmental report (Cour des Comptes (2014)) states that conditional on an audit, the size of penalties among non-CAC members is larger than among CAC members of comparable size (around 26,000 euros versus 7,000 euros). It adds that the discrepancy between taxes due and taxes actually paid comes more often from genuine accounting mistakes and delays in payments and less often from outright tax evasion among CAC members than among non-CAC members.

Beyond the financial incentives they entail, the simpler regimes require fewer administrative

tasks and proofs of sound fiscal accounting. In the simpler regimes, individuals do not need to report purchases, sales, or costs, only total revenues, and are not required to comply with rigorous accounting practices. They are nevertheless required to keep documentation and receipts, in case an audit takes place, much like any regular tax payer would do, e.g. to claim itemized deductions.

Having to keep various types of accounts involves more hassle in the standard regime than in the simplified regimes. But the various regimes also differ in how easy it is to file taxes. 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, thus requiring two separate filings. In the super simplified regime, tax and social insurance payments are due monthly or quarterly, based on actual realized revenues (cash in hand), and all are being processed at the same time, thereby minimizing filing and hassle costs.[6]

Ease of misreporting in the simpler regimes. The lighter accounting and reporting requirements for the simplified regimes likely make it much easier to misreport. Although the French tax authority is aware of the risk of cheating and misreporting involved by the introduction of the simpler regimes,[7] auditing individuals in the simpler regimes is not a top priority in light of scarce auditing resources, given the low revenues and incomes of these taxpayers and the hassle involved in accessing their place of residence. Nevertheless, the tax authority carried out two audit programs in 2011, directed at 1162 randomly-selected individuals in simpler regimes. The findings from the first audit show that 30% of taxpayers were under-reporting income, on average by €580, whereas less than 1% of them were over-reporting. The second audit study focused on around 1000 individuals in the Parisian region and uncovered under-reporting of €710 on average. Overall, the tax authority concluded by extrapolating from these two audit programs that it would have recovered about 400 million euros had all the self-employed been audited. These amounts are likely to be lower bounds given that very few among the audited individuals were near the eligibility threshold where under-reporting likely even stronger.

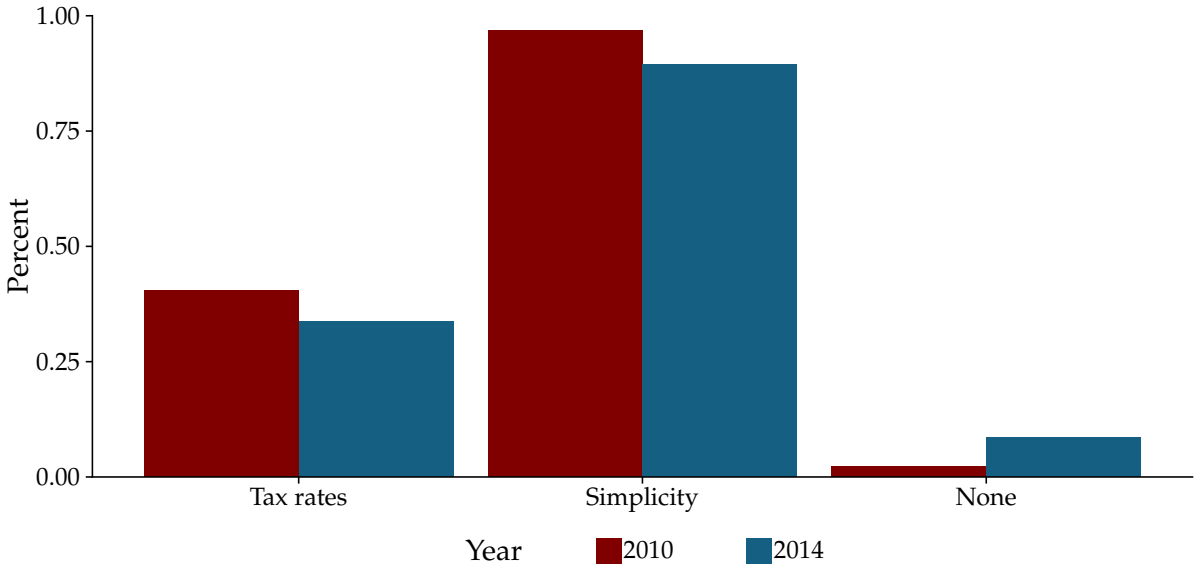
Survey evidence on motives for being in the simplified regimes. Figure 3 depicts the responses from a survey of individuals in simpler regimes for years 2010 and 2014. The survey offers individuals the option to select any or all of the three choices when asked about their perception of the benefits of being in a simpler regime: (i) favorable tax rates, (ii) ease of accounting, reduced costs related to social security payments, registration procedures, and reporting, or (iii) neither of the two aforementioned options.

We see that almost all individuals report simplicity and the concern for hassle costs as being a key motive for choosing a simpler regime. Then tax incentives also play an important role (between 30 and 50% of individuals in the survey mention it as a main motivation for choosing

[6] In addition, the standard regime is the only one subject to the Value Added Tax: self-employed in this regime charge VAT on their products sold and claim VAT on their inputs.

[7] Deprost et al. (2013) from the tax auditing body write "The simplicity of the system and the weakness of the accounting obligations make the (misreporting) risk high."

Figure 3: Advantages of simpler regimes



Notes: The figure plots the answered advantages of simpler regimes (simplified and super simplified) from a survey for year 2010 and 2014.

a simpler regime). Naturally, asking about evasion or avoidance behavior is unlikely to yield truthful answers in such a survey.

2.2 Data and Descriptive Statistics

Data. Our longitudinal data is based on the universe of French tax returns over the period 2006-2015 from the French Internal Revenue Service.[8] 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. Importantly, it allows us to follow individuals over time. We supplement the taxpayer panel with survey data from the French National Statistics Institute[9] available for 2010 and 2014, which asks entrepreneurs about their experience during their first years.

Sample. Our benchmark sample consists of all individuals who are French fiscal residents in mainland France and are between 30 and 59 years of age. We only consider individuals that are primary or secondary taxpayers, excluding dependants such as children. Finally, we keep individuals with self-employment income that are uniquely defined in a regime and activity.[10] Further details about data construction are available in Appendix C.

[8] *Direction Générale des Finances Publiques* (DGFIP).
[9] The survey is called *New enterprises information system*.
[10] It excludes for example individuals with self-employment in different regimes and activity.

Descriptive statistics. Table 1 shows summary statistics by regime (simplified, super simplified and standard) and activity (*I&C Retail* (1), *I&C Services* (2) and *Non Commercial* (3)) for our sample. The average age is around 45. A significant share of those in non-commercial activities live in Paris. Average revenues are higher for retail than for Services and Non-Commercial activities and higher in the standard regime.

Figure 4 plots the number of self-employed over time by regime and status, normalized to 100 in 2010. Panel (a) plots the total number of self-employed individuals, which reached 2,4 millions by 2015. Panel (b) shows the number of entrepreneurs who stay in the same regime as in the previous year ("stayers"). We can see that this number is significantly lower than the number in Panel (a) for entrepreneurs in the standard regime, suggesting that they start switching to the super-simplified regime after it is introduced in 2009. Panel (c) shows the number of new entrepreneurs each year, which amounts to around 15% of those in the standard regime, 20% of those in the simplified, and almost a third of those in the super-simplified. Panel (d) the number who exit self-employment. These numbers are very close to the entry rates for each regime, suggesting that by 2014, the system may have reached a steady-state.

Table 1: Summary statistics, by regime and activity

	Simplified			Super simplified			Standard	
Variable	Retail	Services	Non Commercial	Retail	Services	Non Commercial	Retail and Services	Non commercial
Mean age	46	45	44	44	43	43	47	46
Married	0.54	0.56	0.52	0.55	0.56	0.53	0.66	0.65
Has children	0.54	0.57	0.54	0.57	0.59	0.59	0.59	0.68
Living in Paris	0.02	0.07	0.11	0.02	0.03	0.07	0.02	0.08
SE revenues	14632	9315	9564	11890	9314	9095	–	–
Taxable SE income	4243	4658	6312	3448	4657	6002	25825	59234
Observations	658497	1144606	990993	219161	332724	372843	3028307	2570586

Notes: This table shows summary statistics by regime and activity, for the post-reform period (2009-2015). Note that for the standard regime, Retail and Services activities are pooled together in the tax returns and cannot be distinguished. Self-employed revenues for the standard regime are not observed. The variable living in Paris is in percentage points. Self-employed revenues and taxable income are in thousand euros.

3 Bunching at the Eligibility Thresholds for Simpler Regimes

In this section, we provide evidence of bunching at the eligibility thresholds for the simplified and super simplified regimes, and we perform some comparative analysis on the magnitude of this bunching. We start by describing the different incentives that can generate responses at the threshold. We then describe the methodology to quantify such responses, and we finally provide graphical and estimation evidence of individuals' behavioral response at the eligibility thresholds.

Notches created by the eligibility thresholds. The eligibility threshold for the simpler regimes (simplified and super simplified) and the standard regime can be considered a notch, where average payoffs change discontinuously. When an agent crosses this threshold, they experience the following changes, described in Section 2.1: (i) Their average tax rate changes (monetary incentives). The discontinuity in tax rates and, hence, monetary incentives at the threshold depends on the regime, activity type (which also affects operating costs and the rebate), family income, and other characteristics. Therefore, two agents with the same self-employed revenues can face disparate tax incentives. (ii) Agents' hassle cost of reporting income and filing taxes increases. (iii) It likely becomes more difficult to misreport revenues.

Because of this notch, we expect individuals to strategically locate below the threshold. Furthermore, we expect the effective notch to be larger for agents in the super simplified regime, which has a low flat rate, even lower hassle, and likely higher ease of evasion.

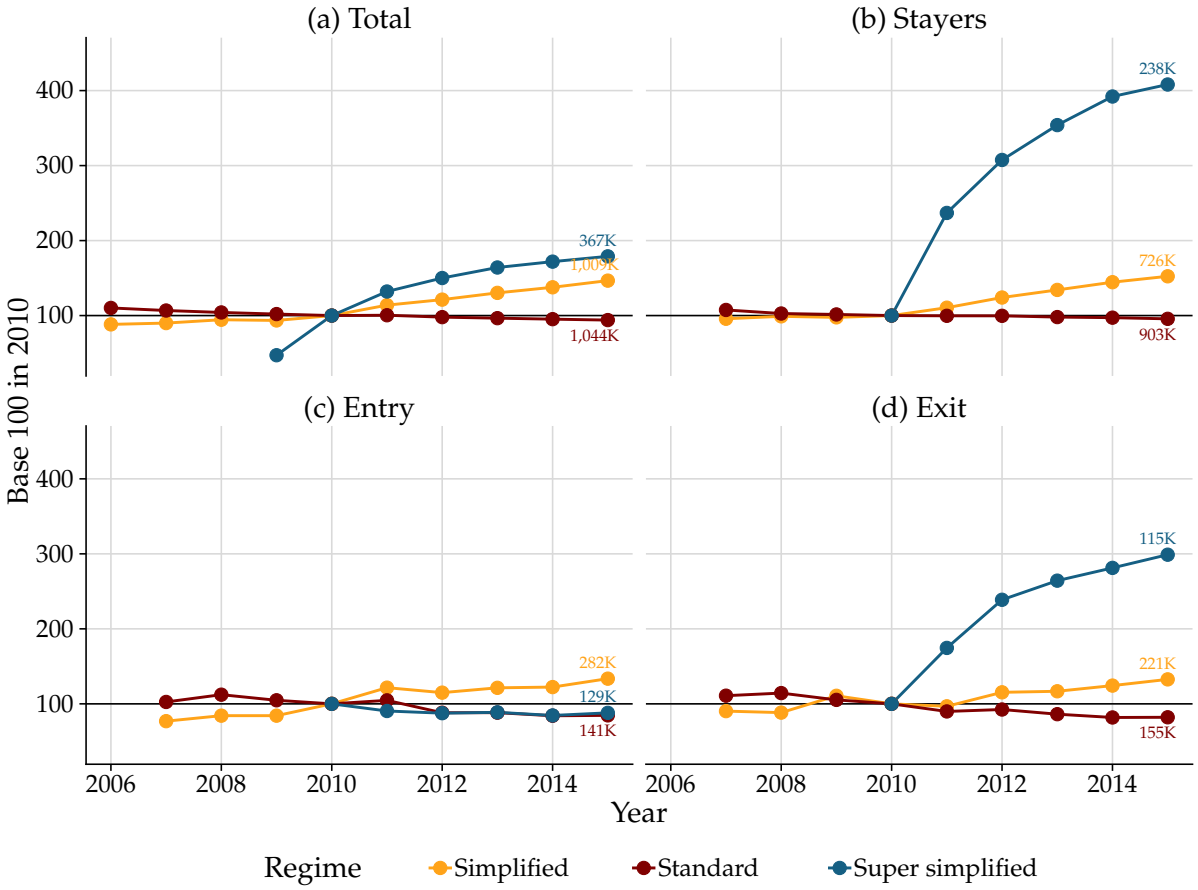
3.1 Quantifying Behavioral Responses with Bunching

Method. Using classic bunching methods we can identify and assess the importance of the behavioral responses at the eligibility threshold between simpler regimes and the standard regime (Saez (2010), Chetty et al. (2011), Kleven and Waseem (2013)). Recall that we do not consistently observe revenues for agents above the eligibility threshold in simpler regimes, hence the empirical distribution at the right of the threshold cannot be used to estimate the counterfactual distribution absent the notch. Here, we show how commonly used estimation procedures in the bunching literature can still be adapted to our setting.

Let $B = B(D)$ denote the extra number of individuals at the left of D following the introduction of a simpler tax regime with eligibility threshold D. To measure B, we estimate the counterfactual income distribution by fitting a smooth polynomial based on the empirical density. Revenues are centered around the eligibility threshold, and grouped by bins of size B_S: bin j contains all individuals with self-employed income in the interval $]B_j - B_S, B_j]$, so that all individuals reporting revenues at exactly the threshold belong to B_D. To estimate a counterfactual distribution to the left of the threshold, we run the following regression:

$$C_j = \sum_p \beta_p \cdot (B_j)^p + \sum_{d=D^-}^{D} \gamma_d \cdot \mathbf{1}[B_j = d] + \sum_r \alpha_r \cdot \mathbf{1}[r \in B_j] + \epsilon_j \tag{1}$$

Figure 4: Number of self-employed by regime over time



Notes: The figure plots the total number of self-employed individuals (panel (a)), the number of stayers in the same regime relative to the previous year (panel (b)), the number of entry (panel (c)) and the number of exit (panel (d)) by regime over time. All series are normalized to 100 in 2010, and raw count in 2015 (in thousand observations) are reported for each of them. An individual is consider entering if he is not observed in the previous year but observed in the current year. On the contrary, an individual is considered exiting if he is observed the previous year but not the current year.

where C_j stands for the number of individuals in income bin B_j, p is a set of polynomial integer exponents; $\mathbf{1}[B_j = d]$ are dummies equal to 1 for bins in the bunching zone (between D^- and D); $\mathbf{1}[r \in B_j]$ is a dummy equal to 1 if bin B_j contains r and r is a multiple of round numbers (for example multiples of 1000, 5000, etc.).

The counterfactual distribution absent the notch is predicted by $\hat{C}_j = \sum_p \beta_p \cdot (B_j)^p + \sum_r \alpha_r \cdot \mathbf{1}[r \in B_j]$ so that the bunching coefficient is equal to $B = \sum_j (C_j - \hat{C}_j)$, for bin j in the bunching zone. Finally, we define the excess mass as $b = B/C_D$, where C_D is defined as the average count of individuals across bins in the bunching zone. Because we do not observe the density distribution to the right side of the threshold, we cannot use the formal method in Kleven and Waseem (2013) to determine the bunching region. To compute standard errors, we generate earnings distributions and excess mass estimates by re-sampling the residuals in (1) using a bootstrap procedure.

Earnings response. The excess mass b is informative about the earnings response Δy^* at the eligibility threshold of the simpler regime. Individuals in the bunching area would have declared income in the interval $[y^*, y^* + \Delta y^*]$ absent the notch. We can express the bunching coefficient B as a function of the counterfactual density at the notch $h_0(y^*)$ and the marginal buncher located at $y^* + \Delta y^*$:

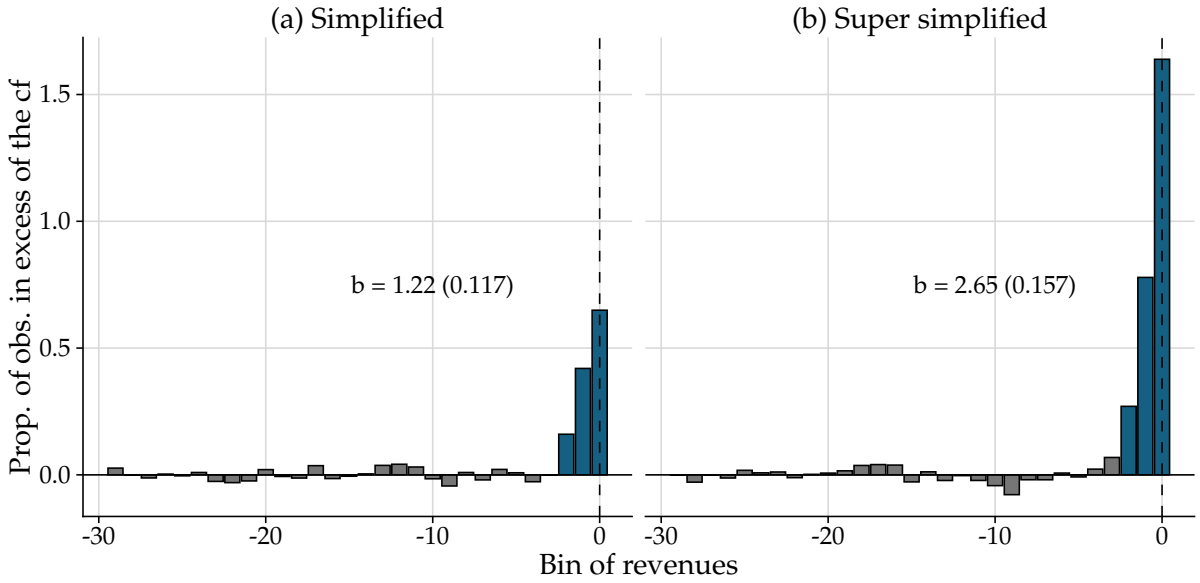
$$B = \int_{y^*}^{y^* + \Delta y^*} h_0(y) dy \approx h_0(y^*) \Delta y^*$$

Let us also define the counterfactual number of individuals located at the threshold by $\hat{\beta}_0$, where $\hat{\beta}_0$ is estimated using equation (1). The estimated density at the threshold is then equal to: $\hat{h}_0(y) = \hat{\beta}_0 / B_S$. From there, we can express Δy^* as a function of the bunching coefficient and the estimated density: $\Delta y^* = (B/\hat{\beta}_0) \times B_S$. The empirical equivalent of $\hat{\beta}_0$ is C_D, so that $\Delta y^* = b \times B_S$.

3.2 Results

Bunching. Figure 5 shows the estimated behavioral responses to the introduction of the simpler regimes for the period 2009-2015. We pool self-employed across all activities and center their revenues around the threshold applicable to them (represented by the dashed vertical line). We split taxpayers into bins of 500 euros for I&C Services and Non Commercial activities, such that $B_j = \{\ldots,]-5000, -4500],]-4500, -4000], \ldots,]-500, 0]\}$. For I&C Retail activity we use bins of 1500 euros, such that $B_j = \{\ldots,]-7500, -6000],]-6000, -4500], \ldots,]-1500, 0]\}$. We report for each bin j the difference between the actual and counterfactual counts, relative to the counterfactual counts by bin: $(C_j - \hat{C}_j)/\hat{C}_j$. For example, 0.5 means that the number of self-employed above the counterfactual is half the number of self-employed under the counterfactual distribution. The bunching region is colored in blue. This representation has two main advantages. First, we can easily compare the observed and counterfactual distributions of self-employed revenues. If the two distributions are similar, the difference must be close to zero. Second, we can easily compare these unit-free numbers across the two simpler regimes. A visual inspection of the distribution

Figure 5: Bunching estimation by regime



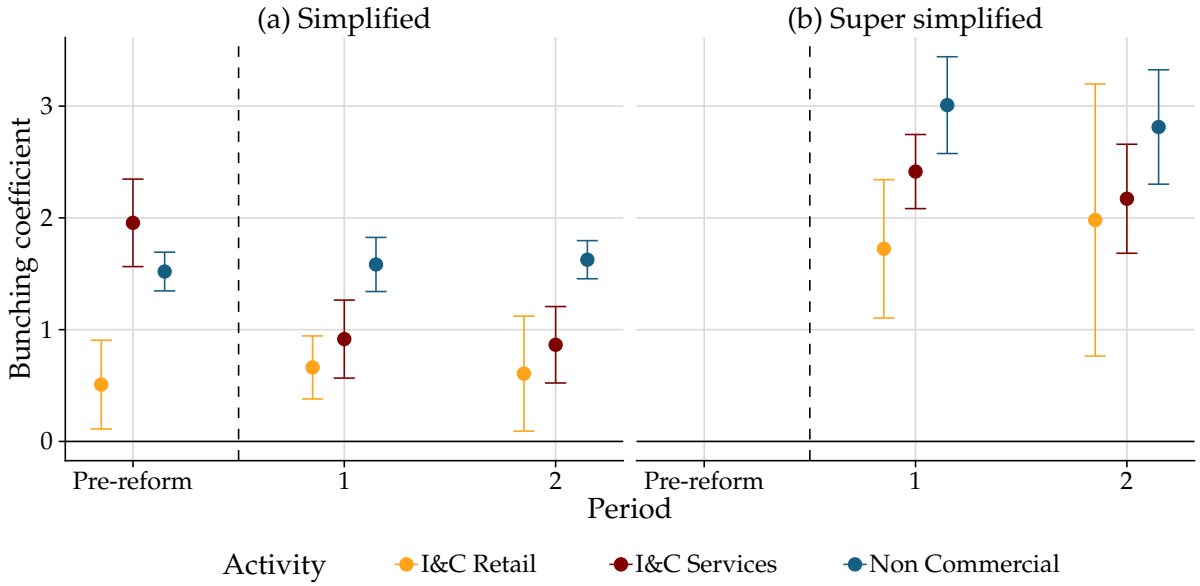
Notes: The figure represents the frequency of revenues, by bins of revenues centered around the eligibility threshold (the vertical dashed line). Results are using the pooled population for 2009-2015 and all agents in the simplified and super simplified regimes. The figure plots $(C_j - \hat{C}_j)/\hat{C}_j$: the difference between the actual and counterfactual counts, relative to the counterfactual counts by bin. The counterfactual distribution is fitted using a smooth polynomial as explained in subsection 3.1. The bunching area is in blue. There is significant bunching, equal to 122% of the average counterfactual frequency within the bunching area for the simplified regime and 265% for the super simplified regime. Standard errors are calculated using a bootstrap procedure in by random resampling (n = 400) of the residuals.

suggests the bunching behavior begins three bins away from the threshold (-1500€ for I&C Services and Non Commercial activities, and -4500€ for I&C Retail). We perform detailed robustness checks below.

The distribution of taxpayers exhibits a sharp spike right below the threshold for the simplified and super-simplified regimes. The difference between the actual and counterfactual distributions is close to zero before the bunching region (in grey) and starkly increases in the bunching region (in blue). The increase is larger for the super simplified (panel (b)) regime compared to the simplified regime (panel (a)), translating into sizeable and significant excess masses, respectively equal to 1.22 for the simplified and 2.65 for the super simplified. This is in line with the stronger incentives to remain in the super simplified regime than the simplified regime, highlighted above.

Heterogeneity in bunching by activity and time period. The incentives to remain in the simpler regime are likely to differ according to the type of activity. Figure 6 shows the excess mass b by activity and period. Panels (a) and (b) respectively report results for the simplified and super simplified regimes. All bunching estimates are large and significant. Similar to the pooled estimations, the behavioral responses for the super simplified regime are larger than for the simplified

Figure 6: Bunching estimation by regime, activity and period



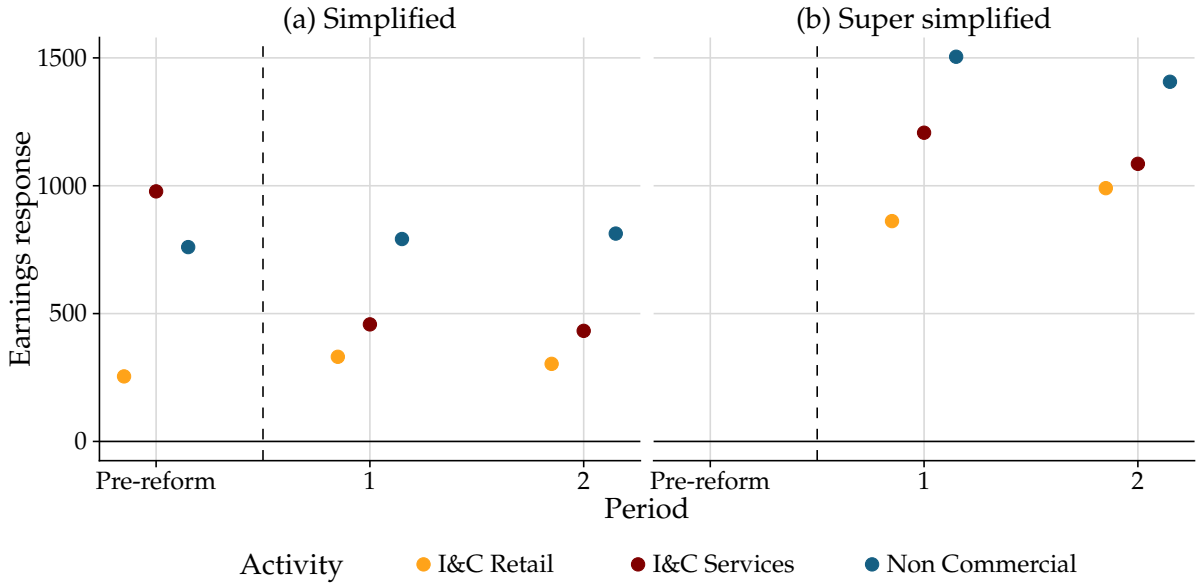
Notes: The figure plots the bunching coefficients b from subsection 3.1 by regime, activity and period. The pre-reform period is 2006-2008, period 1 is 2009-2013 and period 2 is 2014-2015. The counterfactual distribution is fitted using a smooth polynomial as explained in subsection 3.1. There is significant bunching for all categories, but systematically higher for the super simplified regime. Standard errors are calculated using a bootstrap procedure in by random resampling (n = 400) of the residuals.

regime. We also notice that bunching is generally more pronounced in the Non Commercial Activities than in the I&C Services. I&C Retail activities have the lowest bunching estimates. This in turn may reflect the fact that individuals in Non Commercial Activities have more flexibility to adjust their income. Restricting the sample to individuals with only self-employed earnings (and no additional labor earnings) yields similar results (see Figure A.1).

Empirical earnings responses. Figure 7 shows the earning responses implied by the bunching coefficients in Figure 6. They lie between 250 and 1000 euro for the simplified regime and between 800 and 1500 euro for the super simplified regime. Moreover, these earning responses remain very similar across periods, which in turn suggests that variation in tax rates over time have little effect on the earning responses.

Robustness tests. Figure A.2 shows the results of robustness tests on the estimation of the excess mass b. More specifically, we run variants of the above regressions where we both, modify the functional form by changing the degree of the polynomial or by running a Poisson regression, and allow for changes in the number of bins in the bunching area (the number of excluded bins in the plot). We perform this robustness exercise for each of the two simpler regimes and for the various types of activities separately. Each time the average excess mass is represented by

Figure 7: Earnings responses estimation by regime, activity and period



Notes: The figure plots the earnings responses $\Delta y^* = b \times B_S$ from subsection 3.1, by regime, activity and period. The pre-reform period is 2006-2008, period 1 is 2009-2013 and period 2 is 2014-2015.

the dashed line. Our preferred estimate is in red: namely, we see that it is always close or non statistically different from the average across all specifications. This in turn is reassuring about our basic estimation being a robust estimate of the true excess mass.

4 Evidence on Tax Evasion

In this section, we provide direct evidence on tax evasion and misreporting using dynamic bunching methods that take advantage of the panel structure of our data.

4.1 Methodology

To estimate the distortions introduced by the eligibility threshold on various outcomes, we need the counterfactual of these outcomes in the absence of the notch. Bunching estimates from section 3 showed that entrepreneurs manipulate their earnings in response to the threshold, making it invalid to compare individuals in the bunching area to those outside.

Intent to treat design. To circumvent this selection bias into the bunching area, we build on the method developed by Diamond and Persson (2016) to estimate the treatment effect of the discontinuity at the notch using an intent to treat design (ITT henceforth). Our design and empirical implementation also follow Chen et al. (2021). Intuitively, these methods allow us to compare

the observed average outcome of individuals in the bunching area to a potential outcome had the threshold not been implemented. The ITT estimator for any outcome X is given by:

$$ITT(X) = \mathbb{E}[X|\text{Notch}, Y \in (D^-, D)] - \mathbb{E}[X|\text{No Notch}, Y \in (D^-, D)] \tag{2}$$

where Y denotes self-employed earnings, D^- denotes the lower bound of the bunching area and D denotes the eligibility threshold. The first component in Equation 2 is the average X across individuals in the bunching area, which we directly observe in the data. The second component is the counterfactual average X which we need to estimate. This estimator is an ITT effect because the interval (D^-, D) includes both the self-employed that respond to the program and other self-employed individuals who do not respond to the program but happen to be in that area for other reasons.

We now describe the procedure for the estimation of the counterfactual average outcome $\mathbb{E}[X|\text{No Notch}, Y \in (D^-, D)$. By definition, this quantity is itself the combination of the counterfactual density in self-employed earnings $\hat{h}_0(y)$ and the counterfactual average outcome conditional on those earnings $\mathbb{E}[X|\text{No Notch}, Y = y]$:

$$\mathbb{E}[X|\text{No Notch}, Y \in (D^-, D)] = \int_{y=D^-}^{D} \hat{h}_0(y)\mathbb{E}[X|\text{No Notch}, Y = y]\text{d}y \tag{3}$$

To compute an empirical counterpart of these two quantities, we bin self-employed earnings following the same procedure as in section 3. It allows us to estimate $\hat{h}_0(\cdot)$ using the bunching method. For the second quantity, we fit a polynomial regression on binned outcome X_j, excluding the bunching region:

$$X_j = \sum_p \beta_p \cdot (B_j)^p + \sum_{d=D^-}^{D} \gamma_d \cdot \mathbf{1}[B_j = d] + \sum_r \alpha_r \cdot \mathbf{1}[r \in B_j] + \epsilon_j \tag{4}$$

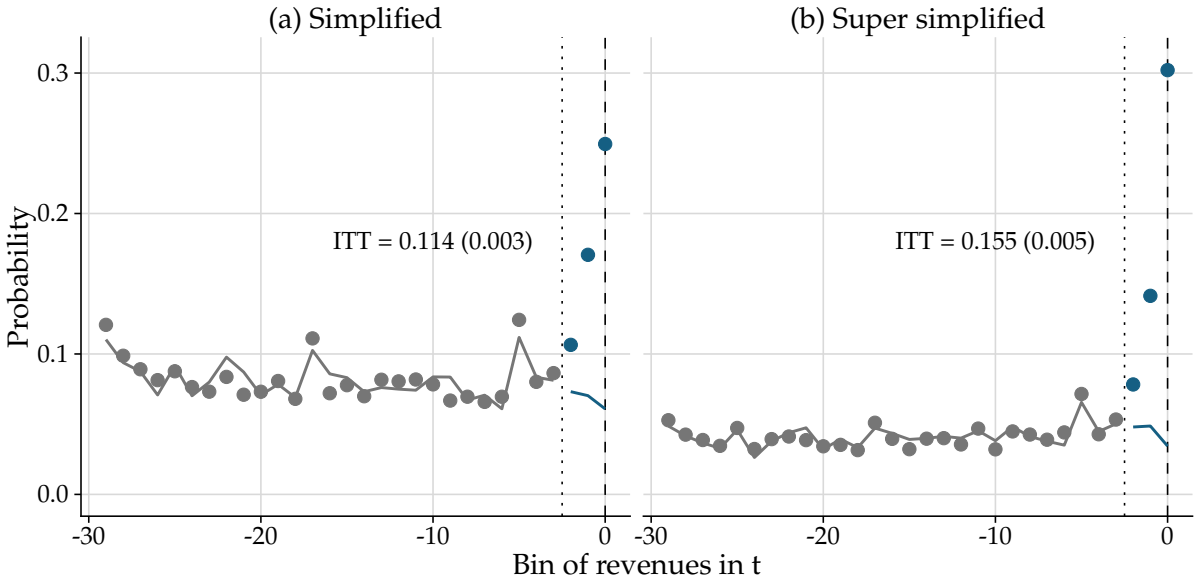
and use as an estimator: $\mathbb{E}[X_j|Y_j, \text{No Notch}] = \sum_p \beta_p \cdot (B_j)^p + \sum_r \alpha_r \cdot \mathbf{1}[r \in B_j]$.

4.2 Results

Self-employed revenue dynamics. Figure 8 shows that the likelihood of remaining at the same distance from the threshold, is particularly high in the bunching regions, 11.4 percentage points higher in the threshold for the simplified regime and 15.5 percentage points higher for the super simplified regime. This is indicative of individuals underreporting their true increase in income from one year to the next.

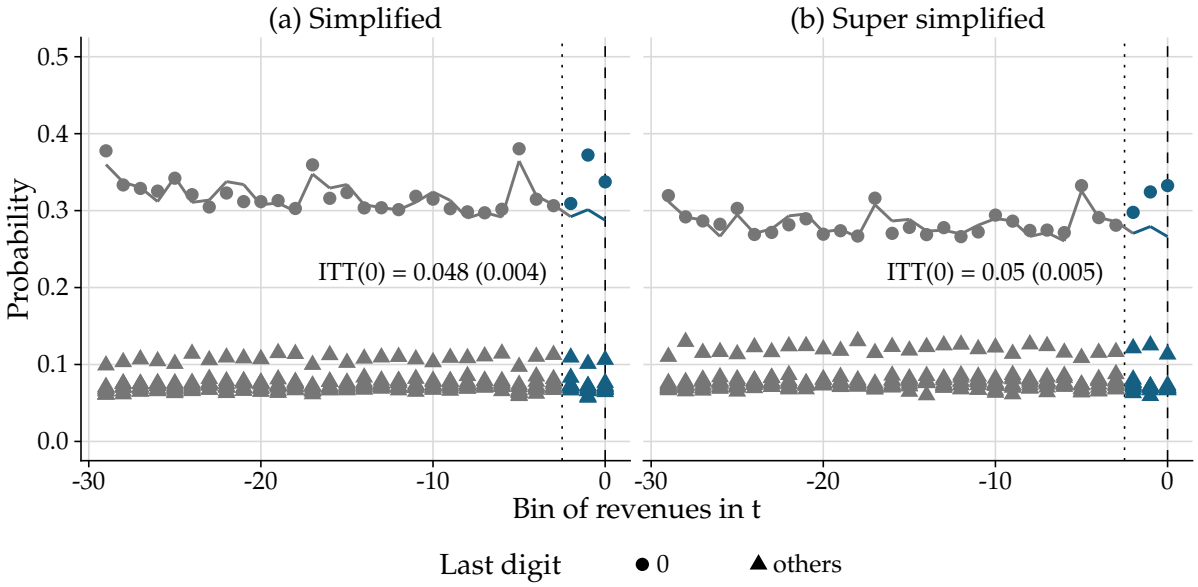
Bunching at specific digits. Absent incentives to evade taxes, we expect the probability to report a given number as last digit to be the same in the bunching area as anywhere else in the distribution of revenues. And under the assumption that the last digit follows Benford's law, we expect

Figure 8: Probability of remaining in the same bin between t and $t+1$



Notes: The figure plots the probability to remain in the same bin between year t and year $t+1$, by initial (year t) bins of revenues centered around the eligibility threshold (the vertical dashed line). Results are using the pooled population for 2009-2015 and all agents in the simplified and super simplified regimes. The counterfactual distribution is fitted using a smooth polynomial as explained in subsection 4.1. The bunching area is in blue, between the dotted and the dashed vertical lines. A discontinuity in the average growth rate is shown in the bunching area, but only significant for the super simplified regime. The probability is on average 11.4 percentage points higher for the simplified regime compared to the counterfactual situation without the threshold, and 15.5 percentage points higher for the super simplified regime. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

Figure 9: Probability to report a specific last digit for self-employed revenues



Notes: The figure plots the probability to report a specific last digit between 0 and 9, by bins of revenues centered around the eligibility threshold (the vertical dashed line). Dots represent the probability to report 0 as last digit and triangles the probability to report other digits. Results are using the pooled population for 2009-2015 and all agents in the simplified and super simplified regimes. The counterfactual distribution is fitted using a smooth polynomial as explained in subsection 4.1. The bunching area is in blue, between the dotted and the dashed vertical lines. A discontinuity in the average growth rate is shown in the bunching area. The probability is on average 4.8 percentage points higher for the simplified regime compared to the counterfactual situation without the threshold, and 5 percentage points higher for the super simplified regime. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

any number in the set {1,2,3,...9} to be reported with equal probability.[11]

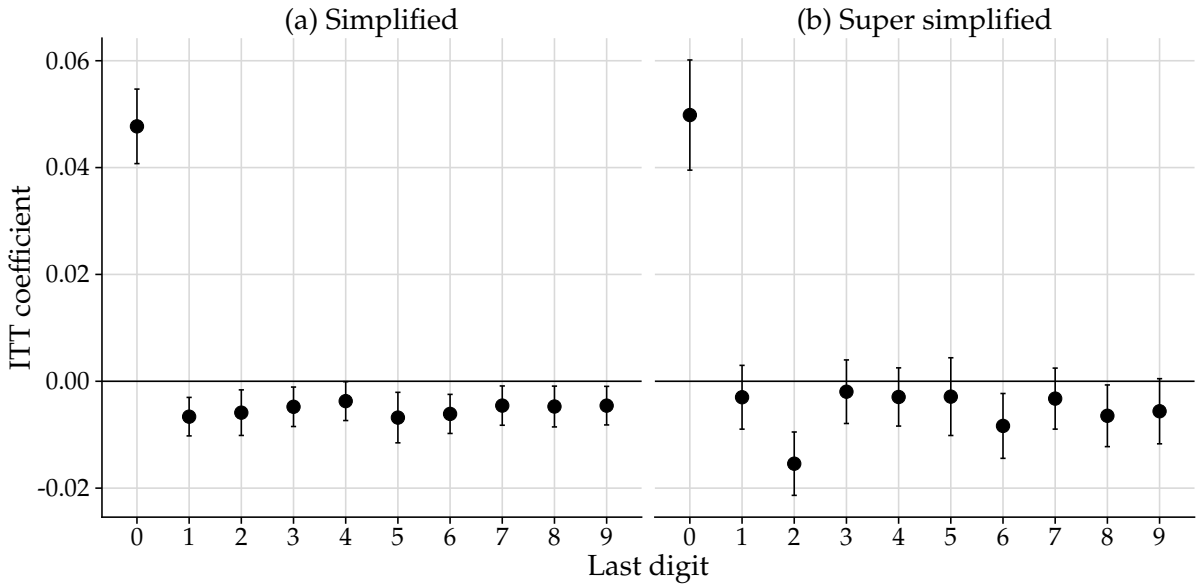
Figure 9 shows the distribution of the probability to report 0 or any other digit by knoner bin and regime. We see that individuals in both the simplified and the super simplified regimes are more likely to report zero no matter where they lie in the revenue distribution, but individuals in the bunching areas for the two regimes are around 5 percentage points more likely to do so.

Figure 10 plots the ITT coefficients for each digit by regime. We see that individuals disproportionately report 0 as the last digit in the bunching regions, which strongly speaks to strategic reporting as a way to evade taxation.

Round numbers bunching. We now dig further into the possibility for strategic reporting, by looking more closely at the numbers individuals actually fill in. Figure 11 shows the probability to report a multiple of 100 euros as a function of both, the individual's distance to the threshold and her activity. Consistent with our analysis in the previous subsection, we find that individuals in

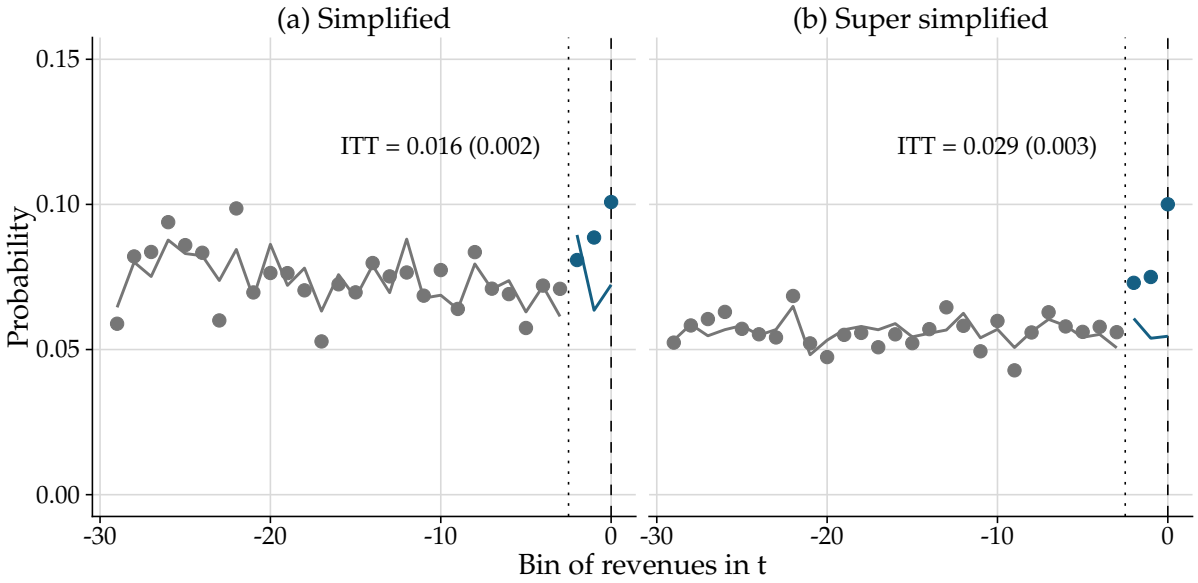
[11] See Varian (1972)).

Figure 10: ITT coefficients for specific last digit for self-employed revenues



Notes: The figure plots the ITT coefficients for the probability to report a last digit between 0 and 9. Results are using the pooled population for 2009-2015 and all agents in the simplified and super simplified regimes. The coefficient is computed using the method in subsection 4.1. For example, the probability is on average 4.8 percentage points higher for the simplified regime compared to the counterfactual situation without the threshold, and 5 percentage points higher for the super simplified regime. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

Figure 11: Probability of reporting a multiple of 100



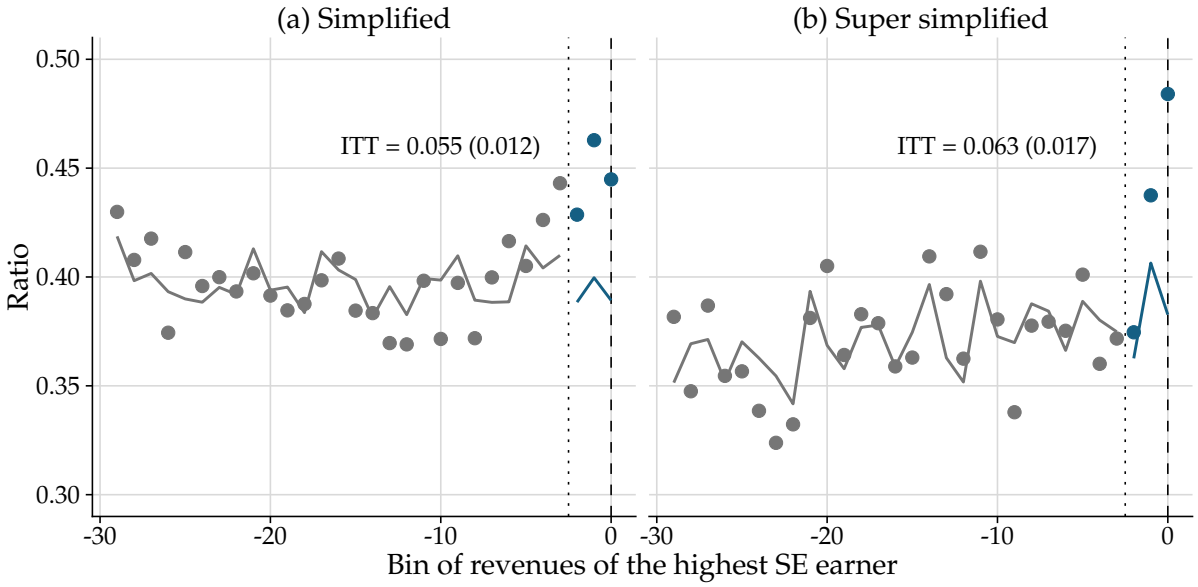
Notes: The figure plots the ITT coefficients for the probability to report a multiple of 100 euros in t (in grey) and in $t+1$ (in blue), for two populations: for individuals not in the bunching area in t and for all individuals in t. Results are using the pooled population for 2009-2015 and all agents in the simplified and super simplified regimes. The coefficient is computed using the method in subsection 4.1 and ranking individuals in bin of centered revenues in $t+1$. The bunching area is in blue, between the dotted and the dashed vertical lines. A discontinuity in the average growth rate is shown in the bunching area. The probability is on average 1.6 percentage points higher for the simplified regime compared to the counterfactual situation without the threshold, and 2.9 percentage points higher for the super simplified regime. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

the bunching regions disproportionately report multiples of 100 euros compared to those outside the bunching region. The probability is on average 1.6 percentage points higher for the simplified regime, and it is on average 2.9 percentage points higher for the super simplified regime.

Income shifting within the household. Another signal for misreporting and avoidance comes from income shifting within the household. The eligibility thresholds apply to individual income, which means that if an individual with self-employed income lives with another individual with self-employed income, they can to some extent relabel their revenues and shift them between the two businesses to remain below the threshold.

We find strong evidence that this is indeed the case by studying couples who both have self-employed earnings in one of the simplified regimes. Our sample for the 2009-2015 period contains 89457 such households. First, on the intensive margin, Figure 12 shows the ratio of the self-employed earnings of the lowest earner to those of the highest earner in the household. We can clearly see that, as the higher earner's self-employed earnings approach the threshold, there is a significant and large jump in the earnings of the lower earner as well. Furthermore, there is

Figure 12: Ratio lowest/highest gross incomes within household



Notes: The figure plots the ratio between the gross income of the lowest self-employed earner and the gross income of the highest self-employed earners, by bins of revenues of the highest self-employed earner, and centered around the eligibility threshold (the vertical dashed line). Results are using the pooled population for 2009-2015 and all agents in the simplified and super simplified regimes. It implies that both members of the household are in one of this two regimes. Panel (a) (resp. panel (b)) plots the result for the highest earner being in the simplified (resp. super simplified) regime. The counterfactual distribution is fitted using a smooth polynomial as explained in subsection 4.1. The bunching area is in blue, between the dotted and the dashed vertical lines. A discontinuity in the ratio is shown in the bunching area. The ratio is on average 5.5 percentage points for the simplified regime compared to the counterfactual situation without the threshold, and 6.3 percentage points higher for the super simplified regime. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

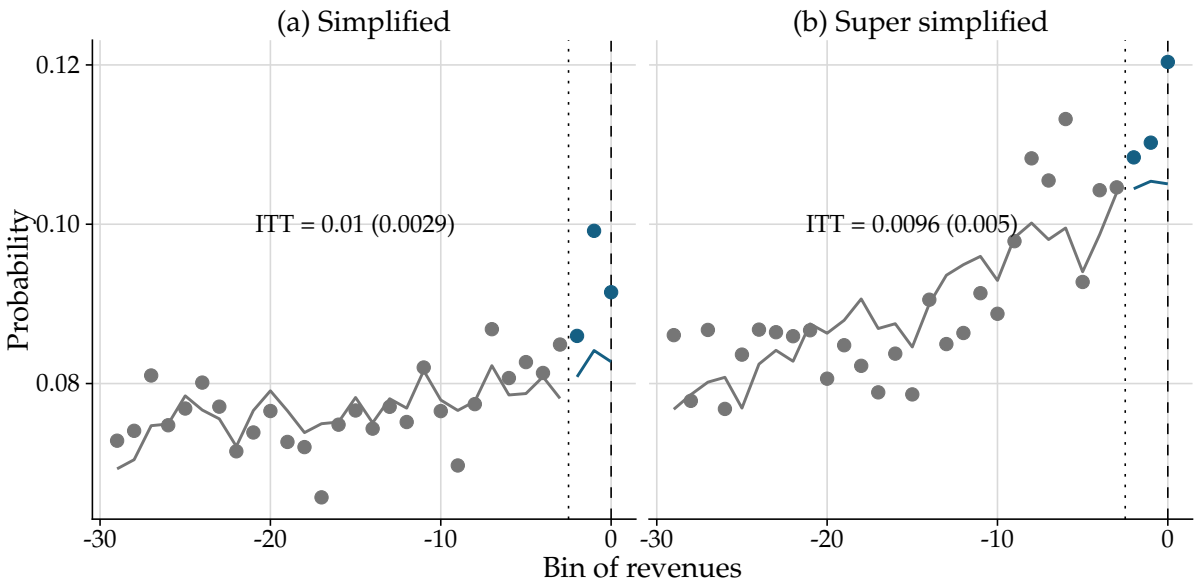
evidence of responses on the extensive margin as well. Figure 13 plots the probability to have a spouse that is also reporting (any) self-employed revenues, by bins of revenues centered around the eligibility threshold. While that probability is increasing overall (which can itself be due to assortative matching by activity or income type), there is a significant discontinuity just in and right below the bunching region.

Labor earnings and "hidden employment." One of the concerns raised in the policy debate about the simplified regimes was that they may lead to "hidden employment," whereby employers would fire employees and hire them again as contractors. This would allow employers to circumvent costlier standard labor contracts and regulations. We therefore consider how employed labor income varies around the eligibility threshold.

Figure 14 shows the average labor earnings reported by self-employed individuals by bins of self-employed revenues centered around the eligibility threshold (the vertical dashed line).[12] La-

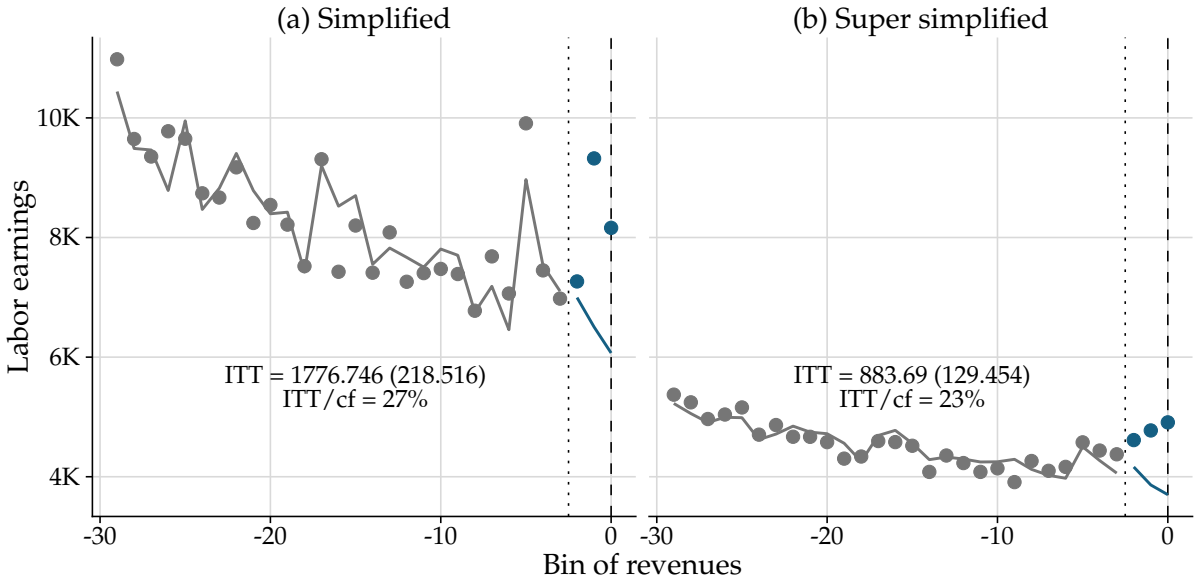
[12]We include individuals with zero earnings.

Figure 13: Probability to have a spouse in a simpler regime



Notes: The figure plots the probability to have a spouse that is also reporting self-employed revenues, by bins of revenues centered around the eligibility threshold (the vertical dashed line). Results are using the pooled population for 2009-2015 and all agents in the simplified and super simplified regimes. It implies that both members of the household are in one of this two regimes. Panel (a) (resp. panel (b)) plots the result for individuals being in the simplified (resp. super simplified) regime. The counterfactual distribution is fitted using a smooth polynomial as explained in subsection 4.1. The bunching area is in blue, between the dotted and the dashed vertical lines. A discontinuity in the probability is shown in the bunching area. The probability is on average 1 percentage points for both the simplified and super simplified regime compared to the counterfactual situation without the threshold. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

Figure 14: Average labor earnings



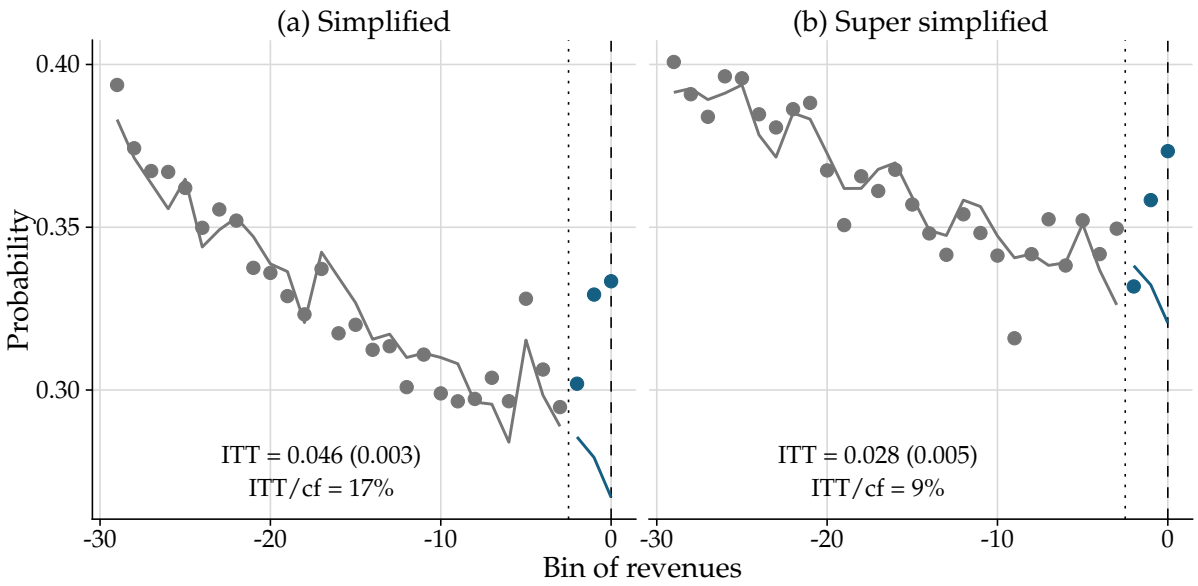
Notes: The figure plots the average labor earnings reported by self-employed individuals, by bins of self-employed revenues centered around the eligibility threshold (the vertical dashed line). Results are using the pooled population for 2009-2015 and all agents in the simplified and super simplified regimes. Panel (a) (resp. panel (b)) plots the result for individuals being in the simplified (resp. super simplified) regime. The counterfactual distribution is fitted using a smooth polynomial as explained in subsection 4.1. The bunching area is in blue, between the dotted and the dashed vertical lines. A discontinuity in the average labor earnings is shown in the bunching area. Labor earnings are on average 1777 euros higher for the simplified regime compared to the counterfactual situation without the threshold, and 884 euros higher for the super simplified regime. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

bor earnings generally decline with the level of self-employed revenues, which suggests that there is substitution between self-employed and employed work, potentially because of time constraints or because of hidden employment. However, there is a discontinuous sharp and significant increase in labor earnings just in the bunching region. This provides some evidence in favor of the hidden employment hypothesis: as hidden employees are about to cross the eligibility threshold, their employers transfer some of their pay in the form of regular salary. Figure 15 plots the probability of reporting any labor earnings in addition to self-employed revenues, which is an extensive margin response. This figure similarly shows that there is a discontinuous increase in the likelihood of reporting labor earnings just before the threshold.

5 Estimating the Value of Tax Simplicity

The previous sections showed evidence that self-employed regimes differ in their simplicity, that there is significant and sharp bunching at the eligibility thresholds, and that tax evasion and mis-

Figure 15: Probability to report labor earnings



Notes: The figure plots the probability to report labor earnings in addition to self-employed revenues, by bins of self-employed revenues centered around the eligibility threshold (the vertical dashed line). Results are using the pooled population for 2009-2015 and all agents in the simplified and super simplified regimes. Panel (a) (resp. panel (b)) plots the result for individuals being in the simplified (resp. super simplified) regime. The counterfactual distribution is fitted using a smooth polynomial as explained in subsection 4.1. The bunching area is in blue, between the dotted and the dashed vertical lines. A discontinuity in the probability is shown in the bunching area. The probability is on average 4.6 percentage points higher for the simplified regime compared to the counterfactual situation without the threshold, and 2.8 percentage points higher for the super simplified regime. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

reporting are likely channels through which individuals remain in the bunching regions. The motivation to remain in the simpler regimes is threefold, as explained above: i) Financial and monetary incentives, which depend on the total income, family situation, and regime; ii) A preference for simplicity. By remaining in the simpler regimes, individuals save on hassle costs and reduce their administrative burdens; iii) Ease of misreporting, for which we showed evidence. In fact, evasion is both part of what might motivate individuals to remain in simpler regimes and what allows them to remain below the eligibility thresholds for these regimes in the first place. Tax simplicity and simplicity of evasion go hand in hand as it is the simplification itself that both eases misreporting and provides an intrinsic value too. A reduced-form identification of these various motives and behaviors is not feasible.

We therefore develop a simple and tractable structural model to disentangle monetary incentives from the simplicity motive and from the ease of misreporting. The model also allows us to quantify the value of tax simplicity. Our estimation method relies on the idea of using the observed bunching for different activities, regimes, and years as the targeted data moments that we want the model to match. More specifically, we jointly estimate the evasion elasticities and the value of tax simplicity, based on the observed bunching at the eligibility thresholds.

Key to this estimation is that we can measure the potential monetary gain (change in tax from one regime to the other) directly for each taxpayer. Taking advantage of the large variability in the incentives faced by taxpayers with different family situations, other incomes, and occupations across the various regimes, and of the fact that those incentives have also changed over time, we derive enough 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 and years, imply a large preference for tax simplicity, a sizable evasion elasticity, and a negligible real income elasticity.

5.1 Model

Preferences. Each agent chooses one of the three regimes: the super simplified, simplified or standard regime ($i = f, m, r$). They then generate actual revenues y_i and report revenues $\tilde{y}_i$. For the ease of exposition, we omit the activity dimension until later. Each agent has a type θ that captures their productivity: higher θ agents have lower utility costs of producing a given level of revenues. The disutility of generating revenues y_i for an agent of type θ is denoted by $h(y_i; \theta)$, increasing in y_i and decreasing in θ. The cost of misreporting revenues is denoted by $g(y_i, \tilde{y}_i)$, increasing in y_i and decreasing in $\tilde{y}_i$. An agent's utility from earning revenue y_i and reporting $\tilde{y}_i$ is thus:

$$u_i(y_i, \tilde{y}_i) = y_i(1 - c_i) - T_i(\tilde{y}_i) - h(y_i, \theta) - g(y_i, \tilde{y}_i) - a_i$$

where c_i is the cost[13] of producing y_i, $T_i(\tilde{y}_i)$ is the total tax liability as a function of reported

[13] We think about this parameter as effective operating costs.

revenues and a_i is a hassle cost. The latter reflects the tax reporting and compliance costs (e.g. administrative accounting requirements, costs of keeping track and complying with the tax procedure). We assume it to be decreasing with the simplicity of the regime: $a_r > a_m > a_f$. Consistent with the earlier tax bunching literature, we assume the following functional forms for $h(y_i;\theta)$ and $g(y_i, \tilde{y}_i)$:

$$h(y_i;\theta) = \frac{\theta}{1+\frac{1}{\varepsilon}}\left(\frac{y_i}{\theta}\right)^{1+\frac{1}{\varepsilon}} \quad \text{and} \quad g(y_i,\tilde{y}_i) = \frac{\kappa_i}{1+\frac{1}{\eta}}\left(\frac{y_i-\tilde{y}_i}{\kappa_i}\right)^{1+\frac{1}{\eta}}$$

where ε is the real income elasticity, η is the evasion elasticity and κ_i a scaling parameter. In accordance with evidence from section 2 and section 4, agents in the simpler regimes can endogenously misreport their income while agents in the standard regime cannot easily misreport their income. Therefore, in the standard regime, the reporting cost is implicitly infinite because of institutional constraints making it hard to misreport.

Modeling the tax discontinuity. The tax liability depends on both the tax base and the tax rate, which can both differ across regimes. In the simplified regime, the taxable income of agents is $(1-\mu)\tilde{y}_m$ with μ a rebate on reported income $\tilde{y}_m$. In the super simplified regime, taxes are directly levied on $\tilde{y}_f$. Finally, the taxable income in the standard regime is $(1-c_r)y_r$, where c_r is the cost of producing gross income y_r. The agent's effective average income tax rate τ_i, which is itself a combination of social contributions and income taxes.[14] We summarize the combination of effective rates and tax bases in each regime in the following table:

Standard regime:	τ_r	is levied on net income	$z_r = (1-c_r)y_r$
Simplified regime:	τ_m	is levied on taxable (reported) income	$z_m = (1-\mu)\tilde{y}_m$
Super simplified regime:	τ_f	is levied on gross (reported) revenues	$z_f = \tilde{y}_f$

5.2 Responses under the Notch

We now describe agents' behaviors' at the eligibility threshold for the simplified regime. Derivations for the super simplified regime are similar. We provide further details in the Appendix B.

Without the notch. Let us start by describing the interior solution. The optimal choices of actual and reported revenues of an agent in the simplified regime are:

[14] In practice, an agent's effective average income tax rate and their social insurance contribution rate depend on their total income (self-employed income, wages and salaries, ordinary capital income, etc.), household composition, activity type, and occupation, as explained in section 2. As a result, both rates could be different across regimes and activities. Further details about the computation of these average tax rates are available in the subsection C.3.

$$y_m = \theta[(1-c_m) - \tau_m(1-\mu)]^{\varepsilon} \quad \text{and} \quad \tilde{y}_m = \theta[(1-c_m) - \tau_m(1-\mu)]^{\varepsilon} - \kappa_i[\tau_m(1-\mu)]^{\eta}$$

For the standard regime, we assume no misreporting cost (i.e $g(\cdot) = 0$) such that an agent reports truthfully its revenues $y_r = \tilde{y}_r$. Then, the interior solution is:

$$y_r = \theta\left[(1-c_r) - (1-c_r)\tau_r\right]^{\varepsilon}$$

Introducing the notch. With the eligibility threshold, there is a marginal agent $y^* + \Delta y^*$ who reports revenues exactly at the threshold y^* but would have reported revenues at $y^* + \Delta y^*$ absent the notch. If they were unconstrained by the notch, their choice would be characterized by reported revenues:

$$y^* + \Delta y^* = (\theta^* + \Delta\theta^*)[(1-c_m) - \tau_m(1-\mu)]^{\varepsilon} - \kappa_i[\tau_m(1-\mu)]^{\eta} \tag{5}$$

and actual revenues:

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

With the notch, this agent reports revenues at the notch, but their actual revenues are $y_m^* = y_m(y^*)$, where y_m^* is given by:

$$\max_{y_m^*} u(y_m^*; y^*) = y_m^*(1-c_m) - \tau_m(1-\mu)y^* - h(y_m^*, \theta^* + \Delta\theta^*) - g(y_m^* - y^*) - a_m$$

which implies:

$$(1-c_m) - \underbrace{\left(\frac{y_m^*}{\theta^* + \Delta\theta^*}\right)^{\frac{1}{\varepsilon}}}_{h'(y_m^*;\theta^*+\Delta\theta^*)} - \underbrace{\left(\frac{y_m^* - y^*}{\kappa_i}\right)^{\frac{1}{\eta}}}_{g'(y_m^*,y^*)} = 0 \tag{6}$$

Let's denote by y_r^I the indifference point in the standard regime, such that the agent is indifferent between earning revenues y_m^* and reporting revenues exactly equal to the threshold y^* or earning y_r^I (which is actual revenues, since there is no misreporting in the standard regime). y_r^I 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 $u_r^I = u_m^*$ gives:

$$y_r^I(1-c_r)(1-\tau_r) - h(y_r^I, \theta^* + \Delta\theta^*) - a_r = \\ y_m^*(1-c_m) - \tau_m(1-\mu)y^* - h(y_m^*, \theta^* + \Delta\theta^*) - g(y_m^* - y^*) - a_m \quad (8)$$

Assumptions. The sets of equations (11)-(14) is a highly non-linear problem and proves to be complicated to solve. We make three further assumptions in order to reduce the dimensions of the problem. First, we take the limit case $\epsilon \to 0$, corresponding to the absence of real response to taxes. This assumption is justified by our results from section 3 where we showed that the bunching pattern cannot be explained by real responses to taxes. Second, we assume $c_m = c_r = c$, which simply means that the underlying business or self-employed activity remains the same from a production function standpoint regardless of the choice of regime. Finally, we assume that the average tax rate is locally uniform around the threshold for a given regime $\times$ activity. Put another way, differences in tax incentives between the simpler regimes and the standard regime only comes from difference in their average tax rates at the threshold. The system of nonlinear equations simplifies to a single equation:

$$[y^*(1-\mu)\tau_m - y_m^*(1-c)\tau_r] + \frac{\kappa}{1+\frac{1}{\eta}}\left(\frac{y_m^* - y^*}{\kappa_i}\right)^{1+\frac{1}{\eta}} - \Delta a = 0 \quad (9)$$

5.3 Identification and Estimation

Solution. For given primitives $\Omega = (\eta, c, a_r, a_m, \kappa)$ and policy parameters $\Phi = (y^*, \tau_m, \tau_r, \mu)$, we can solve for y_m^* using (9). Ultimately, the solution yields a model predicted $\Delta y^*(\Omega, \Phi)$ as a function of the primitives and policy parameters. We can do the equivalent exercise for the super simplified regime. Intuitively, a given observed bunching can be generated by (i) a taste for simplicity (embodied by Δa) and (ii) an evasion response (embodied in the misreporting elasticity η and scaling factor κ). The hassle cost by itself creates a "pure notch" –an increase in the average tax without a change in the marginal tax–even in the absence of monetary incentives and thus has a distortionary effect.

Identification. We now discuss the identification of the different parameters Ω. We use the large variability in the incentives faced by different types of agents at the same eligibility threshold because of their difference in taxes for different regimes and activity.

First, consider the taste for simplicity as defined by the hassle cost. This cost is regime specific (Δa_m, Δa_f) and we expect it to be higher in the super simplified regime compared to the simplified regime. Individuals are more likely to bunch in the former compared to the latter. The difference in average excess masses across the two regimes disciplines the hassle cost.

Second, we consider the evasion elasticity (η) to be homogeneous across regimes and activity. This is motivated by the fact that the opportunity to misreport revenues is similar for these groups

(as they both face the same simplified accounting rules for example (see section 2)). However, the scaling parameter is activity specific ($\kappa_{\text{I\&C Retail}}$, $\kappa_{\text{I\&C Services}}$, $\kappa_{\text{Non Commercial}}$). Intuitively, it may be easier to misreport services or non commercial revenues compared to retail revenues, because the first two are more labor intensive. Given Δa_m and Δa_f, the set of scaling parameters and the evasion elasticity are identified using difference in earnings response across activity.

Estimation. We now explain how we structurally estimate the model using a simulated method of moments. Different agents face widely different incentives because they are in different regimes and activity (because income taxes, social security contribution rates and rebates). As a result, we have many empirical moments Δy^*, which we can target in order to find the parameters that fit best.

We run the estimation for the 2009-2013 period and the 2014-2015 period separately. Our baseline results is based on individuals reporting zero labor earnings. Formally, let i index the regime (super simplified or simplified) and k index the activity (I&C Retail, I&C Services, Non Commercial). For simplicity, we do not index all parameters by the time period. For each regime, activity, and time period, there is a model-predicted bunching interval Δy^*_{ik}. Its empirical counterpart in the data is $\hat{\Delta} y^*_{ik}$.

Remember that the parameters we seek to estimate are the hassle costs, the evasion elasticity and the set of scaling parameters $\Omega = (\Delta a_m, \Delta a_f, \eta, \kappa_{\text{I\&C Retail}}, \kappa_{\text{I\&C Services}}, \kappa_{\text{Non Commercial}})$. We thus have six parameters to estimate and six data moments ($M = 2$ regimes $\times$ 3 activities). The loss function we minimize is $L(\Omega)$:

$$L(\Omega) = g(\hat{\Delta} y^*, \Delta y^*(\Omega))' \times W \times g(\hat{\Delta} y^*, \Delta y^*(\Omega)) \tag{10}$$

where $g(\hat{\Delta} y^*, \Delta y^*(\Omega))$ is a vector of differences between the empirical and observed earnings responses. The weighting matrix W is diagonal and contains the inverse of the empirical earnings responses. Finally, we test the validity of our exercise by running out-of-sample checks using the estimated structural parameters for the full population, including individuals reporting positive labor earnings.

5.4 Estimation Results

Structural parameters. Table 2 shows the result from the estimation. There are several key findings.

First, the evasion elasticity is sizable, slightly above 1. Figure 16 reports the implied evaded amount due to under reporting. It ranges between 500 and 750 euros for the simplified regime and 1150 and 1800 euros for the super simplified regimes, implying substantial levels of misreporting. These numbers are reassuringly in line with the evidence discussed in section 4. Remember that Deprost et al. (2013) find that the average adjustment is between 500 and 700 euros for individuals in the simpler regimes. We focus out attention on individuals that are at the margin between the

simpler regimes and the standard regime, where incentives to evade are much larger than for agents far away from the thresholds.

Second, the value of simplicity is much higher in the super simplified regime, as can be expected. The average hassle cost for the simplified regime relative to the standard one is between 1100 euros and 1600 euros (depending on the period) and for the super simplified regime it is between 6800 euros and 7200 euros per year and per self-employed. These are sizable values in light of the average hourly gross wage of 18.70 euros and a hourly gross minimum wage of 9.31 in 2012 in France.[15] These values, however, are much lower in terms of taxable income. For the simplified regime, the average hassle cost in terms of taxable income is around 400 euros for I&C Retail, around 650 euros for I&C Services and around 900 euros for Non Commercial. For the super simplified regime, the average hassle cost in terms of taxable income is around 2000 euros for I&C Retail, around 3500 euros for I&C Services and around 4600 euros for Non Commercial.

To put these numbers into perspective, we can draw on the existing evidence on hassle costs. Pitt and Slemrod (1989) find that individual itemization entails a cost equal to 0.12% of adjusted gross income. Benzarti (2020) finds a cost of itemizing that is around 0.7% of gross income, which he shows amounts to around 10 to 15 working hours per year. The hassle cost we estimate is somewhat larger because of our specific context. First, we focus our attention on self-employed individuals, a population that is more responsive to tax incentives and able to adjust. Second, the hassle cost also reflects the opportunity to evade. We have seen that the evasion elasticity is large because of the ease of misreporting implied by the simpler regimes. Benzarti (2020) shows that the evasion channel can explain at most 25% of the foregone benefits for joint filers in the 28% marginal tax bracket. In the context of the VAT in Finland, Harju et al. (2019) shows that reporting requirements (compliance costs) are more important than tax variations to explain output response for small firms and entrepreneurs. Using bunching methods, they find that the tax elasticity is small (0.016) and that the compliance cost is large (up to 19% of the value added at the threshold).

Model fit. Figure 17 also shows the percentage deviation between the simulated moments and the empirical moments $(\Delta y^* - \hat{\Delta} y^*)/\hat{\Delta} y^*$. They are very close to the observed earnings responses in the data, suggesting our model describes well the behavior of entrepreneurs. In Figure 18, we perform an out-of-sample check for the goodness-of-fit of our model. We use the parameters estimated on the population with only self-employed revenues to simulate the implied earnings responses for individuals with positive labor earnings.

Discussion. The welfare implications of the existence of these simpler regimes are far from clear and our analysis cannot give a definite answer. In particular, the government may or may not be losing valuable revenues. On the one hand, these regimes were explicitly introduced to facilitate

[15] Information on the gross hourly minimum wage and average hourly wage can be found at these links: https://www.insee.fr/fr/statistiques/serie/000883671 and https://www.insee.fr/fr/statistiques/2508166.

Table 2: Structural Parameter Estimates

Parameter	Description	Period 1	Period 2
η	Elasticity of evasion	1.1	1.2
κ	Scaling parameter for evasion ...		
$\kappa_{\text{I\&C Retail}}$	... in I&C Retail	120	160
$\kappa_{\text{I\&C Services}}$	... in I&C Services	70	60
$\kappa_{\text{Non Commercial}}$	... in Non Commercial	120	130
Δa	Difference in hassle cost between the standard ...		
Δa_m	... and simplified regimes (in euros)	1600	1100
Δa_f	... and super simplified regimes (in euros)	7200	6800

Notes: This table shows the results from the structural estimation, based on the data moments for the 2009-2013 period and the 2014-2015 period and for individuals reporting zero labor earnings. This simulation apply the tax rates based on taxable income.

the creation of firms that would otherwise not exist, and to shift work from the informal to the formal sector. Consistent with this, Barruel et al. (2012) shows that three quarters of the firms created in the first semester of 2010 under the super simplified regime would not have been created without the introduction of this regime. On the other hand, it is unclear exactly whether this additional firm creation generates completely new income (that is taxed at lighter rates and in part evaded) or represents shifts away from employment (which is taxed and is harder to evade). It is possible that the revenue losses induced by tax evasion, which happen at the margin of the thresholds, may turn out small compared to the overall gains of having more self-employed businesses.

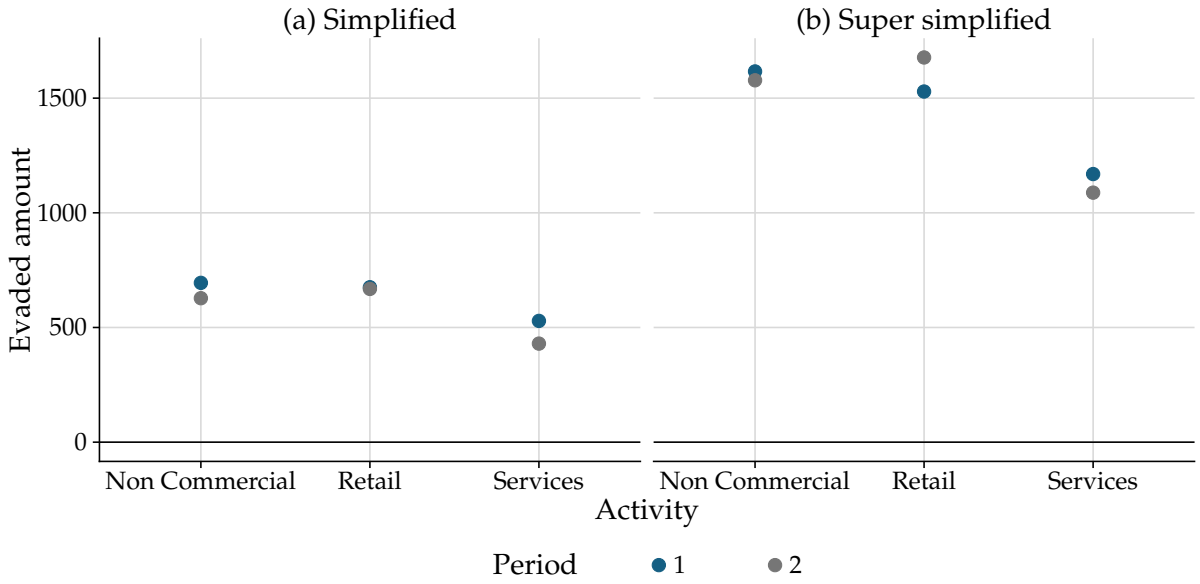
To sum up, evasion is very important, both as an additional value from the simplified regimes and as the means to remain below the threshold. The hassle costs act like a notch and provide a strong incentive to remain below the eligibility threshold.

6 Conclusion

We studied how French self-employed respond to the creation and incentives of simplified tax and reporting regimes. The self-employed bunch massively below the eligibility ceilings for the simplified and super simplified regimes. We started by providing evidence suggesting that at least some of this bunching comes from tax evasion. First, we observe a salient discontinuity in the self-employed earnings dynamic and in the probability to remain close to threshold. Second, the tax returns are more often round numbers close to the threshold than far from it, an indication that the reported figure is more likely to be forged. Third, there is evidence for income shifting within the household. Fourth, we can uncover some level of "hidden employment," whereby employers prefer contracting out work previously done in-house so the employees can benefit from the tax advantages and potentially be able to evade more taxes.

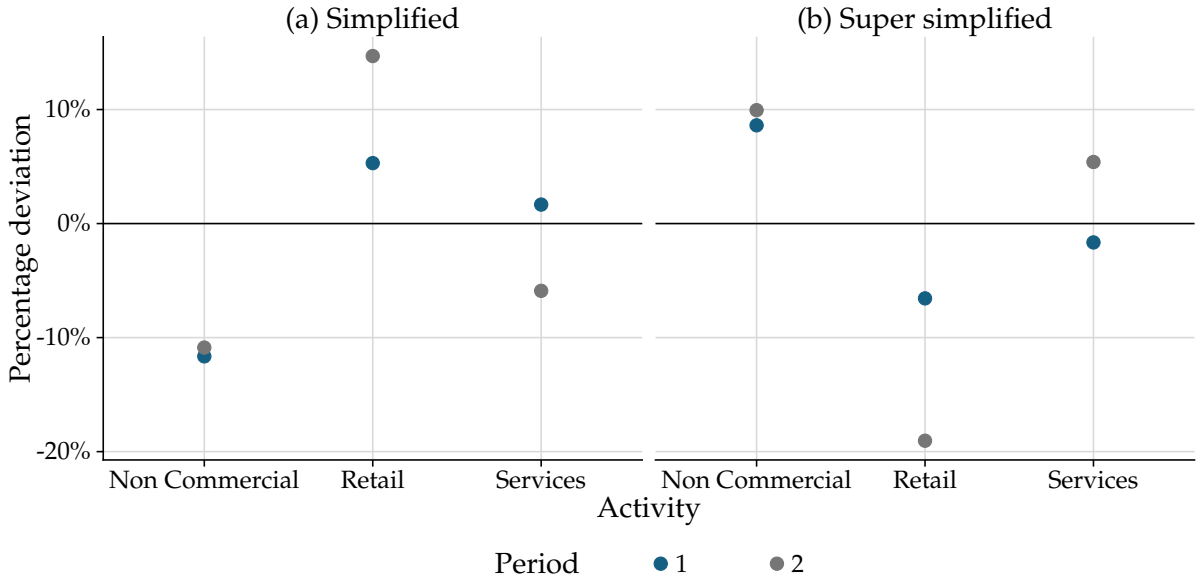
We then used our reduced form bunching estimates as data moments to be matched by a

Figure 16: Evaded amount of the structural estimation



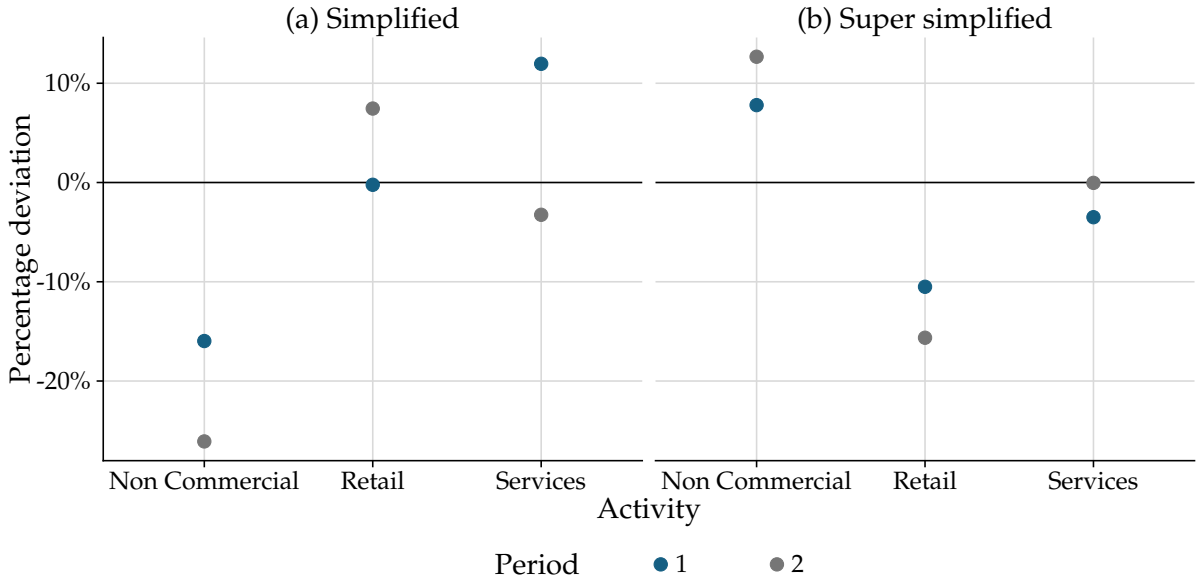
Notes: This table shows evaded amount in euros from the structural estimation based on the period 2009-2013 (period 1) and the period 2014-2015 (period 2), for individuals reporting zero labor earnings. Results are reported separately for the simplified (panel (a)) and the super simplified (panel (b)) regimes.

Figure 17: Empirical fit of the structural estimation



Notes: This table shows percentage deviation between the simulated moments and the empirical moments, $(\Delta y^* - \hat{\Delta} y^*)/\hat{\Delta} y^*$, from the structural estimation based on the period 2009-2013 (period 1) and the period 2014-2015 (period 2), for individuals reporting zero labor earnings. Results are reported separately for the simplified (panel (a)) and the super simplified (panel (b)) regimes.

Figure 18: Out-of-sample fit of the structural estimation



Notes: This table shows percentage deviation between the simulated moments and the empirical moments, $(\Delta y^* - \hat{\Delta} y^*)/\hat{\Delta} y^*$, from the structural estimation based on the period 2009-2013 (period 1) and the period 2014-2015 (period 2), for the full population. Results are reported separately for the simplified (panel (a)) and the super simplified (panel (b)) regimes.

structural model to disentangle the motives for individuals to remain in these simpler regimes. We found that the structural parameters that can best explain the observed bunching across different regimes and activities, turn out to display a large preference for tax simplicity and a sizeable evasion elasticity.

Our analysis could be extended in several interesting directions. A first avenue for future research would be to study whether tax simplicity improves the chances of success of a self-employed activity: do the self-employed individuals who understand tax incentives better end up doing better even in the long-run? Do they become true "entrepreneurs" and ultimately job creators? A second avenue would be to evaluate the general equilibrium effects of the existence of the simplified and super simplified regimes and their impact on public finances and welfare.

References

Abeler, J. and S. Jäger (2015, August). Complex Tax Incentives. American Economic Journal: Economic Policy 7(3), 1–28.

Alinaghi, N., J. Creedy, and N. Gemmell (2021, July). Elasticities of taxable income and adjustment costs: Bunching evidence from New Zealand. Oxford Economic Papers 73(3), 1244–1269.

Almunia, M. and D. Lopez-Rodriguez (2018, February). Under the Radar: The Effects of Monitoring Firms on Tax Compliance. American Economic Journal: Economic Policy 10(1), 1–38.

Barruel, F., S. Thomas, S. Darriné, and H. Mariotte (2012). Trois auto-entrepreneurs sur quatre n'auraient pas créé d'entreprise sans ce régime. Insee Première 1388, 1–4.

Bastani, S. and H. Selin (2014, January). Bunching and non-bunching at kink points of the Swedish tax schedule. Journal of Public Economics 109, 36–49.

Benzarti, Y. (2020, November). How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs. American Economic Journal: Economic Policy 12(4), 38–57.

Bergolo, M., G. Burdin, M. De Rosa, M. Giaccobasso, and M. Leites (2021, October). Digging Into the Channels of Bunching: Evidence from the Uruguayan Income Tax. The Economic Journal 131(639), 2726–2762.

Bhargava, S. and D. Manoli (2015, November). Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment. American Economic Review 105(11), 3489–3529.

Bíró, A., D. Prinz, and L. Sándor (2022, November). The minimum wage, informal pay, and tax enforcement. Journal of Public Economics 215, 104728.

Blumenthal, M. and J. Slemrod (1992, June). The compliance cost of the U.S. individual income tax system: A Second look after tax reform. National Tax Journal 45(2), 185–202.

Boning, W. C., J. Guyton, R. Hodge, and J. Slemrod (2020, October). Heard it through the grapevine: The direct and network effects of a tax enforcement field experiment on firms. Journal of Public Economics 190, 104261.

Brockmeyer, A., S. Smith, M. Hernandez, and S. Kettle (2019, August). Casting a Wider Tax Net: Experimental Evidence from Costa Rica. American Economic Journal: Economic Policy 11(3), 55–87.

Carrillo, P., D. Pomeranz, and M. Singhal (2017, April). Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement. American Economic Journal: Applied Economics 9(2), 144–164.

Chen, Z., Z. Liu, J. C. Suárez Serrato, and D. Y. Xu (2021, July). Notching r&d investment with corporate income tax cuts in china. American Economic Review 111(7), 2065–2100.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011, May). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records *. The Quarterly Journal of Economics 126(2), 749–804.

Chetty, R., J. N. Friedman, and E. Saez (2013, December). Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings. American Economic Review 103(7), 2683–2721.

Coles, J. L., E. Patel, N. Seegert, and M. Smith (2022). How Do Firms Respond to Corporate Taxes? Journal of Accounting Research 60(3), 965–1006.

Cour des Comptes (2014). Les organismes de gestion agréés, 40 ans apres. communication à la Commission des finances, de l'économie générale et du contrôle budgétaire de l'Assemblée nationale.

Craig, A. and J. Slemrod (2022). Tax Knowledge and Tax Manipulation: A Unifying Model. NBER Working Paper No. 30151.

de Paula, Á. and J. A. Scheinkman (2010, October). Value-Added Taxes, Chain Effects, and Informality. American Economic Journal: Macroeconomics 2(4), 195–221.

Deprost, P., P. Laffon, and D. Imbaud (2013). Évaluation du régime de l'auto-entrepreneur. Rapport de l'Inspection générale des finances et de l'Inspection générale des affaires sociales.

Devereux, M. P., L. Liu, and S. Loretz (2014, May). The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records. American Economic Journal: Economic Policy 6(2), 19–53.

Diamond, R. and P. Persson (2016). The long-term consequences of teacher discretion in grading of high-stakes tests. Technical report, National Bureau of Economic Research.

Engström, P. and B. Holmlund (2009, August). Tax evasion and self-employment in a high-tax country: Evidence from Sweden. Applied Economics 41(19), 2419–2430.

Farhi, E. and X. Gabaix (2020, January). Optimal Taxation with Behavioral Agents. American Economic Review 110(1), 298–336.

Feldman, N. E., P. Katuščák, and L. Kawano (2016, March). Taxpayer Confusion: Evidence from the Child Tax Credit. American Economic Review 106(3), 807–835.

Gelber, A. M., D. Jones, and D. W. Sacks (2020, January). Estimating Adjustment Frictions Using Nonlinear Budget Sets: Method and Evidence from the Earnings Test. American Economic Journal: Applied Economics 12(1), 1–31.

Grottke, M. and J. Lorenz (2017). Tax consultants' incentives: A game-theoretic investigation into the behavior of tax consultants, taxpayers, and the tax authority in a setting of tax complexity. Passauer Diskussionspapiere - Betriebswirtschaftliche Reihe, No. B-30-17.

Harju, J. and T. Matikka (2016, August). The elasticity of taxable income and income-shifting: What is "real" and what is not? International Tax and Public Finance 23(4), 640–669.

Harju, J., T. Matikka, and T. Rauhanen (2019). Compliance costs vs. tax incentives: Why do entrepreneurs respond to size-based regulations? Journal of Public Economics 173, 139–164.

Johns, A. and J. Slemrod (2010, September). The distribution of income tax noncompliance. National Tax Journal 63(3), 397–418.

Katz, L. F. and A. B. Krueger (2017). The role of unemployment in the rise in alternative work arrangements. American Economic Review 107(5), 388–92.

Katz, L. F. and A. B. Krueger (2019). The rise and nature of alternative work arrangements in the United States, 1995-2015. ILR Review [Internet] 72(2), 382–416.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. Econometrica 79(3), 651–692.

Kleven, H. J. and M. Waseem (2013, May). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan*. The Quarterly Journal of Economics 128(2), 669–723.

LaLumia, S., J. M. Sallee, and N. Turner (2015, May). New Evidence on Taxes and the Timing of Birth. American Economic Journal: Economic Policy 7(2), 258–293.

le Maire, D. and B. Schjerning (2013, November). Tax bunching, income shifting and self-employment. Journal of Public Economics 107, 1–18.

Mortenson, J. A. and A. Whitten (2020, August). Bunching to Maximize Tax Credits: Evidence from Kinks in the US Tax Schedule. American Economic Journal: Economic Policy 12(3), 402–432.

Naritomi, J. (2019, September). Consumers as Tax Auditors. American Economic Review 109(9), 3031–3072.

Parker, S. C. (2003). Does Tax Evasion Affect Occupational Choice?*. Oxford Bulletin of Economics and Statistics 65(3), 379–394.

Pirttilä, J. and H. Selin (2011). Income Shifting within a Dual Income Tax System: Evidence from the Finnish Tax Reform of 1993*. The Scandinavian Journal of Economics 113(1), 120–144.

Pitt, M. M. and J. Slemrod (1989). The compliance cost of itemizing deductions: Evidence from individual tax returns. The American Economic Review 79(5), 1224–1232.

Pomeranz, D. (2015, August). No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. American Economic Review 105(8), 2539–2569.

Saez, E. (2010, August). Do Taxpayers Bunch at Kink Points? American Economic Journal: Economic Policy 2(3), 180–212.

Slemrod, J. (2005, May). The Etiology of Tax Complexity: Evidence from U.S. State Income Tax Systems. Public Finance Review 33(3), 279–299.

Tazhitdinova, A. (2018, September). Reducing evasion through self-reporting: Evidence from charitable contributions. Journal of Public Economics 165, 31–47.

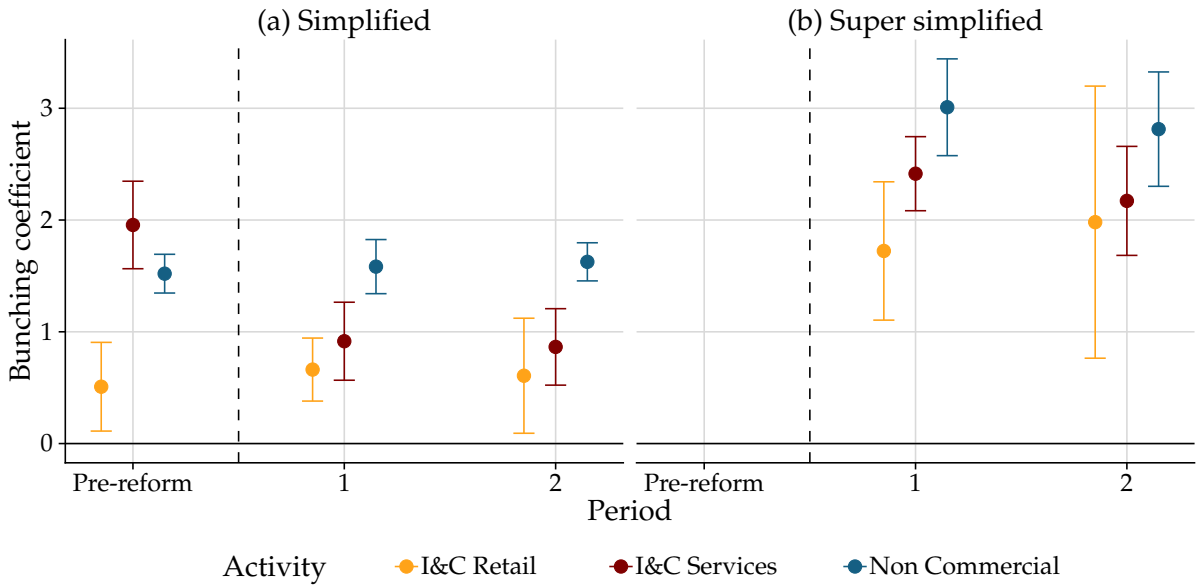
Tazhitdinova, A. (2020, April). Do only tax incentives matter? Labor supply and demand responses to an unusually large and salient tax break. Journal of Public Economics 184, 104162.

Varian, H. R. (1972). Letters to the editor. The American Statistician 26(3), 62–66.

Warskett, G., S. L. Winer, and W. Hettich (1998, May). The Complexity of Tax Structure in Competitive Political Systems. International Tax and Public Finance 5(2), 123–151.

Zwick, E. (2021, May). The Costs of Corporate Tax Complexity. American Economic Journal: Economic Policy 13(2), 467–500.

Figure A.1: Bunching estimation by regime, activity and period, conditional on having no labor earnings



Notes: The figure plots the bunching coefficients b from subsection 3.1 by regime, activity and period2), for individuals reporting zero labor earnings. The pre-reform period is 2006-2008, period 1 is 2009-2013 and period 2 is 2014-2015. The counterfactual distribution is fitted using a smooth polynomial as explained in subsection 3.1. There is significant bunching for all categories, but systematically higher for the super simplified regime. Standard errors are calculated using a bootstrap procedure in by random resampling (n = 400) of the residuals.

A Additional Figures

B Structural Model

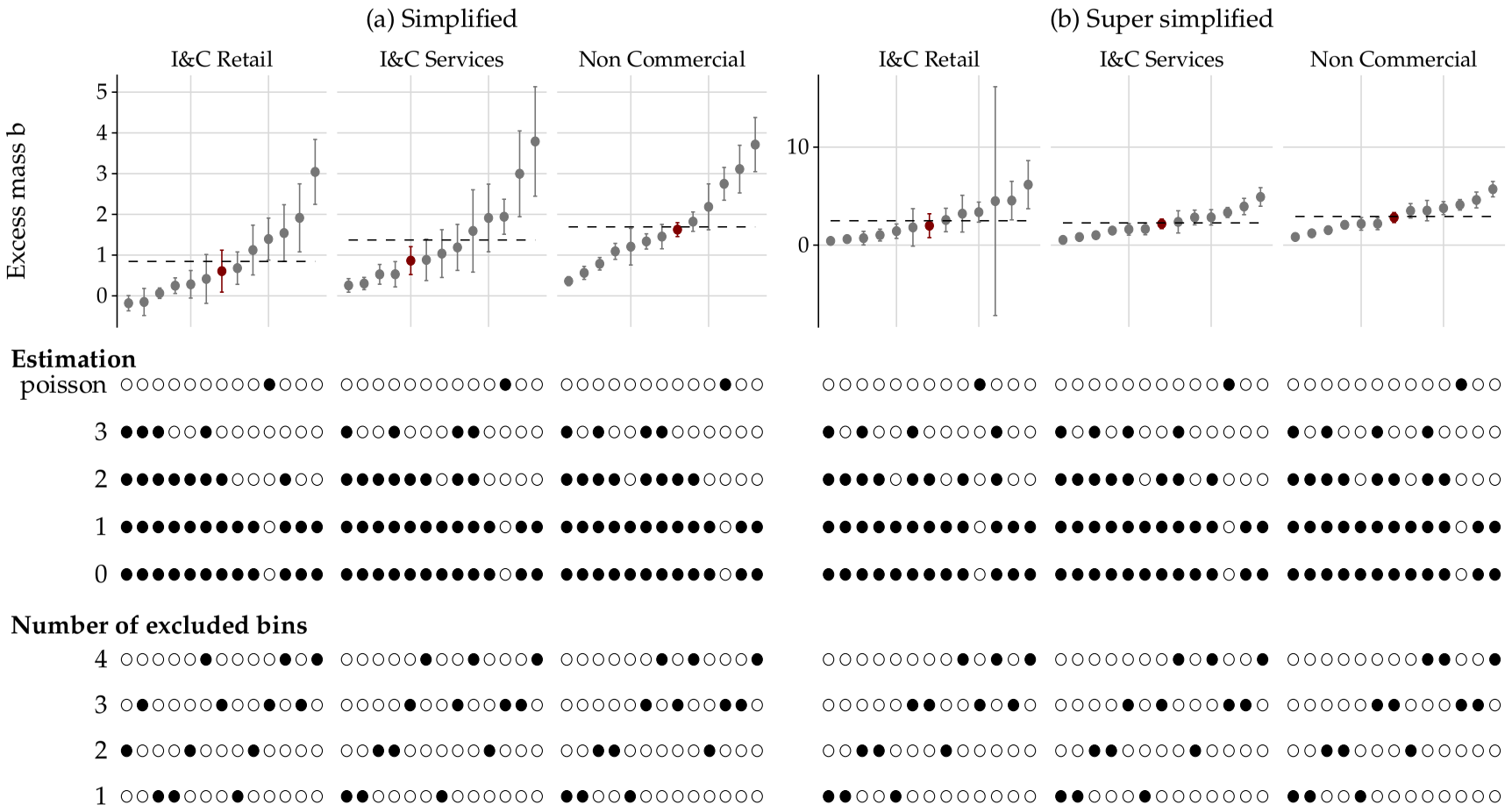
B.1 Setup

Preferences Each agent choose to produce under the super simplified, simplified or standard regime ($i = f, m, r$), which generates real revenues y_i and reported revenues $\tilde{y}_i$. They have a type θ that captures her productivity and that reduces their cost of producing a given level of revenues. The disutility of generating revenues y_i for an agent of type θ is denoted by $h(y_i; \theta)$, increasing in y_i and decreasing in θ. The cost of misreporting revenues is denoted by $g(y_i, \tilde{y}_i)$, increasing in y_i and decreasing in $\tilde{y}_i$. An agent's utility from earning revenue y_i and reporting $\tilde{y}_i$ is thus:

$$u_i(y_i, \tilde{y}_i) = y_i(1 - c_i) - T_i(\tilde{y}_i) - h(y_i, \theta) - g(y_i, \tilde{y}_i) - a_i$$

where c_i is the cost of producing y_i, $T_i(\hat{y}_i)$ is the total tax liability as a function of reported revenues and a_i a hassle cost. This hassle cost is decreasing with the simplicity of the regime: $a_r > a_m > a_f$.

Figure A.2: Bunching estimates by regime, activity and period



Parametric assumptions Consistent with previous the literature, we assume the following functional forms for $h(y_i;\theta)$ and $g(y_i,\tilde{y}_i)$:

$$h(y_i;\theta) = \frac{\theta}{1+\frac{1}{\varepsilon}}\left(\frac{y_i}{\theta}\right)^{1+\frac{1}{\varepsilon}} \quad \text{and} \quad g(y_i,\tilde{y}_i) = \frac{\kappa}{1+\frac{1}{\eta}}\left(\frac{y_i-\tilde{y}_i}{\kappa}\right)^{1+\frac{1}{\eta}}$$

where ε is the real income elasticity, η is the evasion elasticity and κ a scaling parameter mapping $h(\cdot)$ and misreporting. In the standard regime, the reporting cost is implicitly infinite because of institutional constraints making it hard to misreport.

B.2 Full model

B.2.1 Simplified regime

Without the notch Let us start by describing the interior solution (without the notch) for the simplified regime. Now, the total tax liability is defined has $T_i(\tilde{y}_m) = \tilde{y}_m(1-\mu)\tau_m$. The optimal choice of real and reported revenues of an agent are (from the first order conditions on y_m and $\tilde{y}_m$):

$$\begin{aligned}
\frac{\partial u_m}{\partial y_m} &= (1-c_m) - \left(\frac{y_m}{\theta}\right)^{\frac{1}{\varepsilon}} - \left(\frac{y_m-\tilde{y}_m}{\kappa}\right)^{\frac{1}{\eta}} = 0 \\
\frac{\partial u_m}{\partial \tilde{y}_m} &= -(1-\mu)\tau_m + \left(\frac{y_m-\tilde{y}_m}{\kappa}\right)^{\frac{1}{\eta}} = 0
\end{aligned}$$

which implies that:

$$y_m = \theta[(1-c_m)-\tau_m(1-\mu)]^{\varepsilon} \quad \text{and} \quad \tilde{y}_m = \theta[(1-c_m)-\tau_m(1-\mu)]^{\varepsilon} - \kappa[\tau_m(1-\mu)]^{\eta}$$

We now describe the interior solution for the standard regime. We assume no misreporting cost (i.e $g(\cdot)=0$). It means that an agent reports truthfully its revenues $y_r = \tilde{y}_r$. It implies that:

$$y_r = \theta\left[(1-c_r)-(1-c_r)\tau_r\right]^{\varepsilon}$$

Introducing the notch With the eligibility threshold, there is a marginal agent $y^* + \Delta y^*$ who reports revenues exactly at the threshold y^* but would have reported revenues at $y^* + \Delta y^*$ absent the notch. If she were unconstrained by the notch, her choice would be characterized by reported revenues:

$$y^* + \Delta y^* = (\theta^* + \Delta\theta^*)[(1-c_m)-\tau_m(1-\mu)]^{\varepsilon} - \kappa[\tau_m(1-\mu)]^{\eta} \tag{11}$$

and actual revenues:

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

With the notch, this agent reports revenues at the notch, but his actual revenues are $y_m^* = y_m(y^*)$ as a function of the report, where y_m^* is given by:

$$\max_{y_m^*} u(y_m^*; y^*) = y_m^*(1 - c_m) - \tau_m(1 - \mu)y^* - h(y_m^*, \theta^* + \Delta\theta^*) - g(y_m^* - y^*) - a_m$$

which implies:

$$(1 - c_m) - \underbrace{\left(\frac{y_m^*}{\theta^* + \Delta\theta^*}\right)^{\frac{1}{\varepsilon}}}_{h'(y_m^*;\theta^*+\Delta\theta^*)} - \underbrace{\left(\frac{y_m^* - y^*}{\kappa}\right)^{\frac{1}{\eta}}}_{g'(y_m^*,y^*)} = 0 \tag{12}$$

Let's denote by y_r^I the indifference point in the standard regime, such that the agent is indifferent between earning revenues y_m^* and reporting revenues exactly equal to the threshold y^* or earning y_r^I (which is actual revenues, since there is no misreporting in the standard regime). y_r^I is interior, and hence characterized by the optimal (tangency) condition in standard regime:

$$y_r^I = (\theta^* + \Delta\theta^*)[(1 - c_r)(1 - \tau_r)]^{\varepsilon} \tag{13}$$

The indifference condition $u_r^I = u_m^*$ gives:

$$\begin{aligned} y_r^I(1 - c_r)(1 - \tau_r) - h(y_r^I, \theta^* + \Delta\theta^*) - a_r = \\ y_m^*(1 - c_m) - \tau_m(1 - \mu)y^* - h(y_m^*, \theta^* + \Delta\theta^*) - g(y_m^* - y^*) - a_m \end{aligned} \tag{14}$$

B.3 Model with no real elasticity

We now take the limit case $\epsilon \to 0$, corresponding to no real response.

$$\theta^* + \Delta\theta^* = y_m^* \times \left[(1 - c_m) - \left(\frac{y_m^* - y^*}{\kappa}\right)^{\frac{1}{\eta}}\right]^{-\varepsilon} \longrightarrow y_m^* \tag{15}$$

$$y_r^I = (\theta^* + \Delta\theta^*)[(1 - c_r)(1 - \tau_r)]^{\varepsilon} \longrightarrow y_m^* \tag{16}$$

$$\left[y_r^I(1 - c_r)(1 - \tau_r)\right] - [y_m^*(1 - c_m) - \tau_m(1 - \mu)y^* - g(y_m^*, y^*)] - \Delta a = 0 \tag{17}$$

$$y^*+\Delta y^* = (\theta^*+\Delta\theta^*)[(1-c_m)-\tau_m(1-\mu)]^{\varepsilon}-\kappa[\tau_m(1-\mu)]^{\eta} \longrightarrow y_m^* = (y^*+\Delta y^*)+\kappa[\tau_m(1-\mu)]^{\eta} \quad (18)$$

Combining the previous couple of equations, we have a reduced form equation:

$$y_m^*(c_m - c_r) + [y^*(1-\mu)\tau_m - y_m^*(1-c_r)\tau_r] + \frac{\kappa}{1+\frac{1}{\eta}}\left(\frac{y_m^* - y^*}{\kappa}\right)^{1+\frac{1}{\eta}} - \Delta a = 0 \quad (19)$$

If we assume $c_r = c_m$, it reduces to:

$$[y^*(1-\mu)\tau_m - y_m^*(1-c)\tau_r] + \frac{\kappa}{1+\frac{1}{\eta}}\left(\frac{y_m^* - y^*}{\kappa}\right)^{1+\frac{1}{\eta}} - \Delta a = 0 \quad (20)$$

C Data Construction

C.1 Data

POTE. Our main dataset contains the universe of French tax returns over the period 2006-2015, compiled by the French Internal Revenue Service. 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. Importantly, it is a panel data with unique individual identifiers and unique household identifiers (both anonymized).

New entrepreneurs information system. The second dataset we use is a survey provided by the French National Statistics Institute available for 2010 and 2014. It asks entrepreneurs about their experience during the first years of their business activity.

C.2 Construction of the Sample

The following section describes the construction of the sample for the *POTE* dataset, associated with the replication package folder (`0_data_creation`).

Population. Our benchmark sample consists of all individuals who are French fiscal residents in mainland France and are between 30 and 59 years of age in a given year, who are are main filers (i.e we exclude dependants such as children). We also delete individuals with any change in their marital status. More precisely, we do not include years in which individuals divorce or experience the death of the spouse. Moreover, we do not include years in which individuals get marry, for years before 2010 (included). The reason is that the French personal income tax is reported at the household level which generates different reporting obligations for individuals that change their marital status in a given tax year. In addition, we keep only individuals and households that are

uniquely observed in a given year. Finally, we restrict our analysis to the 2006-2015 period. Finally, we limit our analysis

Self-employment restrictions. We start by restricting our sample to individuals that have some self-employed revenues in the simple regime, the super simplified regime or the standard regime in a given year. Then, we keep individuals with self-employment income that are uniquely defined in a regime and activity. It excludes for example individuals with self-employment in different regimes and activity. We also delete households that have any agricultural income because they are subject to specific tax parameters.

In the case of the super simplified regime, there is an additional requirement on the family income as of year $t-2$, which has to be below a year-specific threshold f_t^* that corresponds to the third tax bracket cutoff. We delete individuals filling under the super simplified regime with family income above this threshold. After discussion with the tax administration, this situation should not be possible but is observed in the data. Various explanations are possible: differences in reporting requirements, errors or unobserved changes in tax regimes. Since we have no further information, we focus on individuals in the super simplified regime that are also eligible for it.

C.3 Variables Construction

In this section, we describe the construction of important variables for our analysis.

Self-employed revenues. The full construction of self-employed revenues by regime and activity is available in the SAS file `2_macro_sample_se.sas`.

Labor earnings We adopt a strict definition of labor earnings. It includes wages and salaries in item box **1AJ** for the first filer and in item box **1BJ** for the second filer. It encompasses most sources of labor earnings. An individual is considered to have labor earnings if the reported amount is strictly greater than zero.

Tax rates We consider effective average tax rates that are regime, activity and period specific. The effective average tax rate τ_i is a combination of an income tax τ_{inc} and social contributions levied with rate τ_{sc}. We describe below the system for the period 2009-2015.

For the super simplified regime, it is simple: income taxes and social contributions are computed on gross self-employed revenues y with flat tax rates. Once they are applied, the business is released from the social contributions and income taxes. The flat income tax is equal to 1% for I&C Retail activities, to 1.7% for I&C Services activities and to 2.2% for Non Commercial activities. The flat rate for social contributions is between 12% and 15% for I&C Retail activities, between 21% and 25% for I&C Services activities and between 18% and 24% for Non Commercial activities.

For the simplified regime, two options are available regarding taxes and social contributions. First, to compute both on self-employed taxable income (after applying the rebate), and with pro-

gressive income taxes and social contributions. Second, to have a progressive income tax and a flat rate for social contributions, similar to the super simplified regime. In practice, we cannot observe if self-employed individuals adopt one option or the other. In the paper, we adopt the first option. The advantage of this scenario is that the difference in effective average tax rates between the simplified regime and the standard regime at the threshold is close to 0, which is a limiting case for the identification of hassle costs. Results for the structural model under the alternative computation are qualitatively similar and available upon request. We compute average tax rates close the threshold. The average income tax rate is directly observable is the data. It is between 2% and 8%, depending on the activity and period. The average social contributions tax rate is between 47% and 50% for I&C Retail and I&C Services activities, and around 36% for Non Commercial activities. Note that the computation of the social contribution tax rates is more challenging because it is a collection of multiple taxes, on different sub-activities we do not observe with precision. We compute them based on information available here: https://www.ipp.eu/en/ipp-tax-and-benefit-tables/. The full detail on the computation is available upon request.

Finally, we compute the counterfactual effective average tax rates (if individuals would have chosen the standard regime). To do so, income and social contributions are computed both on self-employed taxable income under the assumption that the cost is equal to the rebate, and with progressive income taxes and social contributions (similar to the simplified regime). We use the average income tax rate directly available in the data, which take into account the self-employed incomes for the super simplified regime. Then, we compute the average social contributions tax rate using the same method as described for the simplified regime. The implicit assumptions is that the effective average tax rate computed for individuals at the threshold is a good approximation for the situation about the threshold in the standard regime.