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

Philippe Aghion Maxime Gravoueille

Matthieu Lequien Stefanie Stantcheva

May 7, 2024

Abstract

A common complaint about tax systems is that they are too complex and several countries offer simplified regimes to subsets of taxpayers. But do such regimes simplify tax reporting or just make it easier to evade taxes? We use panel data on the universe of income tax returns in France and the introduction of simplified tax regimes for the self-employed to assess whether individuals' shift towards these simpler tax regimes–among others, captured by substantial bunching below the eligibility thresholds–is driven by the desire for simplicity or by motives related to tax evasion. Tax evasion plays a significant role in explaining individuals' attraction to simpler tax regimes. We develop a structural model to quantitatively assess the importance of simplicity and evasion motives. Our preferred estimates indicate a significant preference for tax simplicity, increasing in the degree of simplicity, ranging from 162 to 942 euros per year per self-employed individual, along with a sizable evasion elasticity.

JEL codes: H26, H24, H21

Keywords: Taxation, Entrepreneurship, Evasion, Complexity, Bunching

*Philippe Aghion: Collège de France, INSEAD, and London School of Economics (philippe.aghion@insead.edu); Maxime Gravoueille: Monash University (maxime.gravoueille@monash.edu); Matthieu Lequien: INSEE; Stefanie Stantcheva: Harvard, NBER, and CEPR (sstantcheva@fas.harvard.edu). A much earlier version of this paper, not intended for publication, circulated under the title "Tax Simplicity and Heterogeneous Learning" as NBER Working paper 24049. The earlier working paper did not have panel data, which only subsequently became available to researchers, so this new paper provides a different analysis made possible by this data. The original working paper was in collaboration with Ufuk Akcigit, whose invaluable inspiration greatly contributed to shaping it and to whom we are indebted. We thank Salome Baslandze, Antonin Bergeaud, Pierre Boyer, Antoine Bozio, Antoine Ferey, Bertrand Garbinti, Johnathan Goupille-Lebret, Larry Katz, Henrik Kleven, Etienne Lehmann, Anasuya Raj, and participants at the NBER Entrepreneurship Summer Institute, Paris seminar participants for useful comments and feedback on the earlier version. Stantcheva gratefully acknowledges support from NSF CAREER Grant #1654517. Access to confidential data is offered by the CASD – Centre d'Accès Sécurisé aux Données.

1 Introduction

"Simplicity is the ultimate sophistication", wrote Leonardo da Vinci. This is particularly true for tax policy: the best designed tax incentives may turn out to be ineffective if people do not understand them. 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. 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.

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, we exploit individual panel information and the introduction 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 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–each increasing in their degree of tax simplicity. 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.

There are at least three motives for the *quest for simplicity*, by which we mean the desire to choose a simpler tax regime: (1) *monetary incentives* - given the differences in income tax base and rates in the different regimes, taxpayers might 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) *tax evasion* - it is much easier to misreport income in the simpler regimes than in the standard regime. In this paper we argue that 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 driven by both tax simplicity and the ease of tax evasion.

Our focus on the self-employed stems from two main considerations. First, those are typically shown to be less constrained than wage earners and can more easily adjust their incomes to tax incentives (e.g., Kleven & Waseem, 2013; Saez, 2010), 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 wide variety of fiscal "regimes" that differ in their institutional parameters, such as accounting and reporting requirements, eligibility thresholds, tax rates and rebates. This leads to substantial differences in monetary incentives, tax simplicity, and ease of misreporting across regimes. Moreover, institutional parameters vary across activities within fiscal regimes and have undergone changes over time, providing valuable additional policy variation that helps our estimation.

In a nutshell, we find evidence of a willingness to remain in the simpler regimes, i.e. of a quest for simplicity, from observing significant bunching at the eligibility thresholds for the simpler self-employment regimes. Furthermore, we find that this willingness increases with the degree of simplicity of the self-employment regime: we observe more bunching at the eligibility threshold for the super simplified regime than at the one for the simplified regime. We also show that the observed bunching at the eligibility thresholds is to a large extent driven by tax evasion through misreporting, where tax evasion itself is made possible by–and increases the attraction of–simpler tax regimes.

More precisely, we exploit new individual panel data from the French tax authority[1] to analyze individuals' choice of tax regime and then infer the extent to which this choice is driven by tax simplicity versus or 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. 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 preference for staying in simpler regimes. 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

[1]The French tax authority is the called the *Direction générale des Finances publiques* or DGFip.

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, even when there is little or no discontinuity in monetary incentives around the threshold. This in turn reflects the importance of tax simplicity and tax evasion motives above and beyond pure monetary incentives.

Second, we show that indeed tax evasion motives partly explain the desire to remain in the simpler regimes. 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 initially bunch at the eligibility threshold of simpler regimes significantly reduce their self-employed income growth rates to remain below the eligibility threshold in the following year, a behavior we do not observe among individuals in the standard regime. Additional evidence for evasion comes from the fact that revenue statements are more often round numbers and end in non-random digits 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"–a form of tax avoidance or evasion–whereby employers contract out work previously done in-house, effectively circumventing costly labor contracts and relabeling self-employed work as employment.

Finally, we quantitatively assess the importance of tax simplicity and tax 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. The monetary equivalent for the taste of simplicity in our preferred estimation ranges from, depending on the activity, to 162 to 235 euros in the simplified regime and from 651 to 942 euros in the super simplified regime, per year and per self-employed individual. For comparison, the average hourly gross wage in France in 2012 was 18.70 euros.[2] This means that the hassle cost, on average, equates to between 9 and 13 hours of work in the simplified regime and between 35 and 50 hours of work in the super simplified regime.

[2]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.

Our paper lies at the intersection of several strands of the literature. Most closely related to our analysis are the literatures on tax simplicity, on tax evasion, and on bunching and structural estimation.

Tax simplicity: 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 at 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. Additionally, Blesse et al. (2019) and Benzarti (2020) demonstrate, using survey data, that people strongly prefer simpler tax systems. 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 of the institutional context (such as the interplay between public authorities, tax preparers and taxpayers) in shaping tax complexity. Finally, Farhi and Gabaix (2020) develop a theory of optimal taxation considering behavioral agents with misperceptions. We contribute to this literature by exploiting individual panel information on the choice between different tax regimes to provide evidence of a quest for 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.[3] Similarly, Saez (2010) and LaLumia et al. (2015) demonstrate that self-employed earners respond to tax incentives created by the EITC in the US. 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. A growing literature investigates the various mechanisms underlying tax evasion among individuals and firms (e.g., Bergolo et al., 2021; Best, 2014; Bohne & Nimczik, 2018; Fisman & Wei, 2004; Harju & Matikka, 2016). Additionally, a substantial body of research examines the impact of tax enforcement on tax evasion and compliance (e.g., Almunia & Lopez-Rodriguez, 2018; Boning et al., 2020; Brockmeyer et al., 2019; Carrillo et al., 2017; de Paula & Scheinkman, 2010; Kleven et al., 2011; Naritomi, 2019; Pomeranz, 2015; Tazhitdinova, 2018). 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,

[3]Parker (2003) finds no effect of tax incentives on the occupational choice to be self-employed and on tax evasion in Great-Britain.

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. 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.[4] 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.[5] Devereux et al. (2014) and Coles et al. (2022) focus on responses to corporate taxes. 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. 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. 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. 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 2006-2015, 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 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 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.

[4] Tazhitdinova (2020) also documents 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.

[5] 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.

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 in France. 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.[6] These categories are not always intuitive. For instance, developing and selling software pertains to the Non Commercial type, while purchasing and selling equipment goods pertains to the I&C Retail category. Similarly, bakery, butchery, or restaurant businesses are counted as I&C Retail activities, while construction work, 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.

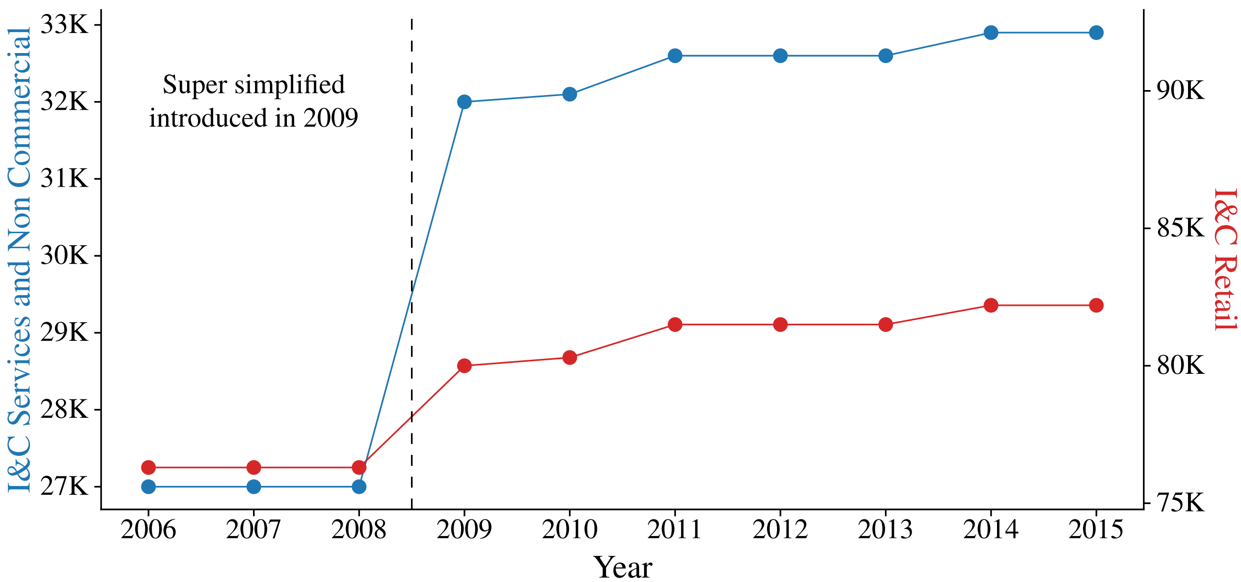
Self-employed regimes. We focus on self-employed businesses that are taxed at the personal income tax schedule. Starting in 2009, the self-employed can choose between one of three regimes: the *super simplified* regime (created in 2009), the *simplified* regime (created in 1999) and the *standard* regime. 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.

[6] 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".

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 $k \in$ {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.[7] 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.[8]

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



Note: The figure plots the evolution of the eligibility threshold by activity over time, in thousand euros. The eligibility threshold corresponds 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 corresponds to the introduction year of the super simplified regime in 2009.

Above the threshold y^*_{kt}, 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. Individuals can remain up to two consecutive years in the tolerance region, after which they have to transition to another regime.

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

[8]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.

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.[9] In the super simplified regime, taxable income is simply equal to revenues (i.e. the rebate $\mu = 0$).[10] 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 D.

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 under this regime.

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

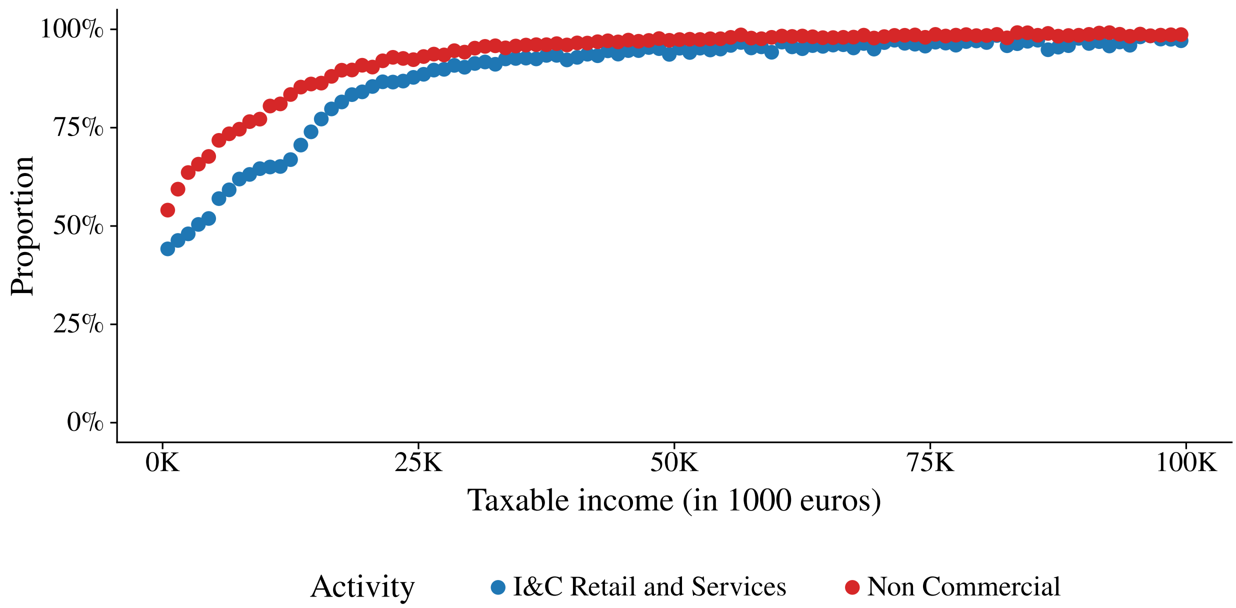
Self-employed individuals in the standard regime have to keep detailed accounts to document their revenues and costs, following standard rigorous accounting practices. Businesses in this regime can call upon a "certified accounting center" (hereafter, CAC), which helps them keep their accounts and also 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), a very large share of 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

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

[10] 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.

members.

Figure 2: Take up of CAC by activity



Note: The figure plots the proportion of agents in the standard regime who are members of a CAC, by activity and bin of taxable income. The data used for the figure spans from 2006 to 2015. The x-axis represents taxable income in the standard regime, i.e., net business income, in bins of 1000 euros. At low income levels, there is a sizable fraction of agents who are not CAC members. This fraction declines rapidly and converges to zero at around 30,000 euros.

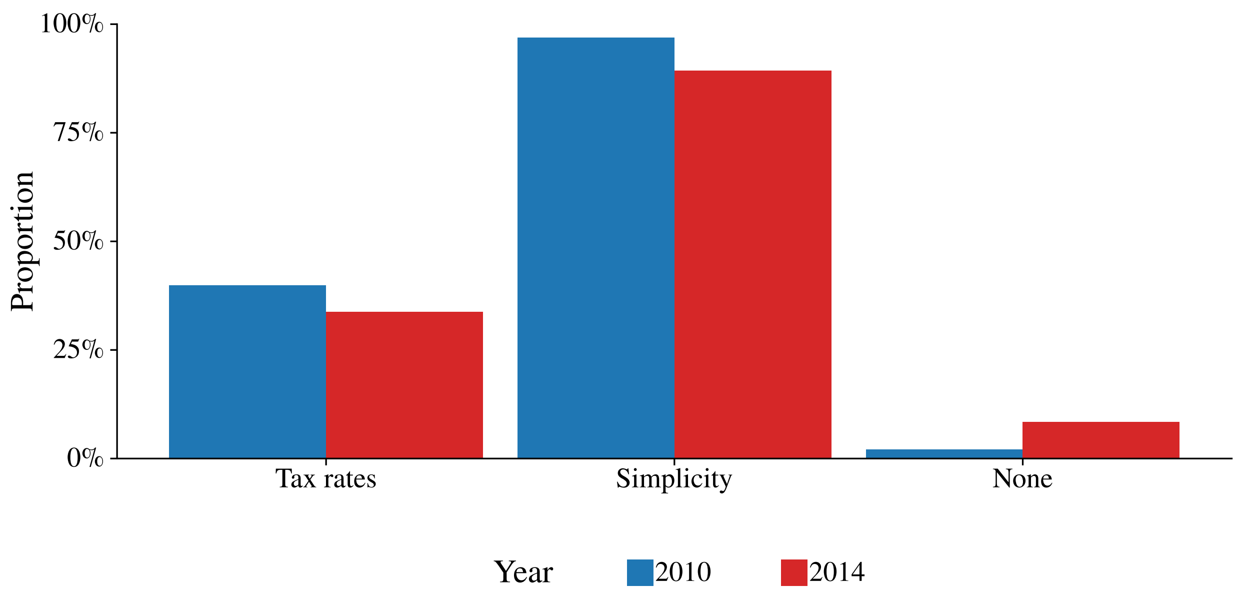
Beyond the financial incentives they entail, the simpler regimes require fewer administrative tasks and proofs of sound fiscal accounting. In the simpler regimes, individuals do not need to report purchases, sales, or costs, only total revenues, and are not required to comply with rigorous accounting 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.[11]

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

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,[12] 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 is likely even stronger.

Figure 3: Advantages of simpler regimes



Note: The figure plots the reported advantages of simpler regimes (simplified and super simplified) from the *New enterprises information system* survey for the years 2010 and 2014.

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

[12] 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."

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

Table 1: Summary statistics, by regime and activity

	Age	Married	Has Children	Live in Paris	Self-emp. revenues	Self-emp. taxable income	N
Panel A: simplified regime							
I&C Retail	46	54%	54%	2%	14,632	4,243	658,497
I&C Services	45	56%	57%	7%	9,315	4,658	1,144,606
Non Commercial	44	52%	54%	11%	9,564	6,312	990,993
Panel B: super simplified regime							
I&C Retail	44	55%	57%	2%	11,890	3,448	219,161
I&C Services	43	55%	59%	3%	9,314	4,657	332,724
Non Commercial	43	53%	59%	7%	9,095	6,002	372,843
Panel C: standard regime							
I&C Retail and Services	47	66%	59%	2%	–	25,825	3,028,307
Non Commercial	46	65%	68%	8%	–	59,234	2,570,586

Note: This table shows summary statistics by regime and activity, for the period spanning 2009 to 2015. Note that for the standard regime, I&C Retail and I&C Services activities are pooled together in the tax returns and cannot be distinguished. Self-employed revenues for the standard regime are not observed.

2.2 Data and Descriptive Statistics

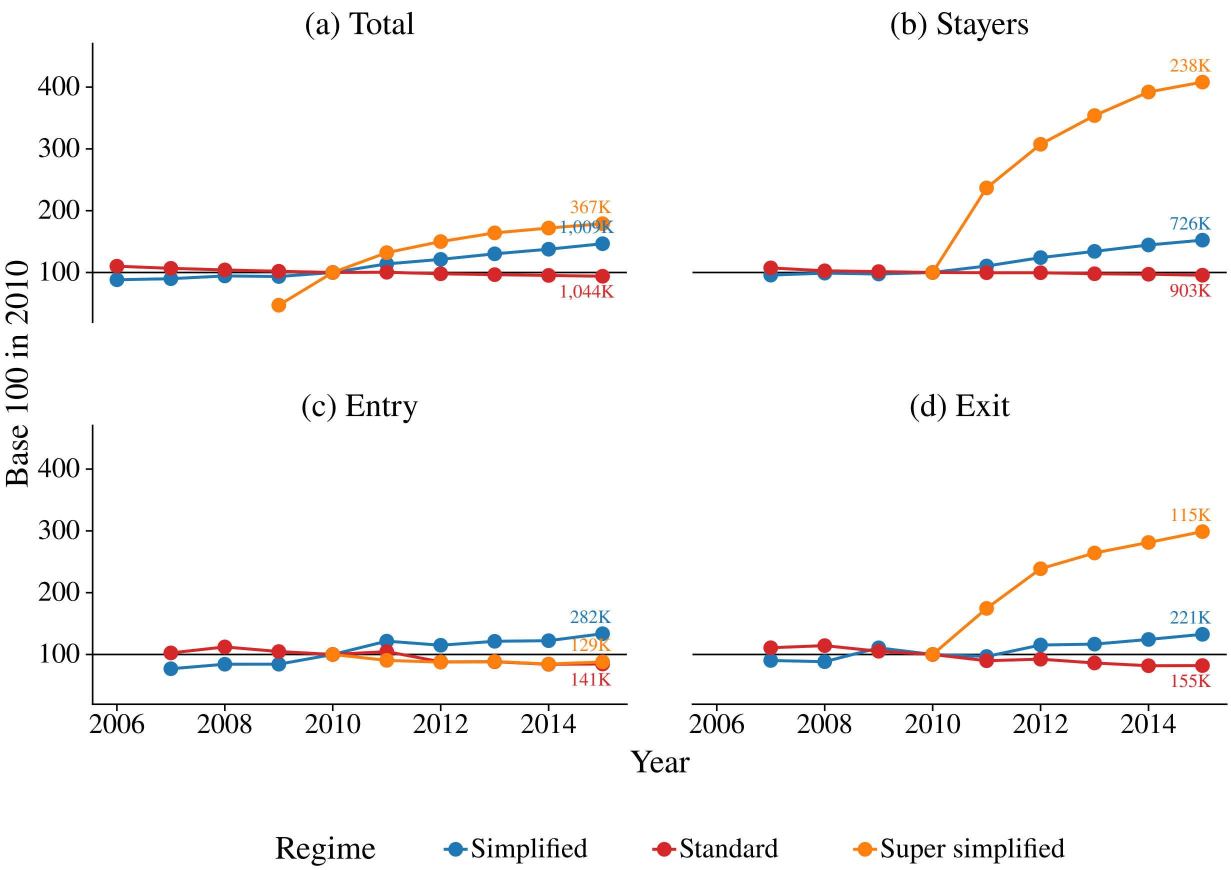
Data. Our longitudinal data is based on the universe of French tax returns over the period 2006-2015 from the French Internal Revenue Service.[13] The income tax returns contain comprehensive income data at the individual and household levels, as well as key demographic information such as household 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 available for 2010 and 2014, which asks entrepreneurs about their experience during their first years.[14]

[13] *Direction Générale des Finances Publiques* (DGFIP).

[14] The survey is called *New enterprises information system*.

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.[15] Further details about data construction are available in Appendix D.

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



Note: 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 entries (panel (c)), and the number of exits (panel (d)) by regime over time. Stayers are individuals who remain in the same regime between two consecutive periods. Entries represent the number of new entrepreneurs each year. Exits represent the number of entrepreneurs who exit self-employment each year. All series are normalized to 100 in 2010, and raw counts in 2015 (in thousand observations) are reported for each of them. An individual is considered entering if they are not observed in the previous year but observed in the current year. On the contrary, an individual is considered exiting if they are observed the previous year but not the current year.

[15]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 was 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) shows 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.

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: (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 (Chetty et al., 2011; Kleven & Waseem, 2013; Saez, 2010). 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 D denote the eligibility threshold for a simplified tax regime. Let $B = B(D)$ denote the extra number of individuals at the left of D following the introduction of a simpler tax regime with this eligibility threshold. To measure B, we estimate the counterfactual income distribution that would apply absent the simplified regime. We do so by fitting a smooth polynomial to the empirical density to the left of the threshold. Revenues are centered around the eligibility threshold by calculating the difference between individuals' revenues and the threshold, and then normalized by bin size B_S, such that $y^{\text{norm}} = (y - D)/B_S$. Bin j, where j is an integer, contains all individuals with self-employed income in the interval $]B_j - 1, B_j]$, so that all individuals reporting revenues exactly at the threshold belong to B_D. 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. Let D^- denote the upper bound of the interval where the empirical and counterfactual distributions begin to differ, i.e., $]D^- - 1, D^-]$. The bunching region includes normalized revenues in the interval $]D^- - 1, D]$. We present a series of robustness tests where we vary the bunching region in Figure A.2 and Figure A.3. To estimate a counterfactual distribution to the left of the threshold, we run the following regression:

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

where C_j stands for the number of individuals in 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 region; $\mathbf{1}[r \in B_j]$ 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 $\hat{C}_j = \sum_p \hat{\beta}_p \cdot (B_j)^p + \sum_r \hat{\alpha}_r \cdot \mathbf{1}[r \in B_j]$ 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. To compute standard errors, we generate normalized 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 region 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 estimate of $\hat{\beta}_0$ is C_D, so that $\Delta y^* = b \times B_S$.

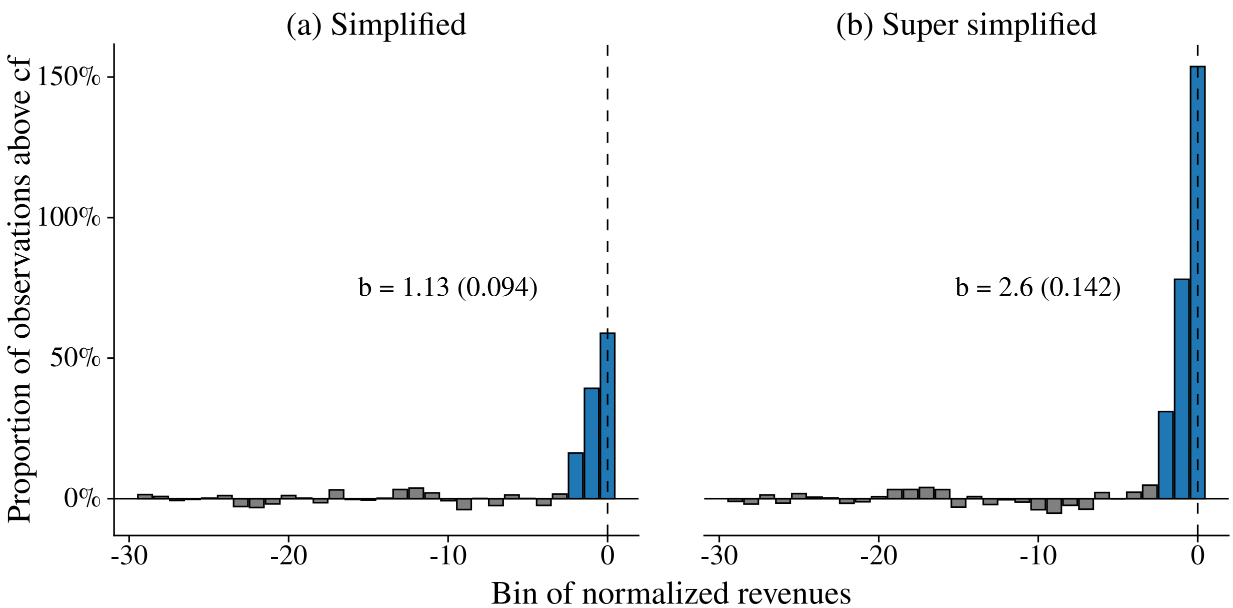
3.2 Results

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, and bins of 1000 euros for the I&C Retail activity, such that $B_j = \{\dots, -9, -8, \dots, 0\}$. 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 suggests the bunching behavior begins three bins away from the threshold (-1500€ for I&C Services and Non Commercial activities, and -3000€ 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

1.13 for the simplified and 2.60 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.

Figure 5: Bunching estimation by regime



Note: The figure represents the frequency of normalized revenues, divided into bins centered around the eligibility threshold (the vertical dashed line). Taxpayers are divided into bins of 500 euros for I&C Services and Non Commercial activities, and bins of 1000 euros for the I&C Retail activity. The results use the pooled population data for 2009-2015 and include all agents in the simplified and super simplified regimes. The figure plots $(C_j - \hat{C}j)/\hat{C}j$, which represents 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 Section 3. The bunching region is indicated in blue. Significant bunching is observed, equal to 113% of the average counterfactual frequency within the bunching region for the simplified regime and 260% for the super simplified regime. Standard errors are calculated using a bootstrap procedure involving random resampling (n = 400) of the residuals.

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. Panel A of 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 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. Panel B of Figure 6 shows the earning responses implied by the excess masses in Panel A. They lie between 400 and 1000 euros for the simplified regime and between 1000 and 2000 euros 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 has little effect on the earning responses.

Robustness tests. Figure A.2 (Figure A.3 respectively) shows the results of robustness tests on the estimation of the excess mass b for the period 2009-2013 (2014-2015 respectively). More specifically, we run variants of the above regressions where we both, allow for changes in the number of bins in the bunching region (the number of excluded bins in the plot) in panel A, and modify the functional form by changing the degree of the polynomial or by running a Poisson regression in Panel B. We perform this robustness exercise for each of the two simpler regimes and for the various types of activities separately. Each time the excess masses follow the same pattern across regimes and activity. Additionally, our preferred specification provides estimates relatively close to other alternative specifications, suggesting that it is a robust estimate of the true excess mass.

4 Evidence on Tax Evasion

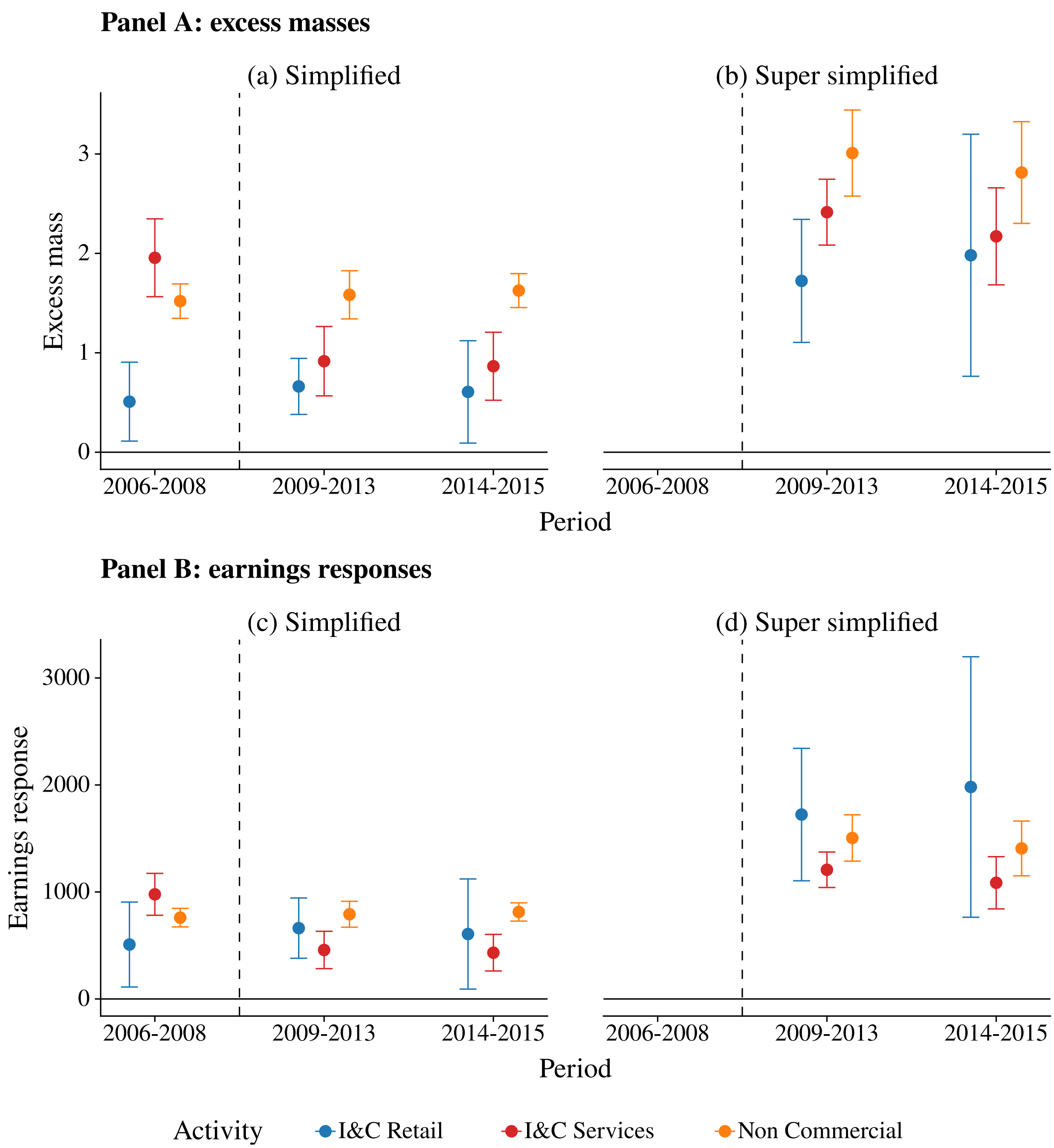
In this section, we present direct evidence on tax evasion and misreporting. We first use dynamic bunching methods that leverage the panel structure of our data, and then employ an intent-to-treat design to understand the channel through which misreporting occurs.

4.1 Self-employment Income Dynamics

We follow the bunching estimation methodology developed by Garbinti et al. (2023) to estimate the dynamic effects of the threshold between the simpler regimes and the standard regime on the distribution of self-employment taxable income growth rates. We estimate: i) the proportion of bunchers with regard to individual growth rates of self-employed taxable incomes, ii) the reduction in the growth rate of self-employment taxable incomes among the treated group (an intent-to-treat or ITT), iii) the growth rate reduction in self-employed taxable income among the bunchers (a local average treatment effect or LATE).

Treatment and control groups. We limit our analysis to the period from 2009 to 2015, during which both the simplified and super simplified regimes were in effect. For each individual i engaging in activity $k = \{\textit{I\&C Retail}, \textit{I\&C Services}, \textit{Non Commercial}\}$ in year t, we define the

Figure 6: Bunching estimation by regime, activity and period



Note: The figure plots the excess masses b (panel A) and the earnings responses Δy^* (panel B) obtained from Section 3, categorized by regime, activity, and period. The counterfactual distribution is fitted using a smooth polynomial, as explained in Section 3. The vertical dashed line corresponds to the introduction year of the super simplified regime in 2009. Standard errors for the excess masses are calculated using a bootstrap procedure with random resampling (n = 400) of the residuals. Standard errors for the earnings responses are determined from the excess masses, with the formula $se(\Delta y^*) = se(b) \times B_S$.

distance between their self-employed taxable income $Z_{i,k,t}$ and the eligibility threshold $\tilde{D}_{k,t}$ for simpler regimes, expressed in taxable income, as $\tilde{Z}_{i,k,t} = Z_{i,k,t} - \tilde{D}_{k,t}$. Note that we omit the regime index for clarity. Using self-employed taxable incomes instead of gross incomes enables us to make direct comparisons among individuals in different regimes, as gross incomes are not reported for individuals in the standard regime.[16] In our analysis, we denote $t' = t + 1$.

Individuals in the treatment group meet the following conditions: They are in one of the simpler regimes in year t (simplified or super simplified), regardless of their regime choice in t'. They are also initially close to the threshold $\tilde{D}_{k,t}$, with a normalized annual self-employed taxable income $\tilde{Z}_{i,k,t}$ falling within the range $]-1000, 0]$.

Individuals in the control group are those in the standard regime in both t and t' and that have a normalized annual self-employed taxable income $\tilde{Z}_{i,k,t}$ falling within the range $]9000, 10000]$. We select the income interval for our control group sufficiently far from the threshold separating the simpler regimes and the standard regime such that there is no discontinuity in their observed growth rates of taxable income, and thus they are not affected by the threshold, yet they remain comparable to individuals in the treated group.[17] Panel A of Figure 7 displays the distribution of self-employment taxable income growth rates for both the treated and control groups. In the control group, there is no noticeable discontinuity, whereas the treated group exhibits a salient spike at the threshold of zero growth.

Normalized growth rates. Below, we describe our empirical strategy, which closely follows the new approach developed by Garbinti et al. (2023).

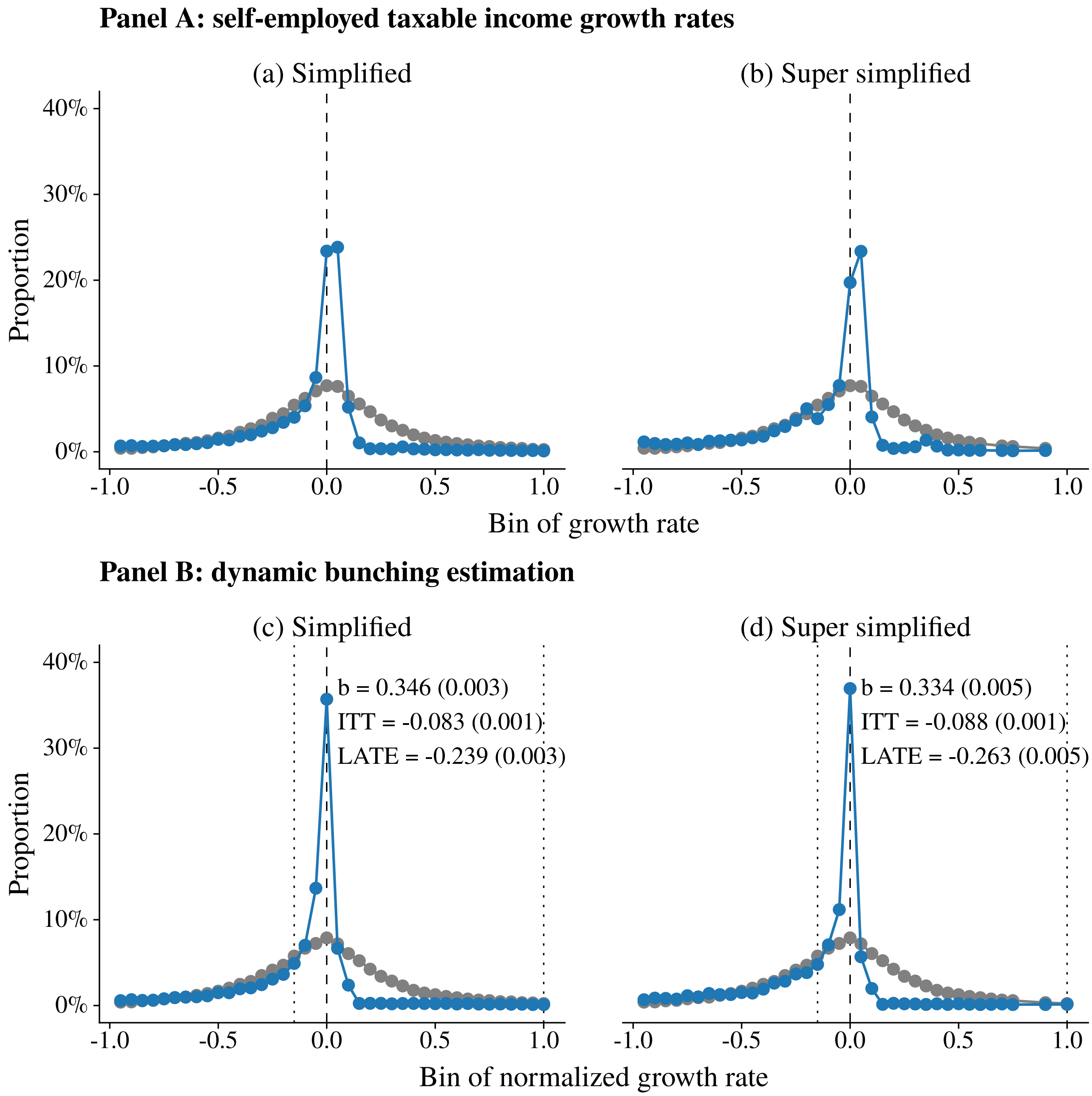
Individuals in the treatment group are directly affected by the eligibility threshold, while those in the control group are not. To compare their growth rate distributions, we normalize both groups. We begin by explaining the normalization process for the treatment group and then describe the normalization for the control group.

For individuals in the treatment group, we compute the growth rate of self-employed taxable income in excess of the growth rates required for an individual to be at the eligibility threshold between the simpler regimes and the standard regime in next period t'. This growth rate is referred to as the "normalized growth rate", as introduced by Garbinti et al. (2023). The normalized growth rate for individual i between year t and the subsequent year t', engaged in activity k in t and activity

[16]The self-employed taxable incomes for the simpler regimes are computed by multiplying the gross incomes by one minus the rebate. The rebate depends on the activity type: 71% for I&C Retail, 50% for I&C Services, and 34% for Non Commercial activities.

[17]Note that for the standard regime, I&C Retail and I&C Services activities are pooled together in the tax returns and cannot be distinguished. Still, the self-employed taxable income thresholds between the two activities are sufficiently distant so that the two control groups are distinct.

Figure 7: Dynamic behavioral responses to the threshold



Note: The figure displays two panels. Panel A plots the distribution of self-employed taxable income growth rates. Panel B plots the distribution of normalized growth rates, as defined in Section 4. The treated group is represented in blue, and the control group is shown in grey. The results are based on the pooled population data for 2009-2015, and separate results are presented for the simplified and super simplified regimes. The population is restricted to individuals that are below the tolerance threshold in the subsequent period. In panel A, the vertical dashed line represents the threshold of zero growth, while in panel B, it represents zero normalized growth. The vertical dotted lines in panel B represent the lower bound and the upper bound of the bunching region with interval $]-0.15, 1]$. Panels (c) and (d) provide the key statistics from our bunching analysis: i) The proportion of bunchers B concerning individual growth rates of self-employed taxable incomes. ii) The reduction in the growth rate of self-employment taxable incomes among the treated group (ITT). iii) The growth rate reduction in self-employed taxable income among the bunchers (LATE). Bins containing less than 13 individuals are not plotted to ensure compliance with French statistical disclosure limitations. Standard errors are calculated using a bootstrap procedure (n=400).

k' in t', is expressed as follows:

$$\tilde{g}_{i,k,t} = \underbrace{\frac{Z_{i,k',t'} - Z_{i,k,t}}{Z_{i,k,t}}}_{\text{Observed growth rate}} - \underbrace{\frac{\tilde{D}_{k',t'} - Z_{i,k,t}}{Z_{i,k,t}}}_{\substack{\text{Growth rate needed}\\ \text{to reach the threshold}}} = \frac{Z_{i,k',t'} - \tilde{D}_{k',t'}}{Z_{i,t}}.$$

If $\tilde{g}_{i,k,t} = 0$, it means that individual i's taxable income has grown precisely at the rate required for them to reach the threshold in the subsequent year, $\tilde{D}_{k',t'}$. If the normalized growth rate is negative (respectively positive), it indicates that individual i ends up below (respectively above) the threshold in the subsequent year.

Next, we construct the normalized growth rate for the control group. Following Garbinti et al. (2023), we calculate the growth rate of self-employed taxable income in excess of the growth rates required for an individual to be at the "placebo" eligibility threshold. The placebo threshold is set at the same distance, in level, from an individual in the control group as the actual eligibility threshold is from a comparable individual in the treatment group:

$$Z_{i,k',t'} - \tilde{D}_{k',t'} = Z^c_{i,k',t'} - \tilde{D}^c_{k',t'},$$

where $Z^c_{i,k',t'}$ is the self-employment taxable income for an individual in the control group and $D^c_{k',t'}$ is the placebo threshold for individuals engaging in activity k' in the subsequent year t'. We have set the upper bound of our control group to be 10,000 euros higher than the eligibility threshold, leading to $Z^c_{i,k',t'} = \tilde{D}_{k',t'} + 10000$ at this particular juncture. In contrast, for the treatment group, the upper bound is located exactly at the threshold, resulting in $Z_{i,k',t'} - \tilde{D}_{k',t'} = 0$ at this point. Thus, the formulation for the placebo eligibility threshold is as follows:

$$\tilde{D}^c_{k',t'} = \tilde{D}_{k',t'} + 10000.$$

We can then compute the corresponding counterfactual growth rates for individuals in the control group as follows:

$$\tilde{g}^c_{i,k,t} = \frac{Z_{i,k',t'} - \tilde{D}^c_{k',t'}}{Z_{i,k,t}}.$$

The validity of the research design outlined above relies on the assumption that in the absence of the reform, the control and treated groups would have the same distribution of normalized growth rates.

Quantifying changes in normalized growth rates. Panel B of Figure 7 shows the distribution of the normalized growth rates $\tilde{g}_i$ for both the treated and control groups. We divide individuals into bins, indexed by a, where each bin has a width of 5 percentage points of normalized growth rates. These bins are defined as follows: $a = \{\ldots,]-0.1, -0.05],]-0.05, 0], \ldots,]0, 0.05]\}$. In our plots, we indicate the upper bound of these intervals. Point 0 on the horizontal axis corresponds to the growth rate of an individual's taxable income, which is required for them to precisely locate at the eligibility threshold in the subsequent period t'.

We define an interval in which the distribution of normalized growth rates between the treatment and control groups diverges. This interval contains bins from a_L to a_U, where a_L represents the lowest bin below the threshold, and a_U represents the highest bin above the threshold. The bin with individuals exactly at the threshold is $a_{\tilde{D}} =]-0.05, 0]$. We visually set the lower bound at the point where the distribution of normalized growth rates for individuals in the treated group starts to differ, specifically $a_L =]-0.15, -0.1]$.

On the left-hand side of this interval, which includes bins from a_L to $a_{\tilde{D}}$, encompassing negative normalized growth rates, a significantly larger fraction of individuals are in the treatment group compared to the control group. In other words, there is a much greater proportion of individuals in the treatment group who manage to keep their taxable income below the eligibility threshold in the next period compared to the control group. We estimate the share of bunchers, or the excess mass, to be equal to:

$$b = \sum_{a=a_L}^{a_{\tilde{D}}} \left[P^{treated}(a) - P^{control}(a)\right],$$

with $P^{treated}(a)$ (respectively $P^{control}(a)$) representing the proportion of the treated (respectively control) population in a given bin a of normalized growth rates. The excess mass between a_L and $a_{\tilde{D}}$ is mirrored by a lower fraction of individuals in the treatment group than in the control group between $a_{\tilde{D}}$ (excluded) and a_U. The upper bound a_U is set such that the excess mass is equal to the missing mass $M = -\sum_{a>a_{\tilde{D}}}^{a_U} \left[P^{treated}(a) - P^{control}(a)\right]$. In practice, we find that the missing mass is approximately equal to the excess mass when we set $a_U =]-0.95, 1]$, that is at the right tail of the normalized growth rate distribution.

Then, we estimate the reduction in the growth rate of self-employment taxable income for individuals in the treatment group compared to individuals in the control group, namely:

$$\Delta E(g) = \sum_{a=a_L}^{a_U} \left[P^{treated}(a) \times g^{treated}(a) - P^{control}(a) \times g^{control}(a)\right],$$

where $g^{treated}$ (respectively $g^{control}$) denotes the average growth rate of individuals in the treated group (respectively control group) in bin a. We interpret $\Delta E(g)$ as an intent-to-treat (ITT) effect.

Finally, we estimate the growth rate reduction amongst bunchers as follows:

$$\Delta E(g)_b = \frac{\Delta E(g)}{b},$$

where $\Delta E(g)$ is the ITT coefficient and b is the share of bunchers, both estimated as previously. Consequently, $\Delta E(g)_b$ is interpreted as a LATE effect. Standard errors are obtained from a bootstrap procedure.

Results Panel B of Figure 7 displays the distributions of the normalized growth rates for both the treated group and the control group, separately for the simplified and super simplified regimes. We also report the key statistics described in the previous paragraph: the share of bunchers b, the growth rate reduction in the treated group (ITT), and the growth rate reduction among bunchers (LATE).

First, we see that the normalized distributions of growth rates of taxable income are similar between individuals in the treated and the control groups, except in the bunching region. The vertical dotted lines represent the lower bound and the upper bound of the bunching region $=]-0.15, 1]$. We observe an excess mass below the threshold and a missing mass above the threshold, suggesting that behavioral responses to the threshold are driven by a significant share of individuals. For the treated group, the share of bunchers is 35% for the simplified regime and 33% for the super simplified regime.

Second, the growth rate reduction in the treated group (ITT) is large. It is equal to 8.3 p.p for the simplified regime and 8.8 p.p for the super simplified regime. This is to be compared to the average growth rate in the control group equal to -3% in the simplified regime and -4.1% in the super simplified regime.[18]

Third, the average reduction in growth rates of taxable income among the bunchers (LATE) in the treated group is equal to 23.9 p.p in the simplified regime and 26.3 p.p in the super simplified. This is to be compared to the counterfactual growth rate in the bunching region equal to 16.7% in the simplified regime and 15.4% in the super simplified regime.

Robustness tests. Table B.1 reports the three keys statistics of the dynamic bunching estimation, varying the distance between the control and treatment group and the range of normalized self-employed taxable income with respect to our main estimation. All coefficients remain close to our preferred estimation procedure.

[18] We compute the weighted average growth rate in the control as: $\bar{g}^{control} = \sum_a \left[N^{control}(a)/(\sum_a N^{control}(a)) \times g^{control}(a)\right]$, where $N^{control}(a)$ is the number of observations in bin a. Similarly, we compute the counterfactual growth rate in the bunching region as: $\bar{g}^{control} = \sum_{a=a_L}^{a_L} \left[N^{control}(a)/(\sum_{a=a_L}^{a_U} N^{control}(a)) \times g^{control}(a)\right]$, where $N^{control}(a)$ is the number of observations in bin a. We exclude two bins from the calculation in the super simplified because they contain less than 13 individuals and cannot be plotted to ensure compliance with French statistical disclosure limitations.

4.2 Evasion through Misreporting

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 region to those outside.

Empirical strategy. To circumvent this selection bias into the bunching region, we build on the method developed by Diamond and Persson (2016) to estimate the treatment effect of the discontinuity at the notch using a static ITT design. Our design and empirical implementation also follows Chen et al. (2021). Intuitively, this method allows us to compare the observed average outcome of individuals in the bunching region to a potential outcome had the threshold not been implemented. The ITT estimator for any outcome X is defined as:

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

where Y denotes self-employed earnings, D^- denotes the lower bound of the bunching region and D denotes the eligibility threshold. The first term in Equation 2 is the average X across individuals in the bunching region, which we directly observe in the data. The second term is the counterfactual average X which we need to estimate. This estimator measures an ITT effect since the interval (D^-, D) includes both the self-employed that respond to the program (i.e., to the threshold) 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 term 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 $\mathbb{E}[X|\text{No Notch}, Y = y]$:

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

To compute an empirical counterpart of these two terms, we bin self-employed earnings following the same procedure as in Section 3. This allows us to estimate $\hat{h}_0(\cdot)$ using the bunching method. For the second term, 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: $\mathbb{E}[X_j|Y_j, \text{No Notch}] = \sum_p \hat{\beta}_p \cdot (B_j)^p + \sum_r \hat{\alpha}_r \cdot \mathbf{1}[r \in B_j]$.

We now consider several outcomes X that are indicative of evasion and misreporting.

Bunching at specific digits of self-employment revenues. Absent incentives to evade taxes, we expect the probability to report a given number as the last digit to be the same in the bunching region as anywhere else in the revenue distribution. If individuals in the bunching region instead report inaccurate and modified numbers, they are unlikely to choose last digits in accordance with their actual distribution. For instance, we might expect digits such as zero to appear relatively more frequently than other digits (e.g., nine) in the bunching region relative to other parts of the revenue distribution.

To illustrate this pattern with one specific choice of digits, panel A of Figure 8 shows the distribution of the probabilities to report 0 or 9 by bin of revenues 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 regions for the two regimes are around 5 percentage points more likely to do so.

Panel B of Figure 8 plots the ITT coefficients for each digit for the different regimes. We see that individuals in simpler regimes disproportionately report 0 as the last digit in the bunching regions. This in turn suggests that the numbers reported in the simplified regimes are more likely to be manipulated, especially around right below the eligibility threshold.

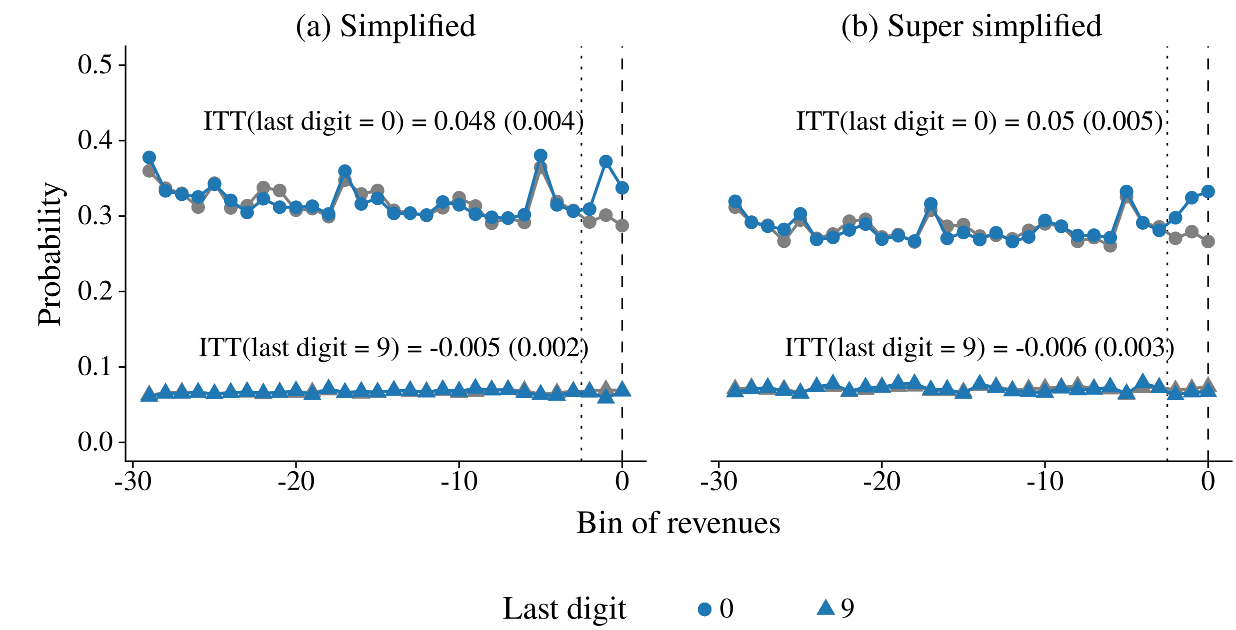
Round numbers bunching of self-employment revenues. We now dig further into the possibility for strategic reporting, by looking more closely at the numbers individuals actually fill in. Figure 9 shows the probability to report a multiple of 100 euros (and excluding multiples of 500 euros) as a function of both, the individual's distance to the threshold and her activity. Consistent with our analysis in the previous section, we find that individuals in 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.

Figure A.4 and Figure A.5 reproduce the same analysis for the probability to report a multiple of 250 euros (and excluding multiples of 500 euros) and a multiple of 50 euros (and excluding multiples of 100 and 250 euros), respectively. Both ITT coefficients are close to zero.

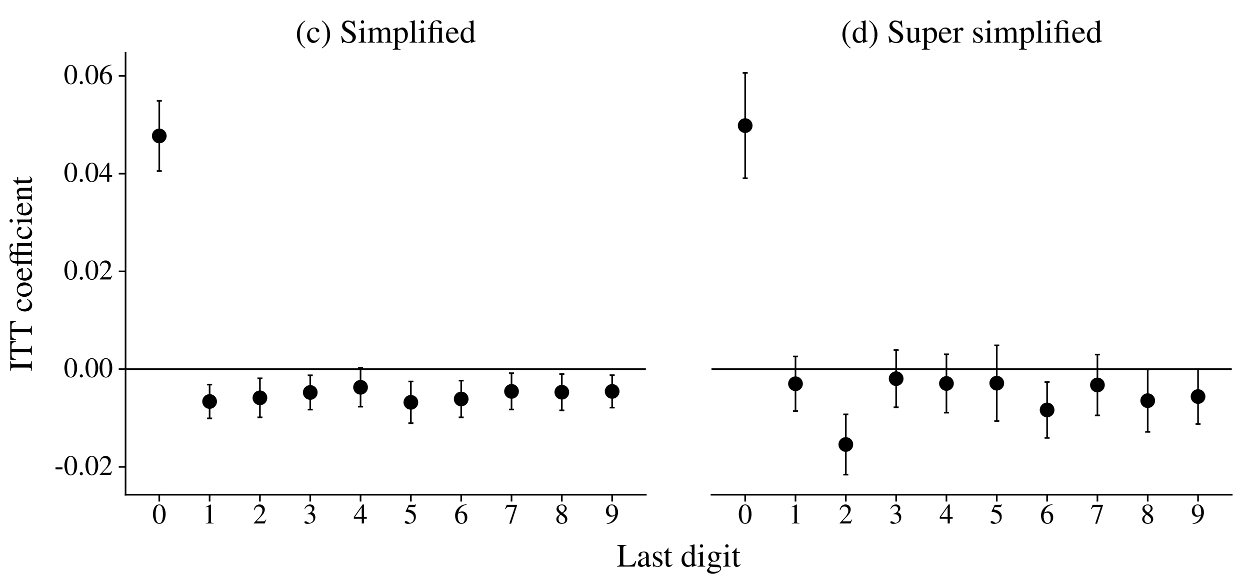
Income shifting within the household. Further evidence of misreporting and avoidance comes from income shifting within the household. The eligibility thresholds apply to individual income, which means that if an individual with self-employed income lives with another individual with self-employed income,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

Figure 8: Last digit reporting behavior for self-employed revenues

Panel A: probability to report 0 or 9 as the last digit of self-employed revenues

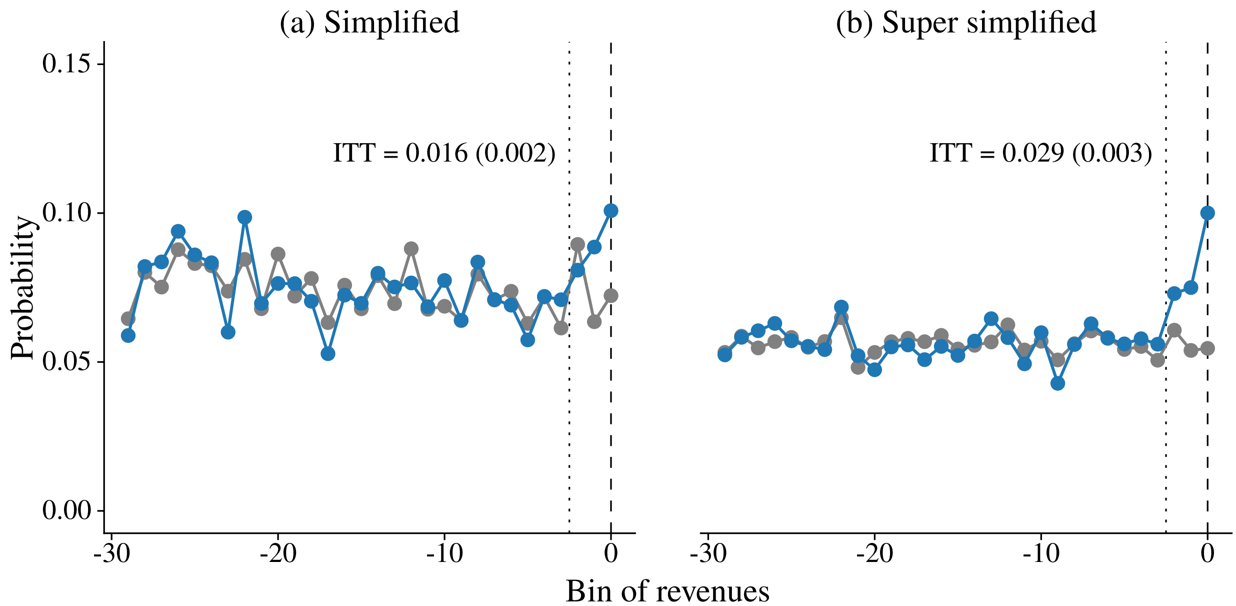


Panel B: ITT coefficients for the last digits of self-employed revenues



Note: The figure displays two panels. Panel A plots the distribution of the probability for the digits 0 (dots) and 9 (triangles), by bins centered around the eligibility threshold (the vertical dashed line). The area between the dotted and dashed vertical lines corresponds to the bunching region. The counterfactual distribution (in grey) is fitted using a smooth polynomial, as explained in Section 4. Comparing the simplified regime to the counterfactual situation without the threshold, the probability is on average 4.8 percentage points higher, and for the super simplified regime, it is 5 percentage points higher. Panel B displays the ITT coefficients for each digit between 0 and 9. The results are based on the pooled population data for 2009-2015, and separate results are presented for the simplified and super simplified regimes. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

Figure 9: Probability of reporting a multiple of 100

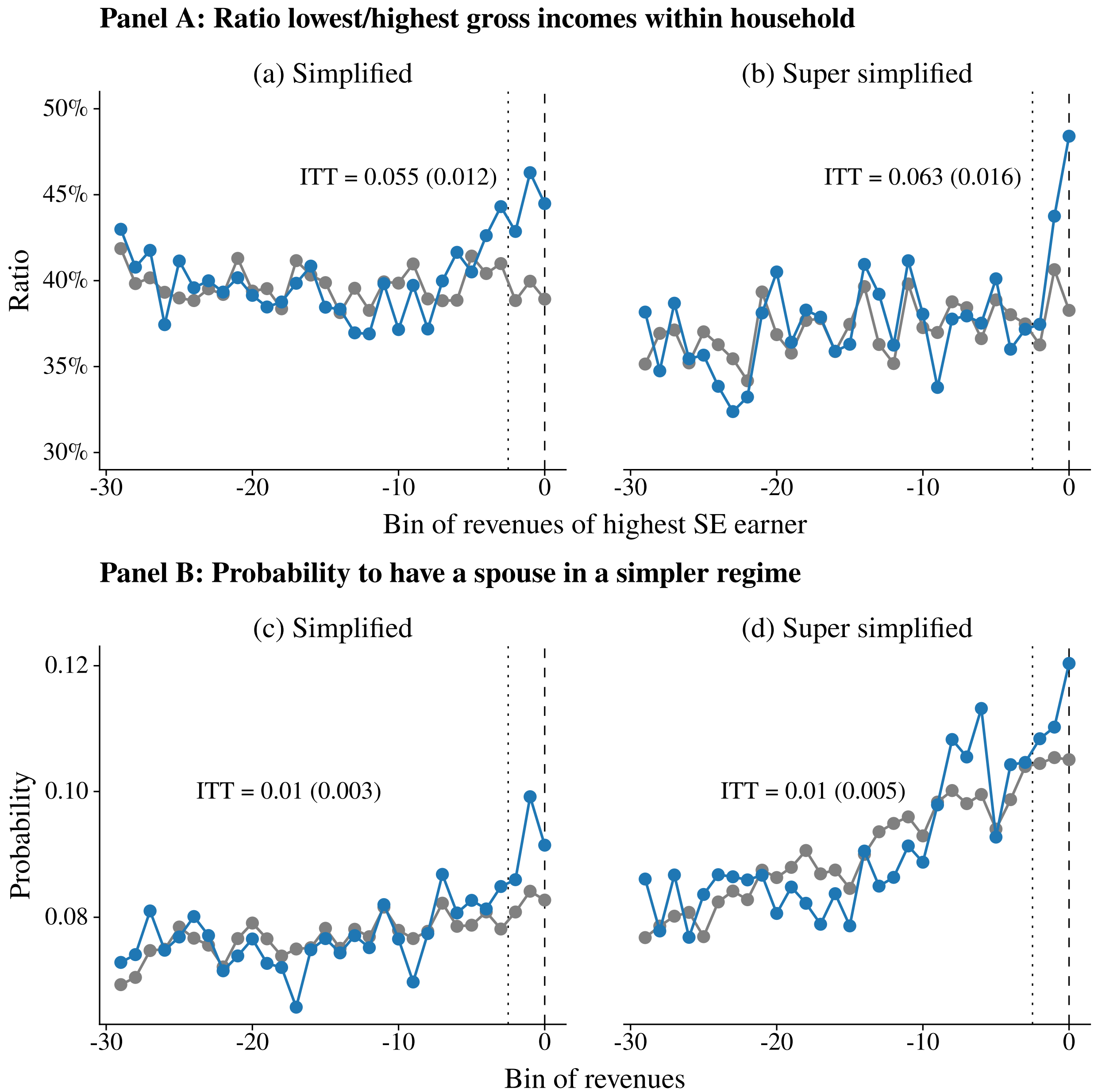


Note: The figure plots the probability of reporting a multiple of 100 euros (and excluding multiples of 500 euros) in self-employed revenues, by bins centered around the eligibility threshold (the vertical dashed line). The bunching region is depicted between the dotted and dashed vertical lines. The counterfactual distribution (in grey) is fitted using a smooth polynomial, following the explanation in Section 4. The results are based on the pooled population data from 2009 to 2015, and they are presented separately for the simplified and super simplified regimes. The ITT coefficient is calculated using the method described in Section 4. Comparing the simplified regime to the counterfactual situation without the threshold, the probability is on average 1.6 percentage points higher, while for the super simplified regime, it is 2.9 percentage points higher. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

We find strong evidence that this is indeed the case by studying couples who both have self-employed earnings in one of the simpler regimes. Our sample for the 2009-2015 period contains 89,457 such households. First, on the intensive margin, Panel A of Figure 10 shows the ratio of the self-employed earnings of the lowest earner to those of the highest earner in the household. We 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 evidence of responses on the extensive margin as well. Panel B of Figure 10 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.

Employer misreporting and "hidden employment". One concern raised in the policy debate on simpler regimes upon their introduction was that they may lead to "hidden employment", whereby

Figure 10: Income shifting within the household



Note: The figure displays two panels. Panel A plots the ratio in gross income between the lowest and highest self-employed earners within a household, by bins of revenues of the highest self-employed earner centered around the eligibility threshold (the vertical dashed line). It implies that both members of the household are in one of these two regimes. The ratio is, on average, 5.5 (6.3 resp.) percentage points higher for the simplified (super simplified resp.) regime compared to the counterfactual situation without the threshold. Panel B plots the probability to have a spouse that is also reporting self-employed revenues, by bins of revenues centered around the eligibility threshold. The probability is, on average, 1 percentage point higher for both the simplified and super simplified regimes compared to the counterfactual situation without the threshold. The bunching region is depicted between the dotted and dashed vertical lines. The counterfactual distribution (in grey) is fitted using a smooth polynomial, following Section 4. The results are based on the pooled population data from 2009 to 2015, and they are presented separately for the simplified and super simplified regimes. The ITT coefficient is calculated using the method described in Section 4. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

employers would fire employees and hire them again as contractors. This in turn would allow employers to circumvent costlier standard labor contracts and regulations. Here, we look at how employed labor income varies around the eligibility threshold.

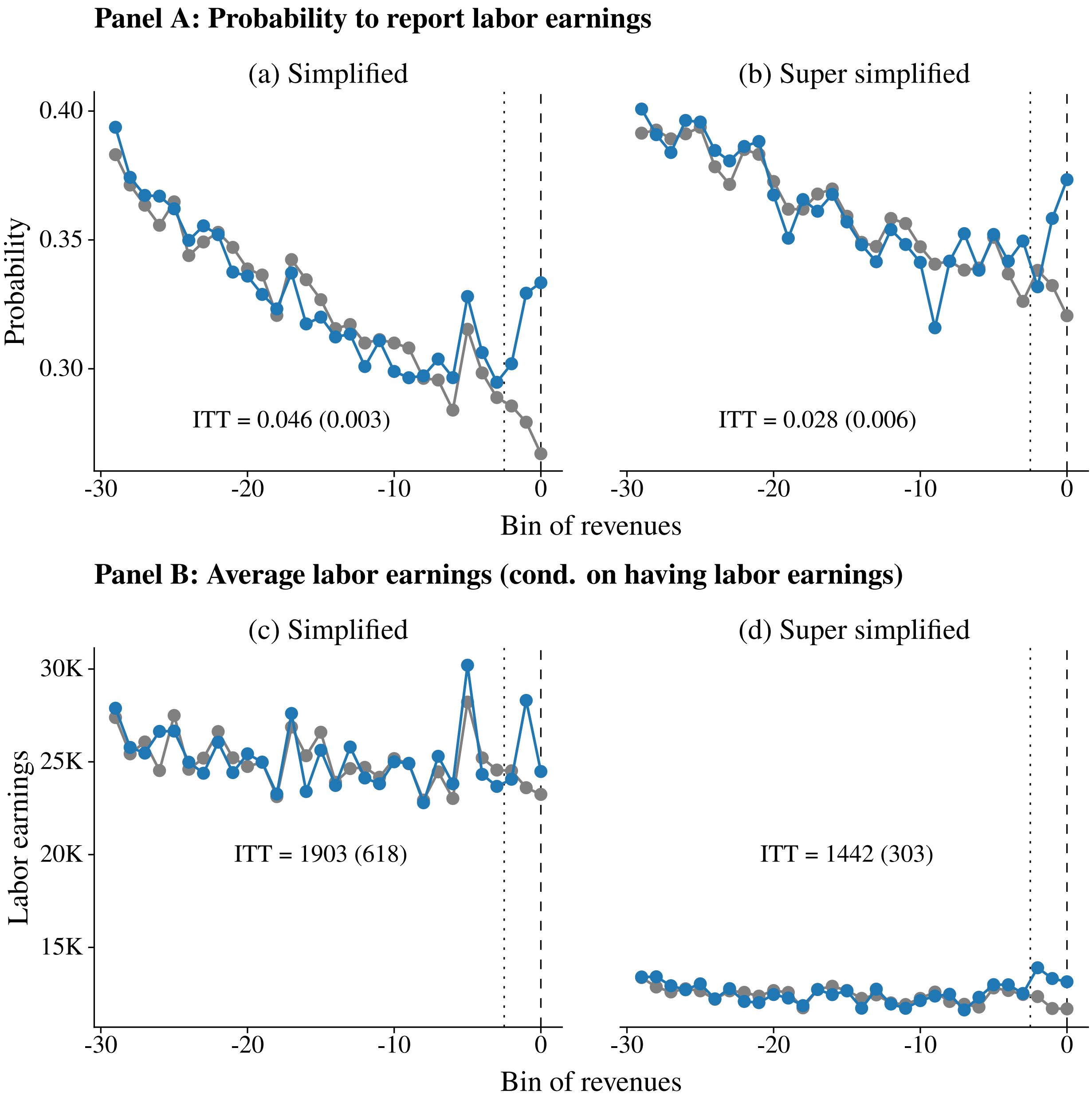
Panel A of Figure 11 plots the probability of reporting any labor earnings in addition to self-employed revenues, which reflects an extensive margin response. This figure shows a discontinuous increase in the likelihood of reporting labor earnings just before the threshold. Panel B of Figure 11 shows the average labor earnings in addition to self-employed revenues, conditional on reporting any labor earnings. This reflects an intensive margin response. There is a discontinuous sharp and significant increase in labor earnings just in the bunching region. Both the probability to report labor earnings and the average labor earnings tend to decline with the level of self-employed revenues, which suggests that there is substitution between self-employed and employed work, potentially due to time constraints or because of hidden employment. However, the sharp discontinuities observed in the bunching region provide empirical support to 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.

5 Estimating the Value of Tax Simplicity

The previous sections showed evidence that self-employed regimes differ in their degree of simplicity, that there is significant and sharp bunching at the eligibility thresholds, and that tax evasion and misreporting are likely channels whereby individuals remain in the bunching regions. The motivation to stay in the simpler regimes is threefold, as explained above: i) Financial and monetary incentives, which depend on total income, family situation, and the tax regime; ii) A preference for simplicity, whereby staying in simpler regimes allows individuals to save on hassle costs and reduce administrative burdens; iii) Tax evasion through misreporting. In fact, evasion both motivates individuals to remain in simpler regimes and allows them to remain below the eligibility thresholds for these regimes in the first place. Tax simplicity and the simplicity of evasion go hand in hand to the extent that simplification makes misreporting easier on top of being intrinsically valuable.

To quantify the value of simplicity, we develop a simplified model of taxpayers' behavior. More specifically, we jointly estimate the evasion elasticity and the preference for simplicity, based on the observed bunching at the eligibility thresholds. Our estimation method is based on the idea of using the observed bunching across activities, regimes, and years as the targeted data moments that our model seeks to match. Key to this estimation is our ability to measure the potential monetary gains (or changes in tax liability) from transitioning between regimes for each group of taxpayers. By leveraging the heterogeneity in the incentives faced by taxpayers across different activities, regimes and years, we can compute enough data moments to inform us about the parameters of interest.

Figure 11: Labor earnings responses



Note: The figure displays two panels. Panel A plots the probability of reporting labor earnings in addition to self-employed revenues and Panel B plots the average labor earnings reported by self-employed individuals conditional on reporting positive labor earnings (intensive margin, by bins of self-employed revenues centered around the eligibility threshold (the vertical dashed line). The probability of reporting labor earnings 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. Labor earnings are, on average, 1903 euros higher for the simplified regime compared to the counterfactual situation without the threshold, and 1442 euros higher for the super simplified regime. The bunching region is depicted between the dotted and dashed vertical lines. The counterfactual distribution (in grey) is fitted using a smooth polynomial, following Section 4. The results are based on the pooled population data from 2009 to 2015, and they are presented separately for the simplified and super simplified regimes. The ITT coefficient is calculated using the method described in Section 4. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

5.1 Model

Preferences. Each agent chooses one among three regimes: the super simplified, simplified or standard regime, indexed by $i \in \{f, m, r\}$. We also refer to the simpler regimes with $s \in \{f, m\}$. Additionally, agents select one among three types of activity, $k \in$ {*I&C Retail, I&C Services, Non Commercial*}. For ease of exposition, we omit the time dimension. The regime-activity pairs (i, k) generate actual revenues y_{ik} and report revenues $\tilde{y}_{ik}$. An agent on regime-activity (i, k) has a type θ_{ik} that captures her productivity: agents with higher θ_{ik} have lower utility costs of producing a given level of revenues. The disutility of generating revenues y_{ik} for an agent of type θ_{ik} is denoted by $h(y_{ik}; \theta_{ik})$, which increases with y_{ik} and decreases with θ_{ik}. The cost of misreporting revenues from y_{ik} to $\tilde{y}_{ik}$ is denoted by $g(y_{ik}, \tilde{y}_{ik})$, increasing in y_{ik} and decreasing in $\tilde{y}_{ik}$. Overall, an agent's utility from earning revenue y_{ik} and reporting $\tilde{y}_{ik}$ is given by:

$$u(y_{ik}, \tilde{y}_{ik}) = y_{ik}(1 - c_{ik}) - T_{ik}(\tilde{y}_{ik}) - h(y_{ik}; \theta_{ik}) - g(y_{ik}, \tilde{y}_{ik}) - a_{ik},$$

where c_{ik} is the cost of producing y_{ik}, $T_{ik}(\tilde{y}_{ik})$ is the total tax liability as a function of reported revenues and a_{ik} is a hassle cost.[19] The latter reflects the tax reporting and compliance costs (e.g., administrative accounting requirements, costs of keeping track and complying with the tax procedure). Given the institutional features, it is to be expected that the hassle cost is lower the simpler the regime, i.e., $a_{rk} > a_{mk} > a_{fk}$. We adopt the following constant-elasticity functional forms for $h(y_{ik}; \theta_{ik})$ and $g(y_{ik}, \tilde{y}_{ik})$:

$$h(y_{ik}; \theta_{ik}) = \frac{\theta_{ik}}{1 + \frac{1}{\varepsilon}} \left(\frac{y_{ik}}{\theta_{ik}}\right)^{1+\frac{1}{\varepsilon}} \quad \text{and} \quad g(y_{ik}, \tilde{y}_{ik}) = \frac{\kappa_{ik}}{1 + \frac{1}{\eta}} \left(\frac{y_{ik} - \tilde{y}_{ik}}{\kappa_{ik}}\right)^{1+\frac{1}{\eta}},$$

where ε denotes the real income elasticity, η represents the evasion elasticity, and κ_{ik} is a scaling parameter. Consistent with evidence from Section 2 and Section 4, agents in the simpler regimes can endogenously misreport their income, whereas agents in the standard regime cannot. Consequently, in the standard regime, the cost of misreporting is effectively infinite due to institutional constraints that hinder misreporting.

Modeling the tax discontinuity. The tax liability depends on both the tax base and the tax rate, both of which may vary across regimes and activities. In the simplified regime, the taxable income of agents is $(1 - \mu_{mk})\tilde{y}_{mk}$, where μ_{mk} is a rebate on reported income $\tilde{y}_{mk}$. In the super simplified regime, taxes are directly levied on $\tilde{y}_{fk}$. Finally, the taxable income in the standard regime is $(1 - c_{rk})y_{rk}$, where c_{rk} is the cost of producing gross income y_{rk}. The agent's effective average

[19] We think about c_{ik} as effective operating costs.

income tax rate τ_{ik} is a combination of social contributions and income taxes.[20] We summarize the combination of effective rates and tax bases in the various regimes in the following table:

Standard regime:	τ_{rk}	is levied on net income	$z_{rk} = (1 - c_{rk})y_{rk}$
Simplified regime:	τ_{mk}	is levied on taxable (reported) income	$z_{mk} = (1 - \mu_{mk})\tilde{y}_{mk}$
Super simplified regime:	τ_{fk}	is levied on gross (reported) revenues	$z_{fk} = \tilde{y}_{fk}$

Table 2 presents the average tax rates from our baseline analysis. Our sample consists of individuals with no labor earnings and in the bunching region, for which the trade-off between the simpler regimes and the standard regime is particularly clear and simple.[21] Details on how the average tax rates are calculated can be found in Appendix D. In the simplified regime, the actual and hypothetical tax rates are essentially the same, so that there are no extra financial incentives apart from the opportunity for evasion and a preference for simplicity.[22]

In the initial period (2009-2013), tax rates mostly favor the super simplified regime. For instance, the average effective tax rates in the super simplified regime are lower compared to those in the standard regime (ranging between 3 and 7 percentage points). However, this reverses in the subsequent period (2014-2015), with tax rates between 1 and 3 percentage points higher in the super simplified regime. This shift is due to a reform that increased the flat rate of social contributions in the super simplified regime to harmonize taxes between the simpler regimes and the standard regime.

5.2 Responses under the Notch

We now describe agents' behaviors' at the eligibility threshold for the simplified regime. We omit the activity type dimension for ease of exposition. Derivations for the super simplified regime are similar, setting $\mu = 0$ and replacing index m by f. Further details are provided in Appendix C.

[20]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 Appendix D.

[21]This condition does not preclude the possibility of a spouse having labor earnings.

[22]A robustness analysis for the calculation of the average tax rates for the simplified regime is provided in Table B.2. The main difference between the baseline and robustness analyses lies in the method for calculating social contributions. In the latter, a flat rate for social contributions in the simplified regime is applied. Deprost et al. (2013) conclude that cost considerations and a preference for simplicity are likely the primary factors influencing the choice of a simpler regime. Our findings align with theirs, showing no sensitivity to alternative definitions of taxation.

Table 2: Average tax rates by regime, activity and period

	Simplified		Super simplified		
Activity	τ_m	τ_r	τ_f	$\tau_f \times (1-\bar{\mu})$	τ_r
Panel A: period 1 (2009-2013)					
I&C Retail	52.08%	52.08%	13.65%	47.07%	50.02%
I&C Services	53.31%	53.31%	24.25%	48.5%	51.42%
Non Commercial	42.7%	42.7%	21.64%	32.79%	39.35%
Panel B: period 2 (2014-2015)					
I&C Retail	51.44%	51.44%	14.77%	50.93%	48.76%
I&C Services	51.18%	51.18%	25.65%	51.3%	49.08%
Non Commercial	41.57%	41.57%	25.49%	38.62%	37.44%

Note: This table reports the average tax rates in the bunching region for individuals reporting zero labor earnings and based on the main method of computation by regime, activity and period. For the simplified regime, τ_m is the average tax rate on taxable self-employed income and τ_r is the counterfactual average tax rate in the standard regime. For the simplified regime, τ_f is the flat rate on gross self-employed revenues, $\tau_f/(1-\bar{\mu})$ is the flat rate expressed in percentage of taxable self-employed revenues and τ_r is the counterfactual average tax rate in the standard regime. The rebate μ depends on the activity type: 71% for I&C Retail, 50% for I&C Services, and 34% for Non Commercial activities. All tax rates are computed with a production cost equal to the rebate. Panel A reports the tax rates for period 1 (2009-2013) and Panel B reports the tax rates for period 2 (2014-2015).

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

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

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

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

Introducing the notch. At the eligibility threshold, there is a marginal agent $y_m^* + \Delta y_m^*$ who reports revenues exactly at the threshold y_m^* but would have reported revenues at $y_m^* + \Delta y_m^*$ otherwise. If the agent was unconstrained by the notch, her choice would be characterized by her reported revenues:

$$y_m^* + \Delta y_m^* = (\theta_m^* + \Delta\theta_m^*)[(1-c_m) - \tau_m(1-\mu)]^{\varepsilon} - \kappa_m[\tau_m(1-\mu)]^{\eta}, \tag{5}$$

and her actual revenues $y_m = (\theta^* + \Delta\theta^*)[(1-c_m) - \tau_m(1-\mu)]^{\varepsilon}$. With the notch, this agent reports revenues at the threshold, but her actual revenues are $y_m^A = y_m(y_m^*)$, where y_m^A is given by:

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

which implies:

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

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

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

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

$$y_r^I(1 - c_r)(1 - \tau_r) - h(y_r^I; \theta_m^* + \Delta\theta_m^*) - a_r = \\ y_m^A(1 - c_m) - \tau_m(1 - \mu)y_m^* - h(y_m^A; \theta_m^* + \Delta\theta_m^*) - g(y_m^A, y_m^*) - a_m. \tag{8}$$

Assumptions. The set of equations (5)-(8) form a non-linear system. We need to make three further assumptions to reduce the dimensionality of the problem and ensure identification. First, we take the limit case $\epsilon \longrightarrow 0$, corresponding to the absence of real response to taxes. This assumption is supported by our results in Section 3 where we showed that the bunching patterns are unlikely to be explained by real responses to taxes. Second, we assume $c_m = c_r = c$, which simply means that the underlying self-employed activity remains the same from a production function standpoint regardless of the choice of regime. Namely:

$$\left[y_m^*(1 - \mu)\tau_m - y_m^A(1 - c)\tau_r\right] + \frac{\kappa_m}{1 + \frac{1}{\eta}}\left(\frac{y_m^A - y_m^*}{\kappa_m}\right)^{1+\frac{1}{\eta}} - \Delta a_m = 0, \tag{9}$$

where the real revenues generated are given by $y_m^A = (y_m^* + \Delta y_m^*) + \kappa_m[\tau_m(1 - \mu)]^{\eta}$ and $\Delta a_m = a_r - a_m$.

5.3 Identification and Estimation

Structural parameters. For a simpler regime s and activity k within a specific time period, we have a set of structural parameters of interest: the elasticity of evasion, the scaling parameter, the hassle cost, and the production cost. We explain below how they vary across regimes and activities.

First, we take the evasion elasticity, η, to be homogeneous across regimes and activities. This is motivated by the fact that the opportunity to misreport revenues is the same across simpler regimes and activities, as they all face the same simplified accounting rules (see Section 2). We express the scaling parameter as a function of the taxable income at the threshold, such that $\kappa_{sk} = \tilde{\kappa} \times (1 - \bar{\mu}_k) \times y^*_{sk}$, where κ is a scaling factor constant across regimes and activities. For a given activity, the threshold in the simplified regime and in the super simplified regime are the same, and $\mu_{mk} = \bar{\mu}_k$ is equal to the rebate in the simplified regime.[23] This normalization allows the scaling parameter to vary across activities in any given simpler regime.

Second, we take the taste for simplicity to be captured by the hassle cost. Δa_{sk} represents the monetary value of simplicity in a regime-activity pair. We expect it to be higher in the super simplified regime compared to the simplified regime, as individuals are more likely to bunch in the former compared to the latter. We compute the value of the hassle cost relative to the taxable income at the threshold, by setting $\Delta a_{sk} = \tilde{\Delta} a_s \times (1 - \bar{\mu}_k) \times y^*_{sk}$. $\tilde{\Delta} a_s$ precisely captures the difference in simplicity across the two simpler regimes, with $\tilde{\Delta} a_m > \tilde{\Delta} a_f$.

Finally, we must consider the production cost c_{ik}, which is not directly observed in the data. Yet, by design the government has set the rebate so as to match c_{ik} with the production costs in the simpler regimes. This, together with our assumption that production costs are similar across regimes (such that $c_{sk} = c_{rk}$) leads us to consider a reference cost level c^*_{sk} and then to investigate whether our results are sensitive to deviations from it. The reference cost c^*_{sk} is taken to equalize taxes in the simpler regime with taxes in the standard regime at the eligibility threshold in the simpler regime.[24] Namely:

$$c^*_{sk} = 1 - (1 - \mu_{sk}) \frac{\tau_{sk}}{\tau_{rk}}.$$

Table 2 shows that taxes in the simpler regimes and the standard regime are close to each other, such that the reference cost is indeed close to the rebate.

Solution. For a simpler regime s and activity k within a specific time period, and given primitives $\Omega_{sk} = (\eta, \tilde{\kappa}, \tilde{\Delta} a_s)$, policy parameters $\Phi_{sk} = (y^*_{sk}, \tau_{rk}, \tau_{sk}, \mu_{sk})$ and cost c_{sk}, we can solve for a model predicted earnings response $\Delta y^*_{sk}(\Omega_{sk}, \Phi_{sk})$ using (9). We iterate the same exercise over each combination of regime and activity.

[23]Note that we also apply the rebate in the simplified regime to the super simplified regime. As a result, we ensure that everything is expressed in the same dimension, the self-employed taxable income.

[24]More precisely, $y^*_{sk}(1 - \mu_{sk})\tau_{sk} = y^*_{sk}(1 - c_{sk})\tau_{rk}$. In the super simplified regime, $\mu_{fk} = 0$.

Identification. We now discuss the identification of the different parameters $\Omega = (\eta, \tilde{\kappa}, \tilde{\Delta}a_m, \tilde{\Delta}a_f)$. We use differences in incentives faced by various types of agents at the eligibility threshold, arising from variations in policy parameters Φ_{sk} across different regimes and activities.

First, consider the taste for simplicity as defined by the hassle cost. This parameter is regime-specific $(\tilde{\Delta}a_{mk}, \tilde{\Delta}a_{fk})$.[25] The difference in average excess masses across the two simpler regimes disciplines the hassle cost. Second, given the hassle costs, 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 different incentives across regimes and activities (i.e., income taxes, social security contribution rates, and rebates). Consequently, we have multiple empirical moments Δy^*_{sk}, which we can target to find the parameters that best fit the data.

We run the estimation for the 2009-2013 period and use the 2014-2015 period for our robustness exercise. Our baseline results are based on individuals reporting zero labor earnings. Let s index the simpler regime (super simplified or simplified) and k index the activity (I&C Retail, I&C Services, Non Commercial). For each combination of regime and activity, there is a model-predicted bunching interval Δy^*_{sk}. Its empirical counterpart in the data is $\hat{\Delta} y^*_{sk}$.

Recall that the parameters we aim to estimate are the hassle costs, the evasion elasticity, and the scaling parameter, denoted as $\Omega = (\eta, \tilde{\kappa}, \tilde{\Delta}a_m, \tilde{\Delta}a_f)$. We have four parameters to estimate and six data moments ($N = 2$ regimes $\times$ 3 activities). The loss function we minimize is denoted by $L(\Omega)$, where:

$$L(\Omega) = \frac{1}{N} \sum_{s,k} \left(\frac{\Delta y^*_{sk}(\Omega, \Phi_{sk}) - \hat{\Delta} y^*_{sk}}{se(\hat{\Delta} y^*_{sk})} \right)^2, \tag{10}$$

where $se(\hat{\Delta} y^*_{sk})$ is the standard error of the earnings response $\hat{\Delta} y^*_{sk}$. Figure 6 shows that some earnings responses are more precisely estimated than others. To accommodate this heterogeneity in the estimation precision, we scale the difference between the simulated and observed earnings responses by the standard error of the estimate, implicitly assigning more weight to more precisely estimated earnings responses.

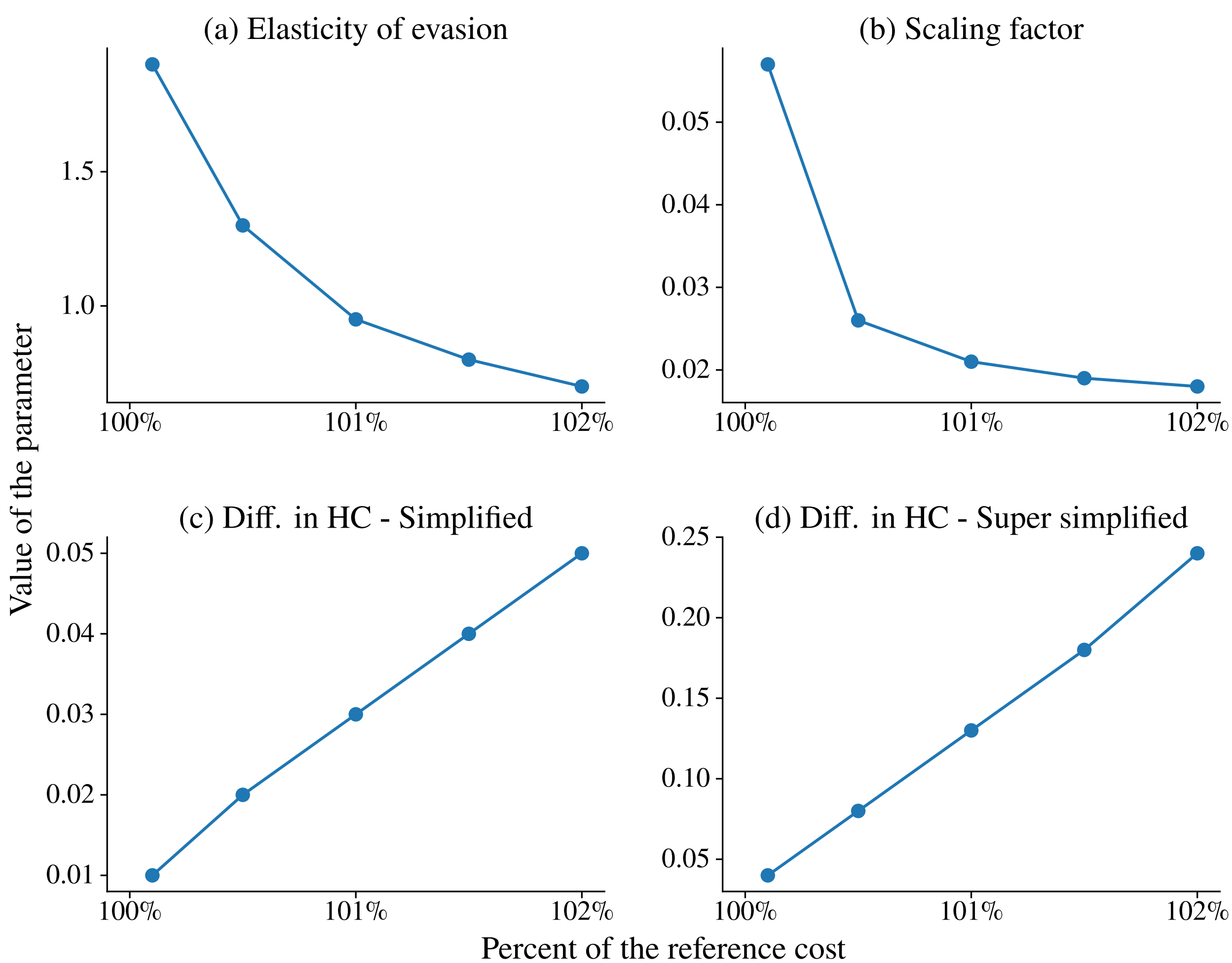
5.4 Estimation Results

Structural parameters. Figure 12 reports the values of the structural parameters for different cost values, defined in percents of the reference cost c^*_{sk}. More precisely, we run the analysis

[25] We expect a positive relationship between the production cost and the hassle costs associated with the simpler regimes. An increase in c_{sk} makes the standard regime more economically appealing compared to the simpler regimes, as the tax base diminishes. Consequently, only a stronger preference for simplicity could justify the observed patterns in earnings response.

separately for 100.1%, 100.5%, 101%, 101.5% and 102% of this cost. Table 3 presents the corresponding estimated evaded amounts, differences in hassle costs and costs in percent of self-employed revenues. There are several key findings.

Figure 12: Structural parameters by cost



Note: This plot shows the results from the structural estimation, based on the data moments for 2009-2013 (period 1) and for individuals reporting zero labor earnings. This simulation applies the main tax definition. The scaling parameter and differences in hassle costs are expressed in percent of the taxable income in the simpler regime.

First, the evasion elasticity is sizable, between 0.7 and 1.9, depending on the cost. The implied evaded amount due to under reporting ranges between 578 and 911 euros for the simplified regime and between 1236 and 1727 euros for the super simplified regime, which implies substantial levels of misreporting. These numbers are stable across scenarios, and 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 our 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 eligibility thresholds.

Second, the value of simplicity is much higher in the super simplified regime. In panels (c) and (d) of Figure 12, we see that the average difference in hassle costs between a simpler regime and the standard regime is increasing with the distance to the reference cost. As the financial incentive to remain in the simpler regimes decreases, only a higher taste for simplicity can explain the observed earnings responses. The difference in hassle costs between the simpler regimes and the standard regime, ranges between 1% and 9% of the taxable income for the simplified regime and between 4% and 24% of the taxable income for the super simplified regime. Converted in monetary equivalents in Table 3, this amounts to between 162 and 1174 euros in the simplified regime (depending on the regime, activity and cost structure) and between 651 euros and 5654 euros in the super simplified regime, per year and per self-employed. These are sizable amounts 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.[26]

To put these numbers into perspective, we can draw on the existing evidence regarding 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 at around 0.7% of gross income, which corresponds to between 10 to 15 working hours per year. Benzarti and Wallossek (2024) find that individuals are willing, on average, to pay 130 dollars in additional taxes to reduce filing costs. The hassle cost we estimate is somewhat larger, which in turn may have to do with the following specificities of our analysis. First, we focus our attention on self-employed individuals, a population which 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 due to the easier 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) show that reporting requirements (compliance costs) are more important than tax variations in explaining output response for small firms and entrepreneurs. Using bunching methods, these authors 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).

Finally, costs expressed as a percentage of the gross income, as implied by our estimation strategy, are only slightly higher than the rebate, except for individuals in the super simplified regime and in the Non Commercial activity. The latter may reflect the difficulty of accurately estimating tax rates in the Non Commercial activity

Model fit. Table B.4 shows the empirical fit of our structural estimation. In particular, it displays the earnings response $\hat{\Delta} y^*$ observed in the data , the earnings response predicted by the model

[26]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 3: Evasion, preference for simplicity and cost implied by the model

Percent of the reference cost	Regime	Activity	Evaded amount	Diff. in HC	Cost
100.1%	Simplified	I&C Retail	664	235	71.1%
		I&C Services	579	162	50.0%
		Non Commercial	863	214	34.0%
	Super simplified	I&C Retail	1652	942	72.8%
		I&C Services	1236	651	52.9%
		Non Commercial	1534	859	45.1%
100.5%	Simplified	I&C Retail	679	469	71.4%
		I&C Services	580	324	50.2%
		Non Commercial	860	428	34.2%
	Super simplified	I&C Retail	1668	1885	73.1%
		I&C Services	1241	1301	53.1%
		Non Commercial	1544	1718	45.2%
101.0%	Simplified	I&C Retail	709	704	71.7%
		I&C Services	601	487	50.5%
		Non Commercial	888	642	34.3%
	Super simplified	I&C Retail	1696	3063	73.4%
		I&C Services	1263	2114	53.4%
		Non Commercial	1573	2791	45.5%
101.5%	Simplified	I&C Retail	726	939	72.1%
		I&C Services	611	649	50.7%
		Non Commercial	900	857	34.5%
	Super simplified	I&C Retail	1713	4240	73.8%
		I&C Services	1273	2928	53.6%
		Non Commercial	1588	3865	45.7%
102.0%	Simplified	I&C Retail	740	1174	72.4%
		I&C Services	620	811	51.0%
		Non Commercial	911	1071	34.7%
	Super simplified	I&C Retail	1727	5654	74.2%
		I&C Services	1282	3904	53.9%
		Non Commercial	1600	5153	45.9%

Note: This table shows the evaded amounts (in euros), differences in hassle costs (in euros) and costs (in percent of self-employed revenues) predicted by the structural parameters from Figure 12 by regime, activity and for period 1 (2009-2013).

Δy^*, the percentage deviation between the simulated moments and the empirical moments $(\Delta y^* - \hat{\Delta} y^*)/\hat{\Delta} y^*$ and the loss. For the in-sample estimation in panel (a), the two sets of moments are close to each other ($L(\Omega) \approx 0.001$ in all scenarios), which in turn suggests that our model does a good job in predicting the behavior of self-employed individuals.

Robustness checks. We conduct two robustness checks to assess the sensitivity of our results to the assumptions in our model.

First, we perform a robustness check of our model's goodness-of-fit in panel (b) of Table B.4. We use the parameters estimated from the 2009-2013 period to simulate the earnings responses for the 2014-2015 period. The loss is slightly higher but only reflects modest absolute differences in earnings responses, as we cannot fully capture dynamic adjustment between periods. Overall, it suggests that our estimation strategy is reasonable.

Second, we perform a similar analysis using an alternative definition for tax rates (as detailed in Table B.2). The structural parameters are presented in Figure A.6. All parameters closely align with those from our main estimation. Specifically, the predicted hassle costs and evaded amounts, as outlined in Table B.3, fall within a similar range. The goodness-of-fit coefficients, in Table B.5, also reflect this robustness. It shows that our results are not too sensitive to alternative tax calculations.

Discussion. The overall welfare implications of introducing the simpler regimes are unclear a priori. In particular, the government may or may not be losing tax revenues. On the one hand, these new tax regimes were explicitly introduced to facilitate 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 whether this additional firm creation generates truly new income (that is taxed at lighter rates and in part evaded) or whether it simply represents a shift away from former employment (which is taxed and is harder to evade). It may well be the case that the revenue losses induced by tax evasion, which happen at the margin of the thresholds, turn out to be small compared to the overall gains of having more self-employed businesses.

Yet, tax evasion plays an important role, both as an incentive for individuals to join the simplified regimes and once in the simplified regime, as an incentive to remain below the threshold. The hassle costs act as a notch and provide a further incentive to remain below the eligibility threshold.

6 Conclusion

We study how French self-employed respond to the creation and incentives of simplified tax regimes. The self-employed bunch substantially below the eligibility ceilings for the simplified and super simplified regimes. We start 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 and non-random digits close to the threshold as compared to further away 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 use our reduced form bunching estimates as data moments to be matched by a 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 feature 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., & Jäger, S. (2015). Complex Tax Incentives. *American Economic Journal: Economic Policy*, *7*(3), 1–28. https://doi.org/10.1257/pol.20130137

Alinaghi, N., Creedy, J., & Gemmell, N. (2021). Elasticities of taxable income and adjustment costs: Bunching evidence from New Zealand. *Oxford Economic Papers*, *73*(3), 1244–1269. https://doi.org/10.1093/oep/gpaa046

Almunia, M., & Lopez-Rodriguez, D. (2018). Under the Radar: The Effects of Monitoring Firms on Tax Compliance. *American Economic Journal: Economic Policy*, *10*(1), 1–38. https://doi.org/10.1257/pol.20160229

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

Bastani, S., & Selin, H. (2014). Bunching and non-bunching at kink points of the Swedish tax schedule. *Journal of Public Economics*, *109*, 36–49. https://doi.org/10.1016/j.jpubeco.2013.09.010

Benzarti, Y. (2020). How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs. *American Economic Journal: Economic Policy*, *12*(4), 38–57. https://doi.org/10.1257/pol.20180664

Benzarti, Y., & Wallossek, L. (2024). Rising income tax complexity. *National Tax Journal*, *77*(1), 000–000.

Bergolo, M., Burdin, G., De Rosa, M., Giaccobasso, M., & Leites, M. (2021). Digging Into the Channels of Bunching: Evidence from the Uruguayan Income Tax. *The Economic Journal*, *131*(639), 2726–2762. https://doi.org/10.1093/ej/ueab002

Best, M. C. (2014). The role of firms in workers' earnings responses to taxes: Evidence from pakistan.

Bhargava, S., & Manoli, D. (2015). Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment. *American Economic Review*, *105*(11), 3489–3529. https://doi.org/10.1257/aer.20121493

Bíró, A., Prinz, D., & Sándor, L. (2022). The minimum wage, informal pay, and tax enforcement. *Journal of Public Economics*, *215*, 104728. https://doi.org/10.1016/j.jpubeco.2022.104728

Blesse, S., Buhlmann, F., & Doerrenberg, P. (2019). Do people really want a simple tax system? evidence on preferences towards income tax simplification. *Evidence on Preferences Towards Income Tax Simplification*, 19–058.

Blumenthal, M., & Slemrod, J. (1992). The compliance cost of the U.S. individual income tax system: A Second look after tax reform. *National Tax Journal*, *45*(2), 185–202. https://doi.org/10.1086/NTJ41788959

Bohne, A., & Nimczik, J. S. (2018). Information frictions and learning dynamics: Evidence from tax avoidance in ecuador.

Boning, W. C., Guyton, J., Hodge, R., & Slemrod, J. (2020). Heard it through the grapevine: The direct and network effects of a tax enforcement field experiment on firms. *Journal of Public Economics*, *190*, 104261. https://doi.org/10.1016/j.jpubeco.2020.104261

Brockmeyer, A., Smith, S., Hernandez, M., & Kettle, S. (2019). Casting a Wider Tax Net: Experimental Evidence from Costa Rica. *American Economic Journal: Economic Policy*, *11*(3), 55–87. https://doi.org/10.1257/pol.20160589

Carrillo, P., Pomeranz, D., & Singhal, M. (2017). Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement. *American Economic Journal: Applied Economics*, *9*(2), 144–164. https://doi.org/10.1257/app.20140495

Chen, Z., Liu, Z., Suárez Serrato, J. C., & Xu, D. Y. (2021). Notching r&d investment with corporate income tax cuts in china. *American Economic Review*, *111*(7), 2065–2100. https://doi.org/10.1257/aer.20191758

Chetty, R., Friedman, J. N., Olsen, T., & Pistaferri, L. (2011). 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. https://doi.org/10.1093/qje/qjr013

Chetty, R., Friedman, J. N., & Saez, E. (2013). Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings. *American Economic Review*, *103*(7), 2683–2721. https://doi.org/10.1257/aer.103.7.2683

Coles, J. L., Patel, E., Seegert, N., & Smith, M. (2022). How Do Firms Respond to Corporate Taxes? *Journal of Accounting Research*, *60*(3), 965–1006. https://doi.org/10.1111/1475-679X.12405
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1475-679X.12405

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

de Paula, Á., & Scheinkman, J. A. (2010). Value-Added Taxes, Chain Effects, and Informality. *American Economic Journal: Macroeconomics*, *2*(4), 195–221. https://doi.org/10.1257/mac.2.4.195

Deprost, P., Laffon, P., & Imbaud, D. (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., Liu, L., & Loretz, S. (2014). The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records. *American Economic Journal: Economic Policy*, *6*(2), 19–53. https://doi.org/10.1257/pol.6.2.19

Diamond, R., & Persson, P. (2016). *The long-term consequences of teacher discretion in grading of high-stakes tests* (tech. rep.). National Bureau of Economic Research.

Engström, P., & Holmlund, B. (2009). Tax evasion and self-employment in a high-tax country: Evidence from Sweden. *Applied Economics*, *41*(19), 2419–2430. https://doi.org/10.1080/00036840701765452
_eprint: https://doi.org/10.1080/00036840701765452

Farhi, E., & Gabaix, X. (2020). Optimal Taxation with Behavioral Agents. *American Economic Review*, *110*(1), 298–336. https://doi.org/10.1257/aer.20151079

Feldman, N. E., Katuščák, P., & Kawano, L. (2016). Taxpayer Confusion: Evidence from the Child Tax Credit. *American Economic Review*, *106*(3), 807–835. https://doi.org/10.1257/aer.20131189

Fisman, R., & Wei, S.-J. (2004). Tax rates and tax evasion: Evidence from "missing imports" in china. *Journal of political Economy*, *112*(2), 471–496.

Garbinti, B., Goupille-Lebret, J., Muñoz, M., Stantcheva, S., & Zucman, G. (2023). *Tax design, information, and elasticities: Evidence from the french wealth tax* (tech. rep.). National Bureau of Economic Research.

Gelber, A. M., Jones, D., & Sacks, D. W. (2020). Estimating Adjustment Frictions Using Nonlinear Budget Sets: Method and Evidence from the Earnings Test. *American Economic Journal: Applied Economics*, *12*(1), 1–31. https://doi.org/10.1257/app.20170717

Grottke, M., & Lorenz, J. (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., & Matikka, T. (2016). The elasticity of taxable income and income-shifting: What is "real" and what is not? *International Tax and Public Finance*, *23*(4), 640–669. https://doi.org/10.1007/s10797-016-9393-4

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

Johns, A., & Slemrod, J. (2010). The distribution of income tax noncompliance. *National Tax Journal*, *63*(3), 397–418. https://doi.org/10.17310/ntj.2010.3.01

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

692. https://doi.org/10.3982/ECTA9113
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA9113

Kleven, H. J., & Waseem, M. (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan*. *The Quarterly Journal of Economics*, *128*(2), 669–723. https://doi.org/10.1093/qje/qjt004

LaLumia, S., Sallee, J. M., & Turner, N. (2015). New Evidence on Taxes and the Timing of Birth. *American Economic Journal: Economic Policy*, *7*(2), 258–293. https://doi.org/10.1257/pol.20130243

le Maire, D., & Schjerning, B. (2013). Tax bunching, income shifting and self-employment. *Journal of Public Economics*, *107*, 1–18. https://doi.org/10.1016/j.jpubeco.2013.08.002

Mortenson, J. A., & Whitten, A. (2020). Bunching to Maximize Tax Credits: Evidence from Kinks in the US Tax Schedule. *American Economic Journal: Economic Policy*, *12*(3), 402–432. https://doi.org/10.1257/pol.20180054

Naritomi, J. (2019). Consumers as Tax Auditors. *American Economic Review*, *109*(9), 3031–3072. https://doi.org/10.1257/aer.20160658

Parker, S. C. (2003). Does Tax Evasion Affect Occupational Choice?*. *Oxford Bulletin of Economics and Statistics*, *65*(3), 379–394. https://doi.org/10.1111/1468-0084.t01-1-00050
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0084.t01-1-00050

Pirttilä, J., & Selin, H. (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. https://doi.org/10.1111/j.1467-9442.2010.01635.x
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9442.2010.01635.x

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

Pomeranz, D. (2015). No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. *American Economic Review*, *105*(8), 2539–2569. https://doi.org/10.1257/aer.20130393

Saez, E. (2010). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy*, *2*(3), 180–212. https://doi.org/10.1257/pol.2.3.180

Slemrod, J. (2005). The Etiology of Tax Complexity: Evidence from U.S. State Income Tax Systems. *Public Finance Review*, *33*(3), 279–299. https://doi.org/10.1177/1091142105275003

Tazhitdinova, A. (2018). Reducing evasion through self-reporting: Evidence from charitable contributions. *Journal of Public Economics*, *165*, 31–47. https://doi.org/10.1016/j.jpubeco.2018.06.001

Tazhitdinova, A. (2020). Do only tax incentives matter? Labor supply and demand responses to an unusually large and salient tax break. *Journal of Public Economics*, *184*, 104162. https://doi.org/10.1016/j.jpubeco.2020.104162

Warskett, G., Winer, S. L., & Hettich, W. (1998). The Complexity of Tax Structure in Competitive Political Systems. *International Tax and Public Finance*, *5*(2), 123–151. https://doi.org/10.1023/A:1008669403754

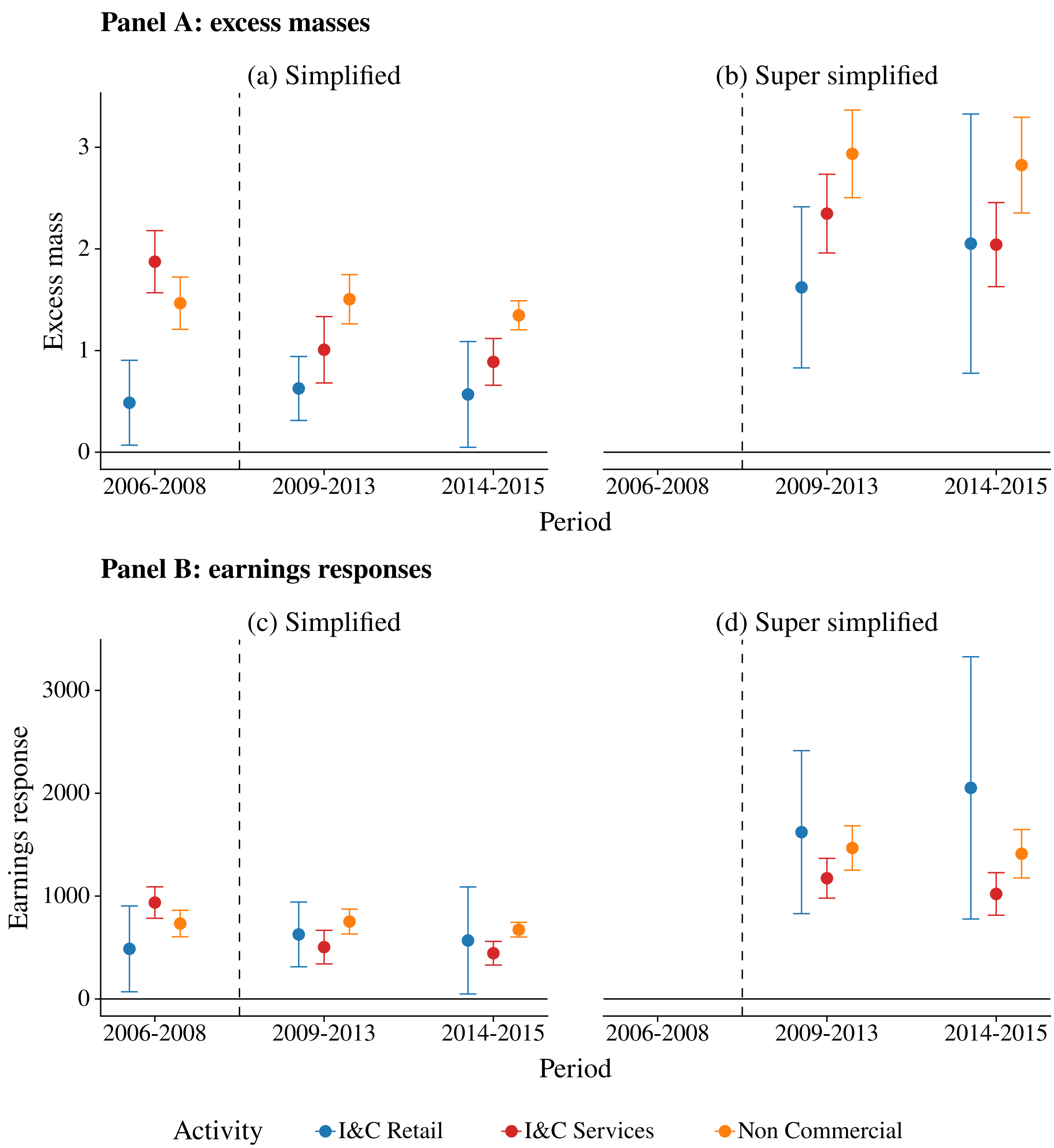
Zwick, E. (2021). The Costs of Corporate Tax Complexity. *American Economic Journal: Economic Policy*, *13*(2), 467–500. https://doi.org/10.1257/pol.20180406

Online Appendix for
"Tax Simplicity or Simplicity of Evasion? Evidence from Self-Employment Taxes in France"

by Philippe Aghion, Maxime Gravoueille, Mathieu Lequien, and Stefanie Stantcheva

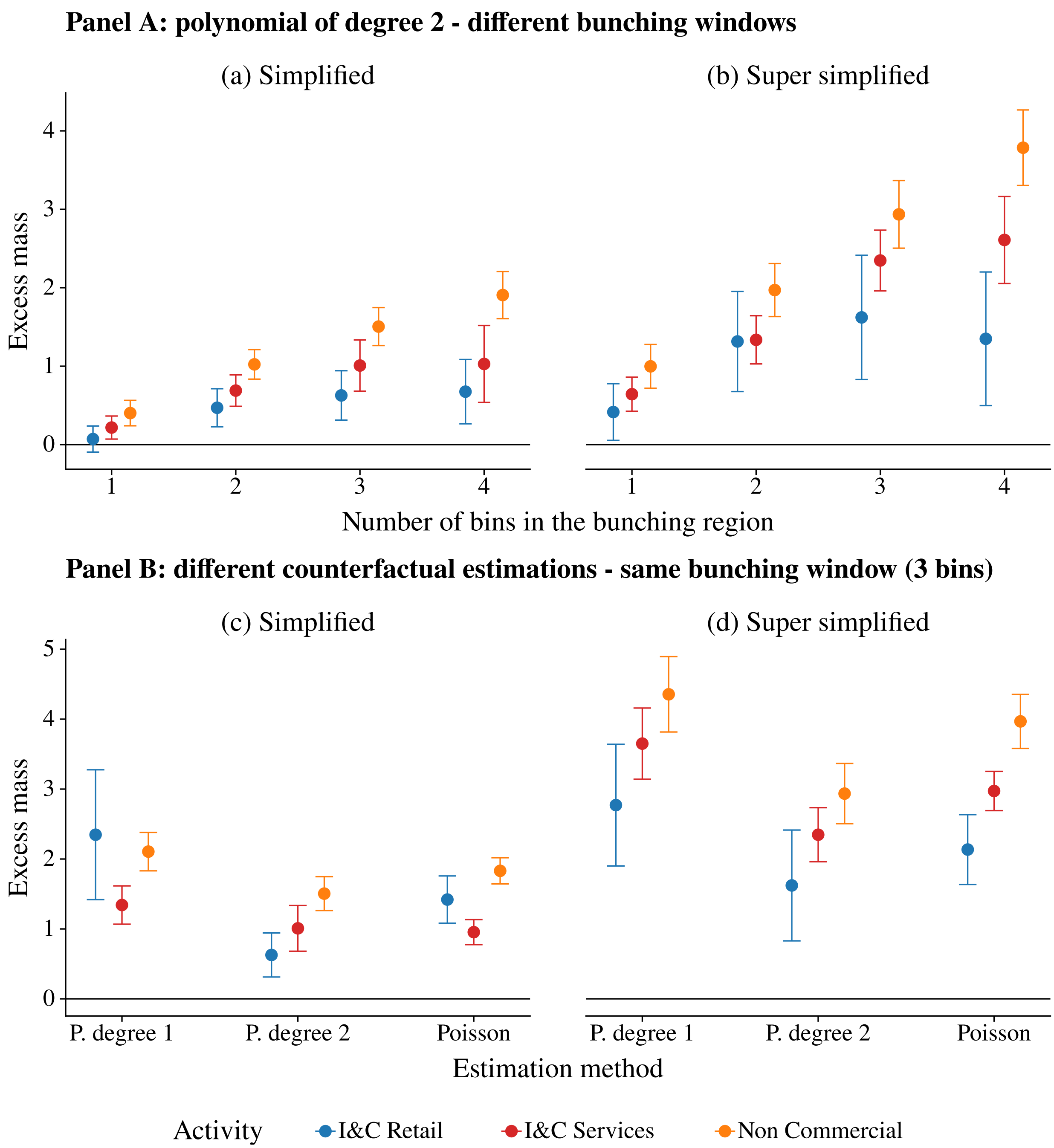
A Additional Figures

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



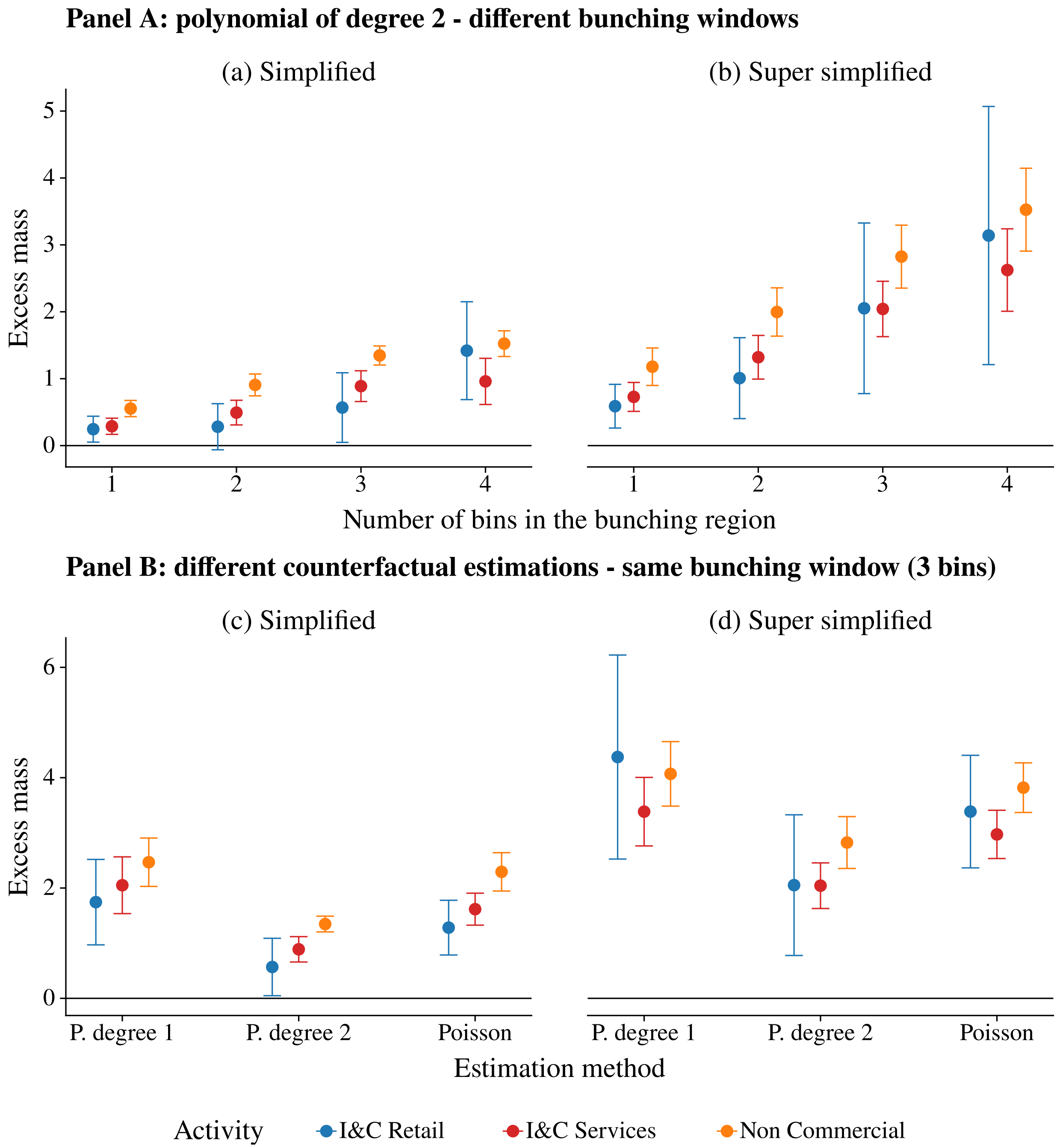
Note: The figure plots the excess masses b (panel A) and the earnings responses Δy^* (panel B) obtained from Section 3, categorized by regime, activity, and period. The population is restricted to individuals reporting zero labor earnings as defined in Appendix D. The pre-reform period spans from 2006 to 2008, period 1 spans from 2009 to 2013, and period 2 spans from 2014 to 2015. The counterfactual distribution is fitted using a smooth polynomial, as explained in Section 3. Standard errors for the excess masses are calculated using a bootstrap procedure with random resampling (n = 400) of the residuals. Standard errors for the earnings responses are determined from the excess masses, with the formula $se(\Delta y^*) = se(b) \times B_S$.

Figure A.2: Bunching estimation robustness tests, 2009-2013



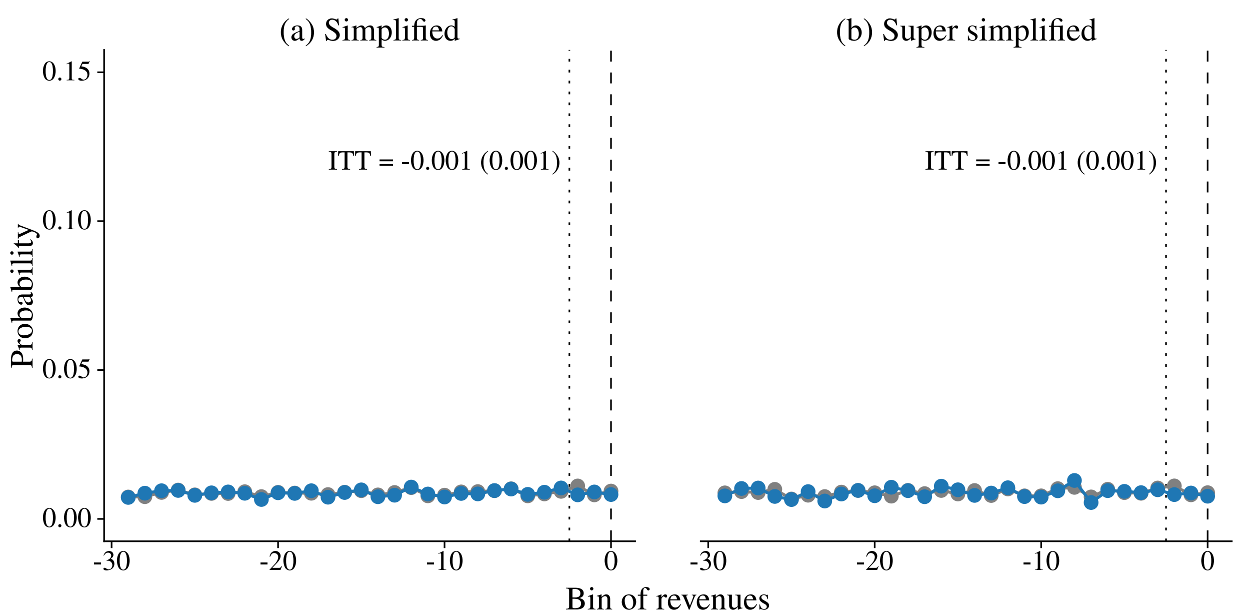
Note: The figure displays the excess masses b obtained from Section 3, by regime. The period of estimation spans from 2009 to 2013. In Panel A, the excess masses are estimated using a counterfactual distribution that is fitted using a smooth polynomial of degree 2, but with a different number of bins in the bunching region (x-axis). In Panel B, the excess masses are estimated using different functional forms for the counterfactual distribution (x-axis), while keeping the bunching window equal to 3 bins. Standard errors for the excess masses are calculated using a bootstrap procedure with random resampling (n = 400) of the residuals.

Figure A.3: Bunching estimation robustness tests, 2014-2015



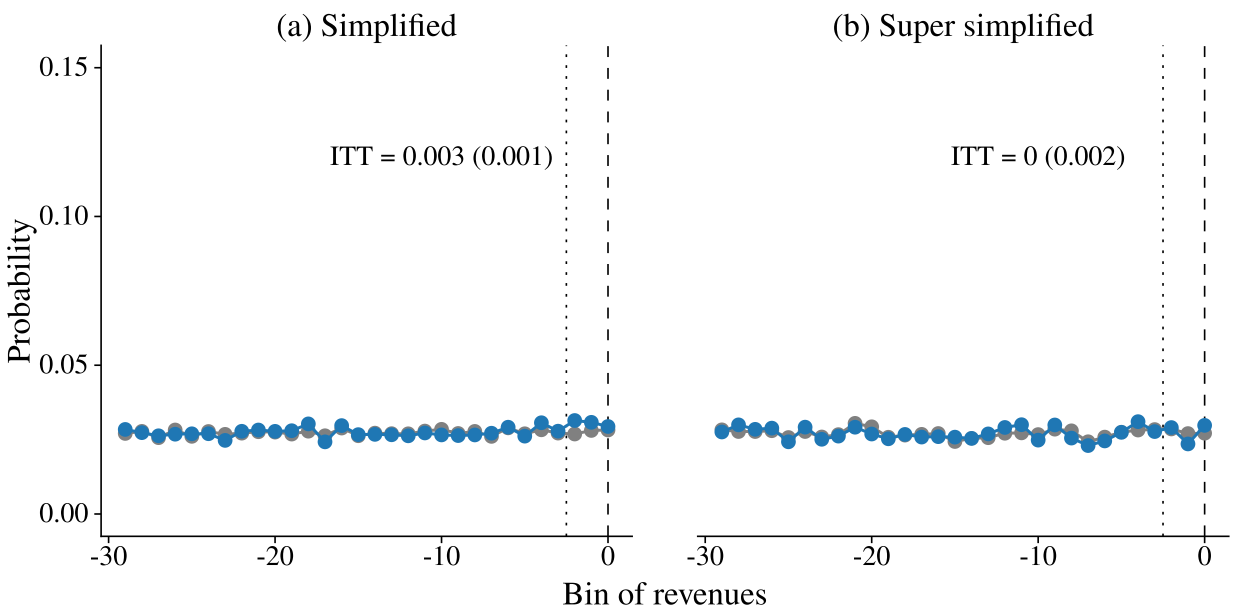
Note: The figure displays the excess masses *b* obtained from Section 3, by regime. The period of estimation spans from 2014 to 2015. In Panel A, the excess masses are estimated using a counterfactual distribution that is fitted using a smooth polynomial of degree 2, but with a different number of bins in the bunching region (x-axis). In Panel B, the excess masses are estimated using different functional forms for the counterfactual distribution (x-axis), while keeping the bunching window equal to 3 bins. Standard errors for the excess masses are calculated using a bootstrap procedure with random resampling (n = 400) of the residuals.

Figure A.4: Probability of reporting a multiple of 250



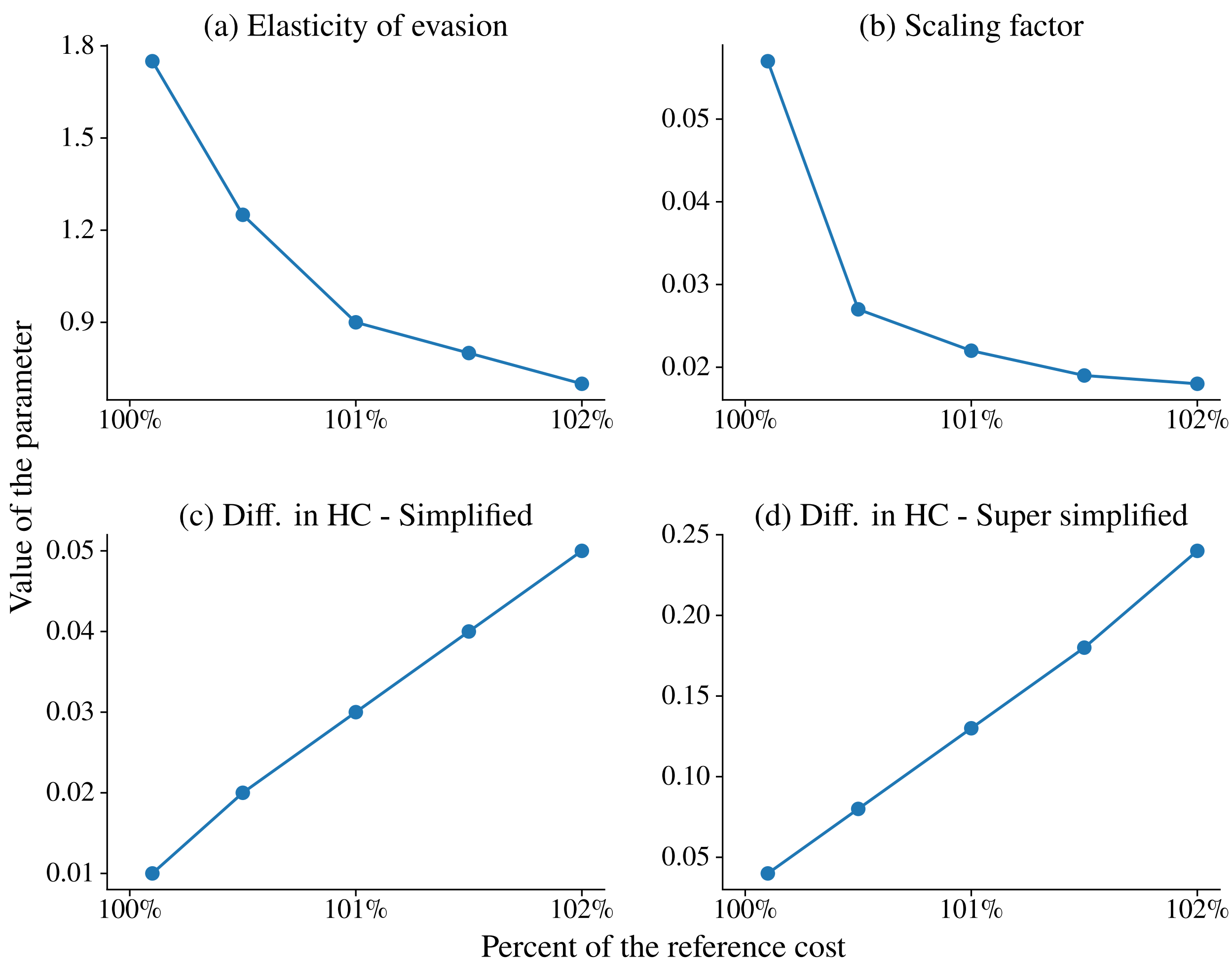
Note: The figure plots the probability of reporting a multiple of 250 euros (and excluding multiples of 500 euros) in self-employed revenues, by bins centered around the eligibility threshold (the vertical dashed line). The bunching region is depicted between the dotted and dashed vertical lines. The counterfactual distribution (in grey) is fitted using a smooth polynomial, following the explanation in Section 4. The results are based on the pooled population data from 2009 to 2015, and they are presented separately for the simplified and super simplified regimes. The ITT coefficient is calculated using the method described in Section 4. For both the simplified and super simplified, there is no difference in the probability to report a multiple of 250 in the bunching region. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

Figure A.5: Probability of reporting a multiple of 50



Note: The figure plots the probability of reporting a multiple of 50 euros (and excluding multiples of 100 and 250 euros) in self-employed revenues, by bins centered around the eligibility threshold (the vertical dashed line). The bunching region is depicted between the dotted and dashed vertical lines. The counterfactual distribution (in grey) is fitted using a smooth polynomial, following the explanation in Section 4. The results are based on the pooled population data from 2009 to 2015, and they are presented separately for the simplified and super simplified regimes. The ITT coefficient is calculated using the method described in Section 4. For both the simplified and super simplified, there is no difference in the probability to report a multiple of 50 in the bunching region. Standard errors are calculated using a bootstrap on the ITT procedure (n = 400).

Figure A.6: Structural parameters by cost, alternative tax definition



Note: This plot shows the results from the structural estimation, based on the data moments for 2009-2013 (period 1) and for individuals reporting zero labor earnings. This simulation applies the alternative tax definition. The scaling parameter and differences in hassle costs are expressed in percent of the taxable income in the simpler regime.

B Additional Tables

Table B.1: Robustness tests for the dynamic bunching estimation

Range of normalized self-empl. taxable income in the treatment group	Simplified B	Simplified ITT	Simplified LATE	Super simplified B	Super simplified ITT	Super simplified LATE
Panel A: distance between the control and treatment groups = 10000						
]-1000,0]	0.346	-0.083	-0.239	0.334	-0.088	-0.263
	(0.003)	(0.001)	(0.003)	(0.005)	(0.001)	(0.005)
]-1500,0]	0.33	-0.081	-0.246	0.308	-0.086	-0.278
	(0.003)	(0.001)	(0.003)	(0.004)	(0.001)	(0.005)
]-2000,0]	0.308	-0.079	-0.257	0.282	-0.084	-0.298
	(0.002)	(0.001)	(0.003)	(0.004)	(0.001)	(0.005)
Panel B: distance between the control and treatment groups = 15000						
]-1000,0]	0.338	-0.075	-0.223	0.326	-0.08	-0.247
	(0.003)	(0.001)	(0.003)	(0.005)	(0.001)	(0.005)
]-1500,0]	0.324	-0.074	-0.228	0.302	-0.078	-0.26
	(0.003)	(0.001)	(0.003)	(0.004)	(0.001)	(0.005)
]-2000,0]	0.301	-0.071	-0.237	0.275	-0.076	-0.278
	(0.003)	(0.001)	(0.003)	(0.004)	(0.001)	(0.005)

Note: The table reports the three key statistics of the dynamic bunching estimation, with different distances between the control and treatment group and different ranges of normalized self-employed taxable income. Panel A takes a distance of 10000 euros between the actual threshold and the placebo threshold, while Panel B takes a distance of 15000 euros. The results are based on the pooled population data for 2009-2015, and separate results are presented for the simplified and super simplified regimes. The population is restricted to individuals that are below the tolerance threshold in the subsequent period. The key statistics from our bunching analysis: i) The proportion of bunchers B concerning individual growth rates of self-employed taxable incomes. ii) The reduction in the growth rate of self-employment taxable incomes among the treated group (ITT). iii) The growth rate reduction in self-employed taxable income among the bunchers (LATE). Standard errors are calculated using a bootstrap procedure (n=400).

Table B.2: Average tax rates by regime, activity and period, alternative tax definition

	Simplified		Super simplified		
Activity	τ_m	τ_r	τ_f	$\tau_f \times (1-\bar{\mu})$	τ_r
Panel A: period 1 (2009-2013)					
I&C Retail	46.7%	52.08%	13.65%	47.07%	50.02%
I&C Services	48.17%	53.31%	24.25%	48.5%	51.42%
Non Commercial	34.93%	42.7%	21.64%	32.79%	39.35%
Panel B: period 2 (2014-2015)					
I&C Retail	51.41%	51.44%	14.77%	50.93%	48.76%
I&C Services	51.11%	51.18%	25.65%	51.3%	49.08%
Non Commercial	41.17%	41.57%	25.49%	38.62%	37.44%

Note: This table reports the average tax rates in the bunching region for individuals reporting zero labor earnings and based on the alternative method of computation by regime, activity and period. For the simplified regime, τ_m is the average tax rate on taxable self-employed income and τ_r is the counterfactual average tax rate in the standard regime. For the simplified regime, τ_f is the flat rate on gross self-employed revenues, $\tau_f/(1-\bar{\mu})$ is the flat rate expressed in percentage of taxable self-employed revenues and τ_r is the counterfactual average tax rate in the standard regime. All tax rates are computed based on the assumption that the cost is equal to the rebate. Panel A reports the tax rates for period 1 (2009-2013) and Panel B reports the tax rates for period 2 (2014-2015).

Table B.3: Evasion, preference for simplicity and cost implied by the model, alternative tax definition

Percent of the reference cost	Regime	Activity	Evaded amount	Diff. in HC	Cost
100.1%	Simplified	I&C Retail	668	235	74.1%
		I&C Services	581	162	54.9%
		Non Commercial	846	214	46.1%
	Super simplified	I&C Retail	1663	942	72.8%
		I&C Services	1251	651	52.9%
		Non Commercial	1552	859	45.1%
100.5%	Simplified	I&C Retail	679	469	74.4%
		I&C Services	578	324	55.1%
		Non Commercial	845	428	46.2%
	Super simplified	I&C Retail	1675	1885	73.1%
		I&C Services	1248	1301	53.1%
		Non Commercial	1553	1718	45.2%
101.0%	Simplified	I&C Retail	713	704	74.7%
		I&C Services	603	487	55.4%
		Non Commercial	878	642	46.5%
	Super simplified	I&C Retail	1708	3063	73.4%
		I&C Services	1274	2114	53.4%
		Non Commercial	1587	2791	45.5%
101.5%	Simplified	I&C Retail	717	939	75.1%
		I&C Services	603	649	55.6%
		Non Commercial	878	857	46.7%
	Super simplified	I&C Retail	1713	4240	73.8%
		I&C Services	1273	2928	53.6%
		Non Commercial	1588	3865	45.7%
102.0%	Simplified	I&C Retail	731	1174	75.5%
		I&C Services	612	811	55.9%
		Non Commercial	891	1071	46.9%
	Super simplified	I&C Retail	1727	5654	74.2%
		I&C Services	1282	3904	53.9%
		Non Commercial	1600	5153	45.9%

Note: This table shows the evaded amounts (in euros), differences in hassle costs (in euros) and costs (in percent of self-employed revenues) predicted by the structural parameters from Figure 12 by regime, activity and for period 1(2009-2013).

Table B.4: Goodness-of-fit of the model

Percent of the reference cost	Regime	Activity	(a) In-sample earnings responses Model	Data	Diff.	Loss	(b) Out-of-sample earnings responses Model	Data	Diff.	Loss
100.1%	Simplified	I&C Retail	634	627	1%	12.4e-04	644	568	13%	259e-04
		I&C Services	525	504	4%		527	444	19%	
		Non Commercial	719	752	-4%		723	674	7%	
	Super simplified	I&C Retail	1550	1622	-4%		1584	2052	-23%	
		I&C Services	1172	1174	0%		1199	1021	17%	
		Non Commercial	1525	1468	4%		1588	1412	12%	
100.5%	Simplified	I&C Retail	632	627	1%	8.4e-04	643	568	13%	228.2e-04
		I&C Services	520	504	3%		527	444	19%	
		Non Commercial	723	752	-4%		732	674	9%	
	Super simplified	I&C Retail	1584	1622	-2%		1612	2052	-21%	
		I&C Services	1150	1174	-2%		1169	1021	14%	
		Non Commercial	1524	1468	4%		1559	1412	10%	
101.0%	Simplified	I&C Retail	629	627	0%	7.3e-04	642	568	13%	221.2e-04
		I&C Services	521	504	3%		530	444	19%	
		Non Commercial	729	752	-3%		740	674	10%	
	Super simplified	I&C Retail	1613	1622	-1%		1637	2052	-20%	
		I&C Services	1145	1174	-2%		1161	1021	14%	
		Non Commercial	1528	1468	4%		1550	1412	10%	
101.5%	Simplified	I&C Retail	629	627	0%	7.2e-04	642	568	13%	221.7e-04
		I&C Services	523	504	4%		533	444	20%	
		Non Commercial	735	752	-2%		746	674	11%	
	Super simplified	I&C Retail	1612	1622	-1%		1634	2052	-20%	
		I&C Services	1133	1174	-3%		1148	1021	12%	
		Non Commercial	1517	1468	3%		1534	1412	9%	
102.0%	Simplified	I&C Retail	629	627	0%	7.1e-04	643	568	13%	219.1e-04
		I&C Services	523	504	4%		534	444	20%	
		Non Commercial	736	752	-2%		748	674	11%	
	Super simplified	I&C Retail	1622	1622	0%		1643	2052	-20%	
		I&C Services	1132	1174	-4%		1146	1021	12%	
		Non Commercial	1516	1468	3%		1530	1412	8%	

Note: This table shows the empirical fit from the structural estimation in Figure 12, based on individuals reporting zero labor earnings, and by regime, activity and period. This simulation uses the main tax definition. The deviation is computed as the percentage deviation between the simulated moments and the empirical moments, $(\Delta y^*_{sk} - \hat{\Delta} y^*_{sk})/\hat{\Delta} y^*_{sk}$. The in-sample results are based on period 1 (2009-2013) and the out-of-sample results are based on period 2 (2014-2015).

Table B.5: Goodness-of-fit of the model - Alternative tax definition

Percent of the reference cost	Regime	Activity	(a) In-sample earnings responses Model	Data	Diff.	Loss	(b) Out-of-sample earnings responses Model	Data	Diff.	Loss
100.1%	Simplified	I&C Retail	642	627	2%	18.7e-04	665	568	17%	296.9e-04
		I&C Services	521	504	3%		539	444	21%	
		Non Commercial	692	752	-8%		736	674	9%	
	Super simplified	I&C Retail	1564	1622	-4%		1598	2052	-22%	
		I&C Services	1173	1174	0%		1199	1021	17%	
		Non Commercial	1529	1468	4%		1588	1412	12%	
100.5%	Simplified	I&C Retail	633	627	1%	10.5e-04	655	568	15%	252.2e-04
		I&C Services	518	504	3%		534	444	20%	
		Non Commercial	711	752	-5%		740	674	10%	
	Super simplified	I&C Retail	1594	1622	-2%		1622	2052	-21%	
		I&C Services	1153	1174	-2%		1173	1021	15%	
		Non Commercial	1530	1468	4%		1563	1412	11%	
101.0%	Simplified	I&C Retail	630	627	0%	7.9e-04	652	568	15%	232.7e-04
		I&C Services	519	504	3%		534	444	20%	
		Non Commercial	723	752	-4%		744	674	10%	
	Super simplified	I&C Retail	1608	1622	-1%		1631	2052	-21%	
		I&C Services	1136	1174	-3%		1152	1021	13%	
		Non Commercial	1518	1468	3%		1538	1412	9%	
101.5%	Simplified	I&C Retail	619	627	-1%	7.4e-04	642	568	13%	221.3e-04
		I&C Services	518	504	3%		533	444	20%	
		Non Commercial	727	752	-3%		746	674	11%	
	Super simplified	I&C Retail	1612	1622	-1%		1634	2052	-20%	
		I&C Services	1133	1174	-3%		1148	1021	12%	
		Non Commercial	1517	1468	3%		1534	1412	9%	
102.0%	Simplified	I&C Retail	619	627	-1%	7.1e-04	642	568	13%	218.8e-04
		I&C Services	519	504	3%		534	444	20%	
		Non Commercial	730	752	-3%		747	674	11%	
	Super simplified	I&C Retail	1622	1622	0%		1643	2052	-20%	
		I&C Services	1132	1174	-4%		1146	1021	12%	
		Non Commercial	1516	1468	3%		1530	1412	8%	

Note: This table shows the empirical fit from the structural estimation in Figure 12, based on individuals reporting zero labor earnings, and by regime, activity and period. This simulation uses the alternative tax definition. The deviation is computed as the percentage deviation between the simulated moments and the empirical moments, $(\Delta y^* - \hat{\Delta} y^*)/\hat{\Delta} y^*$. The in-sample results are based on period 1 (2009-2013) and the out-of-sample results are based on period 2 (2014-2015).

C Structural Model

C.1 Setup

Preferences. Each agent choose to produce under the super simplified, simplified or standard regime ($i = f, m, r$). We also refer to simpler regimes with $s \in \{f, m\}$. Each agent is also in one of the three following activities indexed by $k \in$ {I&C Retail, I&C Services, Non Commercial}. An agent generates real revenues y_{ik} and reported revenues $\tilde{y}_{ik}$. She also has a type θ_{ik} that captures her productivity. The disutility of generating revenues y_{ik} for an agent of type θ_{ik} is denoted by $h(y_{ik}, \theta_{ik})$, increasing in y_i and decreasing in θ_{ik}. The cost of misreporting revenues is denoted by $g(y_{ik}, \tilde{y}_{ik})$, increasing in y_{ik} and decreasing in $\tilde{y}_{ik}$. An agent's utility from earning revenue y_{ik} and reporting $\tilde{y}_{ik}$ is thus:

$$u(y_{ik}, \tilde{y}_{ik}) = y_{ik}(1 - c_{ik}) - T_{ik}(\tilde{y}_{ik}) - h(y_{ik}; \theta_{ik}) - g(y_{ik}, \tilde{y}_{ik}) - a_{ik},$$

where c_{ik} is the cost of producing y_{ik}, $T_{ik}(\tilde{y}_{ik})$ is the total tax liability as a function of reported revenues and a_{ik} a hassle cost. This hassle cost is decreasing with the simplicity of the regime: $a_{rk} > a_{mk} > a_{fk}$.

Parametric assumptions. Consistent with the previous literature, we assume the following functional forms for $h(y_{ik}; \theta_{ik})$ and $g(y_{ik}, \tilde{y}_{ik})$:

$$h(y_{ik}; \theta_{ik}) = \frac{\theta_{ik}}{1 + \frac{1}{\varepsilon}} \left(\frac{y_{ik}}{\theta_{ik}} \right)^{1+\frac{1}{\varepsilon}} \quad \text{and} \quad g(y_{ik}, \tilde{y}_{ik}) = \frac{\kappa_{ik}}{1 + \frac{1}{\eta}} \left(\frac{y_{ik} - \tilde{y}_{ik}}{\kappa_{ik}} \right)^{1+\frac{1}{\eta}},$$

where ε is the real income elasticity, η is the evasion elasticity and κ_{ik} a scaling parameter. In the standard regime, the reporting cost is implicitly infinite because of institutional constraints preventing misreporting.

C.2 Full model

C.2.1 Simplified regime

Without the notch. Let us start by describing the interior solution (without the notch) for the simplified regime ($i = m$). Now, the total tax liability is defined as $T_{mk}(\tilde{y}_{mk}) = \tilde{y}_{mk}(1 - \mu_{mk})\tau_{mk}$. The optimal choice of real and reported revenues of an agent are (from the first order conditions on

y_{mk} and $\tilde{y}_{mk}$):

$$\begin{aligned} \frac{\partial u_{mk}}{\partial y_{mk}} &= (1-c_{mk}) - \left(\frac{y_{mk}}{\theta_{mk}}\right)^{\frac{1}{\varepsilon}} - \left(\frac{y_{mk}-\tilde{y}_{mk}}{\kappa_{mk}}\right)^{\frac{1}{\eta}} = 0, \\ \frac{\partial u_{mk}}{\partial \tilde{y}_{mk}} &= -(1-\mu_{mk})\tau_{mk} + \left(\frac{y_{mk}-\tilde{y}_{mk}}{\kappa_{mk}}\right)^{\frac{1}{\eta}} = 0, \end{aligned}$$

which implies that:

$$y_{mk} = \theta_{mk}[(1-c_{mk})-\tau_{mk}(1-\mu_{mk})]^{\varepsilon} \quad \text{and} \quad \tilde{y}_{mk} = \theta_{mk}[(1-c_{mk})-\tau_{mk}(1-\mu_{mk})]^{\varepsilon}-\kappa[\tau_{mk}(1-\mu_{mk})]^{\eta}.$$

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

$$y_{rk} = \theta_{rk}\left[(1-c_{rk}) - (1-c_{rk})\tau_{rk}\right]^{\varepsilon}$$

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

$$y^*_{mk} + \Delta y^*_{mk} = (\theta^*_{mk} + \Delta\theta^*_{mk})[(1-c_{mk}) - \tau_{mk}(1-\mu_{mk})]^{\varepsilon} - \kappa_{mk}[\tau_{mk}(1-\mu_{mk})]^{\eta}, \tag{11}$$

and actual revenues:

$$y_{mk} = (\theta^*_{mk} + \Delta\theta^*_{mk})[(1-c_{mk}) - \tau_{mk}(1-\mu_{mk})]^{\varepsilon}.$$

With the notch, this agent reports revenues at the notch, but her actual revenues are $y^A_{mk} = y_{mk}(y^*_{mk})$ and are a function of the reported revenues. y^A_{mk} is given by:

$$\max_{y^A_{mk}} u(y^A_{mk}; y^*_{mk}) = y^A_{mk}(1-c_{mk}) - \tau_{mk}(1-\mu_{mk})y^*_{mk} - h(y^A_{mk}; \theta^*_{mk} + \Delta\theta^*_{mk}) - g(y^A_{mk}, y^*_{mk}) - a_{mk},$$

which implies:

$$(1-c_{mk}) - \underbrace{\left(\frac{y^A_{mk}}{\theta^*_{mk} + \Delta\theta^*_{mk}}\right)^{\frac{1}{\varepsilon}}}_{h'(y^A_{mk};\theta^*_{mk}+\Delta\theta^*_{mk})} - \underbrace{\left(\frac{y^A_{mk} - y^*_{mk}}{\kappa_{mk}}\right)^{\frac{1}{\eta}}}_{g'(y^A_{mk}, y^*_{mk})} = 0. \tag{12}$$

Let us denote by y^I_{rk} the indifference point in the standard regime, such that the agent is indifferent between earning revenues y^A_{mk} and reporting revenues exactly equal to the threshold y^*_{mk} or earning y^I_{rk} (which is actual revenues, since there is no misreporting in the standard regime). y^I_{rk} is interior, and hence characterized by the optimal (tangency) condition in standard regime:

$$y^I_{rk} = (\theta^*_{mk} + \Delta\theta^*_{mk})[(1 - c_{rk})(1 - \tau_{rk})]^{\varepsilon}. \tag{13}$$

The indifference condition $u^I_{rk} = u^*_{mk}$ gives:

$$\begin{aligned} y^I_{rk}(1 - c_{rk})(1 - \tau_{rk}) - h(y^I_{rk}, \theta^*_{mk} + \Delta\theta^*_{mk}) - a_{rk} = \\ y^A_{mk}(1 - c_{mk}) - \tau_{mk}(1 - \mu_{mk})y^*_{mk} - h(y^*_{mk}; \theta^*_{mk} + \Delta\theta^*_{mk}) - g(y^A_{mk}, y^*_{mk}) - a_{mk}, \end{aligned} \tag{14}$$

C.3 Model with no real elasticity of income

We now take the limit case $\epsilon \to 0$, corresponding to no real elasticity of income:

$$\theta^*_{mk} + \Delta\theta^*_{mk} = y^A_{mk} \times \left[(1 - c_{mk}) - \left(\frac{y^A_{mk} - y^*_{mk}}{\kappa_{mk}}\right)^{\frac{1}{\eta}}\right]^{-\varepsilon} \longrightarrow y^A_{mk}, \tag{15}$$

$$y^I_{rk} = (\theta^*_{mk} + \Delta\theta^*_{mk})[(1 - c_{rk})(1 - \tau_{rk})]^{\varepsilon} \longrightarrow y^A_{mk}, \tag{16}$$

$$\left[y^I_{rk}(1 - c_{rk})(1 - \tau_{rk})\right] - \left[y^A_{mk}(1 - c_{mk}) - \tau_{mk}(1 - \mu_{mk})y^*_{mk} - g(y^A_{mk}, y^*_{mk})\right] - \Delta a_{mk} = 0, \tag{17}$$

$$\begin{aligned} y^*_{mk} + \Delta y^*_{mk} = (\theta^*_{mk} + \Delta\theta^*_{mk})[(1 - c_{mk}) - \tau_{mk}(1 - \mu_{mk})]^{\varepsilon} - \kappa_{mk}[\tau_{mk}(1 - \mu_{mk})]^{\eta} \\ \longrightarrow y^A_{mk} = (y^*_{mk} + \Delta y^*_{mk}) + \kappa_{mk}[\tau_{mk}(1 - \mu_{mk})]^{\eta}. \end{aligned} \tag{18}$$

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

$$y^A_{mk}(c_{mk} - c_{rk}) + \left[y^*_{mk}(1 - \mu_{mk})\tau_{mk} - y^A_{mk}(1 - c_{rk})\tau_{rk}\right] + \frac{\kappa_{mk}}{1 + \frac{1}{\eta}}\left(\frac{y^A_{mk} - y^*_{mk}}{\kappa_{mk}}\right)^{1+\frac{1}{\eta}} - \Delta a_{mk} = 0.$$

If we assume $c_{rk} = c_{mk} = c_k$, it reduces to:

$$\left[y^*_{mk}(1-\mu_{mk})\tau_{mk} - y^A_{mk}(1-c_k)\tau_{rk}\right] + \frac{\kappa_{mk}}{1+\frac{1}{\eta}}\left(\frac{y^A_{mk} - y^*_{mk}}{\kappa_{mk}}\right)^{1+\frac{1}{\eta}} - \Delta a_{mk} = 0. \tag{19}$$

C.4 Estimation methodology

We inject equation (18) into equation (19), reducing the problem to only one equation with one unknown variable Δy^*. For a given vector of structural parameters $\Omega = \{\Omega_{sk}\}$, policy parameters $\Phi = \{\Phi_{sk}\}$ and cost c_{sk}, we solve the problem using the following sequence:

1. We define the number N of moments to use for the estimation, which we impose to be greater than or equal to the number of structural parameters we wish to estimate. In our estimation strategy, we have 6 moments = 2 regimes × 3 activities, and we estimate 4 parameters $\Omega = (\eta, \tilde{\kappa}, \tilde{\Delta}a_m, \tilde{\Delta}a_f)$. It contains two hassle costs that are regime-specific, one evasion elasticity, and one scaling parameter. More details on the choice of these parameters are available in Section 5.

2. We define a grid of potential values for each parameter:

   - η: 59 evenly spaced numbers between 0.1 and 3 (included).
   - $\tilde{\kappa}$: 100 evenly spaced numbers between 0.001 and 0.1.
   - $\tilde{\Delta}a_m$: 11 evenly spaced numbers between 0 and 0.1.
   - $\tilde{\Delta}a_f$: 31 evenly spaced numbers between 0 and 0.3.

3. We loop over the grids of structural parameters and perform the following operations on the set of candidate structural parameters:

   (a) We solve for equation (19) for each moment separately (i.e., combination of regime and activity). We use a non-linear least-squares solver to incorporate bounds on Δy^*. In particular, we impose that $\Delta y^* \in [0, +\infty[$.

   (b) We compute the loss function over all moments used for the estimation:

   $$L(\Omega) = \frac{1}{N}\sum_{s,k}\left(\frac{\Delta y^*_{sk}(\Omega, \Phi_{sk}) - \hat{\Delta} y^*_{sk}}{se(\hat{\Delta} y^*_{sk})}\right)^2,$$

   where $se(\hat{\Delta} y^*_{sk})$ is the standard error of the earnings response $\hat{\Delta} y^*_{sk}$.

   (c) We select the set of structural parameters that minimize the loss.

D Data Construction

D.1 Data

POTE. Our primary dataset comprises the entirety of French tax returns for the period 2006-2015, which is compiled by the French Internal Revenue Service. These income tax returns provide extensive income-related data at both the individual and household levels, along with critical demographic details such as household composition, individual age, and gender. It is crucial to note that this dataset is panel data, featuring distinct individual and household identifiers (both anonymized) that enable tracking over time.

New entrepreneurs information system. The second dataset we use is a survey provided by the French National Statistics Institute, which is available for the years 2010 and 2014. This survey gathers information from entrepreneurs about their experiences during the initial years of their business activities.

D.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 old in a given year. We include only main filers, excluding dependents such as children. We also exclude individuals who experienced changes in their marital status, specifically those who divorced or had their spouse pass away. Additionally, for years before 2010 (inclusive), we do not include years in which individuals got married. This exclusion is due to the French personal income tax being reported at the household level, leading to different reporting requirements for individuals who change their marital status in a given tax year. Furthermore, we retain only individuals and households that are uniquely observed in a given year. Finally, our analysis is restricted to the 2006-2015 period.

Self-employment restrictions. We begin by limiting our sample to individuals with self-employed revenues in either the simple regime, the super simplified regime, or the standard regime for a specific year. We then retain individuals with self-employment income who can be uniquely classified in a regime and activity, excluding those with self-employment in multiple regimes and activities. Additionally, we exclude households with any agricultural income, as they are subject to specific tax parameters.

In the case of the super simplified regime, there is an additional requirement concerning family income as of year $t-2$. This family income should be below a year-specific threshold denoted as f_t^*, corresponding to the third tax bracket cutoff. We exclude individuals who are under the super simplified regime and have family income above this threshold. It is worth noting that this situation should not occur according to discussions with the tax administration, but it is observed in the data. Possible explanations include differences in reporting requirements, errors, or unobserved changes in tax regimes. Since we lack further information, our focus is on individuals in the super simplified regime who are also eligible for it.

D.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, which includes wages and salaries reported in item box **1AJ** for the first filer and in item box **1BJ** for the second filer. This definition encompasses most sources of labor earnings, and an individual is considered to have labor earnings if the reported amount is strictly greater than zero

Tax rates. We calculate the effective average tax rates by regime (indexed as i), activity (indexed as k), and year (indexed as t). At the eligibility threshold, the effective average tax rate ($\tau_{i,k,t}$) consists of two components: an income tax with a rate of $\tau_{i,k,t}^{inc}$ and social contributions with a rate of $\tau_{i,k,t}^{sc}$. In practice, we compute the tax rates for each individual in the bunching region (i.e., close to the threshold). Below, we provide details about the self-employment tax system during the period from 2009 to 2015.

First, we begin with the super simplified regime. In this regime, both income taxes and social contributions are calculated using flat tax rates applied to gross self-employed revenues denoted as y. The flat income tax rate is set at 1% for I&C Retail activities, 1.7% for I&C Services activities, and 2.2% for Non Commercial activities. As for social contributions, they are subject to flat rates ranging from 12% to 15% for I&C Retail activities, 21% to 25% for I&C Services activities, and 18% to 24% for Non Commercial activities, depending on the specific year.

Second, there are two different methods available to self-employed individuals in the simplified regime for calculating income taxes and social contributions. The first option involves progressive taxes, applied to both income tax and social contributions, which are levied on self-employment taxable income (self-employment revenues minus the applicable rebate). The second option employs a flat tax rate on self-employment revenues for social contributions, similar to the super simplified regime, along with a progressive income tax levied on self-employment taxable income. We do not have data to observe which option individuals choose. In the main estimate of this paper, we adopt the first option. This choice is advantageous because it results in the effective average tax rate difference between the simplified regime and the standard regime at the threshold being close to zero. This setup is ideal for identifying hassle costs. We calculate the average tax rates at the threshold. The average income tax rate is directly observable in the data and ranges from 2% to 8%, depending on the activity and period. The average social contributions tax rate is approximately 47% to 50% for I&C Retail and I&C Services activities, and around 36% for Non Commercial activities. It is worth noting that computing the social contribution tax rates is more complex because it involves multiple taxes on various sub-activities that we do not directly observe in the data. We estimate these rates based on information available at https://www.ipp.eu/en/ipp-tax-and-benefit-tables/.

Finally, we impute the counterfactual effective average tax rates assuming individuals had chosen the standard regime. To do this, we calculate income and social contributions levied on self-employed taxable income, under the assumption that the production cost is equal to the rebate, and with progressive income taxes and social contributions (similar to the first option of the simplified regime). We utilize the average income tax rate directly available in the data, which accounts for self-employed incomes in the super simplified regime. Next, we calculate the average social contributions tax rate using the same method as described for the simplified regime. The implicit assumption is that the counterfactual effective average tax rate computed for individuals at the threshold is a reasonable approximation for the situation further above the threshold in the standard regime. This assumption is justified by the relatively small expected amount of evasion and that the rebate is close to the production cost by design.

Note that we do not take into account potential differences in value-added tax liabilities as we have no further information available on them.

E French Tax Calculation Primer

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

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

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

In brief, the family coefficient serves the same role as the taxable income in the U.S. for determining the tax bracket and total tax paid "per-part."

Tax liability. To get the total tax liability of the household, the "per-part" tax is inflated by the number of parts. The French tax schedule typically looks as follows:

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

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

$$y = \frac{Y}{N}. \tag{20}$$

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

$$T(y, N) = N \times [\sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times (y - \underline{y_{M-1}})]. \tag{21}$$

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

$$T(y, N) = N \times (\tau_1 \times \underline{y_1} + \tau_2 \times (\underline{y_2} - \underline{y_1}) + \tau_3 \times (y - \underline{y_2})). \tag{22}$$

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

Situation 1. If the additional number of parts N_a does not place the household in a higher tax bracket, then:

$$T(y^b, N_b) = N_b \times [\sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times (y^b - \underline{y_{M-1}})]. \quad (23)$$

The difference in taxes is:

$$T(y^b, N_b) - T(y, N) = (N_b - N) \times \sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times (N_b y^b - N_b \underline{y_{M-1}} - N y + N \underline{y_{M-1}}). \quad (24)$$

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

$$T(y^b, N_b) - T(y, N) = (N_b - N) \times \sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times \underline{y_{M-1}}(N - N_b). \quad (25)$$

We can re-arrange the expression to obtain:

$$T(y^b, N_b) - T(y, N) = (N_b - N) \times [\sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) - \tau_M \times \underline{y_{M-1}})]. \quad (26)$$

Situation 2. If the additional number of parts N_a places the household in a higher tax bracket, then:

$$T(y^b, N_b) = N_b \times [\sum_{m=1}^{M} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_{M+1} \times (y^b - \underline{y_M})]. \quad (27)$$

The difference in taxes is:

$$T(y^b, N_b) - T(y, N) = (N_b - N) \times \sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times (N_b \underline{y_M} - N_b \underline{y_{M-1}} - N y + N \underline{y_{M-1}}) + \tau_{M+1} N_b \times (y^b - \underline{y_M}). \quad (28)$$