Skatteforsk
Centre for Tax Research

Working Paper No. 5

Tax Complexity as Price Discrimination

March 2024

Ole Agersnap
Princeton University

Julie Brun Bjørkheim
NHH Norwegian School of Economics

Financial support from The Research Council of Norway (Grant number 341289) is gratefully acknowledged. We thank Annette Alstadsæter, Ron Davies, Henrik Kleven, John Sturm, Damian Vergara, Owen Zidar and seminar participants at Princeton and the DAEiNA annual congress for helpful comments. All errors are our own.

Tax Complexity as Price Discrimination*

Ole Agersnap† Julie Brun Bjørkheim‡

March 18, 2024

Ole Agersnap's Job Market Paper
Most recent version here

Abstract

Most tax systems around the world are highly complex. While several economists have studied the potential costs associated with tax complexity, few have explored if complexity can also have beneficial effects. In a novel model where taxpayers can acquire costly knowledge to reduce their tax burden, we show that when elasticities of taxable income are heterogeneous, a complex tax system can act as a sorting device similar to second-degree price discrimination, where more elastic taxpayers will invest in more tax knowledge. We prove that if elasticities are increasing with income, introducing tax complexity can allow the government to raise higher tax revenues at no efficiency cost. However, we show that complexity primarily benefits the highest earners and thus exacerbates inequality.

In the empirical section of our work, we study a complex tax system in Norway. Using rich register data on business owners, we demonstrate that many taxpayers make accounting decisions that cause them to pay higher taxes than would have been possible, and we quantify the exact size of this tax overpayment at the individual level. We show that overpayment tends to be larger for women, the less wealthy, and immigrants. We validate our model predictions by showing that failure to optimize is associated with a lower estimated tax elasticity.

JEL classification: H20, H24, H31, H26

Keywords: Taxation; Personal Income Taxes; Fiscal Policies and Behavior of Economic Agents: Household; Tax Evasion and Avoidance

*Financial support from The Research Council of Norway (Grant number 341289) is gratefully acknowledged. We thank Annette Alstadsæter, Ron Davies, Henrik Kleven, John Sturm, Damian Vergara, Owen Zidar and seminar participants at Princeton and the DAEiNA annual congress for helpful comments. All errors are our own.

†Princeton University. Email: agersnap@princeton.edu

‡Norwegian School of Economics. Email: julie.bjorkheim@nhh.no

1 Introduction

Why are most modern tax systems so complex? The benefits of simplifying the tax code would seem obvious at a first glance: Fewer compliance costs, less filing requirements and administrative work both for taxpayers and tax authorities, fewer loopholes that can be used for tax avoidance. Many academic researchers have found significant costs associated with complex tax systems (Benzarti, 2020; Pitt and Slemrod, 1989) and have recommended simplifying the tax code (Slemrod, 1989; Fuest, Peichl and Schaefer, 2006). And importantly, simpler tax systems seem to be overwhelmingly popular with voters and the public – for instance, Blesse, Buhlmann and Dörrenberg (2019) find that over 90% of Germans support tax simplification. Nevertheless, complex tax systems continue to be the norm in much of the world.

In this paper, we demonstrate one possible reason why governments might prefer a complex tax system: When taxpayers have heterogeneous elasticities of taxable income (ETI), tax complexity can serve as a form of price discrimination. A monopolist engaging in second-degree price discrimination can extract a higher total profit from its customers by offering bulk deals, limited-time sales or other advantages that cater to more price-sensitive customers, while the less elastic shoppers often pay full price. In a similar way, governments can offer deductions, exemptions and loopholes that can reduce the tax burden for those willing to invest in knowledge about the tax code, while the less knowledgeable end up paying more. In the theoretical section of this paper, we first show that a hypothetical government with perfect information and the power to set individualized tax rates would optimally make taxes lower on more elastic taxpayers. Of course, this is not feasible as governments cannot observe individual elasticities or set individual tax rates, but we go on to demonstrate that a complex tax system can achieve a similar outcome through a self-selection mechanism. We develop a model of tax complexity in which taxpayers can decide to invest in costly knowledge, which will reduce their tax burden. We prove that the more elastic a taxpayer is, the more tax knowledge they will choose to acquire, since a higher elasticity means their individual welfare gains from reducing tax distortions are larger. This demonstrates that tax complexity can serve as a self-selection mechanism akin to price discrimination.

We also show that if elasticities are increasing with income, this self-selection effect can allow the government to raise a higher overall tax revenue than the maximum possible in a "simple" tax system, and without leaving anyone worse off. The reason is that when taxpayers have heterogeneous ETIs, the tax rate that maximizes overall revenues will be inefficiently high for the most elastic individuals. Allowing for complexity gives them a way to reduce their distortionary tax burden and increase their labor supply significantly, while at the same time, the government can continue to extract full tax revenue from those with lower elasticities who will not find it worth the cost to invest in tax knowledge.

While our theoretical model shows that tax complexity can be efficiency-improving under some circumstances, there are of course many negative aspects of tax complexity as well. Generally, complexity benefits the wealthiest and most elastic taxpayers, and therefore increases inequality, which will be at odds with many policymakers' objectives. In a simulation exercise, we demonstrate that when the initial tax rate is substantially below the revenue-maximizing rate, complexity can boost labor output but will generally not lead to increased tax revenues. A government interested in boosting growth might therefore use tax complexity as a tool, but this will come at the expense of increased inequality and somewhat lower tax revenues.

In the second half of this paper, we empirically study the consequences of a complex tax system using de-identified administrative micro-data from Norway. We study sole owners of businesses who, when transferring money from their firm to their personal accounts, can choose to pay it either as wages or as dividends. These two forms of payment differ only in their tax treatment: Dividends are taxed at a flat rate, whereas wages are subject to a progressive tax schedule. For most business owners, the optimal strategy to minimize tax liability will therefore be to pay any transfers in the form of wages up until the point where marginal tax rates cross, and as dividends after that. However, we observe that many taxpayers do not behave optimally, and that this leads some to pay significantly higher effective tax rates than they would have with optimal behavior. The size of tax overpayment differs along demographic dimensions, with women overpaying more than men and immigrants more than native-born Norwegians on average. We also find that complexity is very regressive, as our model predicts: Wealthy taxpayers overpay significantly less than others. The strongest predictor of overpayment that we can see in the data is the amount spent within firms on accountants and financial advisors: Those with high spending make much fewer and smaller mistakes. While we cannot know if these services specifically include tax advice, we think of this spending measure as a proxy for acquisition of tax knowledge, so it is sensible that it would be highly predictive.

Finally, to test the implications of our theoretical model, we investigate if higher tax overpayment is indeed associated with lower elasticity as our theoretical results predict. We study the behavioral effect of an earlier tax reform on our study population, and estimate separate elasticities of taxable income for groups based on the magnitude of their later tax overpayment. Since our measure of overpayment relies on income shifting between tax bases for business owners, this could introduce a problem when attempting to estimate elasticity based on this measure: High tax knowledge can lead to more precise optimization, which might in itself cause estimated tax elasticity to be higher, and so any test of a correlation between overpayment and elasticity could be tautological. To alleviate this concern, we limit our sample for this test to those taxpayers who were not yet self-employed at the time of the earlier tax reform. We do find indicative results that higher tax overpayment is associated with lower elasticity, in line with our theoretical predictions. Further, we find that elasticity estimates are smaller for women, immigrants, people with lower incomes and those below the threshold for the wealth tax. Given that these groups tend to overpay more, this aligns with our theoretical predictions, though we note that many other factors in addition to elasticity differences could be at play.

Taken together, our findings illustrate a possible reason why tax complexity can be advantageous to governments: It can allow them to raise more tax revenues by lowering costly distortions for elastic taxpayers in a more targeted way than through general tax cuts, but it leads to increased inequality and particularly disadvantages less powerful demographic groups. Although maximizing tax revenue is often not the overall goal for governments, we focus on this outcome as it is important in certain contexts. Tax researchers sometimes impose "Rawlsian" social preferences, which aim to simply increase the welfare of the worst-off member of society, assuming that this person is not a taxpayer (Hellwig, 1986). Under such a model, the optimal tax rate would indeed be the one that raises the most tax revenue. In a less extreme case, optimal tax models often assume a welfare weight of zero on the highest earners in an economy (Saez and Stantcheva, 2016; Piketty and Saez, 2013), meaning that the optimal tax rate for this group would be the revenue-maximizing one. If we consider our model in the context of very high earners, it might make sense why tax systems are often especially complex for them. Many of the highest earners can also be highly tax elastic, since they have more opportunities to move their income, assets, or residence into tax havens. A complex tax code can provide hidden benefits that encourage them to remain in their home country, while at the same time raising a higher tax revenue from those high earners who are less elastic. For the wealthiest, a complex tax system with exemptions and loopholes might therefore be a more efficient way of reducing the tax burden than simply lowering their headline tax rates. From the government's point of view, an additional benefit could be that these complex tax incentives are less apparent to the general public than more straightforward tax breaks at the top, which may be unpopular with voters. Therefore, although tax complexity leads to increased inequality, the fact that this is less salient may itself be a desirable feature to policymakers. Our findings highlight the importance of considering possible effects on both tax revenue and inequality when carrying out reforms that significantly alter the complexity of the tax systems.

Our work relates to several strands of the economics literature. Multiple previous studies have developed theoretical models of complex tax systems and on optimal taxation when agents lack information or are not fully rational (Goldin, 2015; Rees-Jones and Taubinsky, 2020; O'Donoghue and Rabin, 2003). Craig and Slemrod (2022) develop one of the most comprehensive models of tax complexity, and briefly touch upon the implications of private costly investments in tax knowledge, like ours. However, their approach involves directly modelling the beliefs and misperceptions of taxpayers, and is therefore less suited to delivering empirical predictions without making explicit assumptions about these beliefs. The model used by Slemrod (1989) is quite similar to ours, also featuring costly tax knowledge in a very simple framework, but his main focus is on the choice between the itemized and standard deduction in a U.S. context. Krause (2000) is also similar in spirit, but models imperfect information by both the taxpayer and the tax authority, and focuses on compliance and audit risk.

Our paper also relates to empirical studies of tax complexity, misinformation, and tax overpayment. Many studies, including Zwick (2021), Chetty, Looney and Kroft (2009), Saez (2010), and Benzarti (2020), demonstrate empirically at a population level that taxpayers do not react to tax changes in a way that is consistent with rationality and full information, but cannot determine if a given individual taxpayer is making a mistake. Aghion et al. (2017) estimate how much taxpayers value simplicity by studying a discrete choice between different tax regimes for small businesses in France, which differ in both tax incentives and simplicity. This is one of the few studies besides ours to also quantify tax misoptimization at an individual level – some taxpayers choose a tax regime which is both more complex and leads to a higher tax liability. Abeler and Jäger (2015) demonstrate misoptimization in an experimental study mimicking a complex tax system. This study also highlights the role of cognitive ability in tax optimization. While our model focuses on the role of elasticities in the acquisition of tax knowledge, it is clear that other factors such as cognitive ability are important. The findings of Bastani and Waldenström (2021) imply that tax elasticity itself may be increasing with cognitive ability. Relatedly, Alstadsæter, Kopczuk and Telle (2019) find that networks play an important role in tax avoidance: People are more likely to take advantage of a particular tax avoidance strategy if others in their close family use it too. While we do not capture cognitive ability and networks in our model, including these would tend to strengthen our theoretical results further, since our main findings are based on an assumption that higher earners are more elastic and hence more likely to invest in tax knowledge. If these people also higher average cognitive ability and more knowledgeable networks, it would further increase their propensity to gain tax knowledge. However, as we highlight in our empirical section, this also means that tax complexity is a highly regressive tool to boost tax revenue.

While not many researchers have explored possible benefits of tax complexity, some have studied a similar idea for welfare transfers. Nichols and Zeckhauser (1982) pointed out the possible use of ordeals as a screening mechanism to discourage the less needy from applying for welfare benefits, and this idea has since been explored further by Alatas et al. (2016) and Sunstein (2018), among others. Kleven and Kopczuk (2011) study the optimal complexity of benefit programs in a context where higher complexity improves the government's ability to detect who is eligible, but also imposes a cost on applicants that discourages some from applying.

The rest of this article is organized as follows: In the next section, we introduce a simple model that formalizes our argument and illustrates how a complex tax system can function like of price discrimination. Section 3 introduces the institutional setting for our empirical results, and Section 4 describes our data sources and details how we calculate tax overpayment at the individual level. Section 5 presents empirical evidence of tax overpayment at the population level and for specific subgroups. Section 6 estimates elasticities using an earlier tax reform and links this to subsequent tax overpayment, while the final section concludes.

2 Model

In this section, we develop a model that illustrates how a complex tax system can act in a similar fashion to price discrimination. By formulating conditions under which a complex tax system can generate higher tax revenues than a simple one, we show that this increase in tax revenues comes at no efficiency cost. We prove that in this model, more elastic taxpayers will acquire more tax knowledge and thus get a lower effective tax rate. However, we start with a more straightforward example to illustrate the basic ideas at play.

2.1 First-best case: Optimal individualized tax rates

Consider an economy consisting of N workers with quasi-linear and isoelastic[1] utility functions of the following form:

$$U_i(C, L) = C - \frac{e^{-\alpha_i/\varepsilon_i}}{1 + 1/\varepsilon_i} L^{1+\frac{1}{\varepsilon_i}}$$
$$C = (1 - \tau) L,$$

where wages are normalized to 1, L is labor supply, and τ is the tax rate, assumed to be a flat tax in this simple example. For convenience, we will use the notation $\nu = 1 - \tau$ to refer to the net-of-tax rate in the following. ε_i is individual i's elasticity of labor supply with respect to net-of-tax rate ν, and α_i is a preference parameter which shifts overall labor supply without affecting elasticity.

A classic optimal tax problem for this model might derive the flat net-of-tax rate ν which maximizes some social welfare function W subject to a government tax revenue requirement R. We will consider a slight variation of this problem. Suppose the government has perfect information and knows ε_i for every citizen, and that it has the ability to set a separate tax rate τ_i for every single person in the economy based on this information. What will the set of optimal tax rates be in this case? Letting $V_i(\nu)$ and $L_i(\nu)$ denote, respectively, the indirect utility function and the labor supply function that follows from the optimal solution to the individual's problem, we could write this variant problem as:

$$\max_{\nu_i, \lambda} W\left(V_1(\nu_1), \ldots, V_N(\nu_N)\right) + \lambda \left(\sum_{i=1}^{N} (1 - \nu_i) L_i(\nu_i) - R \right). \tag{1}$$

[1] We see that preferences are isoelastic by substituting in the budget constraint and taking the first order condition. Doing so and rearranging yields:

$$L = \left(1 - T'(L)\right)^{\varepsilon_i} e^{\alpha_i}.$$

Taking the first order condition with respect to ν_i for $i = 1, \ldots, N$, we get:

$$\frac{\partial W}{\partial V_i} \cdot \frac{dV_i}{d\nu_i} + \lambda \left(-L_i\left(\nu_i\right) + \left(1 - \nu_i\right) \frac{dL_i}{d\nu_i} \right) = 0. \tag{2}$$

By the envelope theorem, we have

$$\begin{aligned} \frac{dV_i}{d\nu_i} &= \frac{\partial U_i\left(L_i\left(\nu_i\right), \nu_i\right)}{\partial \nu_i} \\ &= L_i\left(\nu_i\right). \end{aligned}$$

If we let $\psi_i = \frac{\partial W}{\partial V_i}$ denote person i's welfare weight, we can rewrite (2) as:

$$\psi_i L_i\left(\nu_i\right) = -\lambda \left(-L_i\left(\nu_i\right) + \left(1 - \nu_i\right) \frac{dL_i}{d\nu_i} \right).$$

Inserting $L_i\left(\nu_i\right) = \nu_i^{\varepsilon_i} e^{\alpha_i}$ in this and rearranging, we can ultimately reduce this equation to:

$$\nu_i = \frac{\varepsilon_i}{1 - \frac{\psi_i}{\lambda} + \varepsilon_i}.$$

This is the optimal net-of-tax rate for person i. It is increasing in elasticity ε_i, and approaches 1 as the elasticity goes to infinity – in other words, the more elastic someone is, the less they should be taxed.[2] This result is probably not surprising – from classical optimal tax theory, we know that the higher the aggregate elasticity in an economy, the lower the optimal tax will be, and our example here simply shows that a similar result will hold on an individual level.

We can think of this example as analogous to first-degree price discrimination. When a monopolist is assumed to have perfect information and the ability to charge a separate price for each customer, the optimal price will be higher for less elastic customers. Although the government in this example isn't motivated by profit maximization, a similar result holds.

Of course, the assumptions made in the first-degree price discrimination model are extremely strong: The monopolist must have perfect information and the ability to set unique

[2] This is of course assuming that everything else is kept equal. If those with lower elasticities have higher welfare weights, it would be possible that more elastic taxpayers have a higher optimal tax rate overall.

prices for everyone. The analogous assumptions in our tax example may be slightly less unrealistic: Many governments have vast information on taxpayers, and unlike private companies, governments usually do not face legal restrictions that would in principle prevent them from individualizing tax rates to some extent. Nevertheless, an individual's tax elasticity is not directly observable, and even if the government could estimate it, setting individualized tax rates would likely run counter to common notions of horizontal tax fairness.

2.2 Complexity as second-degree price discrimination

While first-degree price discrimination functions primarily as an illustrative device in theoretical models, second-degree price discrimination is very often seen in practice. The core idea of second-degree price discrimination is to offer different "menus" such that customers will self-select into different options based on their price elasticity and other characteristics, ultimately allowing the monopolist to extract a higher revenue. This can be done in multiple ways, for instance through bulk discounts, coupons, or limited time promotions. At a very high level, these pricing practices create a trade-off between price and convenience: If you are willing to do the requisite planning to buy in bulk, or to cut out coupons to bring to the store, you are rewarded with a lower overall price. Ngwe (2017) finds that the location choices of outlet stores works by a similar logic: They are often placed in comparatively remote locations to avoid cannibalizing sales at the "flagship" stores too much, while simultaneously allowing for price-sensitive customers to travel the extra distance and save money. Essentially, these strategies work by creating an *ordeal* that requires some effort to overcome in order to access the lower price.

We can think of a complex tax system as functioning in a similar way. Those who are highly sensitive to taxes will spend time and effort on understanding the tax system – or pay someone else to do this work for them – and can thus lower their total tax rate, for example by taking advantage of obscure deductions, optimizing the balance between wages and dividends, optimally timing realizations of capital gains and losses, or setting up holding companies or other legal structures. A complex tax system allows the government to extract as much tax revenue as possible from those less sensitive to taxes, while creating an ordeal that requires some wasteful effort to overcome, but which lets the more elastic taxpayers reduce their tax burden. It may even be the case that this complexity can lead to efficiency gains, if the cost of learning about the tax system is outweighed by the reduced distortionary effects of taxation for the more elastic taxpayers. To answer this question, we formulate a mathematical model of tax complexity.

We now write the taxpayer's problem as:

$$\begin{aligned} \max_{C,L,S} U_i(C,L) &= C - D_i(L) \\ \text{s.t. } C &= w_i L \cdot N(S) - Q(S) \\ L, S &\geq 0, \end{aligned}$$

where we think of S as time spent studying the tax system. $D_i(L)$ can be interpreted as the disutility of labor, $Q(S)$ as the disutility or cost of acquiring tax knowledge, and $N(S)$ as the net-of-tax rate associated with knowledge level S.[3] We assume that $N(S)$ has a weakly negative first derivative with respect to S everywhere, as knowledge about the tax system may help the taxpayer find strategies to reduce the overall tax burden. The government's problem will then be to decide on the complexity of the tax system by choosing the shape of the function $N(S)$.

The solution to the problem is pinned down by the first order conditions. In the case where $S \geq 0$, these can easily be found to be:

$$\begin{aligned} wN(S) &= D'(L) \\ wLN'(S) &= Q'(S). \end{aligned} \tag{3}$$

Without making further assumptions on functional forms, we cannot analytically derive a

[3] Since we are considering an economy consisting of multiple taxpayers with differing elasticities, a completely correct notation would include taxpayer-specific subscripts on the choice variables: L_i, C_i, S_i. However, we omit these for readability whenever it does not lead to ambiguities.

closed-form solution. However, we can characterize how introducing a marginal amount of complexity into the model will affect outcomes. To do this, we make the following definitions:

Definition 1. *We define a* simple problem *under this model as a consumer utility maximization problem with no complexity, i.e. where $N(S) = 1 - \tau_0$ is a simple net-of-tax rate that does not depend on S. Trivially in the simple problem, any optimal solution will have $S = 0$ since any positive S will incur a cost and yield no tax benefits.*

Definition 2. *Given a simple problem, we define a corresponding* marginally complex problem *as one which introduces a small degree of complexity, replacing $N(S) = 1 - \tau_0$ with $N_\delta(S)$ such that*

- $N_\delta(0) = 1 - \tau_0$
- $N'_\delta(S) > 0$ *for* $S < \delta$
- $N'_\delta(S) = 0$ *for* $S \geq \delta$

for small δ.

This definition helps us to analyze how complexity will affect decisions at the margin without having to worry about discrete responses, since only the initial marginal amount of tax knowledge will reduce the tax burden. We are now ready to find necessary and sufficient conditions for marginal complexity to generate increased tax revenue.

Proposition 1. *Consider a taxpayer i with wage w_i, elasticity of taxable income ε_i and labor output L_i^0 at the optimal solution to the simple problem. If we introduce marginal complexity to the model, i.e. replace the simple net-of-tax rate $1 - \tau_0$ with $N_\delta(S)$ that satisfies $N_\delta(0) = 1 - \tau_0$, $N'_\delta(S) > 0$ for $S < \delta$ and $N'_\delta(S) = 0$ for $S \geq \delta$ with $\delta > 0$ sufficiently small, then tax revenues generated by taxpayer i will increase relative to the simple problem if and only if*

$$w_i L_i^0 N'_\delta(0) > Q'(0)$$

and

$$\varepsilon_i > \frac{1-\tau_0}{\tau_0}.$$

Proof. We first note that the first condition, $w_i L_i^0 N_\delta'(0) > Q'(0)$, is a necessary condition for i to optimally choose a positive level of tax knowledge. If this does not hold, then for a sufficiently small δ, i will simply prefer a corner solution with $S = 0$, and therefore tax revenue will be unchanged relative to the simple problem.

Now, to see how optimal labor supply will change as a function of tax knowledge, consider the first-order condition associated with the labor variable around $S = 0$:

$$D'(L) = w_i N_\delta(S)$$

Notice that the above expression will hold with equality both at $L = L_i^0$ and $S = 0$ for the optimal solution to the simple problem, and at any (L, S) that solves the complex problem. We can therefore find the derivative of labor as a function of tax knowledge around L_i^0 by taking the total derivative of this expression with respect to S around $S = 0$ and rearranging. This yields:

$$\left.\frac{\mathrm{d}L}{\mathrm{d}S}\right|_{S=0} = \frac{w_i N_\delta'(0)}{D''\left(L_i^0\right)},$$

which we notice is positive whenever D is strictly convex. Now we turn to tax revenue. This can be written as $R = w_i L_i \left(1 - N_\delta(S)\right)$, which we can again totally differentiate around $S = 0$ to get

$$\mathrm{d}R = w_i \left(1 - N_\delta\left(0\right)\right)\mathrm{d}L - N_\delta'\left(0\right) w_i L_i^0 \mathrm{d}S.$$

Rearranging yields

$$
\begin{aligned}
\left.\frac{\mathrm{d}R}{\mathrm{d}S}\right|_{S=0} &= w_i\left(1-N_\delta(0)\right)\left.\frac{\mathrm{d}L}{\mathrm{d}S}\right|_{S=0} - N_\delta'(0)\, w_i L_i^0 \\
&= w_i\tau_0 \frac{w_i N_\delta'(0)}{D''\left(L_i^0\right)} - N_\delta'(0)\, w_i L_i^0 \\
&= w_i L_i^0 \tau_0 \frac{w_i N_\delta'(0)}{D''\left(L_i^0\right) L_i^0} - N_\delta'(0)\, w_i L_i^0 \\
&= w_i L_i^0 N_\delta'(0)\left[\frac{\tau_0 w_i}{D''\left(L_i^0\right) L_i^0} - 1\right]. \qquad (4)
\end{aligned}
$$

Introducing a marginal amount of complexity to the simple model results in an increase in tax revenue from person i if and only if the expression above is greater than zero. To rewrite this in terms of the ETI, notice from the first-order condition to the simple problem, we have that

$$
L_i^0 = \left[D'\right]^{-1}\left(w_i\left(1-\tau_0\right)\right).
$$

Letting $g = D'$ and $n = 1-\tau$ for ease of notation, and taking the derivative of this with respect to the net-of-tax rate n, we get:

$$
\begin{aligned}
\left.\frac{\mathrm{d}L_i}{\mathrm{d}n}\right|_{\tau=\tau_0} &= w_i\left[g^{-1}\right]'\left(w_i\left(1-\tau_0\right)\right) \\
&= \frac{w_i}{g'\left(g^{-1}\left(w_i\left(1-\tau_0\right)\right)\right)} \\
&= \frac{w_i}{D''\left(L_i^0\right)}.
\end{aligned}
$$

We can therefore write the elasticity of taxable income at L_i^0 as

$$
\begin{aligned}
\varepsilon_i &= \frac{\mathrm{d}(w_i L_i)}{\mathrm{d}n}\cdot\frac{1-\tau_0}{w_i L_i^0} \\
&= \frac{\mathrm{d}L_i}{\mathrm{d}n}\cdot\frac{1-\tau_0}{L_i^0} \\
&= \frac{w_i\left(1-\tau_0\right)}{D''\left(L_i^0\right) L_i^0}.
\end{aligned}
$$

Using this, we can now write (4) as

$$\left.\frac{\mathrm{d}R}{\mathrm{d}S}\right|_{S=0} = w_i L_i^0 N_\delta'(0)\left[\frac{\tau_0}{1-\tau_0}\varepsilon_i - 1\right],$$

which is positive if and only if $\varepsilon_i > \frac{1-\tau_0}{\tau_0}$. □

In the above result, note that the first condition, $w_i L_i^0 N_\delta'(0) > Q'(0)$, is simply a participation constraint which states that the benefit of the first marginal unit of tax knowledge – income times the slope of the net-of-tax function – must exceed the cost. If this condition isn't satisfied, the individual will optimally choose no tax knowledge, and the tax revenue derived from them will trivially be the same as in the simple problem. The second condition, $\varepsilon_i > \frac{1-\tau_0}{\tau_0}$, can be rewritten as $\tau_0 > \frac{1}{1+\varepsilon_i}$. This simply states that the tax rate, τ_0, is above the revenue-maximizing rate for this particular individual. The result is therefore very intuitive: Introducing complexity allows the taxpayer to obtain a lower tax rate, and beyond the peak of the Laffer curve, this lower tax rate is associated with higher revenues. Note that even though tax knowledge is costly, the increased tax revenue does not come at an overall efficiency cost, as the taxpayer's opportunity set has only increased. Remaining at the optimum from the simple model is still an option, and the taxpayer must therefore be made better off by choosing to obtain tax knowledge.

Corollary 1. *Starting from an economy with a single flat tax rate, introducing tax complexity can increase the maximum tax revenue that the government can collect if there exists some income threshold $\tilde{I}$ such that those with incomes above $\tilde{I}$ have a larger aggregate ETI than those with incomes below $\tilde{I}$.*

Proof. Let $\bar{\varepsilon}$ be the aggregate ETI of the full population. It is a standard result from the optimal tax literature that the revenue-maximizing flat tax rate in this economy will be $\tau^* = \frac{1}{1+\bar{\varepsilon}}$. Let H denote the set of taxpayers with incomes above $\tilde{I}$. The government can design a complex tax system with functions N_δ and Q determined such that $\tilde{I} = \frac{Q(0)}{N_\delta'(0)}$. From the proof of the previous proposition, we know this ensures that for δ sufficiently small, the set of those who will invest in tax knowledge will approach H. Let ε_H be the aggregate ETI for the group of taxpayers in H. By the definition of the income threshold $\tilde{I}$, the aggregate ETI of those who invest in tax knowledge will be higher than for those who do not, and therefore also higher than the ETI for the total population, so thus $\varepsilon_H > \bar{\varepsilon}$. We show that overall tax revenue is increased by extending the argument from the previous proof to an aggregate population.

Let R_H denote aggregate tax revenue collected for the group H, and let I_H denote their aggregate income. As before using n for the net-of-tax rate, notice that we can write the aggregate elasticity for the high-income group as

$$
\begin{aligned}
\varepsilon_H &= \frac{\mathrm{d}I_H}{\mathrm{d}n} \cdot \frac{n}{I_H} \\
&= \sum_{i \in H} w_i \frac{\mathrm{d}L_i}{\mathrm{d}n} \cdot \frac{n}{I_H} \\
&= \sum_{i \in H} \frac{w_i^2}{D''\left(L_i^0\right)} \cdot \frac{n}{I_H}
\end{aligned}
$$

Using our derivation from the previous proof, we can write the change in aggregate tax revenue when we allow for a marginal amount of tax complexity as:

$$
\begin{aligned}
\left.\frac{\mathrm{d}R_H}{\mathrm{d}S}\right|_{S=0} &= \sum_{i \in H} w_i \left(1 - N_\delta(0)\right) \left.\frac{\mathrm{d}L_i}{\mathrm{d}S}\right|_{S=0} - \sum_{i \in H} N'_\delta(0)\, w_i L_i^0 \\
&= \sum_{i \in H} w_i \tau^* \frac{w_i N'_\delta(0)}{D''\left(L_i^0\right)} - \sum_{i \in H} N'_\delta(0)\, w_i L_i^0 \\
&= N'_\delta(0) \left[\tau^* \sum_{i \in H} \frac{w_i^2}{D''\left(L_i^0\right)} - I_H\right] \\
&= N'_\delta(0) \left[\tau^* \varepsilon_H \cdot \frac{I_H}{1 - \tau^*} - I_H\right] \\
&= N'_\delta(0)\, I_H \left[\varepsilon_H \cdot \frac{\tau^*}{1 - \tau^*} - 1\right],
\end{aligned}
$$

which is positive whenever the term in brackets is greater than zero. But this follows from the definition of τ^* from the beginning of the proof, since

$$
\varepsilon_H \cdot \frac{\tau^*}{1 - \tau^*} > \bar{\varepsilon} \cdot \frac{\tau^*}{1 - \tau^*} = \bar{\varepsilon} \cdot \frac{1}{\bar{\varepsilon}} = 1.
$$

□

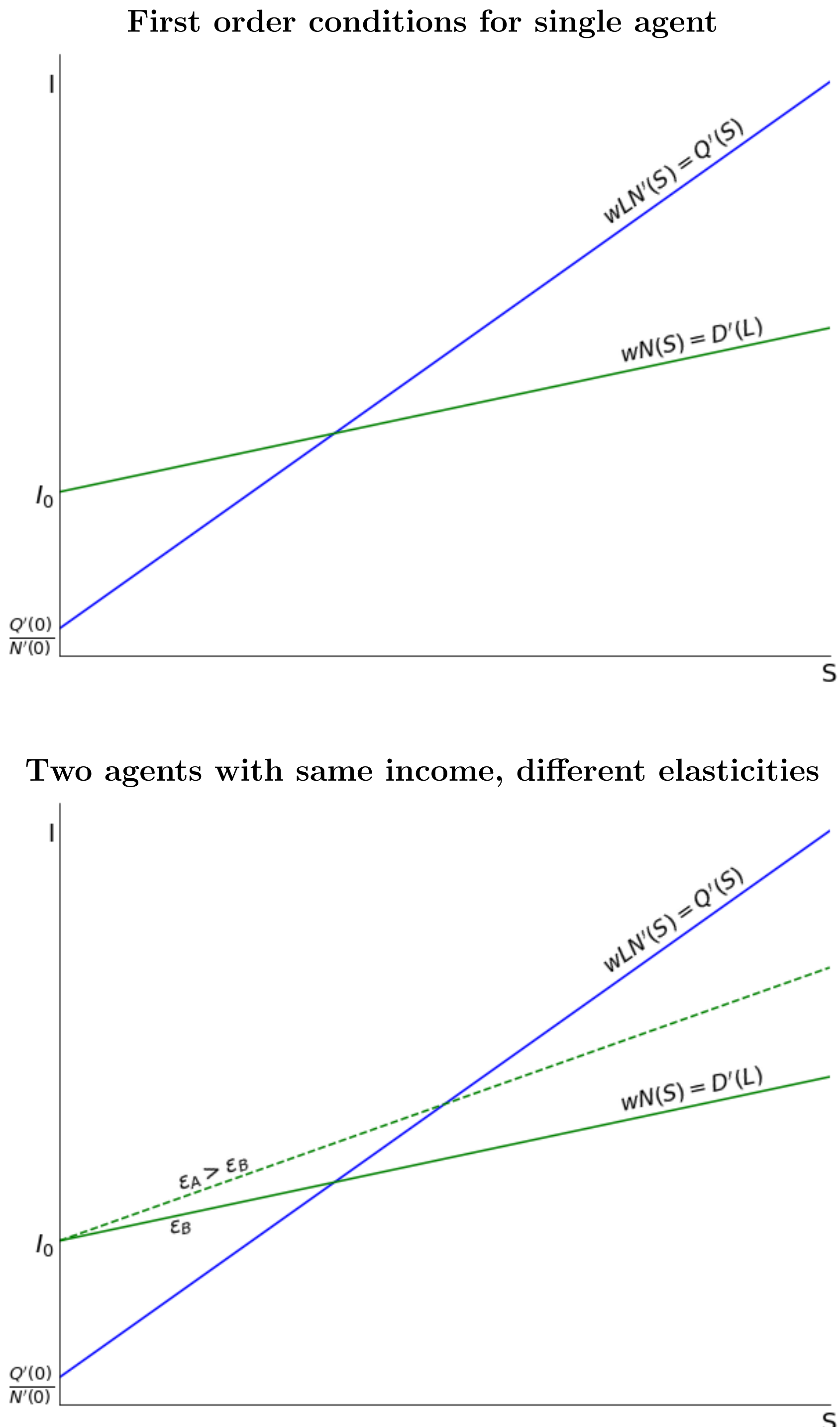
Just like in the previous result, increased tax revenue comes at no efficiency cost since introducing complexity makes the opportunity set larger for everyone, so no one can end up worse off. Thus, in cases where the result applies, complexity actually leads to an efficiency improvement by allowing those who are on the downward-sloping side of the Laffer curve to gain tax knowledge and lower their effective tax rate while increasing their labor supply enough to offset the mechanical loss in tax revenue. If the surplus tax revenue is spent on e.g. a small lump-sum transfer to everyone in the economy, the introduction of complexity will make everyone strictly better off.

Note that the previous result assumed a single, flat tax rate. However, the same logic and arguments would apply in the case of a progressive tax rate: Complexity can generate increased tax revenue whenever there is an income threshold above which the aggregate ETI is large enough that this subpopulation are on the downward-sloping part of the Laffer curve.

The results we have seen so far have relied on more elastic taxpayers investing in tax knowledge while less elastic taxpayers choose not to. However, this has been driven by our assumptions on differences in income: For the initial unit of tax knowledge, elasticity does not even factor into the decision to invest or not, as this is fully determined by the condition $w_i L_i^0 N_\delta'(0) > Q'(0)$. We might therefore worry about what happens outside of the context of marginal complexity, when we allow for larger investments in tax knowledge. The next result shows there is indeed a positive link between elasticity and investment in tax knowledge once we move beyond the marginal case.

Proposition 2. *Consider two taxpayers, A and B, with identical income $I_0 > \frac{Q'(0)}{N'(0)}$ in the simple equilibrium. If A has a higher ETI than B across the entire income spectrum, i.e. $\varepsilon_A > \varepsilon_B$, then under complexity, A will invest in more tax knowledge than B, have a higher overall income and a lower effective tax rate than B. If the two taxpayers have identical elasticities across the income spectrum, they will have the same income and tax knowledge under complexity.*

Figure 1: Graphical representation of first-order conditions in (I, S)-space



Notes: This figure shows a graphical representation of the first-order conditions for the agent's optimization problem under tax complexity. The bottom panel illustrates that if two taxpayers have the same initial income but different tax elasticities, the more elstic taxpayer will have a steeper increase in labor earnings as a function of tax knowledge. See the proof of Proposition 2 for further details.

Proof. In short, this proposition states that if we keep income fixed in the simple equilibrium, investments in tax knowledge will be a monotonous function of a person's elasticity of taxable income. However, note that it's not enough for the ETI at the initial simple equilibrium to be the same, as elasticity may change over the income distribution moving towards the complex equilibrium.

The proof can most easily be understood through a graphical explanation. The first-order conditions for the complex problem in (3) can be represented as curves in (I, S)-space as shown in the first panel of Figure 1. The blue line represents the first-order constraint for tax knowledge, and the green one for labor. Note that while the choice variables in the problem as originally stated are L and S, since $I = wL$ and w is assumed fixed for any given taxpayer, we can also think of I as a choice variable directly. Income in the initial, simple equilibrium is given by I_0 – the condition that this is greater than $\frac{Q'(0)}{N'(0)}$ ensures that at least some investment in S is beneficial once complexity is introduced. The equilibrium under complexity will correspond to the point where the two lines meet.

Consider taxpayers A and B. The proof is done if we can show that the intersection of the two lines will occur at a higher (I, S)-pair for A than for B, as illustrated in the second panel of Figure 1. First, notice that the blue line (the constraint on S) will be identical for the two taxpayers, as it only depends on income and on the functions N and Q, which are assumed to be independent of the identity of the taxpayer. Also notice that the intercept of the green line is simply income in the simple equilibrium, which is I_0 for both taxpayers. Therefore, the proof is complete if we can show that the slope of the green line will be steeper for the more elastic taxpayer. We can see this by totally differentiating the constraint associated with the green line, which yields

$$\frac{\mathrm{d}L_i}{\mathrm{d}S_i} = \frac{w_i N'(S_i)}{D''(L_i)}, \tag{5}$$

as we showed for the special case of $S_i = 0$ in the proof of Proposition 1. Now, recall that the ETI of taxpayer i is defined as $\varepsilon = \frac{\mathrm{d}I_i}{\mathrm{d}n} \cdot \frac{n}{I_i}$ where n is the net-of-tax rate, and where we evaluate the response to an *exogenous* change in the net-of-tax rate – that is, at a given effective net-of-tax rate $N(S_i)$, the ETI is the percentage-wise increase in income that i would earn if $N(S_i)$ were exogenously lowered by 1 percent. Therefore, we can find an expression for the elasticity by differentiating the first-order condition for the *simple* problem at a labor supply level of L_i and net-of-tax rate $N(S_i)$. Recall that in the simple problem, labor supply is pinned down by the first order condition $L_i = [D']^{-1}(w_i n)$, so just as in the proof of Proposition 1, we can express elasticity as

$$\begin{aligned}\varepsilon_i &= \frac{\mathrm{d}(w_i L_i)}{\mathrm{d}n} \cdot \frac{N(S_i)}{w_i L_i} \\ &= \frac{\mathrm{d}L_i}{\mathrm{d}n} \cdot \frac{N(S_i)}{L_i} \\ &= \frac{w_i N(S_i)}{D''(L_i) L_i},\end{aligned}$$

which corresponds to $\frac{w_i}{D''(L_i)} = \frac{\varepsilon_i L_i}{N(S_i)}$. Insert this in equation (5) and use the identity $I_i = w_i L_i$ to find:

$$\frac{\mathrm{d}I_i}{\mathrm{d}S_i} = w_i \frac{\mathrm{d}L_i}{\mathrm{d}S_i} = w_i \frac{\varepsilon_i L_i N'(S_i)}{N(S_i)} = \varepsilon_i I_i \frac{N'(S_i)}{N(S_i)}.$$

This completes the proof: Since both taxpayers are assumed to have the same income at $S_i = 0$, if $\varepsilon_A > \varepsilon_B$, we do indeed see that $\frac{\mathrm{d}I_A}{\mathrm{d}S} = \varepsilon_A I_A \frac{N'(S)}{N(S)} > \varepsilon_B I_B \frac{N'(S)}{N(S)} = \frac{\mathrm{d}I_B}{\mathrm{d}S}$ at every point along the curve, so the constraint for taxpayer A is indeed steeper than that of B as illustrated in the second panel of Figure 1, leading to an equilibrium further up along the blue curve, which involves a higher income, higher S, and a lower effective tax rate. Conversely, if $\varepsilon_A = \varepsilon_B$, the two curves will have the exact same slope at every point, and will hence coincide such that A and B end up with the same equilibrium income and effective tax rate under complexity. □

2.3 Simulation

The theoretical results in this section have proved that under some circumstances, tax complexity can lead to increased tax revenue without sacrificing efficiency. However, those propositions assume marginal complexity and only provide sufficient, but not necessary conditions for the results to hold. Therefore, there is scope to explore in further detail what the model implies under different circumstances – for example, when we consider tax complexity beyond the marginal unit, or when we apply complexity to a tax system at an initial tax rate which is lower than the revenue-maximizing one. This subsection investigates these questions by simulating particular specifications of the model in Python.

To parameterize the model, we assume the following class of taxpayer utility functions:

$$U_i(C, L, S) = C - \frac{e^{-\alpha_i/\varepsilon_i}}{1 + 1/\varepsilon_i} L^{1+\frac{1}{\varepsilon_i}} - 0.05S$$

We model an economy with two types, a *high type* with ETI $\varepsilon_h = 1.25$ and a *low type* with $\varepsilon_l = 0.75$. We assume each of these types makes up half of the population. In this version of the model, we have normalized wages to 1, such that incomes are equal to labor supply. The parameter α_i will shift labor output independently of the elasticity – we initially set $\alpha_h = 1.5$ and $\alpha_l = 1$, which ensures that the high type has higher labor supply and income than the low type, in line with the assumptions in our model. We analyze the difference between a simple model with a fixed tax rate, and a complex model that allows taxpayers to reduce their effective tax rate. In the first two cases we examine, we let the complexity function be exponential:

$$N(S) = 1 - \tau \left(0.2e^{-0.3S} + 0.8\right) \tag{6}$$

This function generates a net-of-tax rate of $1 - \tau$ if $S = 0$, and will decline asymptotically towards $1 - 0.8\tau$ as S grows large, so that the effective tax rate can be reduced by at most 20% of its initial value.

The first case we examine illustrates the general idea behind Corollary 1. We set a baseline tax rate of $\tau = 0.495$, which is the flat tax rate that maximizes government revenue in this case.[4] The results are shown in the first three columns of Table 2.3, labelled "Model A". The first column shows tax revenue, labor output and utility by type under the simple model, i.e. at a flat tax rate of $\tau = 0.495$ with no possibility to reduce this through tax knowledge. The

[4]The aggregate ETI $\bar{\varepsilon}$ will be the income-weighted average of the type-specific ETIs, and the revenue-maximizing tax rate can then be calculated as $\tau^* = \frac{1}{1+\bar{\varepsilon}}$.

Table 1: Simulation results

	Model A			Model B			Model C		
	Simple	Complex	Diff.	Simple	Complex	Diff.	Simple	Complex	Diff.
Tax revenue	0.8752	0.8779	0.0027	0.8021	0.7989	-0.0032	0.8081	0.8091	0.0009
S_l		0			0			0.5	
S_h		0.5369			0.359			1	
τ_l	0.495	0.495	0	0.35	0.35	0	0.5	0.4838	0.0162
τ_h	0.495	0.4803	-0.0147	0.35	0.3429	-0.0071	0.5	0.4690	0.0310
L_l	1.6284	1.6284	0	1.9678	1.9678	0	1.6163	1.6555	0.0392
L_h	1.9079	1.9777	0.0698	2.6157	2.6517	0.0360	1.6163	1.7425	0.1262
Utility, low	0.4699	0.4699	0	0.7309	0.7309	0	0.4618	0.4634	0.0016
Utility, high	0.4282	0.4299	0.0018	0.7556	0.7565	0.0009	0.3592	0.3612	0.0021

Notes: This table shows results of simulations of our theoretical model done in Python. S_i represents tax knowledge, τ_i the effective tax rate and L_i labor output for group $i = h, l$. In all cases, the simulations compare the difference between a simple model with a fixed tax rate and a corresponding model that allows for complexity. The first three columns labelled "Model A" illustrate a case where complexity allows for increased tax revenues and increases efficiency starting from the revenue-maximizing flat tax rate. "Model B" shows the same model starting from a lower tax rate, where complexity no longer causes a tax revenue increase. "Model C" shows a case where two individuals with initially identical incomes but different elasticities will self-separate as a consequence of complexity, once again leading to increased tax revenue and welfare. See the main text for more details.

second column shows the corresponding values after we allow for complexity as described by equation (6), and the third column shows the difference between these two cases. We see that allowing for complexity in this case will cause the high type to invest in $S_h = 0.5369$ units of tax knowledge, which causes their effective tax rate to fall by roughly 1.5 percentage points, while the low type will not invest in tax knowledge. Consequently, the high type ends up with higher labor output and higher utility overall.[5] Tax revenues also increase, confirming the result from Corollary 1 in a simulation that allows for tax knowledge to increase beyond just a marginal unit.

The next three columns, labelled "Model B", repeat the same calculations, but starting from a baseline tax rate of 0.35. This example illustrates the effects of introducing complexity in an environment where the baseline tax rate is below the revenue-maximizing rate. The low type still chooses not to invest in tax knowledge, whereas the high type will invest, but by

[5] The fact that utility increases is true by construction, as allowing for complexity simply expands everyone's opportunity set. No one can be made worse off, as they can still choose not to invest in tax knowledge.

less than in the previous example. Tax revenues now decrease once complexity is introduced – the reduced distortion on the high-type taxpayer is not enough to make up for the mechanical reduction in revenues. However, we do see that labor output increases for the high type. In fact, if the government has a goal of increasing labor output, complexity remains a more efficient way to achieve this. To see this, note that tax revenues in the complex version of Model B are 0.7989, and the average labor output between the two types is 2.3097. A simple model with a tax rate of 0.347 would achieve the same tax revenue, but average labor output in that case is only 2.3030. This illustrates another case for complexity beyond just revenue maximization: Total economic output is a metric that governments are often interested in increasing, and tax complexity may be a tool that helps achieve this aim.

The goal of Model C is to show that the assumption of a positive correlation between income and elasticity is not a necessary condition for tax complexity to lead to increased tax revenue. In this example, we show that starting from the same initial income and with both types investing in positive amounts of tax knowledge, complexity can still generate higher rax revenues and be efficiency-improving. We set the initial tax rate to 0.5 and change the value of α_h to 1.346574, which will cause both types to have the same initial labor supply and income under the simple tax system.[6] We also change the complexity function slightly. Instead of (6), we now use the following piecewise linear function:

$$N(S) = \begin{cases} 1 - \tau\left(1 - 0.065S\right) & \text{if } S \leq \frac{1}{2} \\ 1 - \tau\left(1 - \frac{0.065}{2} - 0.059\left(S - \frac{1}{2}\right)\right) & \text{if } \frac{1}{2} < S \leq 1 \\ 1 - \tau\left(1 - \frac{0.065}{2} - \frac{0.059}{2}\right) & \text{if } S > 1 \end{cases} \cdot \tag{7}$$

The outcome of this model is shown in the last three columns of Table 2.3. We see that low types will invest in tax knowledge up to the first kink point, where $S = \frac{1}{2}$, whereas high types will locate at the second kink point with $S = 1$. Both types now gain utility in the complex model over the simple one, while overall tax revenue and labor output will also increase. This

[6] In fact, the high-elasticity type will have a very slightly lower initial income, so that income and initial elasticity are actually negatively correlated in this example.

result illustrates that complexity can also generate horizontal tax inequality: Two individuals starting at the same initial income may self-separate according to their elasticity and achieve different effective tax rates under complexity.

2.4 Discussion of the model

Within the framework of a simple model, we have shown that complexity can indeed serve a purpose, as it allows the government to extract a higher tax revenue than with a simple flat tax rate, assuming that elasticities are increasing with income. This result would apply in the exact same way if we had instead assumed that the baseline tax is progressive, as is the case in most countries. These findings give us an indication of why a complex tax system might serve a useful function for a government: It allows for a progressive tax schedule at a surface level, giving the appearance of a tax system that redistributes income, while at the same time providing deductions and other loopholes such that top earners in practice pay less than the headline rate. It can thus to some extent bridge the gap between two principles of taxation that are at odds with each other: General notions of fairness and public opinion usually favor progressive taxes (Tolbert, Witko and Wolbers, 2019; Barnes, 2022), and this is also often assumed for papers in the tax literature through a social welfare function that places less weight on high earners (Saez and Stantcheva, 2016; Piketty and Saez, 2013). However, several results in the optimal tax literature have shown that distortions from taxation are higher at the top, and that in the extreme, the marginal tax rate should be 0 for the highest individual earner (Mirrlees, 1971; Seade, 1977). Tax complexity may provide a more covert way to reduce distortions at the top, delivering a tax system which is less redistributional than it seems and therefore perhaps less likely to be unpopular with voters.

Our last result shows that complexity does indeed cause taxpayers to self-separate according to elasticity, such that starting from identical incomes under the simple model, a more elastic taxpayer will obtain a lower effective tax rate. This highlights another possible purpose of tax complexity, namely to allow for horizontal inequity in taxation. Striving for horizontal equity in taxation is a widely accepted principle of fairness (Musgrave, 1990), but if there are heterogeneous tax elasticities at a given income level, horizontal tax inequality may improve efficiency and lead to increased tax revenues, as we demonstrated in the previous subsection. Again, this is a case where complexity can lead to unequal treatment that may be desirable to the government, without this being directly codified in the tax law.

As with any economic model, we have made several simplifying assumptions in order to derive meaningful results. We discuss some of these assumptions here, and how they relate to instances of tax complexity that we might find in the real world.

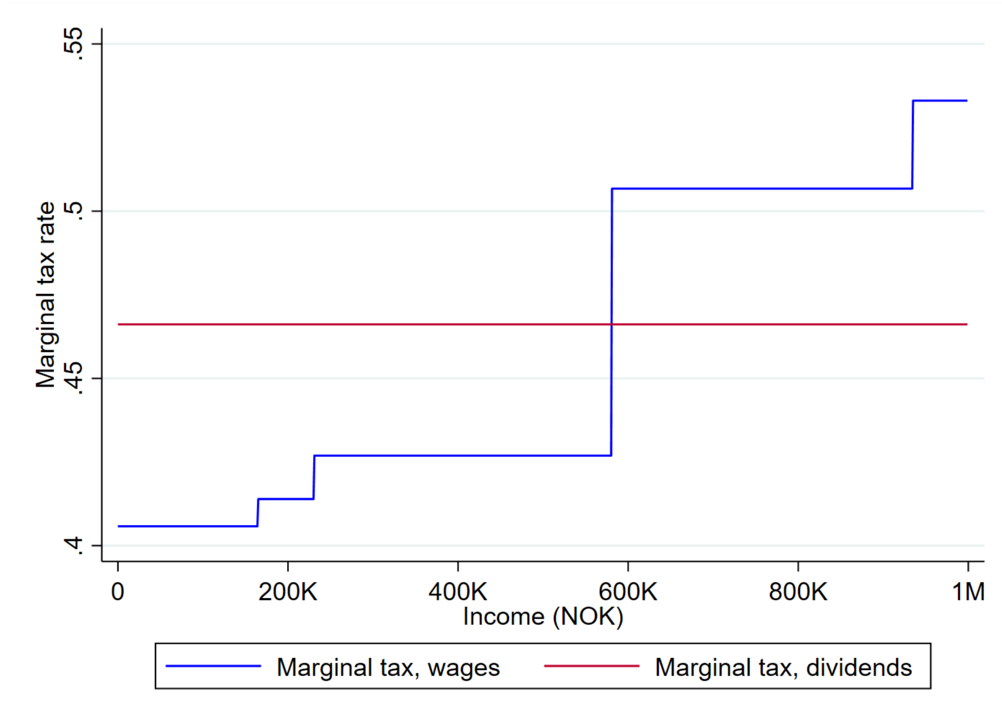
Our most unrealistic assumption is perhaps that the model implies perfect knowledge, in some sense. While taxpayers do not start with full knowledge of the tax system, the model assumes that they "know what they don't know" – i.e., that they know precisely which effective tax rate they would end up with if they had invested in S units of tax knowledge. A more realistic model might assume that taxpayers search over some policy space to find tax provisions which may lead to a lower effective tax rate to a smaller or larger degree. Some taxpayers may end up making sub-optimal decisions in the tax knowledge space – they might not invest in knowledge even though the benefit from doing so would have outweighed the cost, or they might conversely overestimate how much they can reduce their tax burden, and invest too much. While a model incorporating these nuances might be more realistic, it would also lead to many additional questions: If people don't know how much they can lower their effective tax rate by investing in tax knowledge, they need to have some belief about it. If this belief isn't accurate, then what causes the discrepancy, and how should this be modelled? Do taxpayers learn their true tax rate before they make their decision of how much labor to supply? Other papers on tax complexity, such as Craig and Slemrod (2022) do include these richer models of beliefs, but models of this sort are less useful in delivering predictions on e.g. tax revenue unless they impose other, more specific assumptions.

We believe that the assumption of "full knowledge", while not quite realistic, is useful for tractability of the model and in fact makes its predictions more conservative. Allowing for inaccurate beliefs would provide another channel through which complexity could lead to increased government revenue. It is likely that wealthier individuals would on average have a network that gives them access to better high-level information on what deductions are available, what strategies can be used to minimize capital gains taxes, and so on (Alstadsæter, Kopczuk and Telle, 2019). Our results show that the government can increase tax revenues through complexity simply as a result of heterogeneous elasticities, and not because of inaccurate beliefs which would tend to skew the knowledge gap between the rich and poor even further.

Another simplifying assumption in the model is that tax knowledge uniformly lowers the tax rate independently of a taxpayer's identity or initial income. This is clearly not the case in practice: The complexities in the tax code that can be used by low-income wage earners to lower their tax burden are very different than the ones employed by wealthy business owners. This means that the government has much more flexibility to differentiate tax complexity according to individual circumstances than our model implies, so this is again a case where our assumption is relatively conservative.

Finally, we assume that the cost of tax knowledge, $Q(S)$, does not vary by individual. This assumption is debatable: We might assume that more elastic individuals would also have a higher cost of acquiring tax knowledge. If we imagine a model where a fixed time endowment is split into leisure, productive labor, and time spent studying the tax system, we could think of the latter as simply another form of labor which is governed by the same elasticity parameter. In reality though, the time spent on learning about the tax system is minuscule for most people compared to the time spent on normal labor, and it isn't clear that adding these two quantities up into a single "total labor" variable would be a good modelling choice in that case. And on the other hand, if we work under the assumption that income and elasticity are positively correlated – for instance, if more elastic people have higher productivity and wages – it wouldn't be unreasonable to assume that this higher productivity also applies to the ability to understand the tax system. In that case, each "unit" of tax knowledge would be cheaper for the more productive taxpayers to acquire. Ultimately, we have chosen not to assume that one of these opposite effects is stronger than the other, but to keep the cost of tax knowledge independent of the identity of the taxpayer. This would be consistent with a model where tax knowledge is a good that can be bought for a particular price, for instance by hiring an

Figure 2: Tax scheme for labor and dividend income in Norway



Notes: This figure illustrates the tax scheme in Norway, featuring a progressive tax rate for labor income and a flat tax rate on dividends. The marginal tax rate for dividends is calculated as the combined tax rate on corporate profits and dividend payments. The marginal tax rate for labor income reflects personal income taxes, employer social security contributions, and payroll taxes. Payroll taxes are differentiated by geographic location of the employee – the wage tax scheme shown here reflects the payroll taxes for Zone 1, which encompasses the majority of Norway's population and has the highest tax rate. This figure is for the year 2017; tax rates vary slightly over the years, but the overall pattern is similar.

accountant or tax advisor.

3 Evidence from Norway: Overview and Institutional Background

The following sections will explore empirical implications of tax complexity using rich register data from Norway. We are among the first researchers able to quantify the extent to which complexity leads to overpayment or mistakes. In our setting, we define a tax mistake as a situation where, for a given gross income, a taxpayer makes sub-optimal decisions at a pure accounting level, leaving them with a higher effective tax rate than would have been possible.

This section gives a brief overview of our strategy, which we expand on in the following sections.

Norway has a dual income tax system where labor and capital income are taxed separately. Labor income taxes follow a progressive schedule with multiple brackets, while capital income is taxed at a flat rate. Stock-based companies (*aksjeselskap*) are subject to a corporate tax rate on profits in the year they accrue, but are considered separate tax entities, such that individual income taxes are only applied when money is paid out from the firm to an individual.

The specific setting we study involves entrepreneurs who own all shares in their company, and who therefore face a choice for any money that they transfer out of the company to pay themselves: They can pay themselves as an employee in the form of wages, as an owner in the form of dividends, or split the payment between these two types in any way they like. These two payment forms are functionally completely equivalent except for their tax treatment: Wages are taxed progressively, and dividends are subject to a flat tax. For most taxpayers, the marginal tax rates for the two income types cross at a certain point, meaning that there is one particular way to minimize tax liability within a given year: Paying out income as wages up until the point where the marginal tax rates cross, and any remaining income as dividends. Given the total gross amount that the owner transfers from the firm to themselves, the choice between these two income types is purely a question of accounting involving no real costs, and when owners fail to allocate funds optimally, we can therefore unambiguously say that this represents a tax overpayment. We interpret these cases of overpayment as being the result of tax complexity: If taxpayers had complete information about the tax system and were fully rational, everyone should theoretically be able to optimize their tax payment.

We note that this setting is a very specific and quite narrow example of the much broader concept of tax complexity. Not every instance of tax complexity and resulting overpayment comes down to a pure question of costless accounting – in many cases, to take advantage of a complex tax system, people have to pay a real cost, but one which may be outweighed by the tax savings. For example, consider the complexities generated by geographical differences in tax rates across jurisdictions in the U.S. and elsewhere. People can and do take advantage of this complexity by living in areas with lower property tax rates, shopping in states with lower sales taxes, and so on. These decisions can generate tax savings, but also involve a real cost which can be difficult for researchers to quantify. If someone chooses to live in a jurisdiction with high property taxes, we generally do not know if they are unaware that they could pay a lower tax rate by moving elsewhere, or if they simply value their chosen location enough to offset the higher cost of living. Economists tend to assume the latter by imposing rationality in economic models, but in reality, such complexities may cause people to make less-than-optimal choices due to lack of information. The strength of our setting is that there are no real costs of reclassifying income as one type or the other, so we generally avoid ambiguities that could arise when real costs differ across individuals. There is one exception to this, namely a potential ambiguity generated by the public pension system, which we deal with by making conservative assumptions. We explain this further in the following section.

4 Data

We use Norwegian de-identified administrative data obtained through Statistics Norway. Our main data used to calculate tax mistakes cover the period 2011-2017, but for additional analysis in Section 7, we will use income data going as far back as 1996. The data link firms with their owners and employees, and include detailed information on wages and dividends paid, as well as other accounting data. We have information on all owners of shares in Norwegian stock-based companies, which allows us to identify firms that are wholly owned by a single person. Our main dataset thus consists of annual linked firm-owner data, but in order to accurately calculate tax overpayment, we make further sample restrictions as we explain in the following section.

We have rich demographic data on individuals including age, gender, education, immigration status, and geographic location. We also have detailed data on income and wealth, which allows us to calculate the marginal tax rates that apply for each business owner.

4.1 Quantifying Tax Mistakes

This section explains in detail how we calculate the size of tax mistakes, including any assumptions and sample restrictions made to ensure our calculations are as accurate as possible.

Our calculation of tax mistakes is based on the marginal tax rates that apply when full business owners – defined as those who own all shares of a company – transfer money from their firm to their personal accounts. Whether the money is transferred in the form of dividends or wages, taxes are incurred both at the business and the individual level. We assume that owners would prefer to minimize the total tax liability, i.e. the sum of taxes applied on the business and on themselves as individuals.

Dividends are paid out from firm profits, and are thus taxed first at the corporate tax rate τ_{corp} when profits accrue, and subsequently at the dividend tax rate τ_{div} when paid out to the owner. We can therefore calculate the total marginal tax rate on dividends as:

$$1 - (1 - \tau_{corp})(1 - \tau_{div}).$$

However, not all dividends are subject to taxes. The Norwegian tax code features a deduction equivalent to the risk-free return on investments (*skjermningsfradrag*), calculated based on yields on Norwegian government bonds. Only dividends that are in excess of the theoretical return on such a bond are taxed at rate τ_{div}. Our data include detailed information on this deduction, including how much of it is applied to individual dividend payments. In our calculations of tax mistakes, we only consider the taxed portion of dividend payments, leaving out any dividends that are untaxed due to this deduction.

We now turn to the calculation of taxes on wage income. When a business owner receives wages from their own company, they are treated like any other employee from a tax code perspective. The wage is considered a business expense, so the amount spent on wages is not subject to corporate taxes. However, the company does owe payroll taxes (*arbeidsgiveravgift*) on the wage amount, at a rate τ_{pr}. This tax is not taken out of the gross wage paid to workers, but is instead paid directly by the firm in adddition to gross wages, so the marginal tax rate associated with this tax is $\frac{\tau_{pr}}{1+\tau_{pr}}$.[7]

The payroll tax is geographically differentiated into seven zones based on the physical location of the firm, with taxes being lower in more rural areas. Zone 1, which encompasses nearly all major urban areas and contained approximately 78% of the population in 2017, has a payroll tax rate of 14.1%, while the tax rate is 0% in the zone that covers Norway's extreme north. Our data contains information on the location of each firm, so we apply the appropriate payroll tax rate for all firms in our calculations. One zone has a multi-bracket system for the payroll tax, where wage expenses are taxed at a lower rate up to a certain threshold. For firms in this zone, we use data on the firm's total wage expenses to apply the appropriate tax rate, taking wage expenses for all other employees as given.

Received wages are subject to personal income tax τ_{inc} at the individual level.[8] Income tax rates are constant across the country, with the exception of the sparsely populated extreme northern zone, whose residents face slightly lower income tax rates. Using data on the municipality of owner-employees, we apply these lower tax rates for residents of this zone.

Combining payroll and individual income taxes, the effective marginal tax rate on wage income for a business owner can be calculated as:

$$\frac{\tau_{inc} + \tau_{pr}}{1 + \tau_{pr}}$$

However, there is one additional complication when calculating the effective tax rate for wage income, namely the Norwegian pension system. Government pensions in Norway are calculated based on lifetime earnings in a quite similar manner to Social Security in the United States. Importantly, only income from wages and certain other sources counts towards pensions, but

[7]For a simple example demonstrating the logic behind this, consider a hypothetical tax rate of 100%. For a regular income tax, this would mean that the worker would be left with nothing. However, in the case of the payroll tax, it would simply mean that for every Norwegian krone paid out to the worker, the government would receive one krone from the firm as well. The effective marginal tax rate (ignoring subsequent income taxes) would thus be $\frac{\tau_{pr}}{1+\tau_{pr}} = \frac{1}{1+1} = \frac{1}{2}$.

[8]Norway differentiates between personal income taxes and social security taxes (*trygdeavgift*), which are used to finance welfare benefits for e.g. unemployment, maternity leave, and disability, but not old-age pensions. Since social security taxes are levied at the individual level, we use the term personal income tax to refer to the combined rates of these and "actual" income taxes.

not dividend income. Prior to 2011, the system was very intricate, involving a complex calculation with multiple brackets and only considering the highest earning years within a person's career. However, in 2011, the system was simplified for individuals born after 1962. For these people, any wage income in a given year up to a certain threshold will generate pension savings equal to 18.1% of earnings, while any wage income beyond the threshold will not contribute to pensions. Accumulated pension savings for an individual are adjusted upwards every year at a rate equal to average wage growth. Upon retirement, average life expectancy for a cohort is used to convert pension savings to an annuity, which is paid out until death.

Because pension savings are generated from wage earnings but not dividends, we can think of them as a negative tax on wage income. The pre-2011 system is too complex for us to reliably calculate an effective tax rate associated with pension savings, which is why we limit our analysis to the years from 2011 onwards and to business owners born in 1963 or later. For these people, we could in principle think of the pension savings as a negative tax of 18.1% up to the threshold. However, we want to be conservative in our assumptions and rationalize as much as possible of what we observe through potential individual preferences rather than labelling it a mistake. There are several reasons why a hypothetical perfectly rational individual might not value these forced savings at their "nominal" value of 18.1%. For instance, they might be able to obtain a higher average return on their other investments than the growth rate of wages. They might anticipate that they will live fewer years than the average member of their cohort, in which case the expected value of the annuity they receive after retirement will be lower.[9] Or they might simply have intertemporal preferences under which they would prefer saving less than 18.1% of their income. For these reasons, we cannot simply consider pension savings to be a negative tax of 18.1% up to the threshold in our calculation of tax mistakes. If we were to do so, we might overestimate the tax mistake for someone who pays themselves "too much" in dividends relative to wage income. Therefore, we assume that the lower bound of the tax value of pension savings corresponds to half the nominal rate – 9.05% – when calculating

[9] Of course, there is the converse possibility that some taxpayers think they will live longer than average, which would raise their valuation of the pension savings.

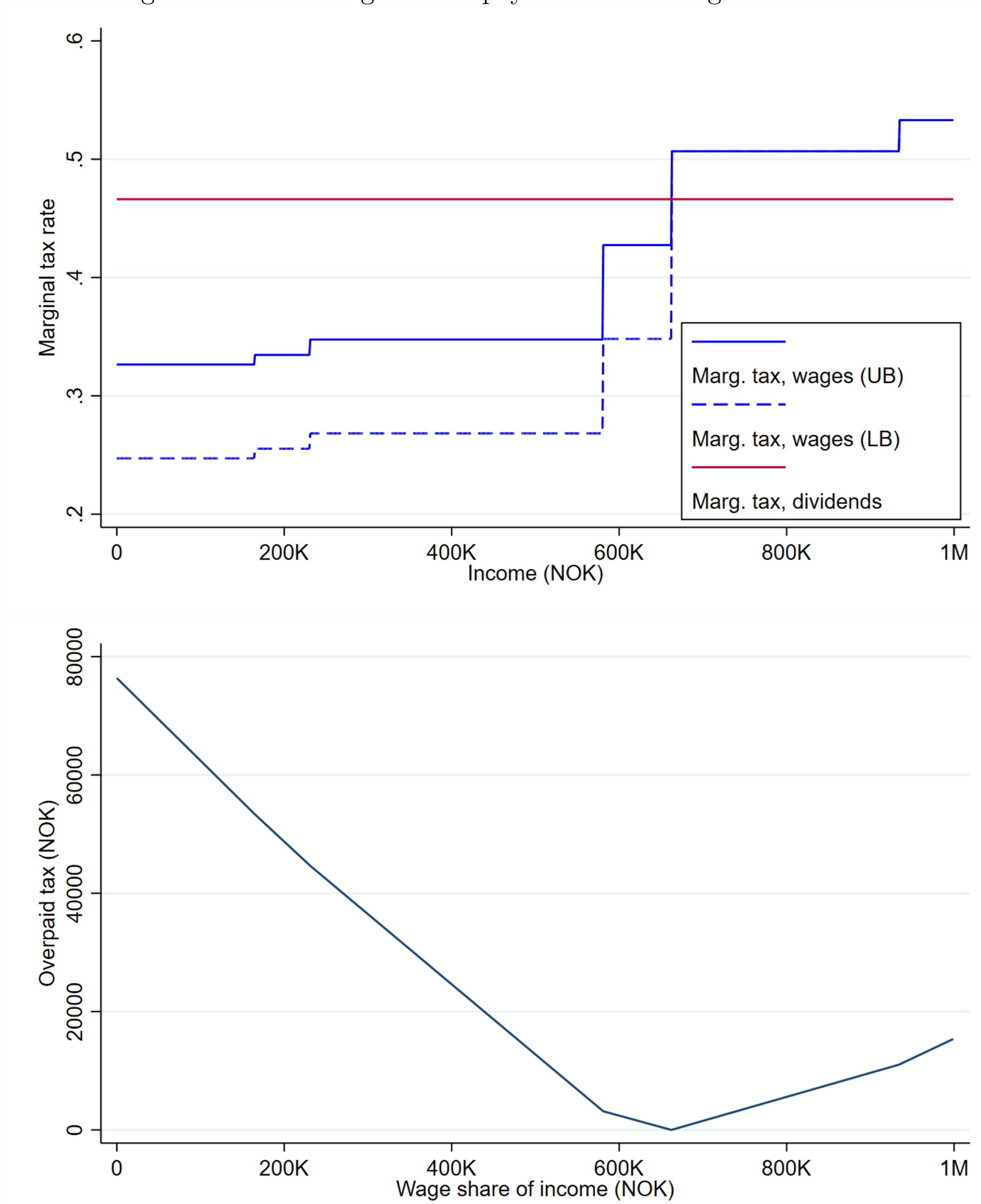
tax mistakes for those who overpay themselves in dividends. We view this as a reasonable and conservative assumption, since for most taxpayers, these pension payments are a significant part of old-age income and must be assumed to have at least some value.

The effect of including pensions in our marginal tax rate calculations is shown in the top panel of Figure 3. This graph corresponds to the tax scheme shown in Figure 2, but with marginal tax rates on wage income adjusted downwards below the pension threshold. This creates an additional bracket at the threshold, and two bounds on marginal tax rates below this point. The dotted line corresponds to the assumption that pension savings are valued at their full nominal value of 18.1%, while the solid line corresponds to a valuation of 9.05%, which leads to a higher effective tax rate on wage income.

The top and bottom panels of Figure 3 together illustrate how we use marginal tax rates to calculate tax overpayment. We can think of the bottom panel as illustrating the tax overpayment associated with possible allocations between wages and dividends for a hypothetical taxpayer transfering a total of NOK 1 million out of their firm. The horizontal axis shows the share of this amount which is paid in the form of wages. Recall that the optimal strategy for a fully rational taxpayer with perfect information would be to pay the transfer in the form of wages up to the point where marginal tax rates cross – at NOK 662,295 in this case – and the rest as dividends. This strategy corresponds to the minimum point of the graph and would generate a tax overpayment of 0. If instead the taxpayer were to pay themselves the entire NOK 1 million in the form of wages, this would correspond to the rightmost point on the graph and generate a tax overpayment of NOK 15,370 - equivalent to the area between the marginal tax rate graphs above the intersection point in the top panel.

If we instead suppose the firm owner pays themselves the NOK 1 million entirely in the form of dividends, this would correspond to the leftmost point on the graph and lead to a tax overpayment of NOK 76,390. This amount is calculated by using the *upper* bound of marginal taxes on labor – i.e. as the area between the solid blue line and the red line below their intersection point – as this is the assumption that gives rise to the lowest estimated overpayment in this case. Note that wage payments below the intersection point only cause

Figure 3: Calculating tax overpayment from marginal tax rates



Notes: The top panel shows tax schedules in 2017 for wages and dividends. Wage incomes up to NOK 662,295 generate pension savings, which reduces the effective marginal tax rate below this point. We assume taxpayers would value these savings at most at their nominal value of 18.1% of income, and at least at half this amount. These assumptions give rise to the two bounds on wage tax rates.

The bottom panel shows the total tax overpayment for a hypothetical business owner transferring money from their business to their personal account. Wage amounts below the optimum (the point where marginal tax rates cross) only lead to tax overpayment when the taxpayer has made corresponding dividend payments, which this graph assumes. The maximum overpayment shown at a wage amount of 0 thus only applies if the taxpayer has made at least NOK 662,295 in dividend payments – lower dividend amounts lead to less tax overpayment. In cases below the optimum, the calculated overpayment is based on the upper bound of the wage tax range, which leads to the smallest overpayment.

tax overpayments to the extent that the owner also pays themselves in the form of dividends. A business owner paying herself NOK 400,000 in the form of wages and nothing in dividends would incur no tax overpayment. If she were to pay herself 400,000 as wages and 100,000 as dividends, her overpayment would be $100,000 \cdot 0.118 = 11,800$, where 0.118 is the difference in marginal tax rates on wages (upper bound) and dividends over the interval from 400K to 500K in the top panel.

The above illustrates the general principles behind our calculations of tax overpayment. However, certain complications apply in specific cases. Some individuals in our sample are full owners of more than one firm, in which case we aggregate up all wage and dividend payments from each firm, and apply the same calculations as if the payments derived from one large firm. Some taxpayers have wage income from other sources than their own firm. In our calculations of tax overpayment, we take this wage income as given – for instance, if a taxpayer has NOK 300,000 of wage income from other sources, our method applies the same calculations as in Figure 3, but starting from an income of 300,000 rather than 0. Finally, we note that in some cases in the extreme northern zone of Norway, the reduced rates for individual income and payroll taxes mean that the marginal tax rate on wage income will never exceed the tax rate on dividend income. In these cases, taxes are optimized when everything is paid as wages. A business owner paying any taxable dividend to themselves will always be making a tax mistake, and we calculate overpayment correspondingly.

5 Tax overpayment: Empirical evidence

We now present evidence on tax overpayment, its distribution within the study population, and how this varies for different population groups. We emphasize that our study population consists of people who are full owners of a business, and who will therefore in many ways be unrepresentative of the general population. The setting we study here deals with one particular form of tax complexity and the dispersion in effective tax rates generated from this, but tax complexity in other areas may lead to very different outcomes. Nevertheless, we believe that some of the evidence presented here can at least be indicative of more general trends.

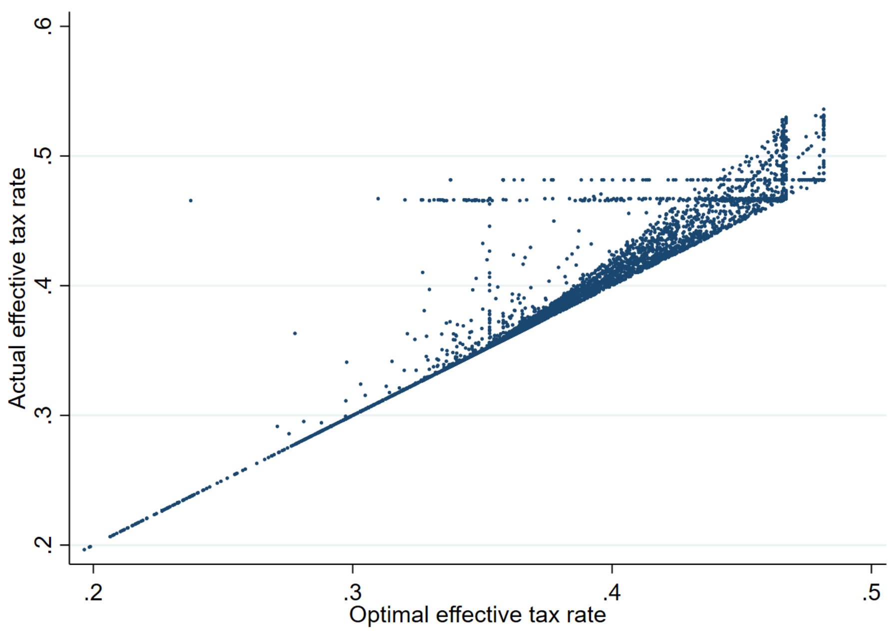
Figure 4 provides an overview of how tax overpayment is distributed in the general population. The top panel plots actual effective tax rates against the theoretical minimized effective tax rates that each individual could have achieved if they had behaved completely optimally. The actual effective tax rate, shown on the vertical axis, is calculated by adding up the total tax amount paid on dividends and wages (including both personal and business taxes) and dividing it by the sum of wages and taxable dividends. The optimal effective tax rate is a result of redoing the same calculation, but using the total tax liability that would have applied if the taxpayer had optimized. In order to more clearly see individual points and patterns in the data, the plot shows a randomly selected sample of 10,000 individual-year observations from our full study population.

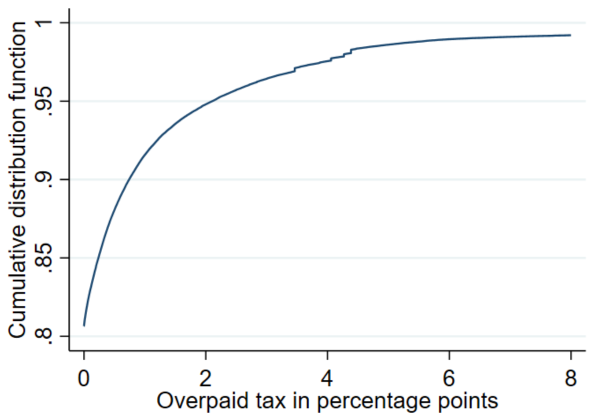
We see a clear pattern here: For those with low wage incomes, who thus also have a low optimal effective tax rate, mistakes are very rare – in the bottom left corner of the scatterplot, nearly everyone is clustered along the 45-degree line. Further up the income distribution, we see much more dispersion in actual effective tax rates. The reason for this pattern is quite simple: Most business owners in most years pay themselves entirely in the form of wages. For taxpayers whose annual wage incomes are low enough, this happens to be the optimal strategy, so there is no mistake made. In the plot, we can clearly see the point where marginal tax rates for wage income start to exceed the tax rate on dividends, since the distribution starts fanning out beyond this point, as many taxpayers pay themselves too much in the form of wages. The two horizontal lines of points at the top of the graph consist of those taxpayers who pay themselves entirely in the form of dividends.[10]

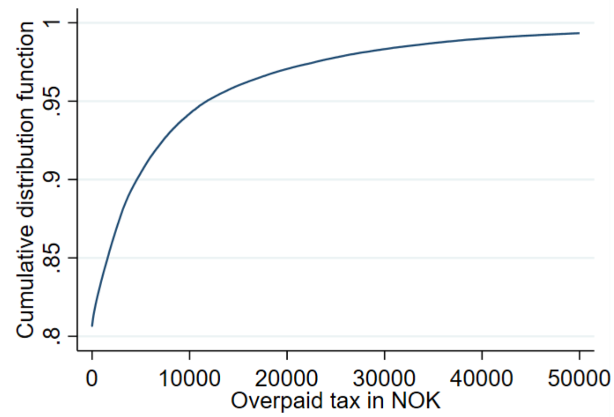
The bottom two panels of Figure 4 show two cumulative distribution functions for tax overpayments. In the bottom left panel, overpayments are expressed as a percentage of total business income – corresponding to the vertical distance between the 45-degree line and each

[10]There are two separate lines since the marginal tax rate on dividends was increased slightly in 2016 and 2017 compared to previous years. The two vertical lines seen near the top of the distribution represents business owners who earn enough income from other sources that they should optimally pay out all business income as dividends, but who nevertheless pay themselves fully or partly in wages.

Figure 4: Distribution of tax overpayment







Notes: The top panel shows a scatterplot that relates actual effective tax rates on business income to the hypothetical effective tax rate that could be attained under optimal behavior, for a random sample of 15,000 taxpayers from our study population. The actual effective tax rate is calculated as the total tax liability on business income divided by the gross amount of this income. The "optimal effective tax rate" assumes the same total gross income from the business, but instead uses the hypothetical tax liability if this income had been optimally split between wages and dividends for the given taxpayer. A small number of taxpayers on the bottom left of this panel have very low effective tax rates due to favorable tax treatment of residents in the extreme northern part of Norway.
The bottom two panels show cumulative distribution functions for tax overpayments. The bottom left panel shows the distribution of surplus tax in percentage points, equivalent to the vertical distances between points in the scatterplot and the 45-degree line. The bottom right panel shows the distribution of tax overpayments in absolute amounts. In the years 2011-2017 which we study, the exchange rate fluctuated roughly between 6 and 8 Norwegian kroner per US dollar. The CDFs start at a value of roughly 0.81 since this is the fraction of taxpayers who do not overpay at all in a given year. Both figures omit some extreme outliers on the top end.

point in the scatterplot. This is essentially a measure of how much higher each taxpayer's effective tax rate becomes as a result of misoptimization. The bottom right panel shows the overpayment distribution in pure money terms. For this panel, note that across the study period, the USD-NOK exchange rate has fluctuated between 6 and 8. For instance, the 95th percentile of this distribution is NOK 11,747, which would roughly correspond to between 1500 and 2000 US dollars.

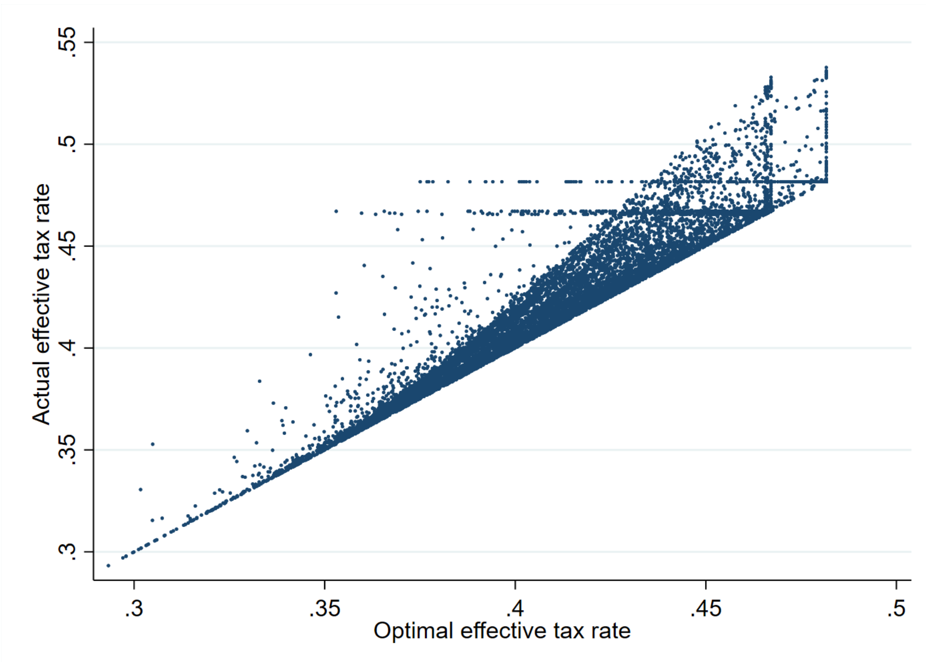
In both of the lower panels, the CDFs start at a value of approximately 0.81, indicating that 81% of taxpayers do not make mistakes. The vast majority of these are people whose total business income does not reach the point in the tax scheme where marginal taxes are higher for wages than for dividends, and who are therefore behaving optimally when they take everything out as wages. Essentially, this means that in this particular case, tax complexity increases with income. Low-income business owners have a very simple optimal strategy, whereas at higher incomes, tax optimization requires splitting income between dividends and wages, which is a more complicated process. Because we are mostly interested in taxpayers who face a substantially complex tax system, much of our subsequent analysis will drop those taxpayers whose total business income is below the crossing point of marginal tax rates.[11] We will refer to the remaining taxpayers as the "high-income" sample. These make up 51,343 of our total 335,704 individual-year observations.[12]

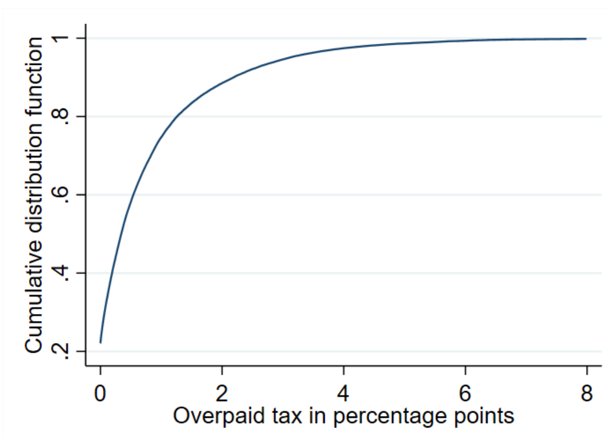
Figure 5 is equivalent to Figure 4, but narrowed down to our high-income sample. The scatterplot now excludes the long "tail" of taxpayers in the lower part of the tax distribution, but otherwise shows all of the same patterns as the scatterplot in Figure 4. Looking at the two

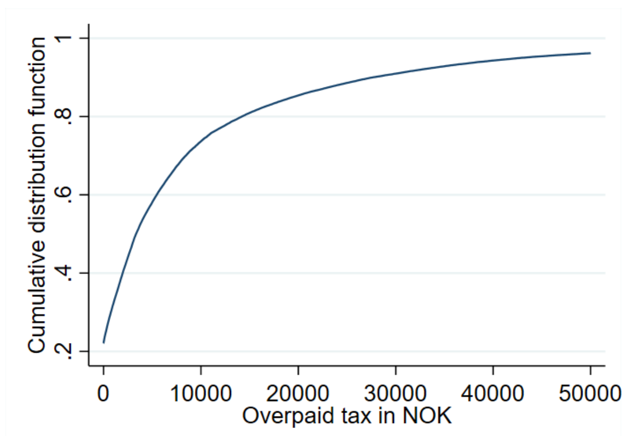
[11] Note that while most of these low-income taxpayers behave optimally by paying only wages, a few do pay some or all of their business income in the form of taxable dividends. Some of these taxpayers can end up making extremely large mistakes – for example, in the top panel of Figure 4, the sample includes one outlier at the top left who overpays their taxes by over 20 percentage points, and ends up paying more than twice as much in taxes as they optimally could have.

[12] The term "high-income" may be misleading in some cases, as business owners can have a high economic income in their business, but only transfer a small amount to their personal accounts. In fact, this can in itself be the outcome of another tax optimization strategy: Keeping retained earnings in the firm and reinvesting them from there will generally incur less capital taxes than paying them out immediately as they are earned. This illustrates once more that tax complexity is much broader than the specific example we study. We note that this additional dimension of tax complexity is not a threat to our strategy: Conditioning on the amount of money that owners transfer out of the firm, we do measure overpayment accurately. It just means that they may be leaving even more money on the table than we can accurately measure.

Figure 5: Distribution of tax overpayment: High-income sample







Notes: This figure corresponds to Figure 4, but reduced to our "high-income" sample of those business owners whose combined wage and dividend income from their business in a given year exceeds the point where marginal tax rate schedules for wages and dividends cross. In other words, this is the subpopulation who, given their total transfer, face a more complex tax situation since they cannot optimally pay the entire amount out as wages. The high-income sample comprises 51,343 individual-year observations out of the total 335,704 in our full study population. To improve readability, the scatterplot is based on 15,000 randomly drawn observations from this sample. See the notes to Figure 4 for further details.

panels with CDFs, we see that unsurprisingly, tax overpayment is much more common within this subgroup, as they face a much more complex tax situation. Only 22% of this group do not overpay any tax at all, and the majority of those are individuals who have so much income from other sources that paying all their business income as dividends is optimal.

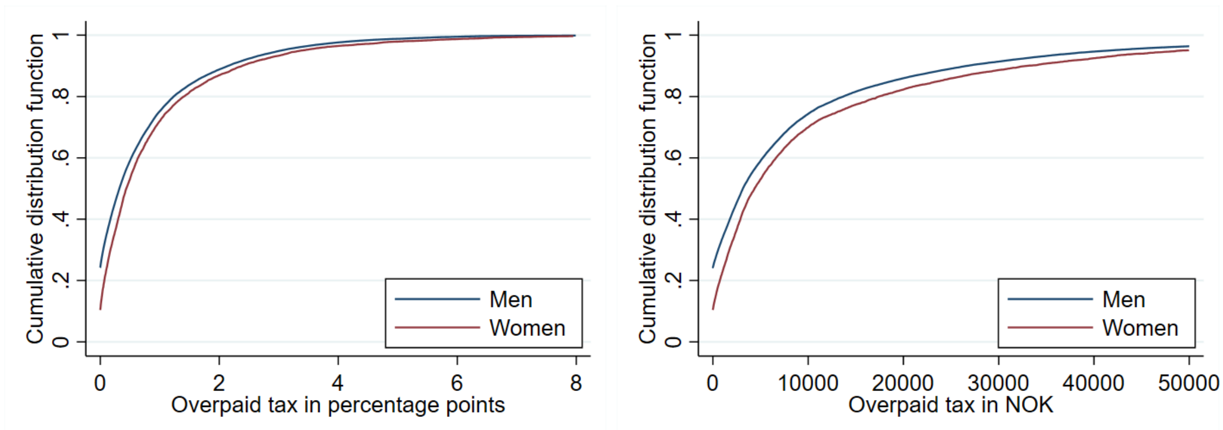
Next, we examine how tax overpayment varies for certain demographic subgroups that we can observe in our data. We may have reason to believe that tax overpayments would differ by demography – either through the mechanism demonstrated in our model, if tax elasticities vary by demographic group, or for other reasons which our model cannot capture. For instance, Alstadsæter, Kopczuk and Telle (2019) have demonstrated that network effects play an important role in tax knowledge acquisition. If some demographic groups have access to more tax knowledge than others through their network, they may require less effort to learn about certain tax minimization strategies.

Figure 6 shows how tax elasticities are distributed within the high-income sample based on gender and immigration status. Immigrants are defined as those born abroad to non-Norwegian parents; native-born Norwegians are defined in this context as those born in Norway to two Norwegian parents. The Norwegian system for classification of immigration background includes other categories for more complex situations, which are dropped from the figure. We clearly see that tax overpayments are larger for women relative to men and immigrants – both when measured in percentage points and in absolute terms, and that this is true across nearly the entire distribution. The mean overpayment for women in this sample is NOK 11092 (0.9 percentage points), versus NOK 9322 (0.75 pp) for men. For immigrants, the mean overpayment is NOK 10998 (0.95 pp), while it is NOK 9307 (0.75 pp) for native Norwegians.

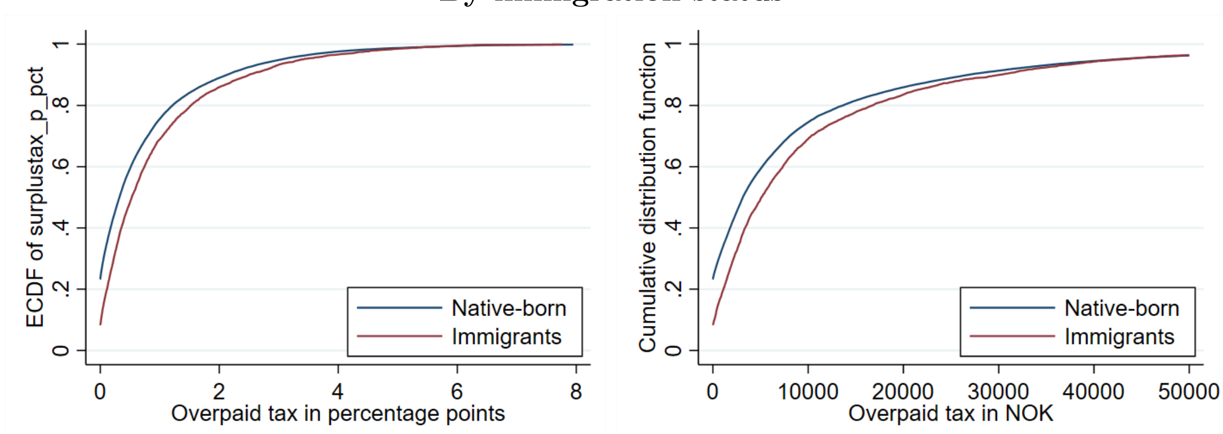
Next, we look at tax mistakes by within-firm advisor spending and by wealth tax liability in Figure 7. We have fairly detailed financial data for all firms, including a line item capturing expenses on outside services such as accounting and financial advisors. We calculate the ratio of advisor spending to total firm assets and assign as "high-spending" those for whom this value is above the median. We can think of this as a proxy for tax knowledge, although we emphasize that it is a crude measure – this spending covers all accounting and financial advice, not

Figure 6: Tax overpayment by demographic groups

By gender

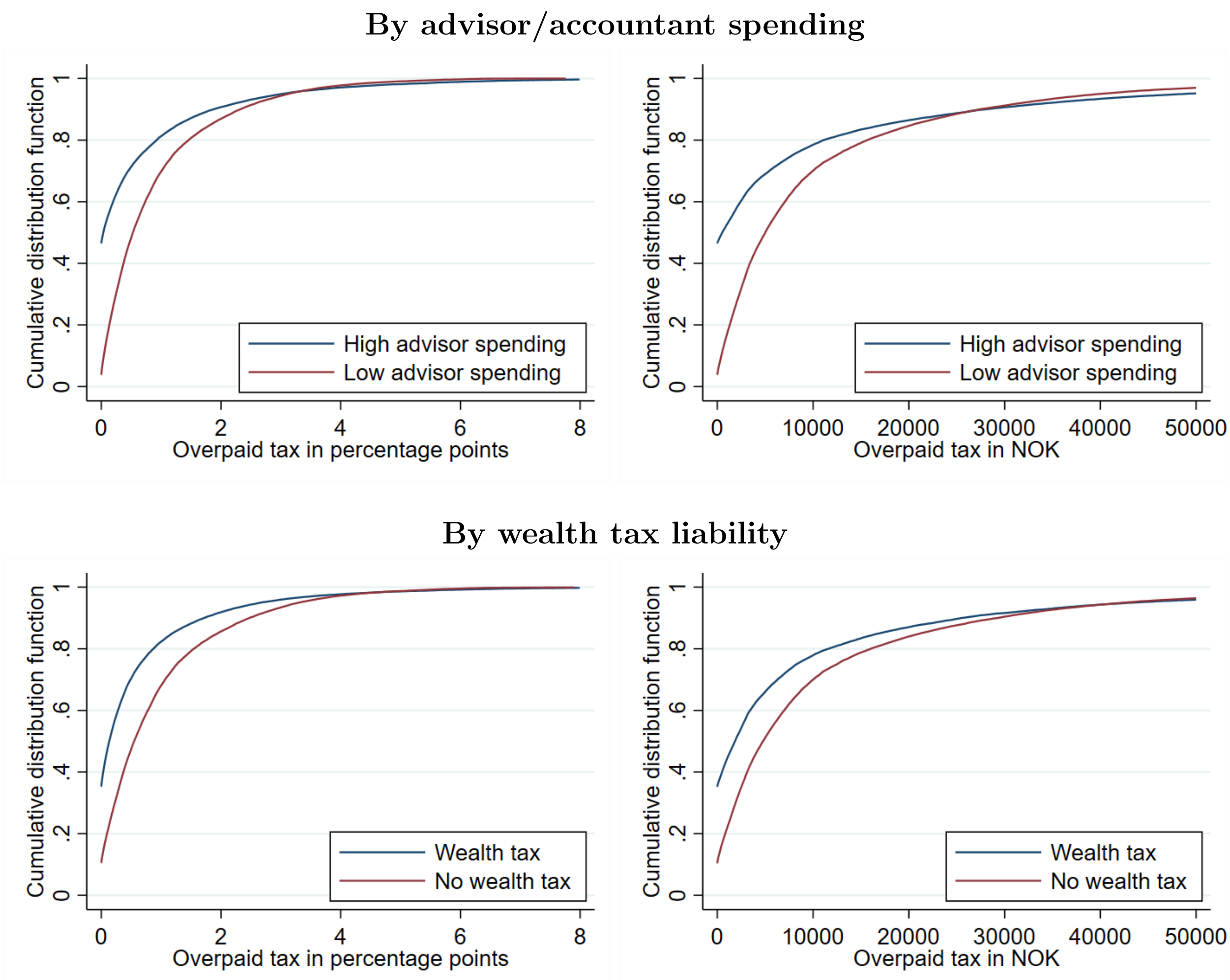


By immigration status



Notes: This figure shows distributions of tax overpayments for demographic subgroups within the high-income sample. In the bottom two panels, immigrants are defined as those born abroad to two non-Norwegian parents, whereas native-born refers to people born in Norway to Norwegian parents. Norwegian immigration classifications includes separate categories for people born abroad with one or two Norwegian parents, and those born in Norway with one or two non-Norwegian parents. These categories are dropped from this figure.

Figure 7: Tax overpayment by wealth and advisor expenditure



Notes: This figure shows distributions of tax overpayments for different subgroups within the high-income sample. In the top two panels, taxpayers are grouped by the size of their within-firm spending on accountants divided by total assets, relative to the median. Wealth tax liability in the bottom two panels is based on total assets owned, including those held indirectly through the business.

necessarily just tax advice. And even if the firm has hired a tax advisor, they might only take the firm's financial situation into account and not the owner's personal taxes. Nevertheless, we see an extremely strong effect: Almost half of those with high advisor spending do not overpay taxes at all, relative to a negligible number for those with low spending. The difference in overpayment is quite marked over most of the distribution, though it does reverse near the top. The average overpayment for those with high advisor spending in this sample is 8310 (0.6 pp), versus 10553 (0.9 pp) for those with low spending.

Norway applies a wealth tax to individuals with net wealth above a certain threshold, which ranges over our study period from NOK 700,000 in 2011 to 1,480,000 in 2017. All assets, including those owned indirectly through a business, are included in this calculation.[13] We think of wealth tax liability as simply a proxy for being wealthy – 28.2% of our full sample and 46.7% of our high-income sample pay wealth tax. Importantly, being liable for the wealth tax does not change any incentives to minimize personal and business taxes when transferring money out of the business. We see that the wealthy overpay significantly less than the rest of the sample, and are less likely to overpay at all. Average overpayment for high-wealth tax payers is 8327 (0.58 pp), while it is 10704 (0.94 pp) for the less wealthy. This makes sense given our model's prediction that tax knowledge is increasing in income, if we use wealth as a proxy for this: People who earn enough to pay wealth taxes typically have more to gain from saving on their taxes, so they are more likely to find it worth the cost to invest in tax knowledge. Also, the wealth tax itself might generate additional complexity which can make these taxpayers more likely to hire a tax advisor for their individual taxes, who may also help them optimize their tax situation in this specific case. However, since we cannot observe any privately hired tax advisors, this is difficult to verify.

We note that our restriction to the high-income sample is important in these figures, because as evidenced by the top panel of Figure 4, tax complexity rises with income in our empirical

[13]Some directly owned assets are applied a discount before being included in the calculation. For example, the tax value of a primary residence is only counted as 25% of its assessed value. Married couples are taxed jointly on their wealth with thresholds being twice as large as those that apply for individuals. The wealth tax rate was 1.1% in 2011-2013, 1% in 2014 and 0.85% from 2015 onwards.

Table 2: Group differences in tax overpayment

	(1)	(2)	(3)	(4)	(5)
Female	3315.2				3378.4
	(257.3)				(265.4)
Immigrant		1010.0			599.6
		(285.5)			(281.5)
Wealth tax			-255.4		148.4
			(246.7)		(240.9)
High advisor spend				-3941.2	-4030.2
				(247.9)	(246.2)

Notes: This table shows Tobit regressions of calculated tax overpayment in NOK on group indicators. The outcome variable is left-censored at 0. Each regression controls for the optimal effective tax rate and a spline over total income deciles, where total income is the sum of taxed dividends and all wage income from the owned business and other sources. Standard errors are clustered at the individual level.

setting. If we simply look at the unrestricted study population, overpayments are larger and more frequent for those who pay wealth tax and for native-born Norwegians, which is due to the fact that these groups on average earn more than their counterparts, and thus are more likely to be in the income region where complexity is high. Restricting to high-income taxpayers is one way to get around this issue, but as a further robustness check, we estimate the difference for our full sample. We regress total tax overpayment at the person-year level on a group indicator and control flexibly for the total of wage and taxed dividend income using a 10-part spline, as well as for the optimal effective tax rate, shown on the horizontal axis in the scatterplot in Figure 4. Because the outcome variable is 0 for over 80% of our full sample, we use a "Tobit" censored regression model, which accounts for the outcome variable being restricted to values weakly greater than zero (Tobin, 1958). Results are shown in Table 2 for each of our indicator variables separately, as well as all together, and roughly confirm what we have seen in the figures. We do note, however, that the estimate for wealth taxes is small and insignificant compared to the relatively large apparent effect in the corresponding graph. This is likely due to some large outliers, as we do see a comparatively fatter tail for the wealth tax payers in Figure 7.

Overall, we have seen that tax complexity leads to dispersion in tax rates and overpayment relative to the optimal tax rate. This overpayment can be very large in some cases and is on average larger for women and immigrants relative to men and native Norwegians, respectively. Wealthier individuals make fewer mistakes, highlighting the adverse effect of complexity on inequality, and those with high in-firm advisor spending make fewer mistakes, which indicates that differences in tax knowledge does explain some of the dispersion.

6 Linking overpayment and elasticity

The previous section demonstrated that in a setting with tax complexity, there is considerable variation in how optimally people behave, and subsequently what effective tax rates are achieved. This section investigates our model's prediction that heterogeneous elasticities can explain some of the variation we observe in effective tax rates.

For this purpose, we use our existing panel of business owners and estimate their behavioral response to an earlier tax reform which was phased in from 2005 onwards. An overview of this reform is shown in Table 6. The reform kept the base tax rate unchanged at 28% while lowering tax rates for higher brackets over the years 2005 and 2006. The middle bracket rate was lowered from 41.5% to 37%, and the highest bracket rate was lowered from 47.5% to 40%. At the same time, the cutoff for the highest bracket was gradually adjusted downwards.[14] Using income data for the taxpayers in our panel for the years around the reform, we can estimate the elasticity of taxable income (ETI). We group taxpayers by the magnitude of the overpayment we observe them making in later years, and test if taxpayers who overpay more are also less elastic when reacting to a tax reform, as our model predicts.

We estimate the ETI using the following panel regression for the years t = 2003-2006:

$$\Delta \log(Y_{it}) = \alpha_0 + \alpha_1 \mathbf{1}_g + \beta_0 \Delta \log(NTR_{it}) + \beta_1 \mathbf{1}_g \Delta \log(NTR_{it}) + \gamma X_{it} + \mu_i + \theta_t + u_{it}, \quad (8)$$

[14]The cutoff for the middle bracket increased slightly over these years, but this happens routinely to account for inflation and avoid bracket creep, even outside of reform years.

Table 3: 2005-2006 tax reform

	2004	2005	2006
Base tax rate	28%	28%	28%
Bracket 1 cutoff (NOK)	354,300	381,000	394,000
Bracket 1 tax rate	41.5%	40%	37%
Bracket 2 cutoff (NOK)	906,900	800,000	750,000
Bracket 1 tax rate	47.5%	43.5%	40%

where Δ represents the two-year difference in values of a variable, Y_{it} represents taxable income, and NTR_{it} is the marginal net-of-tax rate faced by person i in year t. $\mathbf{1}_g$ is an indicator for person i belonging to a particular group, which we will use to test if the ETI differs for taxpayers based on various criteria, e.g. magnitude of tax overpayment in our later study years. The coefficient β_0 provides an estimate of the ETI for those not in group g, whereas β_1 estimates the difference in the ETI for group g relative to non-group members. μ_i and θ_t are individual and year fixed effects, which we use to control for income trends by person and generally over time. In order to consistently estimate the person fixed effects and avoid multicollinearity, we include a few years prior to the 2005 tax reform, which helps establish a baseline income growth trend for each taxpayer.

It is well known that the main regressor of interest, $\Delta \log(NTR_{it})$, will be endogenous, since the net-of-tax rate depends on a taxpayer's bracket, which is a direct function of Y_{it} and hence affected by the error term u_{it} (Saez, Slemrod and Giertz, 2012). To address this issue, we follow Weber (2014), who recommends instrumenting $\Delta \log(NTR_{it})$ with a variable that takes the hypothetical difference in log net-of-tax rates under the tax systems in two different years, but calculated using income from some year prior to the base year. The instrument will thus measure the effect of the change in tax system, without being affected by concurrent random fluctuations in income. In our case, we instrument with the hypothetical change in marginal net-of-tax rates using income two years prior to the base year – e.g. we calculate the difference in net-of-tax rates an individual would face under the 2006 and 2004 tax systems, respectively, if their income were kept fixed at its level in 2002.[15] To instrument for the interaction $\mathbf{1}_g \Delta \log(NTR_{it})$, we use the interaction of the Weber instrument with $\mathbf{1}_g$. Weber (2014) notes that endogeneity will persist if income growth rates are different along the income distribution, and following her, we include as a remedy a income splines for the same year used in calculating the instrument. We drop individuals earning under NOK 100,000 from our sample, and following Saez, Slemrod and Giertz (2012), we trim values of $\Delta \log(Y_{it})$ at 2% on each tail to limit the influence of outliers.

One possible objection to this study design might be that tax knowledge and elasticities are difficult to completely disentangle empirically. An individual with more tax knowledge will be better at tax optimization, and this might empirically manifest as a larger estimated tax elasticity in response to a reform like this one, even if it were just due to income shifting as described in the previous section, rather than a "real" response. To alleviate this concern, we estimate (8) only for those taxpayers whose firms were founded in 2007 or later, such that they are not self-employed during the tax reform.

We look at several groups of taxpayers in our sample to study how their elasticity estimates differ from the rest of the population. First, to investigate our model's prediction that higher tax knowledge is correlated with higher elasticity, we group our sample by the average magnitude of the tax overpayments that we observe them making in later years. We look at two measures: Whether an individual's average tax overpayment over the years 2011-2017 is greater than the median, and whether the average overpayment in percent exceeds 1 percentage point. As we have described in previous sections, there is a strong correlation between income and tax complexity in the full sample, so to avoid accidentally selecting on average income when constructing these samples, we only include in our calculations those observation years where an individual belongs to the high-income sample as defined in the previous section, i.e. the years where total business income is large enough that a simple strategy of paying everything as wages would not be optimal. This limits the sample size quite a lot for these regressions, as

[15] Formally, the exclusion restriction requires using an income year with a sufficiently long lag that random shocks to the income process will be serially uncorrelated with those in the base year. Using data from Denmark, Weber finds that a two-year lag relative to the base year is sufficient to eliminate this serial correlation.

Table 4: Elasticity estimates for 2005-2006 tax reform

$\Delta \log I$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δ log NTR	0.085	0.739	0.267	-0.064	-0.043	0.203	0.107
	(0.160)	(0.904)	(0.651)	(0.172)	(0.164)	(0.171)	(0.167)
Δ log NTR * group		-0.988	-0.485	0.808	0.931	-1.365	-0.059
		(0.815)	(0.627)	(0.233)	(0.599)	(0.619)	(0.838)
Group		> median	>1pp	High inc	WT	Women	Immigr.
Observations	130,858	17,078	16,968	130,858	130,858	130,858	122,558

Notes: This table shows instrumental variable estimates of ETI for various subgroups within our sample. Δ represents a two-year difference. All regressions cover the years 2003-2006 and include individual and year fixed effects. Standard errors are clustered at the individual level. See the main text for further details on the regression specification and group definitions.

anyone who is never part of the high-income sample in 2011-2017 is dropped.

We also study elasticities by income and wealth. Several results from our theoretical section assume a positive correlation between income and elasticity, so we are interested in finding out if this holds empirically. We look at ETI estimates for high-income taxpayers, defined as those earning an average of at least 500,000 annually, and for taxpayers who pay the wealth tax. Basing this on concurrent income or wealth would introduce endogeneity since these are directly influenced by the error term, so we instead group individuals by wealth tax liability in the year 2000 – before our study period begins – and average income in the years 1997-2000. Finally, we look at elasticities for women relative to men, and for immigrants relative to native Norwegians.

Our results are shown in Table 4. The first column just estimates the elasticity for the full sample, omitting any group interaction. The next two columns show results based on the magnitude of later tax overpayment. Column 2 estimates elasticities for those with overpayment above vs. below the median in money terms, and column 3 for those with overpayment of at least 1 percentage point compared to the rest. In both columns, we can see that the estimated effects are numerically quite large in the direction that our model would suggest – taxpayers with large overpayments are on average less elastic – but the small sample size also leads to large standard errors, so neither of these estimates is statistically significant. Although we cannot conclusively prove a link between low elasticity and high tax overpayment based on these regressions, we do consider this to be at least an indication of a possible connection.

Columns 4 and 5 show our results for high-income and high-wealth taxpayers. We see that the estimated elasticity is significantly larger for higher earners, in line with previous estimates by e.g. Kumar and Liang (2020), and that the estimated difference by wealth tax liability is also positive, though not statistically significant. Finally, we see that women have a surprisingly strongly negative elasticity estimate, while immigrants have an estimated very small difference to the native population. The large negative difference estimate for women is statistically significant, and note that the hypothesis that the elasticity for women is 0, is just barely rejected at a 5% level ($F = 3.97, p = 0.046$). The strongly negative estimate is implausible, and could be due to selection issues – our sample of business owners is over 80% male, and if women in our sample are concentrated in certain business areas or fields, results could be influenced by variations in income trends for these fields that happen to covary with the tax reform. However, we also note that a slightly negative elasticity estimate, as we see for some other groups, is not necessarily at odds with standard consumer theory, since the income effect may well dominate the substitution effect around the tax threshold for certain demographics.

Overall, these results lend some credence to our model, though we cannot say for sure that elasticity differences are the main driver of the differences in tax overpayment we observe in the previous section. Many other factors, including differences in cognitive ability or networks, may also influence people's ability to optimize their taxes. However, our theoretical results still hold if differences in tax knowledge are caused by other factors, as long as they are positively correlated with elasticity and income. Even if elasticity differences are not the main driver of overpayment, the fact that they are correlated with overpayment means that complexity may still boost tax revenue and labor output along the lines of what is described in our model.

7 Conclusion

In this paper, we have provided a fresh perspective on the topic of tax complexity by showing that it can function in a similar way to price discrimination, acting as a self-selection mechanism that will cause more elastic taxpayers to learn about tax minimization strategies and achieve a lower effective tax rate. We formalize this insight through a theoretical model and show that this mechanism can in some cases lead to higher tax revenue than under a simple tax system, while at the same time leaving everyone weakly better off. We confirm this finding in a simulation exercise, but also show that benefits tend to accrue to the wealthiest, thereby exacerbating inequality.

In the empirical section of our paper, we study the topic of tax complexity using government microdata from Norway. We look at business owners who face the choice of how to pay themselves when transferring funds out of their business. As full shareholders, they are free to pay themselves any amount in either taxes or dividends, but these two types of compensation have different tax implications. We identify deviations from optimal behavior, which are common for taxpayers with sufficiently high business income and can be very large in some cases. We are able to calculate overpaid tax and show that controlling for income, women and immigrants overpay more, as do those with less wealth. We also see that owners of firms that spend more on advisory services have substantially less tax overpayment, indicating that knowledge is a likely driver of the differences we see. Using a prior tax reform, we estimate tax elasticity by group and find that those who go on to overpay more do appear to have lower elasticities.

Our results show that tax complexity may be advantageous for policymakers, as it allows them to extract more tax revenue through a tax system which is less progressive than it may appear to the general public. However, complexity also leads to increased inequality and puts weaker demographic groups at a disadvantage.

Further research on this topic could help shed light on questions that we have been unable to address. For example, what is the role of heterogeneous elasticities in driving differences in tax knowledge, relative to other possible explanatory factors, such as networks or cognitive ability? How would a tax reform that substantially simplifies or complexifies the tax system impact overpayment overall, and by different demographic and income groups? Can we observe that people learn about the tax system and become more knowledgeable over time? Answering these questions would help us better understand the tradeoffs associated with a simpler tax system.

References

Abeler, Johannes, and Simon Jäger. 2015. "Complex Tax Incentives." *American Economic Journal: Economic Policy*, 7(3): 1–28.

Aghion, Philippe, Ufuk Akcigit, Matthieu Lequien, and Stefanie Stantcheva. 2017. "Tax Simplicity and Heterogeneous Learning."

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi. 2016. "Self-Targeting: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy*, 124(2): 371–427. Publisher: The University of Chicago Press.

Alstadsæter, Annette, Wojciech Kopczuk, and Kjetil Telle. 2019. "Social networks and tax avoidance: evidence from a well-defined Norwegian tax shelter." *International Tax and Public Finance*, 26(6): 1291–1328.

Barnes, Lucy. 2022. "Taxing the rich: public preferences and public understanding." *Journal of European Public Policy*, 29(5): 787–804. Publisher: Routledge _eprint: https://doi.org/10.1080/13501763.2021.1992485.

Bastani, Spencer, and Daniel Waldenström. 2021. "The ability gradient in tax responsiveness." *Journal of Public Economics Plus*, 2: 100007.

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

Blesse, Sebastian, Florian Buhlmann, and Philipp Dörrenberg. 2019. "Do People Really Want a Simple Tax System? Evidence on Preferences Towards Income Tax Simplification."

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99(4): 1145–1177.

Craig, Ashley C., and Joel Slemrod. 2022. "Tax Knowledge and Tax Manipulation: A Unifying Model."

Fuest, Clemens, Andreas Peichl, and Thilo Schaefer. 2006. "Does Tax Simplification Yield More Equity and Efficiency? An Empirical Analysis for Germany."

Goldin, Jacob. 2015. "Optimal tax salience." *Journal of Public Economics*, 131: 115–123.

Hellwig, Martin F. 1986. "The optimal linear income tax revisited." *Journal of Public Economics*, 31(2): 163–179.

Kleven, Henrik Jacobsen, and Wojciech Kopczuk. 2011. "Transfer Program Complexity and the Take-Up of Social Benefits." *American Economic Journal: Economic Policy*, 3(1): 54–90.

Krause, Kate. 2000. "Tax Complexity: Problem or Opportunity?" *Public Finance Review*, 28(5): 395–414. Publisher: SAGE Publications Inc.

Kumar, Anil, and Che-Yuan Liang. 2020. "Estimating taxable income responses with elasticity heterogeneity." *Journal of Public Economics*, 188: 104209.

Mirrlees, J. A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *The Review of Economic Studies*, 38(2): 175–208.

Musgrave, Richard A. 1990. "Horizontal equity, once more." *National Tax Journal*, 43(2): 113–122. Publisher: The University of Chicago Press.

Ngwe, Donald. 2017. "Why Outlet Stores Exist: Averting Cannibalization in Product Line Extensions." *Marketing Science*, 36(4): 523–541. Publisher: INFORMS.

Nichols, Albert L., and Richard J. Zeckhauser. 1982. "Targeting Transfers through Restrictions on Recipients." *The American Economic Review*, 72(2): 372–377. Publisher: American Economic Association.

O'Donoghue, Ted, and Matthew Rabin. 2003. "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes." *American Economic Review*, 93(2): 186–191.

Piketty, Thomas, and Emmanuel Saez. 2013. "Chapter 7 - Optimal Labor Income Taxation." In *Handbook of Public Economics*. Vol. 5 of *handbook of public economics, vol. 5*, , ed. Alan J. Auerbach, Raj Chetty, Martin Feldstein and Emmanuel Saez, 391–474. Elsevier.

Pitt, Mark M., and Joel Slemrod. 1989. "The Compliance Cost of Itemizing Deductions: Evidence from Individual Tax Returns." *The American Economic Review*, 79(5): 1224–1232. Publisher: American Economic Association.

Rees-Jones, Alex, and Dmitry Taubinsky. 2020. "Measuring "Schmeduling"." *The Review of Economic Studies*, 87(5): 2399–2438.

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

Saez, Emmanuel, and Stefanie Stantcheva. 2016. "Generalized Social Marginal Welfare Weights for Optimal Tax Theory." *American Economic Review*, 106(1): 24–45.

Saez, Emmanuel, Joel Slemrod, and Seth H. Giertz. 2012. "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature*, 50(1): 3–50.

Seade, J. K. 1977. "On the shape of optimal tax schedules." *Journal of Public Economics*, 7(2): 203–235.

Slemrod, Joel. 1989. "The Return To Tax Simplification: an Econometric Analysis." *Public Finance Quarterly*, 17(1): 3–27. Publisher: SAGE Publications.

Sunstein, Cass R. 2018. "Sludge and ordeals." *Duke Law Journal*, 68: 1843.

Tobin, James. 1958. "Estimation of Relationships for Limited Dependent Variables." *Econometrica*, 26(1): 24–36. Publisher: [Wiley, Econometric Society].

Tolbert, Caroline J., Christopher Witko, and Cary Wolbers. 2019. "Public Support for Higher Taxes on the Wealthy: California's Proposition 30." *Politics and Governance*, 7(2): 351–364.

Weber, Caroline E. 2014. "Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences." *Journal of Public Economics*, 117: 90–103.

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