INCOME SHIFTING AS INCOME CREATION?

Håkan Selin* and Laurent Simula †

September 20, 2019

Abstract

Income shifting arises as one of the key questions when thinking about the design of a tax system as a whole. We study a simple economy, involving a benevolent policy-maker and a population of agents differing in terms of productivities, labor supply elasticities, and shifting costs. Paying special attention to the cost structure of income shifting, we highlight that when people who shift easily along the extensive margin are also more elastic in labor supply, giving them a lower tax rate is a good thing, and the government should not necessarily combat income shifting. This mechanism may be compared to third-degree price discrimination in industrial organization and works as a form of endogenous tagging. We explore this possibility numerically before showing that our results derived for a policy-maker optimally adjusting two linear tax instruments carry over when two fully non-linear taxes are potentially available. *Keywords*: Income Shifting, Taxation, Optimal Taxation, Labor Income Tax. **JEL Classification**: H21; H24

*Institute for Evaluation of Labour Market and Education Policy (IFAU), Uppsala, Sweden. Email: hakan.selin@ifau.uu.se

†University of Lyon, Ecole Normale Supérieure de Lyon, UMR GATE Lyon Saint-Etienne, France. Email: Laurent.Simula@ens-lyon.fr

1 INTRODUCTION

President Macron's proposal on a low flat capital income tax from 2018 has motivated a vast debate among French citizens, economists, and policy advisers. Its proponents claim that it would boost economic efficiency and, thus, potentially raise tax revenues. By contrast, Berkeley's economist Gabriel Zucman writes in Le Monde:

> "So far, French taxation tried to maintain some neutrality between the taxation of dividends and salaries. Such neutrality is essential to the functioning of the tax system. Because if capital income is taxed less, then anyone who is both an employee and shareholder of her business - owners, entrepreneurs, senior executives and independent - would have interest to see the fruit of their labor in the form of dividends rather than wage income, thus siphoning off revenues from social security and the State." (October 25, 2017)

Such adjustments, entirely driven by tax optimization, are referred to as "income shifting" in the economic literature and are important elements to account for when thinking about how the tax system should be designed, in France, but also elsewhere. Indeed, in comprehensive tax systems like the United States, high-income earners would have an incentive to shift income in response to differences in personal and corporate tax rates as documented by Gordon and Slemrod (2000) while in countries with separate taxation of labor and capital incomes, taxpayers would be inclined to start up closely held corporations and subsequently transfer income between the tax bases (cf., Pirttilä and Selin, 2011, Alstadsæter and Jacob, 2016, and Harju and Matikka, 2016).

In this article, we make the point that, in a well-designed tax system, policy makers may use income shifting as a way to increase both efficiency and equity. Our analysis starts from a simple observation: To shift income, a fixed cost, which differs across individuals, must be incurred. It may correspond to the disutility of gathering information about the tax law, to the time and effort spent to fill in a variety of administrative documents and tax forms, or to the cost of setting up a closely held corporation or entering self-employment. Tazhitdinova (2016) has recently shown that fixed costs of incorporation are empirically important. Heterogeneous fixed shifting costs give rise to a previously neglected *extensive margin* of income shifting.[1] In addition to being empirically relevant, it turns out that fully accounting for this extensive margin – and heterogeneity in labor supply elasticities – has policy implications in sharp contrast with the prevailing view on income shifting; social welfare can be improved by allowing taxing based on the shifting status.

[1]In the previous literature, individuals choose how much labor income to shift, the cost of shifting being smoothly increasing, at an increasing rate. See, e.g., Fuest and Huber (2001); Christiansen and Tuomala (2008); Piketty et al. (2014); Piketty and Saez (2013); and Hermle and Peichl (2015). Convex cost functions are also widely used to analyze the normative implications of tax avoidance in general, see, e.g., Slemrod and Kopczuk (2002), Kopczuk (2001), or Chetty (2009). One of the most powerful conclusions derived in this setting is that governments both increase efficiency and equity when removing incentives to shift labor earnings into more leniently taxed bases (cf. Piketty and Saez, 2013, and Piketty et al., 2014).

To illustrate this mechanism, it is enough to consider an economy in which all income stems from labor effort. Labor incomes can be shifted to an alternative tax base to a resource cost, which is fixed and/or variable. The exact administrative nature of the alternative tax base (say, corporate income tax or specific income tax, etc.) is of no relevance to the analysis. We deliberately neither model capital accumulation nor tax competition to place ourselves in the position which is the least favorable to shifting. It is thus sufficient to consider a static economy. The benevolent social planner designs taxes with the objective to maximize a weighted sum of individual utilities. Agents potentially differ with respect to three characteristics: productivity, labor supply elasticity, and cost of income shifting. Given the tax system, they simultaneously choose how much effort to supply and how much income to shift, if any.

In the spirit of Atkinson and Stiglitz (1980) and Slemrod (1994), we first consider that marginal tax rates are constant (thus focusing on linear taxes) but allow the policymaker to potentially use two tax bases, one for non-shifted earnings and one for shifted earnings, as in Piketty and Saez (2013). When shifting occurs along the extensive margin, the population is usually partitioned into ”shifters” and ”non-shifters” in the social optimum. This partition of the population plays a key part, and shifting status works as a form of ”endogenous tagging”. It implies that some agents with the *same* income determine how much effort to supply based on *different* tax schedules. In the shifting sub-population, the marginal incentives to supply labor is determined by the tax rate on shifted income whilst, in the non-shifting group, by the tax rate on non-shifted income. This mechanism clearly differs from what would be allowed by the introduction of additional tax brackets.[2] It works along the same line as third degree price discrimination in industrial organization. If people who shift easily are also more elastic in labor supply, then giving them a lower tax rate is a good thing. To investigate our analytical results numerically, we provide numerical illustrations, and find circumstances under which welfare gains may be achieved thanks to well-designed income shifting possibilities.

The rest of the article is organized as follows. Section 2 presents the related literature. Section 3 introduces the main blocks of the model. Section 4 presents the main results in a stepwise fashion. Section 5 examines the robustness of our results in two directions; heterogeneity in the convex costs on the one hand, and non-linear taxation on the other hand. Section 6 concludes the article.

2 RELATED LITERATURE

Our work is closely related to Piketty and Saez (2013) who model the cost of income shifting as a convex cost in a linear income tax setting. Similar models are used in Piketty et al. (2014) and Saez and Stantcheva (2017). Considering heterogeneity in skills only, it is shown that governments

[2] Note that this extra degree of freedom allowing the policy-maker to screen more effectively is not exploited when focusing on pure convex shifting costs.

should stop income shifting if it is costless to do so in the hypothetical situation where all income stems from labor effort. With both labor and capital incomes in the model, the optimal tax rates will depend on the elasticities for labor and capital incomes. However, the presence of shifting opportunities lowers the gap between the optimal tax rates on labor and capital incomes (as compared to the tax rate differential arising under the inverse elasticity rule). The same intuition is present in the work by Hermle and Peichl (2015), who derive optimal tax rules in a model with multiple income tax bases. In their model, agents are heterogeneous with respect to skills, shifting abilities and consumption preferences, and may shift income between the tax bases in exchange for a smooth resource cost. The optimal tax formulae differ from the standard ones: they also include a term for the fiscal externalities generated by the cross-elasticities.

Christiansen and Tuomala (2008) examine the role of income shifting in a two-type two-period model along the lines of Stiglitz (1982). They consider that agents can shift income between the two tax bases at a convex cost, but that the government is unable to observe the true amounts of labor and capital income. With heterogeneity in the skill dimension and additively separable preferences, a positive proportional capital income tax is desirable.[3]

Our extensive margin model, where individuals endogenously sort to different tax schedules, relates to a growing body of literature on occupational choices. In this context, Rothschild and Scheuer (2013) consider a model in which all agents face a unique nonlinear tax schedule, whilst Gomes et al. (2017) allow for sector-specific tax schedules. In a related framework, Doligalski and Rojas (2016) analyze the optimal size of the informal economy while considering a model with one sector with taxes and one without taxes. Their model can be seen as a sub-case of our analysis, in which tax differentiation is allowed but constrained to be constant and equal to zero in the second sector. More specifically, the analysis developed in the present article connects to the literature on entrepreneurial income taxation (Parker, 1999, and Scheuer, 2014). Our focus is however different. While the occupational choice literature highlights general equilibrium effects on wages and individual productivity differences in different sectors, we focus on heterogeneity in elasticities and potential welfare gains from sorting into separate tax schedules.[4]

An important contribution of our paper is to bridge two strands of the public finance literature – the tax avoidance literature, on the one hand, and the occupational choice literature, on the other hand. By making a simple change to the tax avoidance technology – by exchanging a variable with a fixed cost, – we enter an interesting setting where individuals sort into different tax schedules.

[3]In the atemporal two-type model of Fuest and Huber (2001), there is a also a convex shifting cost, but agents instead differ with respect to their wealth endowments, and the government imposes non-linear income tax schedules for labor and capital incomes. In the social optimum, wealthy households face the same positive marginal tax rate both for labor and capital incomes. Poor households, on the other hand, face a larger marginal tax rate for capital income than for labor income.

[4]In the context of insurance markets, it has been highlighted by Bond and Crocker (1991) that endogenous categorization, based on consumption of products that are statistically correlated with loss propensities, may relax self-selection constraints. This mechanism relates to the mechanism we are examining, but our setting is very different.

This connection has previously not been emphasized, even though it may be of great importance for tax policy design.

3 A MODEL ALLOWING FOR INCOME SHIFTING

We start by introducing the main blocks of the model that we will specialize in the next sections to focus on the intensive or extensive margin.

3.1 Sources of Heterogeneity in the Population

We consider a population of individuals who are heterogeneous in three dimensions: skills ω, taste for work effort ϵ, [5] and the fixed cost γ incurred when shifting incomes to an alternative tax base (if any).[6] The distribution of ω, ϵ and γ is given by the joint probability density function $f(\omega, \epsilon, \gamma)$ with support included in $\mathbb{R}^3_+$. The policy-maker knows the distribution of types within the population, but is neither able to observe nor recover the type of a specific individual, precluding personalized lump-sum taxes. Without any loss of generality, the size of the population is normalized to 1.

In general, we do not make any restriction on the possible correlations between these three parameters, but we later on pay special attention to a few specific cases. In addition, we define f_i and F_i as the marginal and cumulative density functions of $i = \{\omega, \epsilon, \gamma\}$. We also refer to $F_{\gamma|\kappa}$ as the cumulative density function of γ conditional on $\kappa \equiv (\omega, \epsilon)$.

In this context, we investigate the situation in which a benevolent policy-maker would like to redistribute income within its population. Two tax instruments are available: a tax function T_1 for non-shifted earnings and a tax function T_2 for shifted earnings. In Section 5, we consider the more general situation in which the tax function jointly depends on non-shifted and shifted earnings.

3.2 Shifting Costs

We refer to A as the amount of shifted income. As emphasized in the introduction, shifting earnings involves a cost which might be fixed and/or variable. In our baseline specification, we introduce heterogeneity in the shifting cost through the fixed cost component. We then consider that the cost of shifting an amount A of earnings is equal to:

$$\Gamma(A; \gamma) = C(A) + \gamma \cdot \mathbb{1}_{A>0}, \tag{1}$$

[5] In important specific cases, emphasized below, this parameter corresponds to the labor supply elasticity.

[6] There are strong reasons to believe that shifting costs vary conditional on the earnings ability and preference for work: In some professions, such as lawyers and medical doctors, it is fairly easy to switch from being an employee receiving third-party reported income, whereas in other professions, like teachers, it is probably more costly.

where $\mathbb{1}_{A>0}$ is the indicator function equal to 1 when $A > 0$ and 0 otherwise. Most of the previous literature has focused on the case where $\Gamma(A;\gamma)$ reduces to $C(A)$. We by contrast investigate the implications of a more general – and more empirically relevant– cost structure.

Under this specification, the parameter γ stands for the fixed costs of shifting. The variable cost $C(A)$ is smooth, satisfying $C(0) = 0$, $C'(0) = 0$, $C'(A) \geq 0$, and $C''(A) \geq 0$. It is important to note that it is identical across taxpayers and only depends on shifted income A. An obvious but important implication is that it is independent of non-shifted earnings $wL - A$. There are indeed some evidence that tax avoidance is concentrated on "marginal" incomes (Kleven et al., 2011), and rather on self-reported incomes than on incomes declared by third party such as salary incomes. This would imply that a given amount of income is easier to shift when the amount of non-shifted income becomes larger. In that case, the cost function should include unshifted income $wL - A$ as an additional argument. To highlight the main mechanisms at stake, we below abstract from such a consideration. As an extension, we will in Section 5.1 comment on the case when the variable cost C is heterogeneous within the population.

3.3 Individual Choices

To model individual choices, we use the canonical labor-leisure model, that we augment with a possibility of income shifting. We denote individual consumption (or net income) by Y and labor supplied by L. We allow the disutility of effort to depend on ϵ. More precisely, an individual of skill ω supplying L units of effort receives gross income ωL but incurs a utility loss $v(L;\epsilon)$, with $v'_L > 0$ and $v''_{LL} > 0$. The individual utility function is:

$$U(Y, L) = Y - v(L;\epsilon). \tag{2}$$

Given this specification, there is no income effect on effort L, which only depends on the relevant net-of-tax wage rate. This is in accordance with empirical findings which suggest such income effects to be modest in magnitude (Saez et al., 2012). As previoulsy noted, every individual has the possibility to reduce income that is subject to the first tax instrument T_1, from ωL to $\omega L - A$, at a cost $\Gamma(A;\gamma)$. Overall, an individual thus pays taxes equal to $T_1(\omega L - A) + T_2(A)$ and receives net income:

$$Y = \omega L - T_1(\omega L - A) - T_2(A) - \Gamma(A, \gamma). \tag{3}$$

It should be noted that, because the utility function (2) is quasilinear in net income, we can alternatively interpret the shifting cost $\Gamma(A, \gamma)$ as a loss in utility.

Individual choices proceed from the maximization of the utility function $U(Y, L)$ subject to the budget constraint (3). The indirect utility is therefore defined as:

$$V(\omega,\epsilon,\gamma)=\max_{L\geq 0,A\leq\omega L}\{\omega L-T_1(\omega L-A)-T_2(A)-\Gamma(A,\gamma)-v(L;\epsilon)\}. \tag{4}$$

We refer to $L(\omega,\epsilon,\gamma)$ as the optimal supply of effort and $A(\omega,\epsilon,\gamma)$ as the optimal amount of shifting for an individual of type (ω,ϵ,γ). For later use, we also define:

$$V_1(\omega,\epsilon) = \max_{L\geq 0}\{\omega L-T_1(\omega L)-T_2(0)-v(L;\epsilon)\}, \tag{5}$$

$$V_2(\omega,\epsilon,\gamma) = \max_{L\geq 0}\{\omega L-T_1(0)-T_2(\omega L)-\Gamma(\omega L;\gamma)-v(L;\epsilon)\}. \tag{6}$$

For any given individual, (5) provides the maximum utility $V_1(\omega,\epsilon)$ which can be obtained in the absence of any income shifting. We denote the level of L that maximizes $V_1(\omega,\epsilon)$ by $L_1(\omega,\epsilon)$. (6) instead provides the maximum utility $V_2(\omega,\epsilon,\gamma)$ when the entire earnings are shifted. We denote the level of L that maximizes $V_2(\omega,\epsilon,\gamma)$ by $L_2(\omega,\epsilon,\gamma)$.

3.4 Policy-Maker's Choices

The policy-maker chooses two tax functions, compatible with individuals' incentives, which maximize the social objective function:

$$\iiint g(\omega,\epsilon,\gamma)V(\omega,\epsilon,\gamma)f(\omega,\epsilon,\gamma)d\gamma d\epsilon d\omega, \tag{7}$$

subject to the following revenue constraint:

$$R\leq\iiint[T_1(\omega L(\omega,\epsilon,\gamma)-A(\omega,\epsilon,\gamma))+T_2(A(\omega,\epsilon,\gamma))]f(\omega,\epsilon,\gamma)d\gamma d\epsilon d\omega. \tag{8}$$

R is a tax revenue requirement that does not enter the individuals' utility function. We below focus on purely redistributive tax policies where $R=0$. Regarding social preferences, we should stress the fact that there is a vast literature about how to aggregate individual utilities in the presence of multidimensional heterogeneity. We herein chose to be rather agnostic about this important issue and consider an approach compatible with different views: We simply consider that some unspecified welfare weights $g(\omega,\epsilon,\gamma)\geq 0$ are assigned to various agents conditional on their characteristics, including the preference parameter ϵ. A sub-case is the situation in which those weights only depend on ω and γ, or just on ω. The analytical results derived in the rest of the article are robust to these various specifications.

4 MAIN RESULTS

In this Section, we investigate whether income shifting may be socially desirable when this operation only involves a fixed cost. We thus let the convex cost component $C(\cdot)$ be zero irrespective of how much earnings A are shifted. In other words, $\Gamma(A;\gamma) = \gamma \cdot \mathbb{1}_{A>0}$. To highlight the main mechanisms at stake, our analysis proceeds in steps: (i) we first assume that the fixed cost is uniform and that agents solely differ with respect to skills; (ii) we then introduce unrestricted heterogeneity in both the fixed costs and elasticities. In the next Section, we consider several extensions to test the robustness of our results.

For the sake of simplicity and clarity, we follow Piketty and Saez (2013) and let T_1 be a linear tax and T_2 is a proportional tax. More precisely, $T_1 = -T_1(0) + \tau_1 \cdot (\omega L - A)$ while $T_2 = \tau_2 A$. We thus have $T_2(0) = 0$. Note however that every agent receives the demogrant $-T_1(0)$ if any, even if she decides to shift her entire earnings. An extension addressed in the next Section is to allow both tax functions to be fully non-linear.

4.1 Partition of the Population

When income shifting is associated with a fixed cost and total taxes amount to $-T_1(0) + \tau_1 \cdot (\omega L - A) + \tau_2 A$, a rational individual either shifts nothing ($A = 0$) or her entire labor earnings ($A = \omega L$). Given the definitions of Section 3.3, we see that $A = 0$ in the individual optimum if and only if $V_1 \geq V_2$, i.e.,

$$(1-\tau_1)\omega L_1 + T_1(0) - v(L_1;\epsilon) \geq (1-\tau_2)\omega L_2 + T_1(0) - \gamma - v(L_2;\epsilon), \tag{9}$$

where L_1 and L_2 are endogenously determined as solutions to Problems (5) and (6) respectively. Rearranging (9), we obtain:

$$\gamma \geq [(1-\tau_2)\omega L_2 - (1-\tau_1)\omega L_1] + [v(L_1;\epsilon) - v(L_2;\epsilon)]. \tag{10}$$

This inequality implies that, at a given $\kappa \equiv (\omega, \epsilon)$, the population is divided into two fractions: shifters and non-shifters.

4.2 Uniform Fixed Cost and Heterogeneity in Skills

Most of the articles on optimal income taxation consider that agents only differ with respect to skill levels. We start by addressing the issue of income shifting under this circumstance. This will provide us with a benchmark when we below, in a second variation of our model, introduce unrestricted heterogeneity.

When ϵ and γ are constants, equation (10) implies a cut-off with respect to ω, which partitions the population into two sets. This cut-off level is denoted by $\hat{\omega}$ and equals:

$$\hat{\omega} = \frac{\gamma - [v(L_1;\epsilon) - v(L_2;\epsilon)]}{(1-\tau_2)L_2 - (1-\tau_1)L_1}. \tag{11}$$

The latter is unique for given values of τ_1 and τ_2.[7] Accordingly, there are $\int_0^{\hat{\omega}} f(\omega)d\omega$ non-shifters and $\int_{\hat{\omega}}^{\infty} f(\omega)d\omega$ shifters. Applying the implicit function theorem to (11), it is easy to see that $\partial\hat{\omega}/\partial\tau_1 < 0$ and $\partial\hat{\omega}/\partial\tau_2 > 0$: a higher tax rate τ_1 pushes more agents to shift whilst a larger τ_2 naturally has the opposite effect. The following Proposition provides guidance about the marginal tax rates τ_1 and τ_2. The proof is a direct adaptation of the more general Problem (7)-(8) solved in Appendix A, in which we sum over two distinct uni-dimensional intervals. We define the social welfare weight, expressed in budgetary units, as $b(\omega)$; it is equal to $g(\omega)$ divided by the Lagrange multiplier of the tax revenue constraint (8).

Proposition 1. *Suppose there is heterogeneity in skills only, i.e. both ϵ and γ are constant across agents. In this case, every individual with $\omega > \hat{\omega}$ shift their entire earnings. The optimal marginal tax rate for agents with skill below $\hat{\omega}$ is:*

$$\frac{\tau_1}{1-\tau_1} = \frac{\int_0^{\hat{\omega}}[1-b(\omega)]\omega L_1 f(\omega)d\omega}{\int_0^{\hat{\omega}} \omega L_1 e_1(\omega) f(\omega)d\omega} + \frac{\frac{\partial\hat{\omega}}{\partial\tau_1}\Delta T(\hat{\omega})f(\hat{\omega})}{\int_0^{\hat{\omega}} \omega L_1 e_1(\omega) f(\omega)d\omega}, \tag{12}$$

and the optimal marginal tax rate for more skilled agents is:

$$\frac{\tau_2}{1-\tau_2} = \frac{\int_{\hat{\omega}}^{\infty}[1-b(\omega)]\omega L_2 f(\omega)d\omega}{\int_{\hat{\omega}}^{\infty} \omega L_2 e_2(\omega) f(\omega)d\omega} + \frac{\frac{\partial\hat{\omega}}{\partial\tau_2}\Delta T(\hat{\omega})f(\hat{\omega})}{\int_{\hat{\omega}}^{\infty} \omega L_2 e_2(\omega) f(\omega)d\omega}, \tag{13}$$

where $\Delta T \equiv \tau_2\omega L_2 - \tau_1\omega L_1$.

Agents with $\omega \leq \hat{\omega}$ pay taxes based on the tax rate τ_1 defined in (12) whilst more skilled agents engage in shifting for their entire earnings, thus paying taxes at the rate τ_2 defined in (13). When individuals only differ with respect to skills, all individuals at a certain income level respond to taxes in the same way, so that there is no place for horizontal differentiation. Still, the government may improve social welfare by introducing a second tax bracket, with a lower marginal tax rate. Because $\tau_1 > \tau_2$, the two linear brackets generate a non-convex budget constraint, analogous to that created by two-bracket income tax systems. Previous research has shown that two-bracket systems with a non-convex kink are sometimes welfare improving, as emphasized by Apps et al. (2014) and Slemrod et al. (1994).[8]

[7] Note that $\hat{\omega}$ might be infinity in which case no one would shift earnings.

[8] Note the similarity between our equations (12) and (13) and equations (43) and (44) of Apps et al. (2014), who examine

4.3 Unrestricted Heterogeneity in Fixed Costs and Elasticities

We now cast light on the consequences of allowing for unrestricted heterogeneity in both the fixed costs and elasticities.

Partition of the Population

In contrast with the previous Subsection, we obtain a more complex participation of the population into non-shifters and shifters. This follows from Inequality (10) and is summarized in the following Lemma:

Lemma 1. *Assume* $\Gamma(A;\gamma) = \gamma \cdot \mathbb{1}_{A>0}$*,* $T_1(\omega L - A) = -T_1(0) + \tau_1(\omega L - A)$ *and* $T_2(A) = \tau_2 A$*. For each value of* κ*, define* $\hat{\gamma}(\kappa)$ *as the solution in* γ *to (10) written with equality instead of* $\geq$*.*[9] *Then:*

- *for* $\gamma < \hat{\gamma}(\kappa)$*, the net-of-tax wage rate is* $\omega(1-\tau_2)$*,* $L = L_2(\kappa)$ *and* $A(\kappa,\gamma) = \omega L_2(\kappa)$*;*
- *for* $\gamma \geq \hat{\gamma}(\kappa)$*, the net-of-tax wage rate is* $\omega(1-\tau_1)$*,* $L = L_1(\kappa)$ *and* $A(\kappa,\gamma) = 0$*.*

Moreover, at each κ *such that* $\hat{\gamma}(\kappa) > 0$*,* $\frac{\partial \hat{\gamma}(\kappa)}{\partial \tau_1} = \omega L_1 > 0$ *and* $\frac{\partial \hat{\gamma}(\kappa)}{\partial \tau_2} = -\omega L_2 < 0$*.*

Because the fixed shifting cost enters the individual optimization problem in an additively separable way, it only determines whether or not an agent decides to shift. At a given κ, agents with $\gamma < \hat{\gamma}(\kappa)$ shift their entire earnings. The marginal work incentive is driven by the marginal tax rate on shifted income τ_2. On the contrary, agents with $\gamma \geq \hat{\gamma}(\kappa)$ do not shift any earnings.[10] For each of them the marginal work incentive is entirely driven by τ_1. This partition of the population plays a key role. It implies that, at a given income level, there may be both shifters and non-shifters. Consequently, some agents with the *same* income determine how much effort to supply based on *different* tax rates. This mechanism clearly differs from what is allowed by the introduction of additional tax brackets; in that case, only agents with different incomes may face different tax rates.

Optimal Tax Rates

In order to derive the optimal tax rules, we use a small tax reform perturbation around the optimum. More precisely, we investigate the effects of increasing τ_1, or alternatively τ_2, by a small quantity $\partial\tau > 0$, everything else being equal. Because we contemplate all possible effects starting from a socially optimal allocation, they should add up to zero. We start by considering an increase in the marginal tax rate τ_1 on non-shifted income. This tax variation has the following effects:

---

the optimal two-bracket structure.

[9] If this solution is negative, we set $\hat{\gamma}(\kappa) = 0$.

[10] We make the tie breaking assumption that the κ-agents for whom $\gamma = \hat{\gamma}(\kappa)$ belong to the set of non-shifters. This assumption has no impact in terms of optimal policy, because the set of indifferent agents has measure zero.

- *Net mechanical effect in the non-shifting population:* The rise $\partial\tau$ in τ_1 mechanically increases taxes collected from each agent in the non-shifting population, by an amount $E_1^+ = \omega L_1 \,\partial\tau$. However, given preferences that are quasi-linear in net income, it also reduces each agent's utility by $\omega L_1 \,\partial\tau$, and thus social welfare by $E_1^- = g(\kappa,\gamma)\omega L_1 \,\partial\tau$. Dividing the latter by λ, which refers to the Lagrange multiplier of social government's budget constraint, we obtain the effect on social welfare expressed in units of government revenue: $b(\kappa,\gamma)\omega L_1 \,\partial\tau$ with $b(\kappa,\gamma) = g(\kappa,\gamma)/\lambda$. The net mechanical effect corresponds to the difference between E_1^+ and E_1^-, i.e., $(1 - b(\kappa,\gamma))\,\omega L_1 \,\partial\tau$. Integrating over the set of non-shifters, we obtain:

$$E_1 = \int_\omega \int_\epsilon \int_{\hat{\gamma}(\kappa)}^{\infty} (1 - b(\kappa,\gamma))\,\omega L_1 \,\partial\tau\, f(\kappa,\gamma) d\gamma d\kappa. \tag{14}$$

- *Substitution effect in the non-shifting population:* The increase $\partial\tau$ in τ_1 reduces the net-of-tax wage rates in the non-shifting population. This induces each of them to reduce effort L_1, and thus gross income ωL_1, by an amount:

$$-\frac{\omega L_1 \cdot e_1(\kappa)}{1-\tau_1}\partial\tau, \tag{15}$$

where $e_1(\kappa)$ stands for the labor supply elasticity within the set of non-shifters. As a result, taxes collected from this agent diminish by $\tau_1 \times$ (15). Integrating over the non-shifting population, we obtain:

$$E_2 = -\int_\omega \int_\epsilon \int_{\hat{\gamma}(\kappa)}^{\infty} \frac{\tau_1}{1-\tau_1}\omega L_1 \, e_1(\kappa)\,\partial\tau\, f(\kappa,\gamma)\, d\gamma d\kappa. \tag{16}$$

- *Shifting responses:* At each κ, because of the increase $\partial\tau$ in τ_1, the agents are willing to pay a higher shifting cost; therefore, the cut-off value $\hat{\gamma}(\kappa)$ goes up by $(\partial\hat{\gamma}(\kappa)/\partial\tau_1) \times \partial\tau$. This induces $(\partial\hat{\gamma}(\kappa)/\partial\tau_1) \times \partial\tau \times f(\kappa,\hat{\gamma}(\kappa))$ agents to move from the non-shifting to the shifting population. For each of them, the variation in collected taxes amounts to:

$$\Delta T \equiv \tau_2 \omega L_2 - \tau_1 \omega L_1. \tag{17}$$

This quantity can either be positive or negative, depending on how elastic labor supply is. Integrating over κ, the overall change in collected taxes due to the extensive responses amounts to:

$$E_3 = \int_\omega \int_\epsilon \Delta T \,\frac{\partial\hat{\gamma}(\kappa)}{\partial\tau_1}\,\partial\tau\, f(\omega,\hat{\gamma}(\kappa)) d\kappa = \int_\omega \int_\epsilon \Delta T\, \omega L_1 \,\partial\tau f(\omega,\hat{\gamma}(\kappa)) d\kappa, \tag{18}$$

where $\frac{\partial\hat{\gamma}(\kappa)}{\partial\tau_1} = \omega L_1$ follows from Lemma 1.

A small tax reform perturbation around the social optimum has no first-order effect. Therefore, $E_1 + E_2 + E_3 = 0$. Rearranging, we obtain:

$$\frac{\tau_1}{1-\tau_1} = \frac{\int_\omega \int_\epsilon \int_{\hat{\gamma}(\kappa)}^{\infty} [1 - b(\kappa,\gamma)] \omega L_1 f(\kappa,\gamma) d\gamma d\kappa}{\int_\omega \int_\epsilon \int_{\hat{\gamma}(\kappa)}^{\infty} \omega L_1 e_1(\kappa) f(\kappa,\gamma) d\gamma d\kappa} + \frac{\int_\omega \int_\epsilon \omega L_1 \Delta T(\kappa, \hat{\gamma}(\kappa)) f(\kappa, \hat{\gamma}(\kappa)) d\kappa}{\int_\omega \int_\epsilon \int_{\hat{\gamma}(\kappa)}^{\infty} \omega L_1 e_1(\kappa) f(\kappa,\gamma) d\gamma d\kappa}. \tag{19}$$

We now consider an increase $\partial\tau$ in the optimal marginal tax rate τ_2 on shifted earnings, everything else being equal. This tax reform also has three effects.

- In the population of shifters, it gives rise to a (net) mechanical effect and to a substitution effect. These effects are given by E_1 and E_2, with τ_1 replaced by τ_2, L_1 replaced by L_2, e_1 replaced by the labor supply elasticity e_2 of shifters, and the sum $\int_{\hat{\gamma}(\kappa)}^{\infty}$ replaced by $\int_0^{\hat{\gamma}(\kappa)}$.
- The third effect is the extensive response. At each κ, the increase $\partial\tau$ in τ_2 induces people to leave the shifting population and become non-shifters. By Lemma 1, we know that $\hat{\gamma}(\kappa)$ goes down by ωL_2. All these agents will pay taxes $\tau_1 \omega L_1$ instead of $\tau_2 \omega L_2$, i.e., pay $-\Delta T$ extra in taxes. The net effect on collected taxes is therefore given by:

$$-\int_\omega \int_\epsilon \Delta T \, \omega L_2 \, \partial\tau \, f(\kappa, \hat{\gamma}(\kappa)) d\kappa. \tag{20}$$

Because a tax reform around the social optimum has no first-order effect, the sum of the three effects is equal to zero. Rearranging, we obtain:

$$\frac{\tau_2}{1-\tau_2} = \frac{\int_\omega \int_\epsilon \int_0^{\hat{\gamma}(\kappa)} [1 - b(\kappa,\gamma)] \omega L_2 f(\kappa,\gamma) d\gamma d\kappa}{\int_\omega \int_\epsilon \int_0^{\hat{\gamma}(\kappa)} \omega L_2 e_2(\kappa) f(\kappa,\gamma) d\gamma d\kappa} - \frac{\int_\omega \int_\epsilon \omega L_2 \Delta T(\kappa, \hat{\gamma}(\kappa)) f(\kappa, \hat{\gamma}(\kappa)) d\kappa}{\int_\omega \int_\epsilon \int_0^{\hat{\gamma}(\kappa)} \omega L_2 e_2(\kappa) f(\kappa,\gamma) d\gamma d\kappa}. \tag{21}$$

These results are summarized in the following Proposition, a formal proof of which is provided in Appendix A:

Proposition 2. *Assume $\Gamma(A;\gamma) = \gamma \cdot \mathbb{1}_{A>0}$, $T_1 = -T_1(0) + \tau_1(\omega L - A)$ and $T_2 = \tau_2 A$. The optimal marginal tax rates τ_1 and τ_2 satisfy Equations* (19) *and* (21) *respectively.*

Comments and Discussion of the Role of Heterogeneity

We first see that, in the social optimum, the marginal tax rates τ_1 and τ_2 typically differ.[11] As shown by Equations (19) and (21), a first driving force –captured by the first term on the left-hand side of each formula– is the trade-off between equity concerns (in the numerator) and efficiency (in the denominator). These first terms "look like" the usual optimal linear income tax formula

[11] However, we cannot rule out situations in which there would be no shifting in the optimum. In that case, the cut-off level $\hat{\gamma}(\kappa)$ tends to 0 and the formulae of Proposition 2 collapse into the "usual" optimal income tax rules, with $\tau_1 = \tau_2$.

(cf., e.g., Atkinson and Stiglitz, 1980). However, they are computed as if the total population was restricted respectively to non-shifters, on the one hand, and to shifters, on the other hand. These two sub-populations are of course endogenous to the tax schedule. However, once agents have made their choices, the policy-maker observes, for each agent, whether she belongs to the set of shifters or of non-shifters. In this sense, we may speak of "endogenous" tagging. The second terms on the right-hand side of Equations (19) and (21) are new. They capture extensive margin shifting responses, and their signs depend on the labor supply elasticities of those who are just indifferent between shifting and not shifting. Intuitively, if individuals whom society cares a lot about, and/or whom are more elastic in terms of supplied effort, sort into the tax base for shifted income, it may be optimal for the social planner to differentiate the two tax rates.

The theoretical analysis suggests that individual *heterogeneity* in labor supply elasticities is a key driving force of the optimal taxation policy. It would be important in particular to study whether agents within a given occupation differ depending on their tax status. This point has up to date been addressed only in a study on US physicians (Showalter and Thurston, 1997), which reports that labor supply elasticities are much larger for self-employed physicians than for physicians who are employees. Complementary empirical studies would therefore be of great relevance to provide more general guidance in terms of tax design. It should be pointed out however that such empirical studies are difficult because administrative data typically include information on taxable incomes (that capture both real and avoidance responses), but not on hours of work.

The point we make also has relevance for the discussion on the so-called taxable income elasticity (see Saez et al. (2012) for a survey). It has earlier been highlighted that it is important to distinguish between real and income shifting responses to tax cuts, because the latter can to some extent be controlled by the government. A new insight from Propostion 2 is that the anatomy of the *real response* also matters for optimal tax policy: If some groups of taxpayers are very sensitive to taxes, a lower tax rate on the alternative base may be optimal. Formally, we only modeled labor supply responses. However, "real responses" can readily be given a broader interpretation. Suppose that an individual decides not to migrate if she can get the lower tax τ_2 instead of τ_1.[12] This has analogous implications for the tax base, because if the individual moves to another country no taxes will be collected from tax individual.

An Example with Heterogeneity in Fixed Cost but Uniform Elasticities

To highlight the importance of preference heterogeneity, we now shut down one of the sources of heterogeneity, and assume uniform elasticities of labor supply within the population. This is the

[12] See Simula and Trannoy (2010) and Lehmann et al. (2014) for more specialized treatments of the optimal tax problem in the presence of tax-induced migration.

case when individual preferences are described as follows:

$$U(Y,L) = Y - \alpha \frac{L^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}}, \tag{22}$$

where α is a positive scaling parameter. Under this specification, the disutility of effort is iso-elastic, and the elasticity of labor supply e is constant, equal to ϵ. We further consider that the social objective is the Rawlsian maximin: the social weight function g, and thus the weights b, are thus equal to zero except for the lowest skilled agents. In that case, income shifting is actually socially undesirable as established in the next Corollary.[13]

Corollary 1. *Let the social objective be the Rawlsian maximin and individual preferences be described by* (22)*, with ϵ constant across agents. Then, income shifting is not desirable.*

Proof. In this situation, $e_1 = e_2 = \epsilon$. Hence, Equations (19) and (21) reduce to:

$$\frac{\tau_1}{1-\tau_1} = \frac{1}{e} + \frac{\int_\omega \omega L_1 \Delta T(\omega) f(\omega, \hat{\gamma}(\omega)) d\omega}{e \int_\omega \int_{\hat{\gamma}(\omega)}^{\infty} \omega L_1 f(\omega,\gamma) d\gamma d\omega} \tag{23}$$

$$\frac{\tau_2}{1-\tau_2} = \frac{1}{e} - \frac{\int_\omega \omega L_2 \Delta T(\omega) f(\omega, \hat{\gamma}(\omega)) d\omega}{e \int_\omega \int_0^{\hat{\gamma}(\omega)} \omega L_2 f(\omega,\gamma) d\gamma d\omega}. \tag{24}$$

In addition, $\Delta T(\omega) = \omega^{1+e}[\tau_2(1-\tau_2)^e - \tau_1(1-\tau_1)^e]$, where $\tau(1-\tau)^e$ is increasing if and only if $\tau < \frac{1}{1+e}$. Because τ_1 and τ_2 must both be below the top of the Laffer curve, $\bar{\tau} = \frac{1}{1+e}$, $\Delta T(\omega)$ is negative if and only if $\tau_1 > \tau_2$. Now assume there is some income shifting in the optimum, i.e., $\tau_1 > \tau_2$. As a result, $\Delta T(\omega) < 0$. By (23) and (24), the former implies $\tau_1 < \tau_2$. A contradiction. $\square$

Intuitively, when the government assigns the same social welfare weight (of zero) to all individuals in both sub-populations, the only rationale for differentiating taxes is heterogeneity in elasticities. If not, tax rates should be equalized. In the numerical simulations below, this insight will play an important role.

4.4 Numerical Illustration

We now provide an numerical example in which income shifting is either socially desirable or undesirable depending on the values of the parameters. Our aim is to illustrate how the above optima tax rules may apply in practice and perform sensitivity analysis. We consider a "dual income tax system" in which labor income can either be taxed directly, at a rate τ_1, or transformed into "capital income" at a rate τ_2. To highlight the mechanisms emphasized above, we maintain

[13]We are grateful to an anonymous referee, who suggested this corollary.

the assumption to all earnings originate from labor effort. Regarding individual preferences, we let $U(Y, L)$ be defined by (22), which implies that the elasticity of labor supply is equal to ϵ.

Parameterization

We first need to calibrate the joint distribution of skills and shifting costs. It is well-known that the empirical distribution of hourly wage rates is well approximated by a log-normal distribution, if one abstracts from the top of the distribution. There is considerably less guidance on how to calibrate the distribution of shifting costs. Because we are interested in sensitivity analyses with respect to how ω and γ correlate, it is convenient to assume that these parameters follow a bivariate log-normal distribution. When ω and γ are positively correlated, highly productive agents face a large fixed cost of shifting. We thus expect a low amount of shifting, unless the marginal tax rate τ_2 is sufficiently low compared to τ_1. Empirically, ω and γ are probably negatively correlated, but positive values cannot be ruled out; we thus also provide simulation results for $\rho \geq 0$.

We use Swedish data to calibrate the mean and variance of the wage distribution. The shifting costs are parameterized so that, for the actual average values of T_1 and T_2 in Sweden, the proportion of people deciding to shift incomes roughly reproduces the figure observed in this country as document by Alstadsæter and Jacob (2016).[14]

At each (ω, γ), we consider that there is a normal distribution of ϵ (defined over $\mathbb{R}_+$), with mean $\bar{\epsilon}$ and standard deviation σ_ϵ, which we truncate over 0 and 2. We below explore situations in which σ_ϵ is a fraction of $\bar{\epsilon}$.

Regarding the social objective, we focus on the specific situation in which the policy-maker wants to maximize the well-being of the worst-off agents within the population. In our dataset, the latter are those who do not earn any labor earnings.[15] This corresponds to the "Rawlsian objective", which is extensively used as a benchmark in the optimum income tax literature. Here, the maximin objective is equivalent to maximizing tax revenues, and redistributing the collected amount in a lump-sum manner. In addition, shifting has no direct positive utility effect, through the increased net income of the shifters. We thus deliberately adopt the social perspective that is least favorable to the occurrence of income shifting.

Results

Table 1 shows simulation results for $\bar{\epsilon} = .2$, a reasonable value for Sweden and many developed countries (Saez et al., 2012). The standard deviation σ_ϵ is set equal to $\bar{\epsilon}/2$. Given these values, the bivariable lognormal distribution is not well-defined for ρ below $-.33$, because the correlation

[14] Cf. Appendix B for a more detailed discussion.

[15] Note that heterogeneity with respect to γ is thus irrelevant when considering the well-being of agents without labor earnings.

Table 1: Simulation results for $\bar{e} = .2$ and $\sigma_\epsilon = \bar{e}/2$

ρ	$-.33$	$-.2$	$-.1$	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
τ_1	.806	.794	0.798	.798	.804	.792	.793	.784	.791	.792	0.792	.792	.786
τ_2	.743	.741	.737	.732	.726	.724	.716	.715	.700	.694	.694	.689	.686
$\tau_1 - \tau_2$	.063	.053	.061	.066	.078	.068	.077	.069	.092	.097	.102	.100	.112
% shifters	9.1	6.9	6.6	5.8	6.0	4.4	4.0	2.2	2.7	2.2	1.9	1.7	1.1

Table 2: Simulation results for $\bar{e} = .2$ and various values of σ_ϵ

σ_ϵ	$\bar{e}/2$	$\bar{e}/3$	$\bar{e}/4$	$\bar{e}/6$	$\bar{e}/8$	$\bar{e}/10$
τ_1	.794	.817	.820	.827	.829	.830
τ_2	.741	.775	.796	.810	.818	.830
$\tau_1 - \tau_2$	.053	.042	.024	.017	.011	.000
% shifters	6.9	4.6	2.1	1.5	.8	.000

matrix is then no longer semi-definite positive as required in the definition of a bivariate log-normal distribution. The reader should keep in mind that the empirically most relevant values of ρ are between -0.33 and 0. We see from the Table that the share of shifters tends to decline when ρ increases. Because $\hat{\gamma}(\omega, \epsilon)$, as defined in (10), is increasing in ω, it is intuitive that the fraction of shifters is the largest when ρ is negative and highly skilled individuals on average face low shifting costs. By contrast, the population of shifters is small when ρ is a large positive number. However, when few agents shift, the average elasticity among those who sort into the alternative tax base is larger. Therefore, the tax differential $\tau_1 - \tau_2$ is larger when increasing ρ.

Table 2 reports results from another interesting comparative static exercise, in which we vary the variance of the elasticity. We know from Corollary 1 that the tax rate differential should be zero when the social objective is the maximin and there is no heterogeneity in the elasticities. This analytical result is confirmed in Table 2. We see that the tax rate differential is monotonically increasing in the variance. When the variance is as low as $\bar{e}/10$, the two tax rates are equalized, and the optimal tax rate is equal to the Laffer rate of the standard linear income tax model, $\tau^* = \frac{1}{1+\bar{e}}$. This illustrates a key insight of this article: When elasticities differ in the population, the policy-maker may increase social welfare by differentiating taxes, thus allowing for income shifting.

5 ROBUSTNESS CHECKS AND EXTENSIONS

5.1 Unrestricted Heterogeneity with an Intensive Margin

We first extend the results of Section 4 to a more complex shifting cost structure: in addition to the heterogeneous fixed cost and elasticities, individuals face a variable cost of shifting, and we allow for heterogeneity in this variable cost. Formally, the total shifting cost is given by $\Gamma(A;\gamma,\theta) = C(A;\theta) + \gamma \cdot \mathbb{1}_{A>0}$, where $C(A;\theta)$ is a convex cost. θ is a new dimension of heterogeneity with support over $\mathbb{R}_+$. We let $\partial^2 C(A;\theta)/(\partial A \partial\theta) > 0$, with $\lim_{\theta\to\infty}\partial^2 C(A;\theta)/(\partial A\partial\theta) = +\infty$ and $\lim_{\theta\to\infty}\partial^2 C(A;\theta)/(\partial A^2) < +\infty$. This sorting condition implies that the larger θ the more difficult it is to shift earnings at the margin. The first limit assumption ensures that, at a given (ω,ϵ), agents with a sufficiently large θ do not shift their earnings; the second one is mostly technical and allows us to rely on the implicit function theorem to determine the way in which A varies with θ, everything else being equal. Finally, in contrast to the previous section, we now study a more standard setting in which shifting and labor supply are seperately determined. For simplicity, in line with the earlier literature on "intensive margin shifting", e.g. Chetty (2009) and Piketty and Saez (2013), we focus on cases in which the corner solution $A = \omega L$ does not arise. In this setting, every agent of type $(\omega,\epsilon,\gamma,\theta)$ maximizes her utility $U(Y,L)$ subject to the budget constraint

$$Y = \omega L - \tau_1(\omega L - A) + T_1(0) - \tau_2 A - \Gamma(A;\gamma,\theta). \tag{25}$$

Assuming interior solutions, the first-order conditions with respect to L and A are respectively:

$$v'(L;\epsilon) = \omega(1-\tau_1), \tag{26}$$

$$C'(A;\theta) = \tau_1 - \tau_2. \tag{27}$$

The first of these conditions defines labor supply L_1, which only depends on τ_1 and the heterogeneity parameters ϵ and ω. We call $\tilde{A}(\theta)$ the solution in A to (27). We see that it only depends on the heterogeneity parameter θ and the tax diffferential $\tau_1 - \tau_2$. Applying the implicit function theorem, it follows from (27) that:

$$\frac{d\tilde{A}(\theta)}{d\theta} = -\frac{C''_{A\theta}(A;\theta)}{C''_{AA}(A;\theta)} < 0. \tag{28}$$

At a given (ω,ϵ), we thus formally check that $\tilde{A}(\theta)$ decreases with θ. This implies a possible sorting of the agents with the same (ω,ϵ) into two categories: (i) non-shifters ($A = 0$), because of a too large fixed cost γ and/or a too large θ and (ii) interior shifters ($A = \tilde{A}(\theta)$).

The individual decides not to shift if $V_1 \geq V_3$, where V_3 is indirect utility in the state where the individual shifts $\tilde{A}(\theta)$. This condition can also be written as

$$(1-\tau_1)\omega L_1 + T_1(0) - v(L_1;\epsilon) \geq \omega L_1 - \tau_1[\omega L_1 - \tilde{A}(\theta)] - \tau_2\tilde{A}(\theta) + T_1(0) - C[\tilde{A}(\theta);\theta] - \gamma - v(L_1;\epsilon). \tag{29}$$

Note that the labor supply are the same at a given (ω,ϵ) irrespective of whether the agent chooses

$A = 0$ or $A = \tilde{A}(\theta)$. Rearranging (29), we find that agents shift if and only if $\gamma < \hat{\gamma}(\theta)$ with:

$$\hat{\gamma}(\theta) = (\tau_1 - \tau_2)\tilde{A}(\theta) - C(\tilde{A};\theta). \tag{30}$$

Because shifting does not affect labor supply under such circumstances, shifting activities leave aggregate gross income unaffected. In addition, shifting is not only costly from the individual perspective; it also lowers collected tax revenues. Accordingly, tax differentation can only be optimal if the social planner puts a lot of weight on the individuals who actually benefit from the shifting acitivities. In more formal terms, this can be expressed as:

Proposition 3. *Assume $\Gamma(A;\gamma,\theta) = C(A;\theta) + \gamma \cdot \mathbb{1}_{A>0}$, $T_1 = -T_1(0) + \tau_1(\omega L - A)$ and $T_2 = \tau_2 A$. Suppose agents either shift interior (heterogeneous) amounts, or do not shift at all. Then it is optimal for the social planner to allow for shifting if and only if*

$$\int_{\kappa,\theta}\int_0^{\hat{\gamma}(\theta)} b(\kappa,\theta,\gamma)\tilde{A}(\theta)f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta > \int_{\kappa,\theta}\int_0^{\hat{\gamma}(\theta)} \tilde{A}(\theta)f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta \tag{31}$$

i.e. the social valutation of the aggregate consumption gain from shifting exceeds the aggregate mechanical tax revenue loss. Note that $\int_{\kappa,\theta}$ is a shorthand for $\int_\omega \int_\epsilon \int_\theta$.

Proof. See Appendix C. □

Proposition 3 is related to the point made by Kopczuk (2001): When avoidance behavior is heterogeneous in the population, the optimal tax system may include some seemingly wasteful avoidance opportunities, even if they could be eliminated without any cost. However, this mechanism is of limited interest in the context of income shifting, which typically pertains at the top of the income distribution (among people which typically are assigned low social welfare weights).

To further illustrate the logic, we consider optimal policy under two common social welfare objectives. First, pure utilitarianism implies a constant $b(\kappa,\theta,\gamma) = 1$ for everyone. Hence, under pure utilitarianism the inequality in (31) holds with equality, which in turn implies $\tau_1 = \tau_2$, i.e. no shifting. Second, when the policymaker only cares about the distribution of ω, the maximin objective implies that all weight is given to the individuals with the lowest skill ($\omega = 0$). This implies that the left-hand side of (31) equals zero. Accordingly, $\tau_1 > \tau_2$ cannot be an optimal policy.

5.2 Uniform Convex Cost, Heterogeneous Fixed Cost and Non-Linear Taxes

In Section 4 we emphasized that the tax rate differentiation mechanism at stake differed from the introduction of additional income tax brackets. The aim of this subsection is to demonstrate that the optimal tax rules of Proposition 2 indeed are very similar in an environment in which the government can impose non-linear taxes. For technical reasons we now go back to a setting with less dimensions of heterogeneity: we henceforth assume that individuals only differ with respect to skills, ω, and fixed shifting costs, γ.[16] We allow for a convex cost $C(A)$, which is constant for all. Since we impose restrictions on the heterogeneity parameters we do not consider the results in this section to be a full generalizion of the linear tax model.

The taxation principle implies in the present context that any incentive compatible allocation can be decentralized by a tax schedule: $(\omega \times L - A, A) \mapsto T(\omega \times L - A, A)$ that depends jointly on both unshifted and shifted income.[17] The partial derivatives of the tax function with respect to its first and second arguments are respectively denoted $T_1'(\omega L - A, A)$ and $T_2'(\omega L - A, A)$. They are the natural extensions of τ_1 and τ_2 to a nonlinear tax environment.

Individual Incentives and Elasticities

Because of the interaction between the intensive and extensive shifting margins, it is important to explicitly account for the inequality constraint $A \leq \omega L$ in the utility maximization program (4), that we restate for clarity:

$$V(\omega, \gamma) = \max_{L \geq 0, A \leq \omega L} \left\{ \omega L - T(\omega L - A, A) - C(A) - \gamma - v(L) \right\}. \tag{32}$$

With ϑ referring to the Kuhn-Tucker multiplier of the constraint $A \leq \omega L$, the first-order conditions with respect to L and A are respectively:

$$1 - T_1'(\omega L - A, A) - \vartheta = v'(L)/\omega, \tag{33}$$

$$T_1'(\omega L - A, A) - T_2'(\omega L - A, A) - C'(A) + \vartheta \leq 0 \quad (= if\, A > 0), \tag{34}$$

with $\vartheta \geq 0$ ($= 0$ if $A < \omega L$). To denote the effort function, we below use L_1 for someone who does not shift income at all, and L_2 otherwise. Note that L_1, L_2 and A are functions of ω but not of γ. Provided $A > 0$, combining both equations yields:

$$[1 - (T_2'(\omega L - A, A) + C'(A))]\omega = v'(L) \tag{35}$$

The above condition implies that the labor supply of an agent who at least partially shifts her income is driven by the sum of the marginal tax rate on shifted income $T_2'(\omega L - A, A)$ and the

[16]We believe that this assumption could be relaxed without altering the interpretation of our results. However, such an extension would be non-trivial from the technical viewpoint. Regarding multidimensional screening problems, we refer the reader to Jacquet and Lehmann (2016). This article considers optimal tax rules when agents differ both with respect to a vector of characteristics (e.g. individual skills in various occupations) as well as elasticities; however, in contrast to our article, there is a single non-linear tax function.

[17]Note that the joint tax function $T(\omega L - A, A)$ does not need to be continuous at $A = 0$.

marginal shifting cost $C'(A)$. We thus see that the marginal shifting cost $C'(A)$ plays exactly the same role as the marginal tax rate on shifted earnings. When a shifter decides to increase the amount of shifted income A, she does not only have to pay the marginal tax $T_2'(\omega L - A, L)$, but also the marginal shifting cost $C'(A)$.

An agent decides to set $A > 0$ provided $V(\omega, \gamma) > V_1(\omega)$.[18] Solving this inequality for γ, we obtain a cut-off level $\hat{\gamma}(\omega)$, at each productivity level ω, below which $A > 0$ and above which $A = 0$. More precisely, $\hat{\gamma}(\omega)$ is equal to $max\{0, \hat{\gamma}\}$, with:

$$\hat{\gamma} = \omega\left[L_2(\omega) - L_1(\omega)\right] + T\left(\omega L_1(\omega), 0\right) - T\left(\omega L_2(\omega) - A(\omega), A(\omega)\right) \tag{36}$$

$$-C\left(A(\omega)\right) - v\left(L_2(\omega)\right) + v\left(L_1(\omega)\right). \tag{37}$$

Labour supply elasticities play a key role in the analysis. We now generalize the definitions given in the previous Sections to account for the non-linearity of the tax schedules. For an agent who does not shift earnings at all, the labor supply elasticity e_1 is given by:

$$e_1 \equiv \frac{\partial L_1(\omega)}{\partial[\omega(1 - T_1'(\omega L_1(\omega), 0))]} \frac{\omega(1 - T_1'(\omega L_1(\omega), 0))}{L_1(\omega)} = \frac{v'(L_1(\omega))}{v''(L_1(\omega))L_1(\omega)}. \tag{38}$$

The last equality follows from (5) and the implicit function theorem. We will show below that, in the social optimum, all shifters decide to shift their *entire* earnings, i.e. set $A = \omega L_2$. We therefore define the labor supply elasticity e_2 of such shifters as:

$$e_2 \equiv \frac{\partial L_2(\omega)}{\partial\left[\omega\left(1 - T_2'(0, \omega L_2(\omega)) - C'(\omega L_2(\omega))\right)\right]} \frac{\omega\left[1 - T_2'(0, \omega L_2(\omega)) - C'(\omega L_2(\omega)\right]}{L_2(\omega)}$$

$$= \frac{v'(L_2(\omega))}{v''(\omega L_2(\omega))L_2(\omega)}. \tag{39}$$

This definition accounts for the fact that the marginal shifting cost C' exactly plays the same part as the marginal tax rate on shifted earnings T_2', as emphasized above. The last equality follows from (35) and the implicit function theorem. From the above definitions of e_1 and e_2, we see that two agents of same ability ω may have different labor supply elasticities depending on their shifting status.

[18] As already emphasized, it is innocuous –in terms of policy implications– whether we impose a strict or weak inequality.

The Social Planner's Problem

From now on, it is convenient to define the indirect utility of a shifter (with $A > 0$) gross of the fixed cost γ. We call it $V_2 \equiv V + \gamma$. Crucially, due to the quasi-linearity assumption, it is independent of γ.[19] V_1 is the indirect utility of a non-shifter (with $A = 0$). By the taxation principle, the social planner solves:

Problem 1. *Find V_1, V_2, L_1, L_2 and $A \in [0, \omega L_2]$ maximizing:*

$$\int_0^\infty \int_0^{\hat{\gamma}(\omega)} g(\omega,\gamma)[V_2(\omega) - \gamma] f(\omega,\gamma)\, d\gamma d\omega + \int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty g(\omega,\gamma) V_1(\omega) f(\omega,\gamma)\, d\gamma d\omega \tag{40}$$

with $\hat{\gamma}(\omega) = V_2(\omega) - V_1(\omega)$, subject to (i) the budget-constraint:

$$\begin{aligned} &\int_0^\infty \int_0^{\hat{\gamma}(\omega)} [\omega L_2(\omega) - v(L_2(\omega)) - V_2(\omega) - C(A(\omega))] f(\omega,\gamma)\, d\gamma d\omega \\ &\quad + \int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty [\omega L_1(\omega) - v(L_1(\omega)) - V_1(\omega)] f(\omega,\gamma)\, d\gamma d\omega = 0; \end{aligned} \tag{41}$$

(ii) the first-order conditions for incentive-compatibility:

$$\frac{dV_1(\omega)}{d\omega} = \frac{v'(L_1(\omega))}{\omega} L_1(\omega) \text{ for non-shifters;} \tag{42}$$

$$\frac{dV_2(\omega)}{d\omega} = \frac{v'(L_2(\omega))}{\omega} L_2(\omega) \text{ for shifters;} \tag{43}$$

(iii) and the second-order conditions for incentive compatibility, stating that gross earnings are non-decreasing in skills within each set of agents (non-shifters and shifters);

The conditions for incentive compatibility are derived in Appendix D. Below we adopt the so-called "first-order approach" and do not formally account for the monotonicity constraints (iii) when solving for Problem 1.

Social Optimum

We have seen in Section 4.1 that when marginal tax rates are constant and shifting only involves a fixed cost, a rational agent either shifts her entire earnings or nothing. As stressed above, this is not necessarily the case when taxes are nonlinear and shifting involves a convex cost together with a fixed cost. It turns out however that, *in the social optimum*, rational agents behave in the same dichotomic way.

[19] In the linear income tax model of Section 4, the quasi-linearity assumption can be relaxed. However, in the non-linear model the quasi-linearity assumption is more central, because it ensures that both V_1 and V_2 are independent of γ.

Proposition 4. *Assume (i) $\Gamma(A;\gamma) = C(A) + \gamma \cdot \mathbb{1}_{A>0}$, with $C'(A) > 0$, (ii) $T : (\omega \times L - A, A) \mapsto T(\omega \times L - A, A)$ is nonlinear, and (iii) at a given ω, let ΔT be the net difference in taxes $T_2 - T_1$ collected from the marginal shifter with $\gamma = \hat{\gamma}(\omega)$. In the social optimum:*

(a) $A \in \{0, \omega L_2\}$;

(b) for non-shifters $(A = 0)$,

$$\frac{T_1'(\omega L_1(\omega)}{1 - T_1'(\omega L_1(\omega))} = \left[1 + \frac{1}{e_1(\omega)}\right] \times \frac{\int_{x=\omega}^{\infty} \int_{\gamma=\hat{\gamma}(\omega)}^{\infty} [1 - b(x,\gamma)] f(x,\gamma) d\gamma dx + \int_{x=\omega}^{\infty} \Delta T(x) f(x, \hat{\gamma}(x)) dx}{\omega f_\omega(\omega)[1 - F_{\gamma|\omega}(\hat{\gamma}(\omega))]}; \quad (44)$$

(c) for shifters $(A = \omega L_2)$,

$$\frac{T_2'(\omega L_2(\omega))}{1 - T_2'(\omega L_2(\omega)) - C'(\omega L_2(\omega))} = \left[1 + \frac{1}{e_2(\omega)}\right] \times \frac{\int_{x=\omega}^{\infty} \int_{\gamma=0}^{\hat{\gamma}(\omega)} [1 - b(x,\gamma)] f(x,\gamma) d\gamma dx - \int_{x=\omega}^{\infty} \Delta T(x) f(x, \hat{\gamma}(x)) dx}{\omega f_\omega(\omega) F_{\gamma|\omega}(\hat{\gamma}(\omega))}. \quad (45)$$

Proof. See Appendix E. □

Proposition 4 first establishes that, in this model environment in which the convex cost function C only depends on the shifted amount A, the sorting of agents between "pure shifters" and "pure non-shifters"[20], highlighted in Section 4, is robust both to the introduction of a smooth shifting cost and to the relaxation of the linear tax assumptions. This result relies on the fact that $C(A)$ is uniform across agents, implying that the incentive compatibility constraints are independent of the *level* of A – they only depend on whether A is positive or not.[21] Proposition 4, points (b) and (c), may be seen as an extension of Diamond (1998) to an economy in which shifting is possible. We see that the optimal tax rules have the same structure as those derived, in the previous Section, when shifting only involves an extensive margin (cf. Proposition 2). A new element, in the square

[20] It should be noted that, depending on the parameters, there might be no shifters within the population. This will notably be the case if the shifting costs are quite large.

[21] Another way to understand this result is that it generalizes the result by Piketty et al. (2014) to an environment with non-linear taxes: it is socially optimal to design taxes in such a way that no one shifted income along the intensive margin. The reason is that this activity is a pure waste of real resources in pursuit of tax savings. By contrast, allowing for income shifting along the extensive margin induces a fixed cost at the individual level but may boost income creation through tax rate differentiation. Of course, point a of Proposition 4 depends on the assumption that there is no heterogeneity in the convex cost.

bracket on the right-hand side of Equation (45), is the marginal shifting cost. Everything else being equal, a higher marginal shifting plays in favor of a lower marginal tax rate on shifting earnings, T_2'.

6 CONCLUSION

Most previous studies of income shifting based on optimal tax theory led to very strong statements: such ”loopholes”, resulting in pure losses in efficiency, should be avoided. At the same time, more macro-oriented analyses often conclude that capital income taxes should not be too large or even zero. These conclusions in favor of differentiating capital and labor income taxes typically rely on a logic of ”capital accumulation”.

This article shows that, even abstracting from classical arguments in favor of tax rate differentiation, there might be scope for income shifting in a well-designed tax system. In the presence of heterogeneity, allowing for income shifting may be used by the policy-maker as a screening device. Indeed, when people who shift easily are also sufficiently elastic in labor supply, then giving them a lower tax rate is a good thing: it generates extra incomes resulting in larger collected taxes and more room for redistribution to the needy ones.

It thus appears that income shifting and income creation are not necessarily orthogonal to one another. This offers a new perspective on the debate in France highlighted in the introduction. The statement should however be qualified: only a ”well-designed” income shifting may be socially profitable. For this reason, we believe that a key message of our analyses and discussions is that it is crucial to make advances towards less aggregated measures of labor supply elasticities, and in particular estimate labor supply elasticities at different segments of the labor market for shifters and non-shifters. In addition, it would be interesting to pay special attention to potential differences in shifting costs (based on the observed employment status). Contributions like those of Tazhitdinova (2016) are key in this respect. Such empirical knowledge should constitute a key input in the discussion on whether tax reductions are targetted to the right set of agents and, thus, rest on sound bases.

Finally, we want to acknowledge that our model framework – in similarity with all optimal tax models – is very stylized. Any empirical attempt to estimate the quantities defined in this article needs to account for country-specific institutional features. E.g. in real-world dual income tax systems, which can be found in the Nordic countries, it is a central policy concern how to tax owner-managers, who are not subject to third-party reporting and whose incomes have aspects of being both labor and capital incomes. In countries like Finland and Sweden the shifting incentives are affected by an imputed ”dividend allowance”, which determines how much income the owner-manager may tax to the lower capital income tax rate. Our model is policy relevant insofar the fixed costs of setting up a closely held corporation are substantial (only a subset of the population shifts) and the marginal incentives to earn income are larger for the shifters. However, since we abstract from the role of risk-taking, entrepreneurship, and capital income generation, we do not claim to cover all aspects of the actual tax policy problem facing real-world governments. Our contribution is to elucidate the tight link between income shifting and potentially welfare-improving endogenous tagging, a link which up to now has been neglected.

ACKNOWLEDGMENTS

We are grateful to co-editor Johannes Spinnewijn and two anonymous referees for valuable feedback. We also thank Spencer Bastani, Sören Blomquist, Pierre Boyer, Christophe Bravard, Katherine Cuff, Marcus Eliason, Peter Egger, Alex Gelber, Aart Gerritsen, Roger Gordon, Thomas A. Gresik, Laurence Jacquet, Kai Konrad, Wojciech Kopczuk, Marko Köthenbürger, Claus Kreiner, Etienne Lehmann, Claire Lelarge, Violaine Louvet, Luca Micheletto, Gareth Myles, Olof Rosenqvist, Dominik Sachs, Emmanuel Saez, Florian Scheuer, Joel Slemrod, Floris Zoutman, and participants at the workshop on Individual and Small Business Taxation in Warsaw, the Workshop on Behavioral Responses to Income Taxation in Mannheim, the Third Taxation Theory Conference in Toulouse, the Meeting of the Italian Association for Public Economics, National Swedish Conference in Economics, CESifo Area Conference on Public Sector Economics, the WIP Workshop in Helsinki, the NoCeT Conference "From Panama to BEPS: Tax Evasion or Tax Avoidance - International and National Policies to Confront Personal and Corporate Tax Strategies", and seminar participants at GATE Lyon St-Etienne, Umeå University, and Uppsala University for insightful comments and suggestions.

REFERENCES

Alstadsæter, A. and M. Jacob (2016). Dividend taxes and income shifting. *The Scandinavian Journal of Economics 118*(4), 693–717.

Alstadsater, A. and M. Jacob (2017). Who participates in tax avoidance? Evidence from Swedish microdata. *Applied Economics 49*(28), 2779–2796.

Apps, P., N. Van Long, and R. Rees (2014). Optimal piecewise linear income taxation. *Journal of Public Economic Theory 16*(4), 523–545.

Atkinson, A. B. and J. E. Stiglitz (1980). *Lectures on Public Economics*. McGraw-Hill Inc, US.

Bond, E. W. and K. J. Crocker (1991). Smoking, skydiving, and knitting: The endogenous categorization of risks in insurance markets with asymmetric information. *Journal of Political Economy 99*(1), 177–200.

Chetty, R. (2009). Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance. *American Economic Journal: Economic Policy 1*(2), 31–52.

Christiansen, V. and M. Tuomala (2008). On taxing capital income with income shifting. *International Tax and Public Finance 15*, 527–545.

Diamond, P. A. (1998). Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates. *American Economic Review 88*(1), 83–95.

Doligalski, P. and L. E. Rojas (2016, August). Optimal redistribution with a shadow economy. Technical report.

Fuest, C. and B. Huber (2001). Labor and capital income taxation, fiscal competition, and the distribution of wealth. *Journal of Public Economics 79*(1), 71–91.

Gomes, R., J.-M. Lozachmeur, and A. Pavan (2017). Differential taxation and occupational choice. *Review of Economic Studies, forthcoming*.

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

Harju, J. and T. Matikka (2016). The elasticity of taxable income and income-shifting: what is “real” and what is not? *International Tax and Public Finance*, 1–30.

Hermle, J. and A. Peichl (2015). Jointly optimal income taxes for different types of income. Technical report.

Jacquet, L. and E. Lehmann (2016). Optimal taxation with heterogeneous skills and elasticities: Structural and sufficient statistics approaches. Technical report.

Kleven, H. J., M. B. Knudsen, C. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica 79*(3), 651–692.

Kopczuk, W. (2001). Redistribution when avoidance behavior is heterogeneous. *Journal of Public Economics 81*(1), 51–71.

Lehmann, E., L. Simula, and A. Trannoy (2014, 09). Tax me if you can! Optimal Nonlinear Income Tax Between Competing Governments *. *The Quarterly Journal of Economics 129*(4), 1995–2030.

Parker, S. C. (1999). The optimal linear taxation of employment and self-employment incomes. *Journal of Public Economics 73*(1), 107–123.

Piketty, T. and E. Saez (2013). Optimal labor income taxation. In A. J. Auerbach, R. Chetty, M. Feldstein, and E. Saez (Eds.), *Handbook of public economics, vol. 5*, pp. 391 – 474. Elsevier.

Piketty, T., E. Saez, and S. Stantcheva (2014). Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy 6*(1), 230–71.

Pirttilä, J. and H. Selin (2011). Income shifting within a dual income tax system: Evidence from the finnish tax reform of 1993. *Scandinavian Journal of Economics 113*(1), 120–144.

Rothschild, C. and F. Scheuer (2013). Redistributive taxation in the Roy model. *The Quarterly Journal of Economics 128*(2), 623–668.

Saez, E., J. Slemrod, and S. 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.

Saez, E. and S. Stantcheva (2017). A simpler theory of optimal capital taxation. *Journal of Public Economics, forthcoming*.

Scheuer, F. (2014). Entrepreneurial taxation with endogenous entry. *American Economic Journal: Economic Policy 6*(2), 126–63.

Showalter, M. H. and N. K. Thurston (1997). Taxes and labor supply of high-income physicians. *Journal of Public Economics 66*(1), 73–97.

Simula, L. and A. Trannoy (2010). Optimal income tax under the threat of migration by top-income earners. *Journal of Public Economics 94*(1), 163 – 173.

Slemrod, J. (1994). Fixing the leak in Okun's bucket optimal tax progressivity when avoidance can be controlled. *Journal of Public Economics 55*(1), 41 – 51.

Slemrod, J. and W. Kopczuk (2002). The optimal elasticity of taxable income. *Journal of Public Economics 84*(1), 91 – 112.

Slemrod, J., S. Yitzhaki, J. Mayshar, and M. Lundholm (1994, February). The optimal two-bracket linear income tax. *Journal of Public Economics 53*(2), 269–290.

Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics 17*(2), 213–240.

Tazhitdinova, A. (2016, December). Income shifting and the cost of incorporation. Available at SSRN: https://ssrn.com/abstract=2902834.

APPENDIX A

Because of Lemma 1, the social planner's problem formulated in Subsection 3.3 is equivalent to:

$$\max_{\tau_1,\tau_2,T_1(0)} \int_\kappa \int_0^{\hat{\gamma}(\kappa)} g(\kappa,\gamma) V_2(\kappa,\gamma) f(\kappa,\gamma)\, d\gamma d\kappa + \int_\kappa \int_{\hat{\gamma}(\kappa)}^\infty g(\kappa,\gamma) V_1(\kappa) f(\kappa,\gamma)\, d\gamma d\kappa, \tag{46}$$

subject to:

$$\tau_2 \int_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega L_2(\kappa) f(\kappa,\gamma) d\gamma d\kappa + \tau_1 \int_\kappa \int_{\hat{\gamma}(\kappa)}^\infty \omega L_1(\kappa) f(\kappa,\gamma) d\gamma d\kappa - R - T_1(0) = 0. \tag{47}$$

We let λ be the Lagrange multiplier of the budget constraint (47). The derivative of (46) with respect to τ_1 is:

$$-\int_\kappa \int_{\hat{\gamma}(\kappa)}^\infty g(\kappa,\gamma) \omega L_1(\kappa) f(\kappa,\gamma) d\gamma d\kappa. \tag{48}$$

We used the fact that $V_1(\kappa) = V_2(\kappa, \gamma)$ for $\gamma = \hat{\gamma}(\kappa)$. We now compute the derivative of the budget constraint (47) with respect to τ_1. We obtain:

$$\int_\kappa \int_{\hat{\gamma}(\kappa)}^\infty \omega L_1(\kappa) f(\kappa, \gamma) d\gamma d\kappa + \tau_1 \int_\kappa \int_{\hat{\gamma}(\kappa)}^\infty \omega \frac{\partial L_1(\kappa)}{\partial \tau_1} f(\kappa, \gamma) d\gamma d\kappa$$
$$+ \int_\kappa \left[\tau_2 \omega L_2(\kappa) - \tau_1 \omega L_1(\kappa)\right] \frac{\partial \hat{\gamma}}{\partial \tau_1} f(\kappa, \hat{\gamma}) d\kappa. \tag{49}$$

From Lemma 1, we know that $\partial\hat{\gamma}/\partial\tau_1 = \omega L_1$. We now write the first-order condition (48) $+ \lambda \cdot$ (49) $= 0$, rearrange, and use the definition of e_1 to obtain (19).

To obtain (21), we first compute the derivative of (46) with respect to τ_2. Using the indifference condition at $\hat{\gamma}$, we obtain:

$$-\int_\kappa \int_0^{\hat{\gamma}(\kappa)} g(\kappa, \gamma) \omega L_2(\kappa) f(\kappa, \gamma) d\gamma d\kappa. \tag{50}$$

We now compute the derivative of (47) with respect to τ_2:

$$\int_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega L_2(\kappa) f(\kappa, \gamma) d\gamma d\kappa + \tau_2 \int_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega \frac{\partial L_2(\kappa)}{\partial \tau_2} f(\kappa, \gamma) d\gamma d\kappa$$
$$+ \int_\kappa \left[\tau_2 \omega L_2(\kappa) - \tau_1 \omega L_1(\kappa)\right] \frac{\partial \hat{\gamma}}{\partial \tau_2} f(\kappa, \hat{\gamma}) d\kappa. \tag{51}$$

From (10), we know that $\partial\hat{\gamma}/\partial\tau_2 = -\omega L_2$. We now write the first-order condition: (50) $+ \lambda \cdot$ (51) $=$ 0, rearrange, and use the definition of e_2 to obtain (21). □

APPENDIX B

Skills ω and shifting costs γ follow a bivariate log normal distribution, i.e.

$$(\omega, \gamma) \sim \ln \mathcal{N}(\mu_\omega, \mu_\gamma, \sigma_\omega^2, \sigma_\gamma^2, \rho),$$

where μ_x and σ_x stand for the mean and standard deviation of $\log(x)$. ρ is the correlation coefficient for the bivariate normal distribution of $\log(\omega)$ and $\log(\gamma)$. We approximate the distribution of skills using wage rates. We observe the mean and standard deviation on micro-data (LINDA) on monthly wages in Sweden (full time equivalents) as of 2009.

We do not, however, observe the moments of the shifting cost distribution. Our strategy is to calibrate the latter by choosing μ_γ and σ_γ in such a way that the actual share of 'shifters' is reproduced, conditional on the actual Swedish wage distribution, the actual Swedish tax system, and a given distribution of elasticities. We set our target, i.e. the actual fraction of shifters, to be 5 %. Alstadsater and Jacob (2017) report that 2.8% of Swedish individuals aged 18-70 are active shareholders in closely held corporations 2000-08. Considering the fact that the share has increased over time and that our wage data covers a younger sample (individuals aged 18-65) we think that 5 % is a reasonable number to use in the calibration.

We calculate marginal labor income tax rates and marginal dividend income tax rates for all individuals in the LINDA sample of 2009. We do not only consider the statutory tax rates, but also the payroll tax rate and the corporate tax rate.[22] In the LINDA wage sample, the average marginal labor tax rate amounted to 0.505, whereas the average (constant) marginal capital tax rate amounted to 0.410. Hence, we set $\tau_1 = 0.505$ and $\tau_2 = 0.410$ when calibrating the model. We then obtain a mean of the log wages of about 10, a mean of the fixed cost of shifting of 11, and for standard deviations we take 10% of the mean values.

APPENDIX C

The social welfare function has the following form:

$$\underbrace{\int_{\kappa,\theta}\int_{\hat{\gamma}(\theta)}^{\infty} g(\kappa,\theta,\gamma)V_1(\kappa)f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta}_{Non-shifters} + \underbrace{\int_{\kappa,\theta}\int_{0}^{\hat{\gamma}(\theta)} g(\kappa,\theta,\gamma)V_3(\kappa,\theta,\gamma)f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta}_{Interior-shifters}, \quad (52)$$

which also can be written as

$$\begin{aligned}&\int_{\kappa,\theta}\int_{0}^{\infty} g(\kappa,\theta,\gamma)[(1-\tau_1)\omega L_1 + T_1(0) - v(L_1;\epsilon)]f(\kappa,\theta,\gamma)d\gamma\kappa d\theta + \\ &\int_{\kappa,\theta}\int_{0}^{\hat{\gamma}(\theta)} g(\kappa,\theta,\gamma)[(\tau_1-\tau_2)A(\theta) - C(A;\theta) - \gamma]f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta.\end{aligned} \quad (53)$$

The government's budget constraint is

$$\int_{\kappa,\theta}\int_{0}^{\infty} \tau_1\omega L_1 f(\kappa,\theta,\gamma)d\gamma\kappa d\theta - \int_{\kappa,\theta}\int_{0}^{\hat{\gamma}(\theta)} (\tau_1-\tau_2)Af(\kappa,\theta,\gamma)d\gamma d\kappa d\theta - T_1(0) = 0. \quad (54)$$

The derivative of (52) with respect to τ_2 is

$$\begin{aligned}&\int_{\kappa,\theta}\frac{\partial\hat{\gamma}(\theta)}{\partial\tau_2}g(\kappa,\theta,\hat{\gamma})[(\tau_1-\tau_2)A(\theta) - C(A;\theta) - \hat{\gamma}(\theta)]f(\kappa,\theta,\hat{\gamma})d\kappa d\theta \\ &\qquad - \int_{\kappa,\theta}\int_{0}^{\hat{\gamma}(\theta)} g(\kappa,\theta,\gamma)A(\theta)f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta \\ &= -\int_{\kappa,\theta}\int_{0}^{\hat{\gamma}(\theta)} g(\kappa,\theta,\gamma)A(\theta)f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta,\end{aligned} \quad (55)$$

[22] If a owner of a closely held corporation distributes profits as wage income, her marginal tax rate is $\frac{\tau_{personal}+\tau_{payroll}}{1+\tau_{payroll}}$. If she distributes profits as dividend income her marginal tax rate is $\tau_{corporate} + \tau_{dividends} - \tau_{dividends} \times \tau_{corporate}$. In 2009 $\tau_{corporate} = 0.263$, $\tau_{dividends} = 0.2$ and $\tau_{payroll} = 0.3142$ were all proportional, whereas $\tau_{personal}$ varied between 0 and 0.565. When calculating $\tau_{personal}$ we accounted for the Swedish central government tax, local tax, basic allowance and the earned income tax credit.

where the first term is zero. This follows from (30). The derivative of (54) with respect to τ_2 is

$$\begin{aligned} -\int_{\kappa,\theta}\frac{\partial\hat{\gamma}(\theta)}{\partial\tau_2}(\tau_1-\tau_2)A(\theta)f(\kappa,\theta,\hat{\gamma})d\kappa d\theta \\ +\int_{\kappa,\theta}\int_0^{\hat{\gamma}(\theta)}A(\theta)f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta \\ -\int_{\kappa,\theta}\int_0^{\hat{\gamma}(\theta)}(\tau_1-\tau_2)\frac{\partial A(\theta)}{\partial\tau_2}f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta. \end{aligned} \tag{56}$$

We write the Lagrangian as (52) + λ (54) = 0. Hence,

$$\begin{aligned} -\int_{\kappa,\theta}\int_0^{\hat{\gamma}(\theta)}b(\kappa,\theta,\gamma)A(\theta)f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta \\ -\int_{\kappa,\theta}\frac{\partial\hat{\gamma}(\theta)}{\partial\tau_2}(\tau_1-\tau_2)A(\theta)f(\kappa,\theta,\hat{\gamma})d\kappa d\theta \\ +\int_{\kappa,\theta}\int_0^{\hat{\gamma}(\theta)}A(\theta)f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta \\ -\int_{\kappa,\theta}\int_0^{\hat{\gamma}(\theta)}(\tau_1-\tau_2)\frac{\partial A(\theta)}{\partial\tau_2}f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta = 0. \end{aligned} \tag{57}$$

Hence,

$$\frac{\int_{\kappa,\theta}\int_0^{\hat{\gamma}(\theta)}[1-b(\kappa,\theta,\gamma)]A(\theta)f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta}{\int_{\kappa,\theta}\int_0^{\hat{\gamma}(\theta)}\frac{\partial A(\theta)}{\partial\tau_2}f(\kappa,\theta,\hat{\gamma})d\gamma d\kappa d\theta+\int_{\kappa,\theta}\frac{\partial\hat{\gamma}(\theta)}{\partial\tau_2}A(\theta)f(\kappa,\theta,\hat{\gamma})d\kappa d\theta}=\tau_1-\tau_2. \tag{58}$$

Applying the implicit function theorem to (27), it follows that $\frac{\partial A(\theta)}{\partial\tau_2}<0$. Moreover, it directly follows from (30) that $\frac{\partial\hat{\gamma}}{\partial\tau_2}<0$. The denominator of (58) will therefore be negative. Hence, a nessecary and sufficient conditition for the right hand side to be positive is that $\int_{\kappa,\theta}\int_0^{\hat{\gamma}}[1-b(\kappa,\theta,\gamma)]A(\theta)f(\kappa,\theta,\gamma)d\gamma d\kappa d\theta$ is negative.

APPENDIX D

To simplify notations, we below drop the functions' arguments. By the envelope theorem, it follows from (32) that $dV(\omega,\gamma)/d\omega=(1-T_1')L-\vartheta L$. Plugging (33) into the latter, we obtain:

$$\frac{dV}{d\omega}=\frac{v'(L)}{\omega}L. \tag{59}$$

Hence, $dV_i/d\omega=v'(L_i(\omega))L_i(\omega)/\omega$, for $i=1,2$. The second-order conditions for incentive compatibility (iii) are standard given that our preference specification satisfies the single-crossing condition.□

APPENDIX E

We form the Lagrangian associated with Problem 1 (without accounting for the monotonicity constraints (iii)). At a given skill level ω, we denote the Lagrange multiplier associated with the incentive compatibility constraints by $\lambda_1(\omega)$ and $\lambda_2(\omega)$ respectively; μ refers to the Lagrange multiplier of the budget constraint (41) and $\lambda_A(\omega)$ stands for the Kuhn-Tucker multiplier of constraint $A \leq \omega L_2$.

To this aim, we first multiply conditions (43) and (42) by λ_i, respectively, and integrate by parts to get:

$$\int_0^\infty \lambda_i(\omega)[\frac{dV_i(\omega)}{d\omega} - \frac{v'(L_i(\omega)}{\omega}L_i(\omega)]d\omega = -\int_0^\infty \lambda_i'(\omega)V_i(\omega)d\omega - \int_0^\infty \lambda_i(\omega)\frac{v'(L_i(\omega)}{\omega}L_i(\omega)d\omega, \tag{60}$$

for $i = \{1, 2\}$. Note that we have used the fact that $\lim_{\omega\to\infty} \lambda_i(\omega)V_i(\omega) - \lambda_i(0)V_i(0) = 0$ because of the transversality conditions: $\lim_{\omega\to\infty} \lambda_i(\omega) = \lambda_i(0) = 0$. The Lagrangian associated with Problem 1 may now be rewritten as follows:

$$\begin{aligned}\mathcal{L} = &\int_0^\infty \int_0^{\hat{\gamma}(\omega)} g(\omega,\gamma)[V_2(\omega) - \gamma]f(\omega,\gamma)\,d\gamma d\omega + \int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty g(\omega,\gamma)V_1(\omega)\,f(\omega,\gamma)\,d\gamma d\omega \\ &- \int_0^\infty \left\{\lambda_2(\omega)\frac{v'(L_2(\omega))}{\omega}L_2(\omega) + \lambda_2'(\omega)V_2(\omega) + \lambda_1(\omega)\frac{v'(L_1(\omega))}{\omega}L_1(\omega) + \lambda_1'(\omega)V_1(\omega)\right\}d\omega \\ &+ \mu\int_0^\infty \int_0^{\hat{\gamma}(\omega)} [\omega L_2(\omega) - v(L_2(\omega)) - V_2(\omega) - C(A)]f(\omega,\gamma)\,d\gamma d\omega \\ &+ \mu\int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty [\omega L_1(\omega) - v(L_1(\omega)) - V_1(\omega)]f(\omega,\gamma)\,d\gamma d\omega + \int_0^\infty \lambda_A(\omega)[A(\omega) - \omega L_2(\omega)]d\omega.\end{aligned} \tag{61}$$

First-Order Conditions

- With respect to $V_1(\omega)$: for almost all values of ω,

$$\int_{\gamma=\hat{\gamma}(\omega)}^{\gamma=\infty} [g(\omega,\gamma) - \mu]f(\omega,\gamma)\,d\gamma - \lambda_1'(\omega) - \mu\Delta T(\omega)f(\omega,\hat{\gamma}(\omega)) = 0, \tag{62}$$

  where $\Delta T(\omega) = [\omega L_2(\omega) - C(A(\omega)) - v(L_2(\omega)) - V_2(\omega)] - [\omega L_1(\omega) - v(L_1(\omega) - V_1(\omega)]$ is *the extra tax paid by the marginal shifter*. When writing down (62), we have used the fact that $\hat{\gamma}(\omega) = V_2(\omega) - V_1(\omega)$, which in turn implies $\partial\hat{\gamma}(\omega)/\partial V_1(w) = -1$.

- With respect to $V_2(\omega)$: for almost all values of ω,

$$\int_{\gamma=0}^{\gamma=\hat{\gamma}(\omega)} [g(\omega,\gamma) - \mu]f(\omega,\gamma)\,d\gamma - \lambda_2'(\omega) + \mu\Delta T(\omega)f(\omega,\hat{\gamma}(\omega)) = 0, \tag{63}$$

  Note that we have used the fact that $\partial\hat{\gamma}(\omega)/\partial V_2(w) = 1$.

- With respect to $L_1(\omega)$: for almost all values of ω,

$$\lambda_1(\omega)[-\frac{v''(L_1(\omega))}{\omega}L_1(\omega) - \frac{v'(L_1(\omega))}{\omega}] + \mu\left[\omega - v'(L_1(\omega)))\right]\left[1 - F_{\gamma|\omega}(\hat{\gamma}(\omega))\right] = 0. \quad (64)$$

- With respect to $L_2(\omega)$: for almost all values of ω,

$$\lambda_2(\omega)[-\frac{v''(L_2(\omega))}{\omega}L_2(\omega) - \frac{v'(L_2(\omega))}{\omega}] - \omega\lambda_A(\omega) + \mu\left[\omega - v'(L_2(\omega))\right]F_{\gamma|\omega}(\hat{\gamma}(\omega)) = 0. \quad (65)$$

- With respect to $A(\omega)$: for almost all values of ω,

$$\mu C'(A(\omega))F_{\gamma|\omega}(\hat{\gamma}(\omega)) = \lambda_A(\omega). \quad (66)$$

Proof of point (a) in Proposition 4

Consider a given ω. From the agent's optimal conditions (33) and (34), we see that all agents choosing $A = 0$ choose the same L, while all other agents choose the same (A, L)-combination. Focus on this second set of agents and assume $0 < A < \omega L$. The right-hand side of (66) is 0 because $\lambda_A(\omega) = 0$ for these agents. In addition, $C'(A(\omega))$ on the left-hand side of (66) is constant and positive (because $C'(A) > 0$ for $A > 0$). Hence, for condition (66) to hold, the fraction of shifters at that ω, namely $F_{\gamma|\omega}(\hat{\gamma}(\omega))$, must be 0. Consequently, if there are shifters at a given ω, they cannot choose $0 < A < \omega L$. The only possibility is that all of them shift their entire earnings.$\square$

Proof of Points (b) and (c)

Combining (65) and (66), we obtain:

$$-\lambda_2(\omega)\frac{v'(L_2(\omega))}{\omega}\left[\frac{v''(L_2(\omega))}{v'(L_2(\omega)}L_2(\omega) + 1\right] + \mu\omega\left[1 - \frac{v'(L_2(\omega))}{\omega} - C'(A(\omega))\right]F_{\gamma|\omega}(\hat{\gamma}(\omega)) = 0. \quad (67)$$

From (35), $1 - \{v'(L_2(\omega))/\omega\} - C'(A(\omega)) = T_2'(\omega L_2(\omega))$. Using this relationship and the definition of $e_2(\omega)$, we obtain:

$$\frac{T_2'(\omega L_2(\omega))}{1 - T_2'(\omega L_2(\omega)) - C'(\omega L_2(\omega))} = \frac{\lambda_2(\omega)}{\omega\mu F_{\gamma|\omega}(\hat{\gamma}(\omega))}\left[1 + \frac{1}{e_2(\omega)}\right]. \quad (68)$$

Using the same steps, the first-order condition with respect to $L_1(\omega)$ can be written as:

$$\frac{T_1'(\omega L_1(\omega))}{1 - T_1'(\omega L_1(\omega))} = \frac{\lambda_1(\omega)}{\omega\mu\left[1 - F_{\gamma|\omega}(\hat{\gamma}(\omega))\right]}\left[1 + \frac{1}{e_1(\omega)}\right]. \quad (69)$$

Following Scheuer (2014), Appendix A.3, we integrate equations (62) and (63) over the whole support of ω, add them, and use the fact that they sum to 0. We in addition use the transversality conditions $\lim_{\omega\to\infty}\lambda_i(\omega)=\lambda_i(0)=0$ for $i=\{1,2\}$ to obtain:

$$\int_0^\infty \int_0^\infty [g(\omega,\gamma)-\mu]f\,(\omega,\gamma)\,d\gamma d\omega=\bar{g}-\mu=0, \tag{70}$$

where $\bar{g}\equiv\int_0^\infty\int_0^\infty[g(\omega,\gamma)]f\,(\omega,\gamma)\,d\gamma d\omega$ is the average social marginal welfare weight in the population. Integrating equations (62) and (63) between 0 and ω, using the relationship given by (70) and the fact that $\lambda_i(\omega)=\int_{x=0}^{x=\omega}\lambda_i'(x)dx$, for $i=\{1,2\}$, we obtain:

$$\lambda_1(\omega)=\int_{x=0}^{x=\omega}\int_{\gamma=\hat{\gamma}(\omega)}^{\gamma=\infty}[g(\omega,\gamma)-\bar{g}]f\,(\omega,\gamma)\,d\gamma dx-\bar{g}\int_{x=0}^{x=\omega}\Delta T(x)f(x,\hat{\gamma}(x))dx=0. \tag{71}$$

$$\lambda_2(\omega)=\int_{x=0}^{x=\omega}\int_{\gamma=0}^{\gamma=\hat{\gamma}(\omega)}[g(x,\gamma)-\bar{g}]f\,(x,\gamma)\,d\gamma dx+\bar{g}\int_{x=0}^{x=\omega}\Delta T(x)f(x,\hat{\gamma}(x))dx=0, \tag{72}$$

Because $\lim_{\omega\to\infty}\lambda_1(\omega)=\lim_{\omega\to\infty}\lambda_2(\omega)=0$, we can rewrite (72) and (71) as:

$$\lambda_1(\omega)=\int_{x=\omega}^{x=\infty}\int_{\gamma=\hat{\gamma}(\omega)}^{\gamma=\infty}[\bar{g}-g(x,\gamma)]f\,(x,\gamma)\,d\gamma dx+\bar{g}\int_{x=\omega}^{x=\infty}\Delta T(x)f(x,\hat{\gamma}(x))dx=0, \tag{73}$$

$$\lambda_2(\omega)=\int_{x=\omega}^{x=\infty}\int_{\gamma=0}^{\gamma=\hat{\gamma}(\omega)}[\bar{g}-g(x,\gamma)]f\,(x,\gamma)\,d\gamma dx-\bar{g}\int_{x=\omega}^{x=\infty}\Delta T(x)f(\omega,\hat{\gamma}(x))dx=0. \tag{74}$$

We then plug in these values in (69) and (68), and use the definition $b(\omega,\gamma)\equiv g(\omega,\gamma)/\mu$.$\square$