Income Shifting as Income Creation?
The Intensive vs. the Extensive Shifting Margins*

Håkan Selin† and Laurent Simula ‡

October 27, 2017

Abstract

Income shifting has been modeled as a decision along the intensive margin even though it involves significant fixed costs, giving rise to an important extensive margin. We show that accounting for this extensive margin has crucial policy implications: the distinction between income creation and income shifting breaks down. We make this point in a simple linear tax setting with a population of agents differing in terms of productivities, labor supply elasticities, and costs of income shifting. In the most empirically plausible scenario when people who shift easily are also more elastic in labor supply, giving them a lower tax rate is a good thing. This mechanism may be compared to third degree price discrimination in industrial organization. Numerical simulations suggest that fixed shifting costs have a large impact on optimal taxes. We further demonstrate that the conclusions derived for linear taxes carry over to non-linear tax schedules. *Keywords*: Income Shifting, Optimal Taxation, Labor Income Tax. *JEL Classification*: H21; H24

*We are grateful to Spencer Bastani, Sören Blomquist, Pierre Boyer, Christophe Bravard, Katherine Cuff, Marcus Eliason, Alex Gelber, Aart Gerritsen, Roger Gordon, Laurence Jacquet, Wojciech Kopczuk, Claus Kreiner, Etienne Lehmann, Claire Lelarge, Violaine Louvet, Luca Micheletto, Gareth Myles, Olof Rosenqvist, Dominik Sachs, Emmanuel Saez, 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, and seminar participants at GATE Lyon, Umeå University, and Uppsala University for insightful comments and suggestions.

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

‡Ecole Normale Supérieure Lyon and GATE-LSE, France. Email: Laurent.Simula@ens-lyon.fr

1 Introduction

Tax rate changes do not only trigger labor supply and savings responses, but also shifting activities when taxpayers move income from highly taxed bases to more leniently taxed ones. In the United States for example, high-income individuals are highly sensitive to differences in the personal and the corporate marginal tax rates, which foster income shifting responses, as documented by Gordon and Slemrod (2000). In countries with dual tax systems, with separate taxation of labor and capital incomes, taxpayers may engage in income shifting by starting up closely held corporations and subsequently transfer income between the tax bases.[1] These shifting activities are often regarded as purely socially wasteful.

In this article, we make the point that policy designers 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, to the cost of entering self-employment, or to the cost of setting up a closely held corporation, etc. Tazhitdinova (2016) has recently shown that fixed costs of incorporation are empirically important. Even at a given level of earnings ability such costs are likely to vary: In some professions, such as lawyers and medical doctors, it might be fairly straightforward to switch from being an employee receiving third-party reported income, whereas in other professions, like teachers, it might be more costly.

Heterogeneous fixed shifting costs give rise to an important *extensive margin* of income shiftting. This margin has been neglected in the previous theoretical literature, which focused on income shifting as a pure decision along an *intensive margin*.[2] In addition to being empirically relevant, it turns out that fully accounting for this extensive shifting margin has policy implications in sharp contrast with previous findings; it indeed casts light on

[1] For example, Pirttilä and Selin (2011), Alstadsæter and Jacob (2016), and Harju and Matikka (2016) document income shifting in Scandinavian dual tax systems.

[2] 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).

a new and simple mechanism through which income shifting may actually contribute to increasing social welfare.

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 place ourselves in the position which is the least favorable to shifting and deliberately neither model capital accumulation nor tax competition. 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. 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. In principle, tax rate differentiation may also occur when agents face various convex shifting costs, providing the social planner puts large weights on individuals with low shifting costs. This mechanism is however very different from the one we highlight, where differences in elasticities are the key driving force. Moreover, it is not empirically relevant because shifting mostly occurs at the top of the income distribution. To investigate our analytical results numerically, we provide numerical illustrations, calibrated using Swedish data, and find that non-negligible welfare gains may be achieved thanks to income shifting.

Extending our simple model, we then show that our results are robust to: (i) the combination of a fixed cost and convex shifting costs as well as (ii) the relaxation of the linear tax assumptions. In this extended framework, we also derive tax revenue maximizing asymptotic tax rates. Revenue maximizing asymptotic tax rates endogenizing income shifting have not earlier been presented in the literature in spite of the extensive focus on top-income taxation.

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 illustrates the intensive marginal logic in the simplest way. Section 5 casts light on the consequences of allowing for income shifting along a pure extensive margin and provides numerical simulations. Section 6 shows that our results are robust to various extensions and discusses the implications for the revenue maximizing tax rates at the top of the skill distribution. Section 7 concludes.

2 Related Literature

Our work is closely related to Piketty and Saez (2013, Section 4). The latter 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 (2016). Considering heterogeneity in skills only, it is shown that governments 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 formulas 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]

Finally, 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 (2012) 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. In addition, we consider a general framework, which in practice may correspond to a variety of situations where shifting might occur in some form.

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,[4] and the propensity to shift incomes to an alternative tax base. The latter is captured through a cost parameter γ. The distribution of ω, ϵ and

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

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.

3.2 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 given by:

$$U(Y, L) = Y - v(L; \epsilon). \tag{1}$$

Every individual has the possibility to reduce the income that is subject to the first tax instrument T_1, from ωL to $\omega L - A$ at a cost $\Gamma(A, \gamma)$. We refer to this as *income shifting*. As emphasized in the introduction, this cost might be a fixed cost and/or variable. A general specification is:

$$\Gamma(A; \gamma) = C(A; \gamma) + \gamma \cdot \mathbb{1}_{A>0}, \tag{2}$$

where the variable cost $C(A; \gamma)$ is non-decreasing and convex in the shifted amount A (i.e., $C'_A \geq 0$ and $C''_{AA} \geq 0$), whereas γ is a heterogeneity parameter determining shifting costs. $\mathbb{1}_{A>0}$ is an indicator function, equal to 1 when $A > 0$ and 0 otherwise. Most of the previous literature has focused on the case where $\Gamma(A; \gamma) = C(A)$, i.e. the case with a pure convex cost that is the same for everyone.[5] By contrast, we investigate the implications of a more general – and more empirically relevant[6]– cost structure that includes fixed costs.

Overall, an individual pays taxes equal to $T_1(\omega L - A) + T_2(A)$ and thus receives net income:[7]

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

The utility function (1) is quasilinear in net income. Consequently, we can alternatively interpret $\Gamma(A, \gamma)$ as the utility loss induced when an individual decides to shift earnings. 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 $(\omega, \epsilon, \gamma)$. For later use, we also define:

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

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

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 $L_2(\omega, \epsilon, \gamma)$.

3.3 Policy-Maker's Choices

The policy-maker chooses two tax functions. By the taxation principle, this is equivalent to designing the *incentive compatible* allocation, which maximizes the social objective

[5] An exception is Kopczuk (2001), who in the context of tax avoidance emphasizes the welfare implications of heterogeneity in the convex costs of avoidance.

[6] As already emphazised, see e.g., Tazhitdinova (2016) whose findings are consistent with the existence of fixed costs.

[7] A more general specification would allow for endogeneous capital income supply, Q, such that the capital tax payment would be $T_2(Q + A)$. However, the idea to allow for income shifting as a consequence of differential taxation of labor incomes and capital incomes is already well known in the literature. In our article, we instead consider the possibility that income shifting is socially desirable even in the situation when all incomes earned generically originate from labor effort.

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:

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

R is a tax revenue requirement that does not enter the individuals' utility function. When it is set equal to zero, the tax policy is purely redistributive. It should be emphasized that there is a vast literature on the issue of aggregating 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 indeed simply consider that some unspecified welfare weights $g(\omega,\epsilon,\gamma)$ are assigned to various agents conditional on their characteristics, including the preference parameter ε. A sub-case is obviously the situation in which the welfare weights only depend on ω and γ, or just on ω. Our analytical results are robust to all these various specifications.

4 The Intensive Shifting Margin

The objective of this short section is to illustrate the policy implications of modeling income shifting as a pure intensive margin phenomenon. Following Piketty and Saez (2013), we assume that non-shifted income is taxed linearly, while shifted income is taxed proportionally. By denoting the marginal tax rates on non-shifted and shifted income τ_1 and τ_2 respectively, we obtain $T_1 = G + \tau_1 \times (\omega L - A)$ and $T_2 = \tau_2 A$. G is a demogrant; when $G < 0$, the policy-maker distributes a basic income to each agent. Any $\tau_2 \ge \tau_1$ is associated with the same outcome, i.e., the absence of shifting; hence, there is no loss of generality in focusing on $\tau_1 \ge \tau_2$. For simplicity, we assume away corner solutionq in the utility maximization problem. To this aim, we first focus on the simplest model without fixed costs, and we let $\gamma = 0$ for everyone. Hence, $\Gamma(A;\gamma) = C(A;\gamma)$.

Given this set of assumptions, any individual's first-order conditions are independent of each other and can be written as:

$$v'(L;\varepsilon) = \omega(1-\tau_1) \tag{9}$$

$$C'_A(A;\gamma) = \tau_1 - \tau_2 \tag{10}$$

As usual, (9) shows that the individual will supply labor effort until the marginal disutility of doing so equates the marginal after-tax wage. (10) implies that the individual will shift income until the marginal gain, given by the difference between the two marginal tax rates, equates the marginal cost.

When there is no heterogeneity in the convex cost (as e.g in Piketty and Saez (2013)), everyone shifts the same amount of income. It is thus clear that no extra redistribution can be obtain by differentiating tax rates. The only effect of setting a lower tax on shifted incomes is to induce all agents to use up real resources in pursuit of tax savings. From a social perspective this is clearly a pure waste.

When there is heterogeneity in the convex cost, it is possible to construct examples where the social planner should differentiate the two tax rates. Intuitively, if it is cheaper for low-skilled individuals to shift income, the social planner can increase social welfare by allowing for income shifting. From a practical perspective, this mechanism is probably not important, given that income shifting typically pertains to the upper part of the income distribution.[8]

5 The Extensive Shifting Margin

We now cast light on the consequences of allowing for income shifting along the extensive margin. To this aim, we let the shifting cost be a pure fixed cost, i.e., $\Gamma(A;\gamma) = \gamma \cdot \mathbb{1}_{A>0}$. Given this specification, the corner solution $A = \omega L$ may play an important part and we do not make any assumption that would rule it out. In line with Section 4, non-shifted income is taxed linearly whilst shifted income is taxed at a proportional rate.

5.1 Partition of the Population

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

$$(1-\tau_1)\omega L_1 + G - v(L_1;\varepsilon) \geq (1-\tau_2)\omega L_2 + G - \gamma - v(L_2;\varepsilon), \tag{11}$$

[8]For further discussions and more precise conditions, see Selin and Simula (2017, section 3).

which is equivalent to:

$$\gamma \geq [(1-\tau_2)\omega L_2 - (1-\tau_1)\omega L_1] + [v(L_1;\varepsilon) - v(L_2;\varepsilon)]. \tag{12}$$

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

Lemma 1. *Assume there is a fixed shifting cost $\Gamma(A;\gamma) = \gamma \cdot \mathbb{1}_{A>0}$, $T_1 = G + \tau_1(\omega L - A)$ and $T_2 = \tau_2 A$. For each value of κ, define $\hat{\gamma}(\kappa)$ as the solution in γ to (12) written with equality instead of $\geq$.[9] Then:*

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

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

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

At a given κ, agents with $\gamma < \hat{\gamma}(\kappa)$ shift their entire earnings. Because the fixed shifting cost enters the individual optimization problem in an additively separable way, each of them provides effort level L_2, independent of γ. L_2 is therefore a function of the parameters κ. Once an agent has decided to shift her entire earnings, the marginal work incentive is independent of τ_1 and 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] Each of them provides effort $L_1(\kappa)$, independent of γ. The marginal work incentive is driven by τ_1 (and thus independent of τ_2).

This partition of the population plays a key part. 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 schedules. This mechanism clearly differs from what is allowed by the introduction of additional tax brackets.

5.2 Optimal Tax Rates

We now use a small tax reform perturbation around the optimum to determine the optimal tax policy, and thus the marginal tax rates τ_1 and τ_2. More precisely, we investigate

[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.

the effects of increasing τ_1, or alternatively τ_2, by a small quantity $\partial\tau > 0$, everything else being equal. We start by considering an increase in the marginal tax rate τ_1 on non-shifted income . This tax variation has the following effects:

- *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, 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:

$$E_1 = \iint_\kappa \int_{\hat{\gamma}(\kappa)}^{\infty} (1 - b(\kappa,\gamma))\,\omega L_1 \,\partial\tau\, f(\kappa,\gamma) d\gamma d\kappa. \tag{13}$$

- *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:

$$-\frac{\omega L_1 \cdot e_1(\varepsilon)}{1-\tau_1} \times \partial\tau, \tag{14}$$

  where $e_1(\varepsilon)$ stands for the labor supply elasticity within the set of non-shifters. As a result, taxes collected from this agent diminish by $\tau_1 \times$ (14). Integrating over the non-shifting population, we obtain:

$$E_2 = -\iint_\kappa \int_{\hat{\gamma}(\kappa)}^{\infty} \frac{\tau_1}{1-\tau_1}\omega L_1\, e_1(\varepsilon)\,\partial\tau\, f(\kappa,\gamma)\, d\gamma d\kappa. \tag{15}$$

- *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:

$$\Delta T \equiv \tau_2\omega L_2 - \tau_1\omega L_1 \tag{16}$$

  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:

$$E_3 = \iint_\kappa \Delta T \, \frac{\partial \hat{\gamma}(\kappa)}{\partial \tau_1} \, \partial\tau \, f(\omega, \hat{\gamma}) d\kappa = \iint_\kappa \Delta T \, \omega L_1 \, \partial\tau f(\omega, \hat{\gamma}) d\kappa. \tag{17}$$

where $\frac{\partial \hat{\gamma}(\kappa)}{\partial \tau_1} = \omega L_1$ follows from (12).

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:

$$\frac{\tau_1}{1-\tau_1} = \frac{\iint_\kappa \int_{\hat{\gamma}(\kappa)}^{\infty} [1 - b(\kappa, \gamma)] \omega L_1 f(\kappa, \gamma) d\gamma d\kappa}{\iint_\kappa \int_{\hat{\gamma}(\kappa)}^{\infty} \omega L_1 e_1(\varepsilon) f(\kappa, \gamma) d\gamma d\kappa} + \frac{\iint_\kappa \omega L_1 \Delta T(\kappa, \hat{\gamma}) f(\kappa, \hat{\gamma}) d\kappa}{\iint_\kappa \int_{\hat{\gamma}(\kappa)}^{\infty} \omega L_1 e_1(\varepsilon) f(\kappa, \gamma) d\gamma d\kappa}. \tag{18}$$

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, $L_1(\omega)$ replaced by $L_2(\omega)$, $e_1(\omega)$ replaced by the labor supply elasticity $e_2(\varepsilon)$ 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:

$$- \iint_\kappa \Delta T \, \omega L_2 \, \partial\tau \, f(\kappa, \hat{\gamma}) d\kappa. \tag{19}$$

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:

$$\frac{\tau_2}{1-\tau_2} = \frac{\iint_\kappa \int_0^{\hat{\gamma}(\kappa)} [1 - b(\kappa, \gamma)] \omega L_2 f(\kappa, \gamma) d\gamma d\kappa}{\iint_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega L_2 e_2(\varepsilon) f(\kappa, \gamma) d\gamma d\kappa} - \frac{\iint_\kappa \omega L_2 \Delta T(\kappa, \hat{\gamma}) f(\kappa, \hat{\gamma}) d\kappa}{\iint_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega L_2 e_2(\varepsilon) f(\kappa, \gamma) d\gamma d\kappa}. \tag{20}$$

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

Proposition 1. *Assume there is a fixed shifting cost* $\Gamma(A;\gamma) = \gamma \cdot \mathbb{1}_{A>0}$, $T_1 = G + \tau_1(\omega L - A)$ *and* $T_2 = \tau_2 A$. *In the social optimum, the marginal tax rates* τ_1 *and* τ_2 *are given by Equations* (18) *and* (20).

In the social optimum, the marginal tax rates τ_1 and τ_2 typically differ.[11] As shown by Equations (18) and (20), a first driving force is the trade-off between equity concerns (in the numerator) and efficiency (in the denominator), captured by the first term on the left-hand side of both formulas. Both of them "look like" the usual optimal linear income tax formula (cf. e.g., Atkinson and Stiglitz, 1980). However, they are computed as if the total population was restricted to non-shifters and shifters, respectively. 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 (18) and (20) 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 therefore suggests that individual *heterogeneity* in these dimensions 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 *real* labor supply elasticities are much larger for self-employed physicians than for physicians who are employees. Further 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.

[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 1 collapse into the "usual" optimal income tax rules, with $\tau_1 = \tau_2$.

5.3 Numerical Illustration

The analysis of a small tax reform perturbation around the social optimum illuminated the mechanisms behind the optimal tax rates formulas of Proposition 1. We now provide an example showing that it may be socially optimal to allow for income shifting for non-realistic calibrations. In the numerical exercise and to reduce the dimensionality of the problem, we let the taste parameter ε depend deterministically on ω, with:

$$\varepsilon = q_1 + q_2\omega. \tag{21}$$

Regarding individual preferences, we let:

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

which implies $e_1(\varepsilon) = \varepsilon$. Hence, the individual's labor supply elasticity is constant at all levels of labor supply, but varies across people. In the baseline simulations, we further assume an increasing elasticity, from 0.1 at the bottom of the skill distribution to 0.5 for the highest skill level.

We further focus on the specific situation in which the policy-maker wants to maximize the well-being of the worst-off in the population. There are agents with productivity 0 in the data we use (see below). The latter do not earn any income but G; hence none of them will ever shift incomes. it follows that the worst-off are those with $w = 0$. A zero social weight is attached to all other agents. This "maximin" benchmark is of particular interest for two reasons: (i) it is widely considered in the optimal tax literature; (ii) shifting has no direct positive utility effect, through the increased net income of the shifters.

We also 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 considerable less guidance on how to calibrate the distribution of shifting costs. Because we want to perform sensitivity analyses with respect to the correlation of (ω, ε) and γ, it is convenient to assume that these two parameters follow a bivariate log-normal distribution. We use Swedish data to calibrate the mean and variance of the wage distribution. The shifting costs are parameterized so that the proportion of people deciding to shift incomes roughly

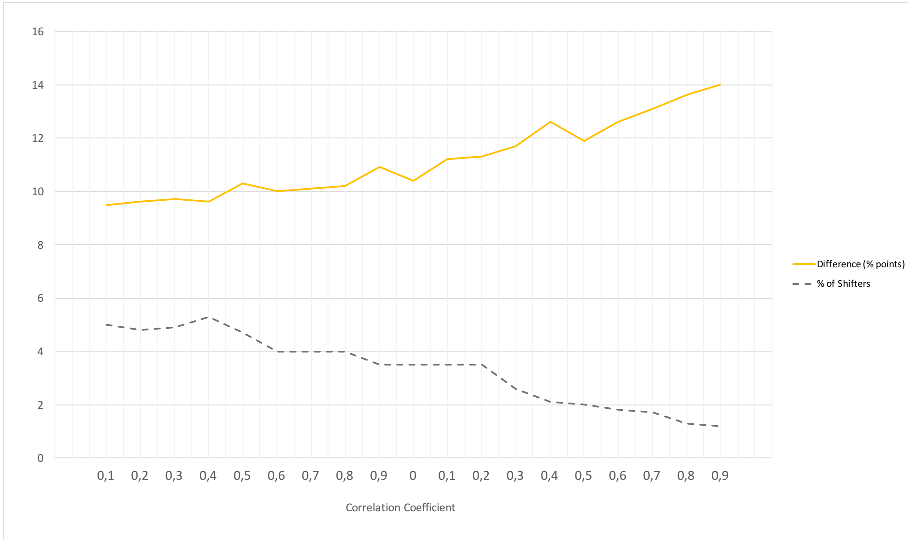

Figure 1: Benchmark Scenario

reproduces the actual figure for Sweden (see Alstadsæter and Jacob, 2016).[12]

In Figure 1, the solid curve shows the gap – in percentage points – between τ_1 and τ_2 for various values of the correlation coefficient for $\log(\omega)$ and $\log(\gamma)$. Additionally, the dashed curve shows the share of the population that chooses to pay the fixed cost and, thereby, shifts their entire labor income into the capital income tax base. The percentage of shifters is declining in the correlation coefficient, from about 6% to 1%. This makes sense since a negative correlation implies that highly skilled individuals (with large elasticities) face low shifting costs. In this numerical illustration, there is always a gap between τ_1 and τ_2, and the tax difference increases with the correlation coefficient.

We now investigate to which extent our results are sensitive to the elasticity range. For three different values of the correlation coefficient ρ (namely -0.99, 0 and 0.99), we examine four different elasticity ranges while keeping the average elasticity in the population constant (at 0.23). The results are reported in Table 1. It appears that the variance of the elasticity is crucial for optimal tax policy. First, when the elasticity is constant in the population, the social planner must set $\tau_1 = \tau_2$. Let us assume that the elasticity does not vary between agents and that there are two subpopulations in the social optimum, one reporting non-shifted income and one reporting shifted income. Given the quasilinear-

[12]We provide a more detailed discussion in Appendix B.

Table 1: Simulation results

Min elasticity	Max elasticity	ρ	τ_1^*	τ_2^*	$\tau_1^* - \tau_2^*$	Shifters %
0	0.725	-0.99	0.79	0.65	0.14	16.3
0	0.725	0	0.78	0.65	0.14	12.7
0	0.725	0.99	0.77	0.64	0.13	8.2
0.1	0.5	-0.99	0.80	0.71	0.09	6
0.1	0.5	0	0.79	0.70	0.09	3.5
0.1	0.5	0.99	0.79	0.68	0.11	1.1
0.15	0.4	-0.99	0.80	0.72	0.08	2.4
0.15	0.4	0	0.80	0.71	0.09	0.6
0.15	0.4	0.99	0.80	0.68	0.11	0.2
0.23	0.23	-0.99	0.81	0.81	0.00	0
0.23	0.23	0	0.81	0.81	0.00	0
0.23	0.23	0.99	0.81	0.81	0.00	0

ity of individual preferences, the top of the Laffer curve would be obtained for the same marginal tax rate in the two subpopulations. Because the social objective that we consider is the maximin, this implies that tax rates should not be differentiated. Second, when the individual with the lowest ability exhibits an elasticity of 0 and the individual with the highest ability has an elasticity of 0.725, elasticities are more dispersed than in our baseline scenario. In this case, the fraction of shifters and the gap in marginal tax rates are much larger.

6 Robustness Checks and Extensions

Sections 4 and 5 illustrated the important distinction between the intensive and extensive margins in the most simple way. In particular, we emphasized that the tax rate differentiation mechanism at stake differed from the introduction of additional income tax brackets. The objective of this section is to show that our results are robust to:

(i) the combination of a fixed cost and convex shifting costs; we therefore consider $\Gamma(A;\gamma) = C(A) + \gamma \cdot \mathbb{1}_{A>0}$, without assuming either $\gamma \equiv 0$ (contrary to Section 4) nor $C(A) \equiv 0$ (contrary to Section 5).

(ii) the relaxation of the linear tax assumptions; hence $T_1(\omega L_1 - A)$ and $T_2(A)$ are non-linear; we simply assume that $T_2(0) = 0$. Depending on the exact form of shifting one has in mind, it might be interesting to think of T_2 as linear.

Multidimensional screening problems are technically challenging. To make the problem sufficiently tractable, we henceforth assume that ω and γ are the only dimensions of heterogeneity within the population. This implies that at a given (ω, γ), all agents have the same ϵ. The three-dimensional screening problem considered above therefore turns into a two-dimensional one. 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.[13]

6.1 Individual Incentives and Elasticities

Because the shifting cost function Γ now combines a smooth cost with a fixed cost, there are potentially three categories of agents, depending on whether agents shift all their earnings, part of them, or nothing. Because of the interaction between the intensive and extensive shifting mechanisms, it is important to explicitly account for the inequality constraint $A \leq \omega L$ in the utility maximization program (4). With λ referring to the Kuhn-Tucker multiplier of the latter constraint, the first-order conditions with respect to L and A are respectively:

$$[1 - T_1'(\omega L - A)]\omega - \lambda\omega - v'(L) \leq 0 \quad (= if A > 0), \tag{23}$$

$$T_1'(\omega L - A) - T_2'(A) - C'(A) + \lambda \leq 0 \quad (= if A > 0), \tag{24}$$

with $\lambda \geq 0 \, (= 0 \text{ if } A < \omega L)$. Provided $A > 0$, combining both equations yields:

$$\left[1 - T_2'(A) - C'(A)\right] \omega = v'(L) \tag{25}$$

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 and the marginal shifting cost. In that case, the marginal shifting cost $C'(A)$ plays exactly the

[13]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.

same part as the marginal tax rate $T_2'(A)$.

An agent decides to set $A > 0$ provided $V(\omega, \gamma) > V_1(\omega)$.[14] 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:

$$\begin{aligned} \hat{\gamma} = \omega \left(L(\omega, \hat{\gamma}) - L_1(\omega) \right) + T_1 \left(\omega L_1(\omega, \hat{\gamma}) \right) - T_1 \left(\omega L(\omega) - A(\omega, \hat{\gamma}) \right) \\ - T_2 \left(A(\omega, \hat{\gamma}) \right) - C \left(A(\omega, \hat{\gamma}) \right) - v \left(L(\omega, \hat{\gamma}) \right) + v \left(L_1(\omega) \right). \end{aligned} \tag{26}$$

Intensive Labor Supply Elasticities

As in the previous sections, elasticities play a key part in the analysis. It is necessary to generalize the definitions above to account for the non-linearity of the tax schedules. For an agent who does not shift earnings at all, the labor supply elasticity is given by:

$$e_1(\omega) \equiv \frac{\partial L_1(\omega)}{\partial \left[\omega \left(1 - T_1'(\omega L_1(\omega)) \right) \right]} \frac{\omega(1 - T_1'(\omega L_1(\omega)))}{L_1(\omega)} = \frac{v'(L_1(\omega))}{v''(L_1(\omega)) L_1(\omega)}. \tag{27}$$

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. It is therefore useful to define the labor supply elasticity of an agent who shifts her entire earnings. It is given by:

$$e_2(\omega) \equiv \frac{\partial L_2(\omega)}{\partial \left[\omega (1 - T_2'(\omega L_2(\omega))) \right]} \frac{\omega \left(1 - T_2'(\omega L_2(\omega)) \right)}{L_2(\omega)} = \frac{v'(L_2(\omega)) + \omega C'(\omega L_2(\omega))}{v''(\omega L_2(\omega)) L_2(\omega)}. \tag{28}$$

The last equality follows from (25). When a shifter supplies one extra unit of labor, she does not only have to pay the marginal tax rate, but also the marginal shifting cost. It is clear from the comparison of e_1 and e_2 that *two agents of skill ω may have different labor supply elasticities*.

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

Extensive Shifting Elasticities

As emphasized below, each agent either chooses $A = 0$ or $A = \omega L$ in the social optimum. Consequently, at a given skill ω, the proportion of shifters in the population is equal to:

$$F_{\gamma|\omega}(\hat{\gamma}(\omega)) = F_{\gamma|\omega}[\omega L_2 - \omega L_1 + T_1(\omega L_1) - T_1(0) - T_2(\omega L_2) + v(L_1) - v(L_2)]. \quad (29)$$

The percentage increase in the proportion of *shifters* in response to a percentage rise in the tax paid as a non-shifter is:

$$\eta_2(\omega) = -\frac{\partial F_{\gamma|\omega}(\hat{\gamma}(\omega))}{\partial T_2(0)} \frac{T_1(\omega L_1)}{F_{\gamma|\omega}(\hat{\gamma}(\omega))} = f(\omega, \hat{\gamma}) \frac{T_1(\omega L_1)}{F_{\gamma|\omega}(\hat{\gamma}(\omega))}. \quad (30)$$

We vary $T_2(0)$ (intially equal to 0) as this quantity is independent of the labor supply choice. Similarly, the percentage decrease in the proportion of *non-shifters* in response to a percentage increase in the tax paid as a non-shifter is:

$$\eta_1(\omega) = -\frac{\partial[1 - F_{\gamma|\omega}(\hat{\gamma})]}{\partial T_2(0)} \frac{T_1(\omega L_1)}{1 - F_{\gamma|\omega}(\hat{\gamma})} = -f(\omega, \hat{\gamma}) \frac{T_1(\omega L_1)}{1 - F_{\gamma|\omega}(\hat{\gamma})}. \quad (31)$$

6.2 The Social Planner's Problem

We define $\tilde{V}(\omega) \equiv V(\omega, \gamma) + \gamma$ as the indirect utility of a shifter (with $A > 0$) gross of the fixed cost γ. The social planner's objective function can thus be written as:

$$\int_0^\infty \int_0^{\hat{\gamma}(\omega)} g(\omega, \gamma)[\tilde{V}(\omega) - \gamma] f(\omega, \gamma)\, d\gamma d\omega + \int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty g(\omega, \gamma) V^1(\omega)\, f(\omega, \gamma)\, d\gamma d\omega. \quad (32)$$

The social planner maximizes (32) with respect to $\tilde{V}$, V^1, L, L^1 and A, within the set of budget-balanced and incentive-compatible allocations.

Lemma 2. *An incentive-compatible allocation satisfies the following first-order conditions:*

$$\frac{d\tilde{V}(\omega)}{d\omega} = \frac{v'(L(\omega))}{\omega} L(\omega) \text{ for shifters;} \quad (33)$$

$$\frac{dV^1(\omega)}{d\omega} = \frac{v'(L^1(\omega))}{\omega} L^1(\omega) \text{ for non-shifters.} \quad (34)$$

Proof. See Appendix C. □

In addition to the first-order conditions provided in Lemma 2, incentive compatible allocations must satisfy second-order conditions: namely, that gross earnings are non-decreasing in skills within each set of agents (shifters and non-shifters). Below we adopt the standard "first-order approach" and do not formally account for these monotonicity constraints when writing the policy-maker's optimization problem.[15] Moreover, any optimal allocation must be budget-balanced, and thus satisfy:

$$\begin{aligned} &\int_0^\infty \int_0^{\hat{\gamma}(\omega)} [\omega L(\omega) - v(L(\omega)) - \tilde{V}(\omega) - \gamma] f(\omega, \gamma)\, d\gamma d\omega \\ &+ \int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty [\omega L^1(\omega) - v(L^1(\omega)) - V^1(\omega)] f(\omega, \gamma)\, d\gamma d\omega \geq R. \end{aligned} \tag{35}$$

6.3 Social Optimum

We have seen in Section 5 that when marginal tax rates are constant and shifting only involves a fixed cost, a rational agent either shifts her entire earnings or nothing. 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.

$\Delta T(\omega) = [\omega L^S - v(L^S) - \tilde{V}^S(\omega)] - [\omega L^1 - v(L^1) - V^1(\omega)]$ is *the extra tax paid by the marginal shifter*

Proposition 2. *Assume (i)* $\Gamma(A; \gamma) = C(A) + \gamma \cdot \mathbb{1}_{A>0}$*, with* $C'(A) > 0$ *and (ii)* T_1 *and* T_2 *are nonlinear, with* $T_2(0) = 0$*. In the social optimum:*

(a) every agent either chooses $A = 0$ *or* $A = \omega L$*;*

(b) for non-shifters,

$$\begin{aligned} \frac{T_1'(\omega L^1)}{1 - T_1'(\omega L^1)} = &\left[1 + \frac{1}{e^1(\omega)}\right] \\ &\times \frac{\int_\omega^\infty \int_{\hat{\gamma}(\omega)}^\infty [1 - b(\omega, \gamma)] f(\omega, \gamma)\, d\gamma d\omega + \int_\omega^\infty \Delta T(\omega) f(\omega, \hat{\gamma(\omega)}) d\omega}{\omega f_\omega(\omega)[1 - F_{\gamma|\omega}(\hat{\gamma})]}; \end{aligned} \tag{36}$$

[15]Satisfaction of the second-order conditions for incentive compatibility can be checked ex post in numerical simulations.

(c) for non-shifters,

$$\frac{T_2'(\omega L^S)}{1-T_2'(\omega L^S)} = \left[1+\frac{1}{e^S(L^S,\omega)} - \frac{C'(\omega L^S)}{1-T_2'(\omega L^S)}\right] \times \frac{\int_{\omega}^{\infty}\int_0^{\hat{\gamma}(\omega)}[1-b(\omega,\gamma)]f\,(\omega,\gamma)\,d\gamma d\omega - \int_{\omega}^{\infty}\Delta T(\omega)f(\omega,\hat{\gamma})d\omega}{\omega f_{\omega}(\omega)F_{\gamma|\omega}(\hat{\gamma})}. \quad (37)$$

Proof. In Appendix D, we write down the Lagrangian for the social planner's problem. Assume $A < \omega L$. The first-order condition with respect to A implies:

$$\mu C'(A) f_{\omega}(\omega) F_{\gamma|\omega}(\hat{\gamma}) = 0. \quad (38)$$

The shadow price of the resource constraint μ and the density $f_{\omega}(\omega)$ are strictly positive. Therefore, $C'(A) > 0$ implies $F_{\gamma|\omega}(\hat{\gamma}) = 0$. □

Proposition 2 shows that the sorting of agents between the groups of pure shifters and pure non-shifters, highlighted in Section 5, is robust both to the introduction of a smooth shifting cost and to the relaxation of the linear tax assumptions.

Proposition 3. *Assume $C'(A) > 0$. In the social optimum,*

$$\frac{T_1'(\omega L^1)}{1-T_1'(\omega L^1)} = \left[1+\frac{1}{e^1(L^1,\omega)}\right] \times \frac{\int_{\omega}^{\infty}\int_{\hat{\gamma}(\omega)}^{\infty}[1-b(\omega,\gamma)]f\,(\omega,\gamma)\,d\gamma d\omega + \int_{\omega}^{\infty}\Delta T(\omega)f(\omega,\hat{\gamma})d\omega}{\omega f_{\omega}(\omega)[1-F_{\gamma|\omega}(\hat{\gamma})]} \quad (39)$$

Proof. See Appendix C. □

The optimal tax rules (39) and (37) have the same structure as those presented in Proposition 1. These expressions could in principle be recovered using small tax reform perturbations. The difference is that we now should consider small marginal tax changes *locally* at the two different earnings levels ωL^1 and ωL^S. For example, increasing T_1' has a negative behavioral effect on the labor supply of the $f_{\omega}(\omega)[1-F_{\gamma|\omega}(\hat{\gamma})]$ non-shifters, which are located at that particular income level. On the other hand, tax revenues are gained from all individuals with earnings in excess of that level, and these will also experience utility losses (but no additional labor supply distortion). Finally, the policy-maker has to take into account that a fraction of individuals will shift incomes to the other tax base when their tax bill as non-shifters increases. Note that the optimal income tax formula derived by Diamond (1998) is nested as a special case of equation (39), with $F_{\gamma|\omega}(\hat{\gamma}) = 0$ and $b(\omega, \gamma) = b(\omega)$.

A new element of (37) is the marginal shifting cost. If the shifters, in addition to the marginal tax rate, has to pay a positive marginal shifting cost it appears from (37) that this motivates a lower marginal tax rate than otherwise. However, if the total shifting cost (fixed cost + variable cost) becomes sufficiently large it will not be optimal for the social planner to allow for shifting. Recall that (37) is informative on the marginal tax rate on shifted income conditional on that there is a positive mass of shifters. We cannot rule out the possiblity that $F_{\gamma|\omega}(\hat{\gamma}) \equiv 0$; it is however beyond the scope of this article to numerically simulate the extended model of Section 6.

A simplification in our model is that shifting costs are exogenous from the government's point of view. In reality, shifting costs are partly endogenous to policy. However, to some extent we account for this policy endogeneity in this non-linear setting, since the government may affect the fixed shifting cost by varying $T_2(0)$ (undetermined sign), i.e. the lump sum component of the tax function for shifted income, which the shifter has to pay regardless of the labor supply choice. One could, of course, imagine other ways in which the government may affect shifting costs (e.g. by changing the legal requirements for corporations). In principle, policy endogeneity of this kind could be incorporated in the analysis by adding new choice variables to the social planner's maximization problem.

6.4 Revenue Maximizing Asymptotic Marginal Tax Rates

We now derive expressions for the revenue maximizing tax rates at the very top of the skill distribution. We therefore let $b(\omega, \gamma) = 0$ as $\omega \to \infty$. In words, this means that the policy-maker places no social value on the indirect utility of top-income earners.

The top of the skill distribution is approximated by a Pareto distribution of coefficient $a \geq 1$. In addition, we assume that the percentage change in the extra tax paid as a shifter converges to $\Delta T / T$. This quantity may either be positive or negative. Moreover, we let the extensive elasticities $\eta^S(\omega)$ and $\eta^1(\omega)$ converge to $\eta^S \leq 0$ and $\eta^1 \geq 0$ respectively. Note that $\eta^1 = -\eta^S \frac{F_\gamma}{1-F_\gamma}$ and F_γ is constant.

Proposition 4. *Assume* $\frac{f_\omega(\omega)\omega}{1-F_\omega(\omega)} \to a$, $\frac{\int_\omega^\infty \frac{\Delta T(\omega)}{T_1(\omega L_1)}\eta^S(\omega)F_{\gamma|\omega}d\omega}{1-F_\omega(\omega)} \to \frac{\Delta T}{T}\eta^S F_\gamma$, $e^1(\omega) \to e^1$, $e^S(\omega) \to e^S$ *and* $C'(\omega L^S) \to c$ *when* $\omega \to \infty$. *The revenue maximizing asymptotic tax rates* τ_P^* *and* τ_C^* *are then given by:*

$$\tau_P^* = \frac{1 - \frac{\Delta T}{T}\eta^1}{1 + a\frac{e^1}{1+e^1} - \frac{\Delta T}{T}\eta^1} \tag{40}$$

and

$$\tau_C^* = \frac{1 - \frac{\Delta T}{T}\eta^S - ca\frac{e^S}{1+e^S}}{1 + a\frac{e^S}{1+e^S} - \frac{\Delta T}{T}\eta^S}, \tag{41}$$

Proof. Because $\lim_{\omega\to\infty} b(\omega,\gamma) = 0$ and $\lim_{\omega\to\infty} F_{\gamma|\omega}(\omega) = F_\gamma$, the double integral in (39) is equal to $\int_\omega^\infty \int_{\hat{\gamma}(\omega)}^\infty f(\omega,\gamma)\, d\gamma d\omega = [1-F_\gamma][1-F(\omega)]$. Therefore, when ω tends to infinity, (39) yields:

$$\begin{aligned} \frac{T_1'}{1-T_1'} &= \{1 + \frac{1}{1-F_\gamma}\lim_{\omega\to\infty}\frac{\int_\omega^\infty \Delta T(\omega,\hat{\gamma})f(\omega,\hat{\gamma})d\omega}{1-F_\omega(\omega)}\}\frac{1}{a}(1+\frac{1}{e^1}) \\ &= \frac{1}{a}(1-\frac{\Delta T}{T}\eta^1)(1+\frac{1}{e^1}). \end{aligned} \tag{42}$$

Solving for T_P', we obtain (40). Similarly, when ω tends to infinity, (37) yields:

$$\frac{T_2'}{1-T_2'} = \frac{1}{a}(1-\frac{\Delta T}{T}\eta^S)(1+\frac{1}{e^S} - \frac{c}{1-T_2'}). \tag{43}$$

Solving for T_C', we obtain (41). □

It should be emphasized that the top marginal tax rates in Proposition 4 are expressed as functions of the *skill* distribution and not of the realized earnings distribution. Indeed, when there is only one (personal) tax base as in Diamond (1998) or Saez (2001), the Pareto parameter of the realized earnings distribution equals $a/(1+\epsilon^1)$. In the present context, this straightforward relationship does no longer necessarily hold. In general, the shape of the right tails of the non-shifted and shifted income distributions are likely to be endogenous to the tax policy.

We see that the revenue maximizing personal income tax rate τ_P^* is negatively related to the real labor supply elasticity of non-shifters e^1 and the Pareto coefficient a. This is in accordance with the results derived in the standard "single tax base" model. In our more general framework, the novelty is that τ_P^* also depends on income shifting along the extensive margin, which is captured by the term $\frac{\Delta T}{T}\eta^1$. If the tax payment as a shifter is larger than the tax payment as a non-shifter, ΔT is positive. Because $\eta^1 \leq 0$, this implies $\frac{\Delta T}{T}\eta^1 \leq 0$. Intuitively, if an increase in the personal marginal tax rate leads to larger tax revenues from the alternative tax base, there is a rationale for setting τ_P^* to a larger value than in the standard "single tax base" model. In particular, this implies that the labor supply elasticity is no longer a sufficient statistic to determine top marginal tax rates.

The optimal tax rate on shifted income, τ_C^*, depends negatively on the labor supply of shifters, through e^S. Once more, heterogeneous elasticities of shifters and non-shifters are important for the optimal rate structure. Since the extensive margin shifting elasticities η^1 and η^S have opposite signs, a positive ΔT will motivate a smaller tax rate on shifted income. The term $ca\frac{e^S}{1+e^S}$ is an additional feature of (41). A positive c will reduce τ_C^* as the total labor supply distortion of shifters is given by $\tau_C^* + c$. Of course, the revenue-maxizing top marginal tax rate (41) is derived in a setting that does not account for capital income supply. Capital income supply considerations are expected to lead to further reductions in τ_C^*. Consequently, the value of τ_C^* provided in Proposition 4 may be regarded as an upper bound for the top marginal tax rate on shifted incomes in an even more general framework.

It is already well-known that the so-called taxable income elasticity, which reflects the percentage change in personal income to a percentage change in the personal net-of-tax rate, falls short of being a valid sufficient statistic for the efficiency cost of earnings taxation in the presence of income shifting (see, e.g., Slemrod, 1998; Saez et al., 2012; Chetty, 2009; and Doerrenberg et al., 2015). The taxable income elasticity, estimated in the spirit of Feldstein (1995), typically encompasses both shifting responses and real responses. However, if an increase in the personal tax rate leads to an increase in corporate tax revenues it is not sufficient to consider the response in the personal income tax base only. Interestingly, Saez et al. (2012) derive an expression for the revenue maximizing personal tax rate for an exogenous share of income shifted (p.11, equation 11). Proposition 4 formalizes the potential importance of fiscal externalities in a novel way, which endogenizes income shifting.

7 Concluding discussion

The optimal tax literature has modelled income shifting as a decision along the intensive margin. However, income shifting involves significant fixed costs, which give rise to an important extensive margin. In this article, we show that the distinction between the intensive and extensive margins has crucial policy implications. We consider a population of agents differing in terms of productivities, labor supply elasticities, and abilities to shift income. In the extensive margin model the distinction between income creation and income shifting breaks down and the social planner should not in general combat shifting. In particular, numerical simulations of a linear tax model suggest that the social planner should allow for income shifting if elasticities are heterogeneous in the population. We demonstrate that the qualitative conclusions drawn from the simple linear tax model carry over to a model with two fully non-linear tax schedules.

Needless to say, tax policy design includes considerations that we abstracted from, such as capital income accumulation and horizontal equity concerns. Still, our model has strong policy relevance because it casts a new light upon the highly controversial issue of tax rate differentiation. The highlighted mechanisms should be kept in mind when thinking about recent policy trends. For example, the present gap between the labor income marginal tax rate of high income earners and the dividend tax rate of owners of closely held companies is 32 percentage points in Sweden (70% vs. 38% when accounting for payroll taxes and the corporate tax). By contrast, other countries, like Norway, are closing the gap motivated by income shifting concerns.

Our analysis has important implications for future empirical work in the area. First, empirical evidence on the nature of the shifting costs and the correlation between earnings abilities, elasticities and shifting costs is desirable. Shifting costs will crucially depend on the institutional setting in place. Moreover, there are intertemporal aspects of the shifting decision, such that individuals may face a large fixed cost in the first year of shifting, but smaller fixed costs in future years. To keep the analysis sufficiently tractable and highlight the main forces at stake, we have abstracted from such issues in our article. Static models can indeed always be regarded as reduced forms of dynamic models, in which utilities would be computed along the life cycle. However, intertemporal aspects must explicitly be addressed in empirical work. A second implication of our results is that empirical researchers should focus less on *the* (homogenous) labor supply elasticity, and pay more attention to *heterogeneity* across individuals and groups.

References

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

Alstadsæter, A. and M. Jacob (2017). Who participates in tax avoidance? Evidence from Swedish microdata. *Applied Economics, forthcoming*.

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

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.

Doerrenberg, P., A. Peichl, and S. Siegloch (2015). The elasticity of taxable income in the presence of deduction possibilities. *Journal of Public Economics*.

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

Feldstein, M. (1995). The effect of marginal tax rates on taxable income: A panel study of the 1986 tax reform act. *Journal of Political Economy 103*(3), 551–72.

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, THEMA Working Paper n2016-04. Universit de Cergy-Pontoise, France.

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

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 (2012). Redistributive taxation in the Roy model. *The Quarterly Journal of Economics*.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies 68*(1), 205–29.

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 (2016). A simpler theory of optimal capital taxation. Working Paper 22664, National Bureau of Economic Research.

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

Selin, H. and L. Simula (2017). Income shifting as income creation? The intensive vs. the extensive shifting margin. Technical report, IFAU working paper 2017:10.

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

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. (1998). Methodological issues in measuring and interpreting taxable income elasticities. *National Tax Journal 51*(n. 4), 773–88.

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

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 : Proof of Proposition 1

The social planner solves the following problem:

$$\max_{\tau_P,\tau_C,G} \int_\kappa \int_0^{\hat{\gamma}(\kappa)} g(\kappa,\gamma) V_2(\kappa) f(\kappa,\gamma)\, d\gamma d\kappa + \int_\kappa \int_{\hat{\gamma}(\kappa)}^{\infty} g(\kappa,\gamma) V_1(\kappa,\gamma) f(\kappa,\gamma)\, d\gamma d\kappa, \tag{44}$$

subject to:

$$\tau_2 \int_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega L_2(\kappa) f(\kappa,\gamma) d\gamma d\kappa + \tau_1 \int_\kappa \int_{\hat{\gamma}(\omega,\varepsilon)}^{\infty} \omega L^P(\kappa) f(\kappa,\gamma) d\gamma d\kappa - R - G = 0. \tag{45}$$

We let λ be the Lagrange multiplier of the budget constraint (45). The derivative of (44) with respect to τ_1 is:

$$\int_\kappa \int_{\hat{\gamma}(\kappa)}^{\infty} g(\kappa,\gamma) \omega L^P(\kappa) f(\kappa,\gamma) d\gamma d\kappa. \tag{46}$$

We used the fact that $V_1(\kappa) = V^C(\kappa, \gamma)$ for $\gamma = \hat{\gamma}(\kappa)$. We now compute the derivative of the budget constraint (45) with respect to τ_1. We obtain:

$$\int_\kappa \int_{\hat{\gamma}(\kappa)}^{\infty} \omega L_1(\kappa) f(\kappa, \gamma) d\gamma d\kappa + \tau_1 \int_\kappa \int_{\hat{\gamma}(\kappa)}^{\infty} \omega \frac{\partial L_1(\kappa)}{\partial \tau_1} f(\kappa, \gamma) d\gamma d\kappa \\ + \int_\kappa [\tau_2 \omega L_2(\kappa) - \tau_1 \omega L_1(\kappa)] \frac{\partial \hat{\gamma}}{\partial \tau_1} f(\kappa, \hat{\gamma}) d\kappa. \tag{47}$$

From Lemma 1, we know that $\partial\hat{\gamma}/\partial\tau_1 = \omega L_1$. We now write (46) $-$ λ(47) $= 0$, rearrange and use the definition of $e(\varepsilon)$ to obtain (18).

To obtain (20), we compute the derivative of the social objective with respect to τ_2. Using the indifference condition at $\hat{\gamma}$, we obtain:

$$\int_\kappa \int_0^{\hat{\gamma}(\kappa)} g(\kappa, \gamma) \omega L_2(\kappa) f(\kappa, \gamma) d\gamma d\kappa. \tag{48}$$

We now compute the derivative of the budget constraint (45) with respect to τ_2:

$$\int_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega L_2(\kappa) f(\kappa, \gamma) d\gamma d\kappa + \tau_2 \int_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega \frac{\partial L_2(\kappa)}{\partial \tau_2} f(\kappa, \gamma) d\gamma d\kappa \\ + \int_\kappa [\tau_2 \omega L_2(\kappa) - \tau_1 \omega L_1(\kappa)] \frac{\partial \hat{\gamma}}{\partial \tau_2} f(\kappa, \hat{\gamma}) d\kappa. \tag{49}$$

From (12) we know that $\frac{\partial\hat{\gamma}}{\partial\tau_2} = -\omega L_2$. We now write (48) $-$ λ(49) $= 0$, rearrange and use the definition of $e(\varepsilon)$ to obtain (20).

Appendix B: Calibration of the Numerical Illustration

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; they must be calibrated somehow. Our strategy is to calibrate the shifting cost distribution by choosing μ_γ and σ_γ in such a way that the actual share of 'shifters' is reproduced, conditional

Table 2: Parameter values used in the simulations

	$\log(\omega)$	$\log(\gamma)$
μ	10.194	11.795
σ	0.302	0.302

Note: Moments of $\log(\omega)$ have been picked from LINDA data as of 2009, whereas the moments of $\log(\gamma)$ have been calibrated.

on the actual Swedish wage distribution, the actual Swedish tax system, and a given distribution of elasticities. Two parameters are unknown to us. For convenience, we assume that the variances of $\log(\omega)$ and $\log(\gamma)$ are the same.[16] Ultimately, we therefore solely calibrate μ_γ.

We set our target, i.e. the actual fraction of shifters, to be 5 %. Alstadsæter 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.[17] 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 impose our baseline assumption regarding the labor supply elasticities; the elasticity is 0.1 for the lowest-skilled individual and 0.5 for the highest-skilled individual, and the elasticity is linearly increasing in ω. Then we find that the fraction of shifters is 5% when $\mu_\gamma = 11.795$. The parameters used in the simulations are summarized in Table 2.

[16] Denoting by e the natural exponential function, the correlation coefficient for the transformed distributions is given by $(e^{\rho\sigma_\omega\sigma_\gamma} - 1)/\sqrt{[e^{\sigma_\omega^2} - 1][e^{\sigma_\gamma^2} - 1]}$. When $\sigma_\omega = \sigma_\gamma$, the correlation coefficient for the transformed distributions is always relatively close to ρ, and identical for $\rho = 0$ and $\rho = 1$.

[17] If an 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.

Appendix C: Proof of Lemma 2

To simplify notations, we below drop the functions' arguments. For agents choosing $A > 0$, it follows from (4) that $dV/d\omega = (1 - T_1')L$ which, combined with (23), yields $1 - T_1' = v'/\omega + \lambda$. Hence,

$$\frac{dV}{d\omega} = \left[\frac{v'(L)}{\omega} + \lambda\right] L, \tag{50}$$

where λ is the Kuhn-Tucker multiplier associated with $A \leq \omega L$. When $A < \omega L$, $\lambda = 0$ and (50) reduces to:

$$\frac{dV}{d\omega} = \frac{v'(L)}{\omega} L. \tag{51}$$

When $A = \omega L$ instead,

$$V = \omega L - T_1(0) - T_2(\omega L) - C(\omega L) - \gamma - v(L), \tag{52}$$

from which:

$$\frac{dV}{d\omega} = \left[1 - T_2' - C'\right] L. \tag{53}$$

Plugging (25) into the latter, we obtain (51) again. By definition of $\tilde{V}(\omega) \equiv V(\omega, \gamma) + \gamma$, $d\tilde{V}/\omega = dV/d\omega$. We thus obtain (33). For non-shifters, (34) follows directly by differentiation of (5).

Appendix D

Problem 1. *Find $\tilde{V}^S(\omega)$, $V^1(\omega)$, $L^S(\omega)$, $L^1(\omega)$, and $A(\omega)$ which maximizes the social objective* (32) *subject to (i) the incentive compatibility conditions* (33) *and* (34)*, (ii) the tax revenue constraint* (35) *with $R = 0$, and (iii) the inequality constraint*

$$A(w) \leq \omega L_2(w). \tag{54}$$

Lagrangian and First-Order Conditions

We form the Lagrangian from the objective (32) and the two sums of incentive compatibility constraints defined by (33) and (34) and the resource constraint (35). At a given skill level ω, we denote the Lagrange multiplier associated with the incentive compatibility constraints by $\lambda(\omega)$ and $\lambda^1(\omega)$ respectively. μ refers to the Lagrange multiplier of the budget constraint (35). $\lambda^A(\omega)$ denotes the Lagrange multiplier of constraint (54).

By integration by parts, we obtain:

$$\int_0^\infty \lambda(\omega)[\frac{d\tilde{V}^S(\omega)}{d\omega} - \frac{v'(L^S)}{\omega}L^S]d\omega = \lim_{\omega\to\infty} \lambda(\omega)\tilde{V}^S(\omega) - \lambda(0)\tilde{V}(0) \\ - \int_0^\infty \lambda'(\omega)\tilde{V}^S(\omega) - \int_0^\infty \lambda(\omega)[\frac{v'(L^S)}{\omega}L^S]d\omega, \quad (55)$$

$$\int_0^\infty \lambda^1(\omega)[\frac{dV^1(\omega)}{d\omega} - \frac{v'(L^1)}{\omega}L^1]d\omega = \lim_{\omega\to\infty} \lambda^1(\omega)V^1(\omega) - \lambda^1(0)V^1(0) \\ - \int_0^\infty \lambda^1(\omega)V^1(\omega) - \int_0^\infty \lambda^1(\omega)[\frac{v'(L^1)}{\omega}L^1]d\omega, \quad (56)$$

where $\lim_{\omega\to\infty} \lambda(\omega)\tilde{V}^S(\omega) - \lambda(0)\tilde{V}^S(0) = \lim_{\omega\to\infty} \lambda^1(\omega)V^1(\omega) - \lambda^1(0)V^1(0) = 0$ due to transversality conditions $\lim_{\omega\to\infty} \lambda(\omega) = \lambda(0) = \lim_{\omega\to\infty} \lambda^1(\omega) = \lambda^1(0) = 0$. Combining the social objective (32), the reformulated conditions for incentive compatibility (55) and (56), together with the resource constraint (35), the Lagrangian may be rewritten as:

$$\mathcal{L} = \int_0^\infty \int_0^{\hat{\gamma}(\omega)} g(\omega,\gamma)[\tilde{V}^S(\omega) - \gamma]f(\omega,\gamma)\,d\gamma d\omega + \int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty g(\omega,\gamma)V^1(\omega)f(\omega,\gamma)\,d\gamma d\omega \\ - \int_0^\infty \lambda'(\omega)\tilde{V}^S(\omega) - \int_0^\infty \lambda(\omega)\frac{v'(L^S)}{\omega}L^S d\omega - \int_0^\infty \lambda^1(\omega)V^1(\omega) - \int_0^\infty \lambda^1(\omega)\frac{v'(L^1)}{\omega}L^1 d\omega \\ + \mu\int_0^\infty \int_0^{\hat{\gamma}(\omega)} [\omega L^S - v(L^S) - \tilde{V}^S(\omega)]f(\omega,\gamma)\,d\gamma d\omega \\ + \mu\int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty [\omega L^1 - v(L^1) - V^1(\omega)]f(\omega,\gamma)\,d\gamma d\omega + \lambda^A(\omega)[A - \omega L^S]. \quad (57)$$

Note that we can write $\mathcal{L} = \int_0^\infty \mathcal{L}(\omega)d\omega$. Accordingly, we can differentiate $\mathcal{L}(\omega)$ with respect to $\tilde{V}^S(\omega)$, $V^1(\omega)$, $A(\omega)$, $L^S(\omega)$ and $L^1(\omega)$ to arrive at necessary conditions that hold at given levels of ω. The first-order condition with respect to $\tilde{V}^S(\omega)$ is:

$$\int_0^{\hat{\gamma}(\omega)} [g(\omega,\gamma) - \mu]f(\omega,\gamma)\,d\gamma - \lambda'(\omega) + \mu\Delta T(\omega)f(\omega,\hat{\gamma}) = 0, \quad (58)$$

where $\Delta T(\omega) = [\omega L^S - v(L^S) - \tilde{V}^S(\omega)] - [\omega L^1 - v(L^1) - V^1(\omega)]$ is *the extra tax paid* *by the marginal shifter*. When writing down (58), we have used the fact that $\hat{\gamma}(\omega) = \tilde{V}^S(\omega) - V^1(\omega)$, which in turn implies $\partial\hat{\gamma}(\omega)/\partial\tilde{V}(w) = 1$. In a similar way, the first-order condition with respect to $V^1(\omega)$ reads:

$$\int_{\hat{\gamma}(\omega)}^{\infty} [g(\omega,\gamma) - \mu] f(\omega,\gamma)\, d\gamma - \lambda^{1\prime}(\omega) - \mu \Delta T(\omega) f(\omega,\hat{\gamma}) = 0. \tag{59}$$

The first-order condition with respect to $L^S(\omega)$ implies that, for all values of ω,

$$\lambda(\omega)[-\frac{v''(L^S)}{\omega}L^S - \frac{v'(L^S)}{\omega}] + \mu \int_0^{\hat{\gamma}} (\omega - v'(L^S)) f(\omega,\gamma) d\gamma - \lambda^A(\omega)\omega = 0. \tag{60}$$

Last, the first-order condition with respect to $A(\omega)$ yields:

$$-\mu C'_A \int_0^{\hat{\gamma}} f(\omega,\gamma) d\gamma + \lambda^A(\omega) = 0. \tag{61}$$

Case (i): Constraint $A \leq \omega L$ is binding

Combining (60) and (61), we obtain:

$$\lambda(\omega)\left[-\frac{v''(L^S)}{\omega}L^S - \frac{v'(L^S)}{\omega}\right] + \mu \int_0^{\hat{\gamma}} [\omega - v'(L^S) - \omega C'(\omega L^S)] f(\omega,\gamma) d\gamma = 0. \tag{62}$$

From (25), $v'(L^S) + \omega C' = (1 - T'_C)\omega$. Using this relationship and dividing (62) by $v'(L^S)$ and rearranging, we obtain:

$$\frac{T'_C(\omega L^S)}{1 - T'_C(\omega L^S) - C'(\omega L^S)} = \frac{\lambda(\omega)}{\omega\mu \int_0^{\hat{\gamma}} f(\omega,\gamma) d\gamma}\left[1 + \frac{1}{e^S(\omega)}\right]. \tag{63}$$

Using the same steps, the first-order condition with respect to L^1 can be written as:

$$\frac{T'_P(\omega L^1)}{1 - T'_P(\omega L^1)} = \frac{\lambda^1(\omega)}{\omega\mu \int_{\hat{\gamma}}^{\infty} f(\omega,\gamma) d\gamma}\left[1 + \frac{1}{\epsilon^1(\omega)}\right]. \tag{64}$$

Case (ii): Constraint $A \leq \omega L$ is not binding

It follows from (61) that:

$$\mu C'_A \int_0^{\hat{\gamma}} f(\omega,\gamma) d\gamma = \mu C'(A) f_\omega(\omega) F_{\gamma|\omega}(\hat{\gamma}) = 0. \tag{65}$$

We have: $\mu > 0$. Suppose first that $C'(A) > 0$ at skill level ω. In this case, the number of shifters at that skill level must be zero in the social optimum; the social planner should set $F_{\gamma|\omega}(\hat{\gamma}) = 0$. Suppose instead that $C'(A) = 0$ at skill level ω. It then follows from (24) that the two marginal tax rates should be equalized; i.e., $T'_P = T'_C$.

Finding expressions for λ and λ^1

Following Scheuer (2014), Appendix A.3, we integrate equations (58) and (59) over the whole support of ω, add them, and use the fact that the sum is equal to 0. Use in addition the transversality condition $\lim_{\omega\to\infty} \lambda(\omega) = \lambda(0) = \lim_{\omega\to\infty} \lambda^1(\omega) = \lambda^1(0) = 0$, we get:

$$\begin{aligned}
&\int_0^\infty \int_0^{\hat{\gamma}(\omega)} [g(\omega,\gamma) - \mu] f(\omega,\gamma)\, d\gamma d\omega - \int_0^\infty \lambda'(\omega) d\omega + \mu \int_0^\infty \Delta T(\omega) f(\omega,\hat{\gamma}) d\omega \\
&\quad + \int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty [g(\omega,\gamma) - \mu] f(\omega,\gamma)\, d\gamma d\omega - \int_0^\infty \lambda^{1\prime}(\omega) d\omega - \mu \int_0^\infty \Delta T(\omega) f(\omega,\hat{\gamma}) d\omega \\
&\qquad = \int_0^\infty \int_0^\infty [g(\omega,\gamma) - \mu] f(\omega,\gamma)\, d\gamma d\omega = \bar{g} - \mu = 0,
\end{aligned} \tag{66}$$

where $\bar{g} = \int_0^\infty \int_0^\infty [g(\omega,\gamma)] f(\omega,\gamma)\, d\gamma d\omega$ is the average social marginal welfare weight in the population. Integrating equations (58) and (59) between 0 and ω, using the relationship given by (66) and the fact that $\lambda(\omega) = \int_0^\omega \lambda'(\omega) d\omega$ and $\lambda^1(\omega) = \int_0^\omega \lambda^{1\prime}(\omega) d\omega$, we obtain:

$$\lambda(\omega) = \int_0^\omega \int_0^{\hat{\gamma}(\omega)} [g(\omega,\gamma) - \bar{g}] f(\omega,\gamma)\, d\gamma d\omega + \bar{g} \int_0^\omega \Delta T(\omega) f(\omega,\hat{\gamma}) d\omega = 0, \tag{67}$$

$$\lambda^1(\omega) = \int_0^\omega \int_{\hat{\gamma}}^\infty [g(\omega,\gamma) - \bar{g}] f(\omega,\gamma)\, d\gamma d\omega - \bar{g} \int_0^\omega \Delta T(\omega) f(\omega,\hat{\gamma}) d\omega = 0. \tag{68}$$

Because $\lim_{\omega\to\infty} \lambda(\omega) = \lim_{\omega\to\infty} \lambda^1(\omega) = 0$, we can rewrite (67) and (68) as:

$$\lambda(\omega) = \int_\omega^\infty \int_0^{\hat{\gamma}(\omega)} [\bar{g} - g(\omega,\gamma)] f(\omega,\gamma)\, d\gamma d\omega - \bar{g} \int_\omega^\infty \Delta T(\omega) f(\omega,\hat{\gamma}) d\omega = 0, \tag{69}$$

$$\lambda^1(\omega) = \int_{\omega}^{\infty} \int_{\hat{\gamma}}^{\infty} [\bar{g} - g(\omega,\gamma)] f(\omega,\gamma)\, d\gamma d\omega + \bar{g} \int_{\omega}^{\infty} \Delta T(\omega) f(\omega,\hat{\gamma}) d\omega = 0. \tag{70}$$

Optimal tax rules in Proposition 3

Combining (64), (66), (70) while using the definition $b(\omega,\gamma) = g(\omega,\gamma)/\mu$ gives (39). Similarly, combining (63), (66), (69) while using the definition $b(\omega,\gamma) = g(\omega,\gamma)/\mu$ gives (37).