Income Creation and/or Income Shifting?
The Intensive vs. the Extensive Shifting Margins[*]

Håkan Selin[†] and Laurent Simula [‡]

January 19, 2017

Abstract

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 carries over to a model with two fully non-linear tax schedules.

Keywords: Income Shifting, 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 Grenoble Alpes (UGA) and Grenoble Applied Econ Lab, UMR GAEL (UGA-CNRS-INRA Grenoble INP), CS 40700, 38058 Grenoble Cedex 09, France. Email: Laurent.Simula@univ-grenoble-alpes.fr

1 Introduction

Income tax systems are often complex with plenty of loopholes and opportunities to shift incomes across tax bases. Accordingly, tax rate changes will not only foster labor supply and savings responses, but also avoidance activities. In countries with separate taxation of labor and capital incomes, taxpayers may start up closely held corporations and transfer income between the two tax bases (see, e.g., Pirttilä and Selin, 2011, Alstadsæter and Jacob, 2016, and Harju and Matikka, 2016). In more "comprehensive" tax systems, like the US, some taxpayers switch between the personal and corporate tax bases (see e.g. Gordon and Slemrod, 1998). The influential taxonomy of tax reform responses of Slemrod (1995, p. 179) sharply distinguishes between 'income creation' and 'income shifting' responses to taxes, where the latter are "not likely to be accompanied by an increase in national income".[1]

In parallel with the established dichotomy between income creation and income shifting it has become standard in the public finance literature to model income shifting as a decision along the *intensive margin* (see, e.g., Fuest and Huber, 2001, Christiansen and Tuomala, 2008, Piketty et al., 2014, Piketty and Saez, 2013 or Hermle and Peichl, 2015). It is typically assumed that the cost of shifting is smoothly increasing, at an increasing rate. Individuals choose how much labor to supply and how much labor income to shift.[2] In this setting, Piketty et al. (2014) as well as Piketty and Saez (2013) forcefully make the point that governments should remove incentives to shift labor earnings into more leniently taxed bases if all income stems from labor effort. Shifting activities are completely wasteful from the society's point of view.

There is, however, practical evidence of a variety of fixed costs associated with income shifting, such as the costs of gathering information about the tax law or setting up a closely held corporation. These fixed costs give rise to an *extensive margin*. When we in the most simple setting replace the convex cost with a fixed cost, we find that the traditional dichotomy between income creation and income shifting breaks down. Moreover, the policy implications are different. In particular, if labor supply elasticities differ in the

---

[1] According to Slemrod's taxonomy, the conceptual difference between the two is that real responses reflect substitution between different consumption goods (among which leisure time) while avoidance activities do not affect the individual's real consumption basket (holding utility constant). However, Slemrod (1995), Slemrod (2001) and Agell and Persson (2000) acknowledge that real responses and avoidance responses also may interact in modern economies.

[2] 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).

population the social planner should potentially allow for some income shifting.

We analyze a simple static economy in which all income stems from labor effort, and the labor income can be shifted to an alternative tax base to a resource cost, which is fixed and/or variable. To focus on the income shifting mechanism *per se*, we deliberately neither model capital accumulation nor tax competition. In other words, we place ourselves in the position which is the least favorable to shifting. Agents potentially differ with respect to three characteristics: skills, taste for effort and cost of income shifting. Given the tax system, they simultaneously choose how much effort to supply and how much income tax shift, if any. The benevolent social planner design taxes with a view of maximizing a weighted sum of individual utilities.

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 respectively, 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 personal income. This mechanism clearly differs from what would be allowed by the introduction of additional tax brackets.

Intuitively, when elasticities differ in the population, the revenue-maximizing tax rates applying to these two populations will also differ, depending on the correlation between skills, shifting costs and elasticities. To investigate our analytical results numerically, we calibrate our model to the Swedish economy, and we find that non-negligible welfare gains can be achieved thanks to income shifting. In our benchmark scenario, it is socially optimal to set the personal tax rate 9 to 15 percentage points higher than the tax rate on shifted income.

A concern, however, could be that our results are contingent on the focus on (i) a pure extensive shifting margin as well as on (ii) the linear restrictions imposed on the tax instruments. We therefore extend our simple model and show that our results are robust to: (i) the combination of a fixed and convex shifting costs; (ii) the relaxation of the linear tax assumptions. In this extended framework, we also derive tax revenue maximizing

asymptotic tax rates. To the best of our knowledge, revenue maximizing asymptotic tax rates, which endogenizes income shifting, have not earlier been presented in the optimal tax literature.

The article is organized as follows. Section 2 introduces the main blocks of the model. Section 3 illustrate the intensive marginal logic in the simplest way. Section 4 casts light in the consequences of allowing for income shifting allong a pure extensive margin and provides numerical simulations. Section 5 shows that our results are robust to various extensions and discuss the implications for the revenue maximizing tax rates at the top of the skill distribution.

Related Literature

As already mentioned, our work is closely related to the textbook model presented by Piketty and Saez (2013, Section 4). They 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). In a model with heterogeneity in skills only, the government should stop income shifting if it is costless to 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 skill, shifting abilities and consumption preferences, and agents may shift income between the tax bases at a smooth resource cost. The optimal tax formulas differ from the standard ones since 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) under the informational assumption that the government cannot observe the true amounts of labor and capital income, and the agents can shift income between the two tax bases at a convex cost. With heterogeneity in the skill dimension and additively separable preferences, a positive proportional capital income tax will be desirable when income shifting opportunities are available in the economy. In the absence of income shifting, there is no such case for a capital income

4

tax when people differ in the skill dimension only.[3]

Finally, our extensive margin model, where individuals endogenously sort to different tax schedules, relates to a growing body of literature on occupational choice and differential tax schedules (Rothschild and Scheuer, 2012 and Gomes et al., 2014). More specifically, it connects the literature on the taxation of entrepreneurial income (Parker, 1999 and Scheuer, 2014). Our focus is different though. 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.

2 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.

2.1 Sources of Heterogeneity in the Population

We consider a population differing with respect to three dimensions of heterogenity: skills ω, taste for work effort ϵ[4], and the propensity to shift incomes from the personal to the corporate income tax base. The latter is captured through a cost parameter γ. 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 unable to observe nor recover the type of a specific individual, precluding personalized lump-sum taxes.

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

---

[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 optimum, wealthy households will face a positive marginal tax rate, which is equal for labor income and capital income. Poor households, on the other hand, will face a marginal tax rate on capital income which is larger than that on labor income.

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

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_P for non-shifted earnings and a tax function T_C for shifted earnings. The first tax base can be thought of as personal income and the second one as corporate income; hence the P and C subscripts.

2.2 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 her income subject to the personal income tax, from ωL to $\omega L - A$ at a cost $\Gamma(A, \gamma)$. We refer to this possibility as *income shifting*. As emphasized in the introduction, this cost may correspond to a fixed cost and/or depend on how much earnings are shifted. A general specification is:

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

where $C(A)$ is a continuous cost, non-decreasing and convex in the shifted amount A (i.e., $C'_A \geq 0$ and $C''_{AA} \geq 0$), whereas γ is a fixed cost of shifting. Most of the previous literature has focused on the case where $\Gamma(A; \gamma) = C(A)$. By contrast, we investigate the implications of a more general –and more empirically relevant– cost structure.

Overall, an individual pays tax liability $T_P(\omega L - A) + T_C(A)$ and thus receives net income:[5]

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

The utility function (1) is quasilinear, linear with respect to net income. Consequently, we

---

[5]A more general specification would allow for endogeneous capital income supply, Q, such that the capital tax payment would be $T_C(Q + A)$. 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.

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:

$$V(\omega, \epsilon, \gamma) = \max_{L, A} \{\omega L - T_P(\omega L - A) - T_C(A) - \Gamma(A, \gamma) - v(L; \epsilon)\}. \tag{4}$$

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

$$V^P(\omega, \epsilon) = \max_L \{\omega L - T_P(\omega L) - v(L; \epsilon)\}, \tag{5}$$

$$V^C(\omega, \epsilon, \gamma) = \max_L \{\omega L - T_P(\omega L - A) - T_C(A) - \Gamma(A; \gamma) - v(L; \epsilon)\}. \tag{6}$$

For any given individual, the first optimization program provides the maximum utility $V^P(\omega, \gamma)$ that can be obtained in the absence of any income shifting. We denote by $L^P(\omega, \epsilon)$ its solution in L. The second optimization program provides the maximum utility $V^C(\omega, \epsilon, \gamma)$ when at least some earnings are shifted. We denote by $L^C(\omega, \epsilon, \gamma)$ its solution in L.

2.3 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:

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

subject to the following revenue constraint:

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

R is a tax revenue requirement that does not enter the individuals' utility function. When it is set equal to zero, the tax policy is purely redistributive.

3 The Intensive Margin

In this section, we illustrate the intensive margin logic in the simplest way and let $\gamma = 0$ for everyone. Hence, $\Gamma(A; \gamma) = C(A)$. All agents in the economy therefore face the same convex shifting cost. At the end of this section, we will comment on the consequences of allowing for heterogeneity in the convex cost. To simplify notations, we drop the parameter γ and write $\kappa = (\omega, \varepsilon)$ and $d\kappa = (d\omega, d\epsilon)$.

To make the analysis more transparent, we assume away the corner solution $A = \omega L$ in this section. This is in line with the previous literature modelling income shifting as a pure intensive margin phenomenon. Following Piketty and Saez (2013), we consider that personal income is taxed linearly, while shifted income is taxed proportionnally. Denoting by τ_P and τ_C the marginal tax rates on personal income and shifted income respectively, we obtain $T_P = G + \tau_P(\omega L - A)$ and $T_C = \tau_C A$. G is a demogrant; when $G < 0$, the policy-maker distributes a basic income to each agent. In our setting, all income primarily stems from labor. Hence, shifting can only occur in one direction, from the personal to the corporate base. Any $\tau_C \geq \tau_P$ is associated with the same outcome, i.e., the absence of shifting; hence, there is no loss of generality in focusing on $\tau_P \geq \tau_C$.

Given this set of assumptions, the individual's first-order conditions are idenpendent and can be written as:

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

$$C'_A(A) = \tau_P - \tau_C \tag{10}$$

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 of doing so, given by the difference between the two marginal tax rates, equates the marginal cost of doing so. Given this structure, we can formulate the following proposition.

Proposition 1. *Suppose $\Gamma(A; \gamma) = C(A)$, $T_P = G + \tau_P(\omega L - A)$ and $T_C = \tau_C A$. In the social optimum, $\tau_P = \tau_C$.*

Proof. The social planner chooses the tax rates (τ_P, τ_C) and the lump-sum income G so as to maximize the social welfare functional (7) subject to the tax revenue constraint (8). Denoting the shadow price of public funds by λ, the Lagrangian of the optimization problem

8

is given by:

$$\iint_\kappa \{g(\kappa)V(\kappa) + \lambda\left[\tau_P \omega L(\omega,\varepsilon) - (\tau_P - \tau_C)A - G - R\right]\}f(\kappa)d\kappa. \tag{11}$$

To simplify notations, we omit the arguments of the different functions. We denote by $b(\kappa) = g(\kappa)/\lambda$ the net social marginal valuation of income of a κ-individual. The first-order conditions with respect to τ_P, τ_C and G are respectively:

$$\iint_\kappa \left[b\frac{\partial V}{\partial \tau_P} + \omega L - A + \tau_P\frac{\partial \omega L}{\partial \tau_P} - (\tau_P - \tau_C)\frac{\partial A}{\partial \tau_P}\right]f(\kappa)d\kappa = 0, \tag{12}$$

$$\iint_\kappa \left[b\frac{\partial V}{\partial \tau_C} + A - (\tau_P - \tau_C)\frac{\partial A}{\partial \tau_C}\right]f(\kappa)d\kappa = 0, \tag{13}$$

$$\iint_\kappa \left[b - 1\right]f(\kappa)d\kappa = 0. \tag{14}$$

From (14), we infer that the average value of b over the population, denoted $\bar{b}$, is equal to 1. The first-order condition with respect to τ_C can be re-written as:

$$\iint_\kappa \left[-bA + A - (\tau_P - \tau_C)\frac{\partial A}{\partial \tau_C}\right]f(\kappa)d\kappa = 0. \tag{15}$$

Using the fact that $\bar{b} = 1$ and the definition of the covariance, we obtain:

$$\tau_P - \tau_C = -\frac{cov(A, b)}{\int_\kappa \frac{\partial A}{\partial \tau_C}f(\kappa)d\kappa}. \tag{16}$$

Because all individuals face the same convex cost function $C(A)$, it follows from (10) that everyone chooses the same A. When A is constant over the population, $cov(A, b) = 0$. By (16), $\tau_P - \tau_C = 0$ in the social optimum. $\qquad\square$

Proposition 1 captures the essence of the prevailing view on income shifting in modern public finance. The intuition underlying the result is the following. Suppose $\tau_P > \tau_C$. It is thus optimal for people to partly shift income. Now, let us investigate the effects of an small increase $\partial \tau_C$ in τ_C.

- To start with, collected taxes increase by $E^+ = A \times \partial \tau_C$ (dollars). This can be referred to as the "mechanical" effect of the tax reform.

- For each individual, the extra taxes paid decrease utility. Given the quasilinear pref-

erences, utility is reduced by $A \times \partial \tau_C$; and thus social utility by $g \times A \times \partial \tau_C$ (expressed in "utils"). The shadow price λ of the budget constraint is the unit of count in welfare. Recalling that $b = g/\lambda$, this loss in social utility can be transformed into a monetary loss for the state, equal to $E^- = b \times A \times \partial \tau_C$ (dollars).

- The increase in τ_C induces a behavioral response. From (9) and (10), we see that the the amount of effort only depends on τ_P. Therefore, total earnings are not impacted. However, the change in τ_C induces each agent to reduce shifted income by $\partial A/\partial \tau_C \times \partial \tau_C$ and increases the personal income tax base accordingly. The gain in terms of collected taxes amounts to $E^{++} = -(\tau_P - \tau_C) \times \partial A/\partial \tau_C \times \partial \tau_C$.

The global impact of the tax reform is given by:

$$\iint_\kappa (E^+ + E^{++} - E^-)\, \partial \tau_C\, f(\kappa)d\kappa = \iint_\kappa \left[A - (\tau_P - \tau_C)\frac{\partial A}{\partial \tau_C} - bA \right] \partial \tau_C\, f(\kappa)d\kappa,$$

$$= -(\tau_P - \tau_C)\partial \tau_C \iint_\kappa \frac{\partial A}{\partial \tau_C} f(\kappa)d\kappa. \quad (17)$$

The second line uses the fact that the social marginal valuation of income b is equal to 1 on average. Because $\partial A/\partial \tau_C < 0$, increasing τ_C unambiguously increases social welfare. In essence, because labor supply is unaffected, the effort spent on tax planning is a pure waste from the society's point of view. A raise in τ_C leaves the total pie to share in the economy unaffected, but induces people to invest less in costly tax planning. Therefore, in the social optimum, the social planner should stop shifting by setting $\tau_P = \tau_C$.

In general, Proposition 1 is no longer valid when all agents do not face the same convex shifting cost function C. To see this, we introduce an additional dimension of heterogeneity θ, which affects the cost of shifting. More precisely, $C(A; \theta)$ with $\partial C(A; \theta)/\partial \theta > 0$. We denote by $f_\theta(\theta)$ the marginal density of θ and allow this parameter to be arbitrarily correlated with the other heterogeneity parameters. Adjusting the steps in the proof of Proposition 1, we obtain:

$$\tau_P - \tau_C = \frac{cov[A(\theta), b(\kappa, \theta)]}{-\iint_\kappa \int_\theta \frac{\partial A}{\partial \tau_C} f(\kappa, \theta)d\kappa d\theta} \quad (18)$$

In this setting, $cov[A(\theta), b(\kappa, \theta)]$ may take on any sign. Remember, however, that the denominator of (18) is always non-negative as $-\partial A/\partial \tau_C \geq 0$. Therefore, the social planner will set $\tau_P > \tau_C$ if $cov[A(\theta), b(\kappa, \theta)] > 0$. In the important special case when the social

marginal welfare weight depends (negatively) on skill only, i.e. $b(\kappa, \theta) = b(\omega)$, one can show that the social planner will set $\tau_P > \tau_C$ if there is a positive dependence of ω and θ.[6] Intuitively, if it is cheaper for low-skilled individuals to shift income, the social planner can increase social welfare by allowing for income shifting. A closely related point was made by Kopczuk (2001) in the context of tax avoidance. We believe that this mechanism is less important, however, in the context of income shifting, which typically is an issue pertaining to the upper part of the income distribution.

4   The Extensive Margin

We now cast light on the consequences of allowing for income shifting along the extensive margin. The shifting cost is a pure fixed cost, i.e., $\Gamma(A; \gamma) = \gamma \cdot \mathbb{1}_{A>0}$. Given this specification, corner solutions may play an important part and we do not make any assumption that would lead to a focus on interior solutions. In other respects, the framework of Section 3 is intact. In particular, individuals differ with respect to three dimensions; skill ω, taste for work effort ε and shifting cost γ. Moreover, personal income is taxed linearly and shifted income is taxed at a proportional rate.

4.1   Partition of the Population

When income shifting is done against a fixed cost and the tax function is $T(\omega L, A) = G + \tau_P(\omega L - A) + \tau_C A$, a rational individual either shifts nothing ($A = 0$) or her entire labor earnings ($A = \omega L$). In the first case, as a non-shifter, her utility amounts to $V^P(\omega, \epsilon)$ and her labor supply is determined by the tax rate on personal income τ_P. In the latter case, as a shifter, her indirect utility is $V^C(\omega, \epsilon, \gamma)$ and the tax rate on shifted income τ_C determines her labor supply. Consequently, she chooses $A = 0$ when $V^P(\omega, \varepsilon) \geq V^C(\omega, \varepsilon, \gamma)$ and

---

[6]Formally, $cov(A, b) = \int_\omega \int_\theta [F_{\omega, \theta}(\omega, \theta) - f_\omega(\omega) f_\theta(\theta)] db(\omega) dA(\theta)$. Given that $db(\omega) < 0$ and $dA(\theta) < 0$, a sufficient condition for $cov(A, b)$ to be positive is that the square bracket inside the double integral be positive, see Cuadras (2002, Theorem 1). Conversely, a sufficient condition for $cov(A, b)$ to be negative is that the square bracket inside the double integral be negative. These conditions on the sign of the cumulative density function relative to the product of the marginal probability density functions of the joint distribution of (ω, θ) correspond to a generalization of correlation, called *quadratic dependence*. In words, positive (negative) quadrant dependence means that the joint probability that both ω and θ are larger than a pair $(\hat{\omega}, \hat{\theta})$ is larger (smaller) than the product of the two independent probabilities for all possible pairs $(\hat{\omega}, \hat{\theta})$.

$A = \omega L$ otherwise. Using (5) and (6), we see that $V^P(\omega, \varepsilon) \geq V^C(\omega, \varepsilon, \gamma)$ if and only if:

$$(1 - \tau_P)\omega L^P + G - v(L^P; \varepsilon) \geq (1 - \tau_C)\omega L^C + G - \gamma - v(L^C; \varepsilon), \tag{19}$$

which is equivalent to:

$$\gamma \geq [(1 - \tau_C)\omega L^C - (1 - \tau_P)\omega L^P] + [v(L^P; \varepsilon) - v(L^C; \varepsilon)]. \tag{20}$$

This inequality implies that, at a given κ, the population can be divided into two fractions: shifters and non-shifters. For each value of κ, we call $\hat{\gamma}(\kappa)$ the solution in γ to Equation (20) written with equality instead of $\geq$.[7] Given this cut-off level:

- People 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 an effort level L^C, which is independent of γ. L^C 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_P and driven by the marginal tax rate on shifted income, τ_C.

- The other agents decide not to shift earnings.[8] Each of them selects effort $L^P(\kappa)$, independent of γ. The marginal work incentive is driven by τ_P (and thus independent of τ_C).

At every κ for which $\hat{\gamma}(\kappa) > 0$, a rise in τ_P increases the incentive to shift; hence the cutoff level $\hat{\gamma}(\kappa)$ goes up. Conversely, if τ_C increases, the incentive to shift diminishes and $\hat{\gamma}(\kappa)$ goes down.[9]

This partition of the population, at a given κ, plays a key part. It implies that, at a given income level, there may be 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. To illustrate this point, we may consider a tax system for which T_P is piece-wise linear, but there is no possibility of income shifting. In that case, at a given income level,

---

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

[8]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.

[9]Formally, this follows from the application of the implicit function theorem to the definition of $\hat{\gamma}(\kappa)$.

all agents face the same tax liability. The results concerning the partition of the population are summarized in the following Lemma.

Lemma 1. *Assume $\Gamma(A; \gamma) = \gamma \cdot \mathbb{1}_{A>0}$ and $T(\omega L, A) = G + \tau_P(\omega L - A) + \tau_C A$. Then:*

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

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

Moreover, at every κ for which $\hat{\gamma}(\kappa) > 0$, $\frac{\partial \hat{\gamma}(\kappa)}{\partial \tau_P} = \omega L^P > 0$ and $\frac{\partial \hat{\gamma}(\kappa)}{\partial \tau_C} = -\omega L^C < 0$.

4.2 Optimal Tax Rates

We use a small tax reform perturbation around the optimum to determine the optimal tax rates τ_P and τ_C. More precisely, we investigate the effects of increasing τ_P, or alternatively τ_C, by a small quantity $\partial \tau > 0$, everything else being equal. We start by considering an increase in the marginal tax rate τ_P on personal income. This tax variation has the following effects:

- *Net mechanical effect in the non-shifting population:* The rise $\partial \tau$ in τ_P mechanically increases taxes collected from each agent in the non-shifting population, by an amount $E_1^+ = \omega L^P \partial \tau$. However, given quasilinear-in-income preferences, it also reduces each agent's utility by $\omega L^P \partial \tau$, and thus social welfare by $E_1^- = g(\kappa, \gamma) \omega L^P \partial \tau$. Diving the latter by λ, we obtain the effect on social welfare expressed in dollars: $b(\kappa, \gamma) \omega L^P \partial \tau$ with $b(\kappa, \gamma) = g(\kappa, \gamma)/\lambda$. The net mechanical effect corresponds to the difference between E_1^+ and E_1^-, i.e., $(1 - b(\kappa, \gamma)) \omega L^P \partial \tau$. Summing over the set of non-shifters, we obtain:

$$E_1 = \iint_\kappa \int_{\hat{\gamma}(\kappa)}^\infty (1 - b(\kappa, \gamma)) \omega L^P \partial \tau \, f(\kappa, \gamma) d\gamma d\kappa. - \frac{\omega L^P \cdot e^P(\varepsilon)}{1 - \tau_P} \times \partial \tau, \quad (21)$$

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

$$(22)$$

where $e^P(\varepsilon)$ stands for the labor supply elasticity within the set of non-shifters. As a result, taxes collected from this agent diminish by $\tau_P \times (21)$. Summing over the non-shifting population, we obtain:

$$E_2 = -\iint_\kappa \int_{\hat{\gamma}(\kappa)}^\infty \frac{\tau_P}{1-\tau_P}\omega L^P e^P(\varepsilon)\, \partial\tau\, f(\kappa,\gamma)\, d\gamma d\kappa. \tag{23}$$

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

$$\Delta T \equiv \tau_C \omega L^C - \tau_P \omega L^P \tag{24}$$

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

$$E_3 = \iint_\kappa \Delta T \frac{\partial\hat{\gamma}(\kappa)}{\partial\tau_P}\, \partial\tau\, f(\omega,\hat{\gamma})d\kappa = \iint_\kappa \Delta T\, \omega L^P\, \partial\tau f(\omega,\hat{\gamma})d\kappa. \tag{25}$$

where $\frac{\partial\hat{\gamma}(\kappa)}{\partial\tau_P} = \omega L^P$ follows from (20).

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:

$$\frac{\tau_P}{1-\tau_P} = \frac{\iint_\kappa \int_{\hat{\gamma}(\kappa)}^\infty [1-b(\kappa,\gamma)]\omega L^P f(\kappa,\gamma)d\gamma d\kappa}{\iint_\kappa \int_{\hat{\gamma}(\kappa)}^\infty \omega L^P e^P(\varepsilon)f(\kappa,\gamma)d\gamma d\kappa} + \frac{\iint_\kappa \omega L^P \Delta T(\kappa,\hat{\gamma})f(\kappa,\hat{\gamma})d\kappa}{\iint_\kappa \int_{\hat{\gamma}(\kappa)}^\infty \omega L^P e^P(\varepsilon)f(\kappa,\gamma)d\gamma d\kappa}. \tag{26}$$

We now consider an increase $\partial\tau$ in the optimal marginal tax rate τ_C 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 respectively given by E_1 and E_2, with τ_P replaced by τ_C, $L^P(\omega)$ replaced by $L^C(\omega)$, $e^P(\omega)$ replaced by the labor supply elasticity $e^C(\varepsilon)$ 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_C induces

14

induces people to leave the shifting population and become non-shifters. By Lemma 1, we know that $\hat{\gamma}(\kappa)$ goes down by ωL^C. All these agents will pay taxes $\tau_P \omega L^P$ instead of $\tau_C \omega L^C$, i.e., $-\Delta T$. The net effect on collected taxes is therefore given by:

$$ -\iint_\kappa \Delta T \,\omega L^C \,\partial \tau_C \,f(\kappa,\hat{\gamma}) d\kappa. \tag{27}$$

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:

$$ \frac{\tau_C}{1-\tau_C} = \frac{\iint_\kappa \int_0^{\hat{\gamma}(\kappa)} [1-b(\kappa,\gamma)]\omega L^C f(\kappa,\gamma)d\gamma d\kappa}{\iint_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega L^C e^C(\varepsilon) f(\kappa,\gamma)d\gamma d\kappa} - \frac{\iint_\kappa \omega L^C \Delta T(\kappa,\hat{\gamma}) f(\kappa,\hat{\gamma})d\kappa}{\iint_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega L^C e^C(\varepsilon) f(\kappa,\gamma)d\gamma d\kappa}. \tag{28}$$

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

Proposition 2. *Assume $\Gamma(A;\gamma) = \gamma \cdot \mathbb{1}_{A>0}$ and $T(\omega L, A) = G + \tau_P(\omega L - A) + \tau_C A$. In the social optimum, the marginal tax rates τ_P and τ_C are given by Equations* (26) *and* (28).

In the social optimum, the marginal tax rates τ_P and τ_C typically differ.[10] As shown by Equations (26) and (28), 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 shifters or non-shifters. In this sense, we may speak of "endogenous" tagging. The second terms on the right-hand side of Equations (26) and (28) 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 are more elastic 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 suggest that individual *heterogeneity* in

---

[10]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 $\tau_P = \tau_C$.

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 in elasticities depending on their tax status. To the best of our knowledge, this point was up to date only addressed in one study devoted to US physicians (Showalter and Thurston, 1997). The latter 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 further guidance in terms of tax design. It should be pointed out however that such empirical studies are difficult because large administrative data sets typically include data on taxable incomes (that capture both real and avoidance responses), but not on work hours.

4.3 Numerical Simulations: Basic Setup

The analysis of a small tax reform perturbation around the social optimum illuminated the mechanisms behind the optimal tax rates formulas of Proposition 2. However, a quantitative analysis is required to see whether it is socially optimal to allow for income shifting for plausible calibrations and, if so, how large the difference between the optimal marginal tax rates τ_P and τ_C should be.

First of all, we consider that the social objective is the maximin, which implies that the social planner attaches social weight zero to all individuals, except from the lowest skilled individual. In this case, the social planner chooses τ_P and τ_C such that tax revenues are maximized. It follows that the social planner would set τ_C lower than τ_P only if this results in larger collected taxes. This benchmark is of particular interest because we place ourselves in the situation which is the least favorable to income shifting (in particular, shifting has no direct positive utility effect, through the increased net income of the shifters).

In the numerical exercise, we let the taste parameter ε depend deterministically on ω. More specifically, we assume that ε is a linear function of ω:

$$\varepsilon = q_1 + q_2\omega. \tag{29}$$

Regarding individual preferences, we consider that the utility function $U(Y, L)$ is given by:

$$U(Y, L) = Y - \alpha\frac{L^{1+\frac{1}{\varepsilon}}}{1 + \frac{1}{\varepsilon}}, \tag{30}$$

16

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

We need to callibrate the joint distribution of skills and shifting costs. It is empirically well-known that the distribution of hourly wage rates is well approximated by a log-normal distribution, if one abstracts from the top of the distribution. We have considerable less guidance regarding how to calibrate the distribution of shifting costs. Because we want to perform sensitivity analysis regarding the correlation of (ω, ε) and γ, it is convenient for us to consider that these two parameters are described by a bivariate log-normal distribution. We use Swedish data to callibrate the mean and variance of the wage distribution. Regarding the shifting costs, we parameterize them so that the proportion of people deciding to shift incomes roughly reproduces the actual figure for Sweden, see Alstadsæter and Jacob (2016). We provide a more detailed discussion in Appendix B.

4.4 Numerical Simulations: Results

In Figure 2, the blue curve shows the gap – in percentage points – between τ_P and τ_C for 21 different values the correlation coefficient for $\log(\omega)$ and $\log(\gamma)$. Additionally, the orange curve shows which share of the population chooses to pay the fixed cost and, thereby, shift their entire labor income into the capital income tax base. The socially optimal allocation has the following features. First, 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. Second, there is always a gap between τ_P and τ_C, which ranges from about 9,5 to 14,6 percentage points, and the tax difference is actually increasing in the correlation coefficient. This also makes sense, because the revenue-maxmizing tax rates in the two subpopulations depend on the distributions of elasticities. Intuitively, when the pool of shifters shrinks, the average earnings elasticity in the subpopulation of shifters will increase.

The optimal marginal tax rates on labor and capital incomes are depicted in Figure 2 for different values the correlation coefficient for $\log(\omega)$ and $\log(\gamma)$. There we see that τ_C is considerably more sensitive to changes in the correlation coefficient than τ_P. Intuitively, since the fraction of shifters is much smaller than the fraction of non-shifters the average labor supply elasticity (which determines the revenue-maximizing tax rate) is more
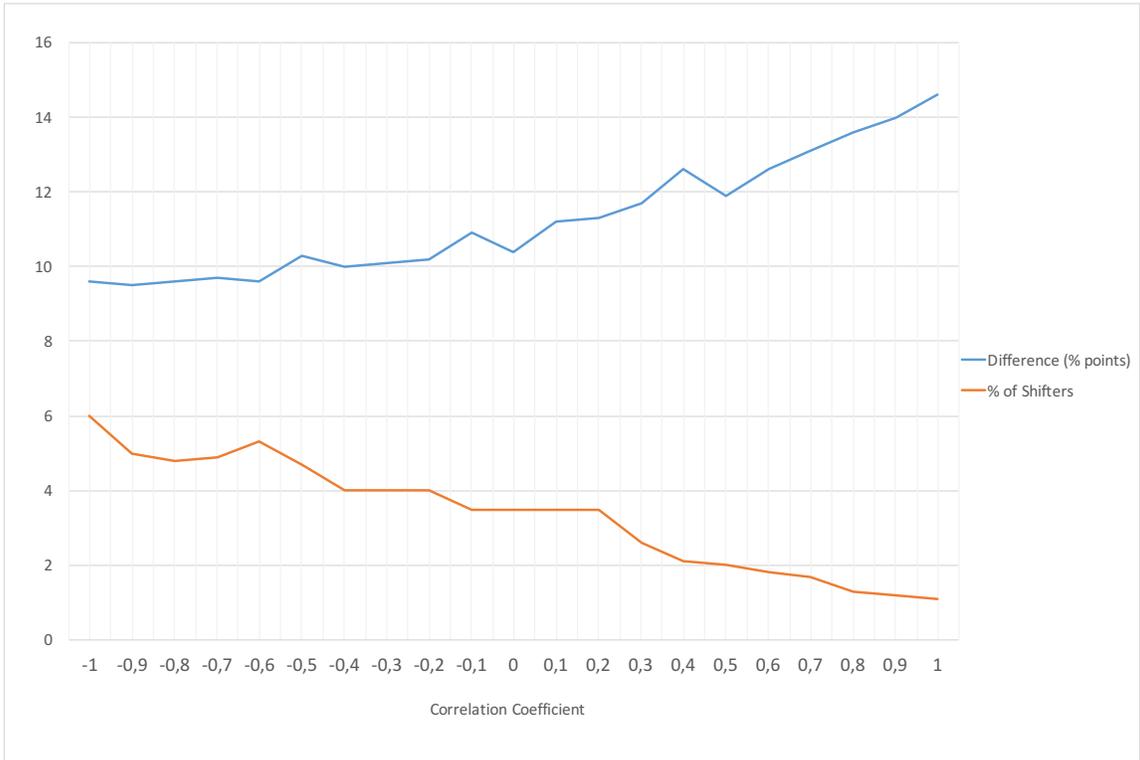
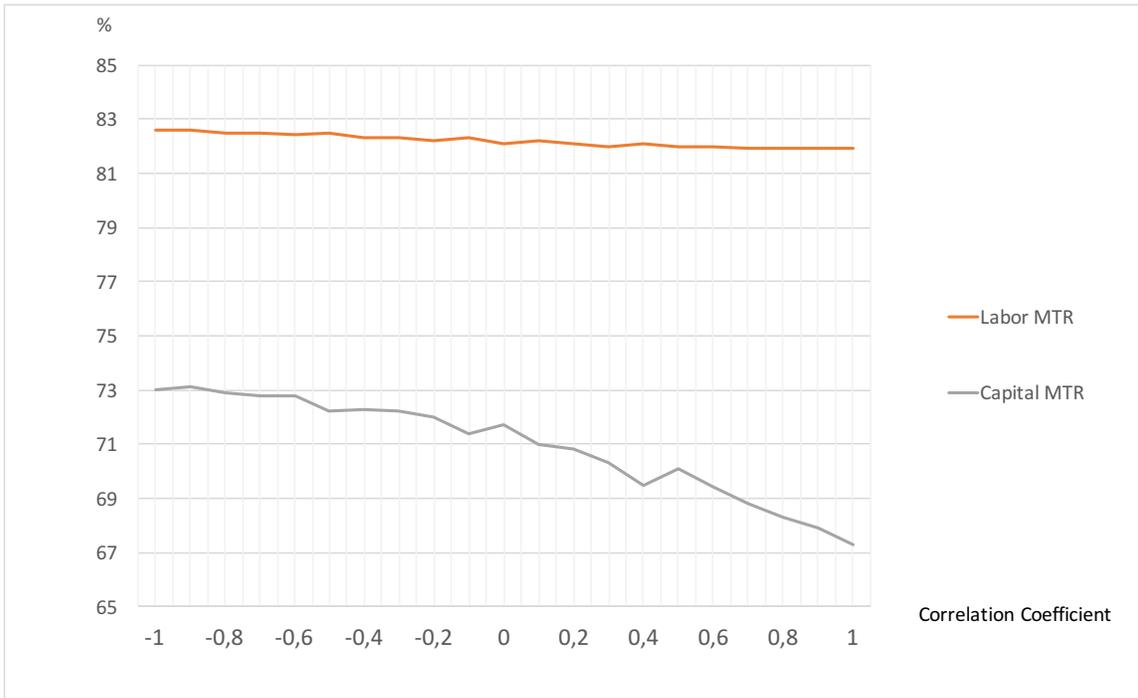Figure 1: Features of the Optimal Allocation (Benchmark Case)

Figure 2: Features of the Optimal Allocation (Benchmark Case)

Table 1: Simulation results

Min elasticity	Max elasticity	ρ	τ_P^*	τ_C^*	$\tau_P^* - \tau_C^*$	Shifters %
0	0.725	-1	0.79	0.65	0.14	16.3
0	0.725	0	0.78	0.65	0.14	12.7
0	0.725	1	0.77	0.64	0.13	8.2
0.1	0.5	-1	0.80	0.71	0.09	6
0.1	0.5	0	0.79	0.70	0.09	3.5
0.1	0.5	1	0.79	0.68	0.11	1.1
0.15	0.4	-1	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	1	0.80	0.68	0.11	0.2
0.23	0.23	-1	0.81	0.81	0.00	0
0.23	0.23	0	0.81	0.81	0.00	0
0.23	0.23	1	0.81	0.81	0.00	0

sensitive to changes in the composition.

We now investigate to which extent our results are sensitive to the elasticity range. For three different values of the correlation coefficient ρ (namely $-1, 0$ and 1), we examine four different elasticity ranges while keeping the average elasticity in the population constant (equal to 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 $\tau_P = \tau_C$. 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 quasilinearity 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 we consider is the maximin, this implies that tax rates should not be differentiated. Second, when the lowest ability individual exhibits an elasticity of 0 and the highest ability individual 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.

5 Robustness Checks and Extensions

Sections 3 and 4 illustrated the important distinction between the intensive and extensive margins in the most simple way. We in particular 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 and convex shifting costs; (ii) the relaxation of the linear tax assumptions. We therefore consider $\Gamma(A; \gamma) = C(A) + \gamma \cdot \mathbb{1}_{A>0}$, without neither assuming $\gamma \equiv 0$ (contrary to Section 3) nor $C(A) \equiv 0$ (contrary to Section 4).

Multidimensional screening problems are technically challenging. To make the problem sufficiently tractable, we from now on assume that ω and γ are the only dimensions of heterogeneity within the population; we thus have $\kappa = \omega$. This implies that at a given (ω, γ), all agents have the same ϵ; the three-dimension screening problem considered before therefore turns into a two-dimension one. We believe that this assumption could be relaxed without modyfing the message of our results. The extension would however not be trivial.[11]

5.1 Individual Choices

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, one part of them, or nothing. The indirect utility V^{NS} of an agent who does not shift anything is now defined as:

$$V^{NS}(\omega) = \max_L \{\omega L - T_P(\omega L) - v(L)\}, \tag{31}$$

yielding the first-order condition:

$$[1 - T_P'(\omega L)]\omega = v'(L). \tag{32}$$

---

[11]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.

For later use, we call $L^{NS}(\omega)$ the solution in L. By contrast, the indirect utility V^S when earnings *at least partially* shifted is:

$$V^S(\omega, \gamma) = \max_{L,A}\{\omega L - T_P(\omega L - A) - T_C(A) - C(A) - \gamma - v(L) \text{ with } A \leq \omega L\}. \quad (33)$$

Because of the interaction between the intensive and extensive mechanisms, it is now important to explicitly account for the inequality constraint $A \leq \omega L$. We let λ^S refer to the Kuhn-Tucker multiplier of the latter. The first-order conditions with respect to L and A are respectively:

$$[1 - T'_P(\omega L - A)]\omega - \lambda^S \omega - v'(L) = 0, \quad (34)$$

$$T'_P(\omega L - A) - T'_C(A) - C'(A) + \lambda^S = 0, \quad (35)$$

with $\lambda^S \geq 0 \, (= 0 \text{ if } A < \omega L)$. Combining the latter, we obtain:

$$\left[1 - T'_C(A) - C'(A)\right]\omega = v'(L) \quad (36)$$

We call $L^S(\omega)$ and $A(\omega, \gamma)$ the solution in L and A respectively. We see from Equation (36) for an individual shifting her earnings –either partly or entirely– the marginal work intensive is driven by the sum of the marginal tax rate on shifted income and the marginal shifting cost. In other words, for a shifter (with $0 < A \leq \omega L$), the marginal shifting cost $C'(A)$ plays exactly the same part as the marginal tax rate $T'_C(A)$.

Before going further, it is important to emphasize that, because γ additively separably enters utility, it only determines the partitioning of the population into non-shifters and shifters. It does not affect optimal values of L^{NS} nor of L^S and A, conditional on the extensive shifting margin decision. In this setting, an agent decides to partly or entirely shift earnings provided $V^S(\omega, \gamma) > V^{NS}(\omega)$.[12] Solving this inequality for γ, we see that –at each skill ω– there is a cut-off level $\hat{\gamma}(\omega)$ below which $A > 0$ and above which $A = 0$. More precisely, $\hat{\gamma}(\omega) > 0$ is equal to $max\{0, \hat{\gamma}\}$, with $\hat{\gamma}$ solution to:

$$\hat{\gamma} = \omega\left(L^S(\omega) - L^{NS}(\omega)\right) + T_P\left(\omega L^{NS}(\omega)\right) - T_P\left(\omega L^S(\omega) - A(\omega, \hat{\gamma})\right)$$
$$- T_C\left(A(\omega, \hat{\gamma})\right) - C\left(A(\omega, \hat{\gamma})\right) - v\left(L^S(\omega)\right) + v\left(L^{NS}(\omega)\right). \quad (37)$$

---

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

5.2 Elasticities

Intensive Labor Supply Elasticities

As in the previous Sections, elasticities play a key part in the analysis. It is necessary to generalize the definitions introduced before 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:

$$e^{NS}(\omega) \equiv \frac{\partial L^{NS}(\omega)}{\partial \left[\omega \left(1 - T'_P(\omega L^{NS}(\omega))\right)\right]} \frac{\omega(1 - T'_P(\omega L^{NS}(\omega)))}{L^{NS}(\omega)} = \frac{v'(L^{NS}(\omega))}{v''(L^{NS}(\omega))L^{NS}(\omega)}. \quad (38)$$

The last equality follows from (32) 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 everything. It is given by:

$$e^S(\omega) \equiv \frac{\partial L^S(\omega)}{\partial \left[\omega(1 - T'_C(\omega L^S(\omega)))\right]} \frac{\omega \left(1 - T'_C(\omega L^S(\omega))\right)}{L^C(\omega)} = \frac{v'(L^S(\omega)) + \omega C'(\omega L^S(\omega))}{v''(\omega L^S(\omega))L^S(\omega)}. \quad (39)$$

As emphasized above, the marginal work incentive of a shifter is not only shaped by the tax function and the wage rate, but also by the marginal shifting cost function. This explains the additional term $\omega C'(\omega L^S(\omega))$ on the right-hand side of Equation (39): When an agent supplies one extra unit of labor, she does not only has to pay the marginal tax rate, but also the marginal shifting cost. It is clear from the comparison of e^{NS} and e^S that *two agents of skill ω may have different labor supply elasticities.*[13]

Extensive Shifting Elasticities

For later use, we also define extensive margin shifting elasticities. We will see below that, in the social optimum, any shifter shifts her entire earnings. Consequently, at a given skill ω, the proportion of shifters is equal to:

$$F_{\gamma|\omega}(\hat{\gamma}) = F_{\gamma|\omega}[\omega L^S - \omega L^{NS} + T_P(\omega L^{NS}) - T_P(0) - T_C(\omega L^S) + v(L^{NS}) - v(L^C)]. \quad (40)$$

---

[13]Another possibility, which we do not explicitly model, is that the $v(L)$ functions are state-dependent (different for shifters and non-shifters), which would generate differential elasticities for shifters and non-shifters even in the absence of continuous shifting costs. It is straightforward to interpret the optimal tax rules presented below in this section in this way.

The percentage change in the number of *shifters* in response to a percentage change in the tax paid as a non-shifter is:

$$\eta^S(\omega) = -\frac{\partial F_{\gamma|\omega}(\hat{\gamma})}{\partial T_C(0)}\frac{T_P(\omega L^{NS})}{F_{\gamma|\omega}(\hat{\gamma})} = f(\omega,\hat{\gamma})\frac{T_P(\omega L^{NS})}{F_{\gamma|\omega}(\hat{\gamma})} \geq 0. \tag{41}$$

We vary $T_C(0)$ as this quantity is independent of the labor supply choice. Similarly, the percentage change in the number of *non-shifters* in response to a percentage change in the tax paid as a non-shifter is:

$$\eta^{NS}(\omega) = -\frac{\partial[1 - F_{\gamma|\omega}(\hat{\gamma})]}{\partial T_C(0)}\frac{T_P(\omega L^{NS})}{1 - F_{\gamma|\omega}(\hat{\gamma})} = -f(\omega,\hat{\gamma})\frac{T_P(\omega L^{NS})}{1 - F_{\gamma|\omega}(\hat{\gamma})} \leq 0. \tag{42}$$

5.3 The Social Planner's Problem

Because there are two dimensions of heterogeneity in the population (skills ω and shifting costs γ), the policy maker faces a multidimensional screening problem. However, conditional on the shifting status, all individuals of skill ω will choose the same labor supply $L(\omega)$ and shifted amount $A(\omega)$. Hence, the problem simplifies. In particular, we will be able to obtain optimal marginal tax rates for given levels of ω. We define $\tilde{V}^S(\omega) \equiv V^S(\omega,\gamma) + \gamma$ as the indirect utility of a shifter gross ($A > 0$) of the shifting cost γ. The social planner's objective function can be written as:

$$\int_0^\infty \int_0^{\hat{\gamma}(\omega)} g(\omega,\gamma)[\tilde{V}^S(\omega) - \gamma]f(\omega,\gamma)\,d\gamma d\omega + \int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty g(\omega,\gamma)V^{NS}(\omega)\,f(\omega,\gamma)\,d\gamma d\omega, \tag{43}$$

where $\hat{\gamma}(\omega) \equiv \tilde{V}^S(\omega) - V^{NS}(\omega)$. The social planner maximizes (43) with respect to $\tilde{V}^S(\omega), V^{NS}(\omega), L^S(\omega), L^{NS}(\omega)$ and $A(\omega)$, within the set of feasible and incentive-compatible allocations.

Incentive Compatilibity

Given the definition of $V^S(\omega,\gamma)$ in (33), we obtain: $dV^S/d\omega = (1 - T_P')L^S$. Combining the latter with (34), we obtain $1 - T_P' = v'/\omega + \lambda^S$. Hence,

$$\frac{dV^S}{d\omega} = \left[\frac{v'(L^S(\omega))}{\omega} + \lambda^S\right]L^S(\omega). \tag{44}$$

24

There are two cases to consider. First, suppose $A < \omega L^S$. Then, $\lambda^S = 0$. Consequently, (44) reduces to:

$$\frac{dV^S}{d\omega} = \frac{v'(L^S(\omega))}{\omega} L^S(\omega). \tag{45}$$

Now, suppose instead that $A = \omega L^S$. We obtain:

$$V^S = \omega L^S - T_P(0) - T_C(\omega L^S) - C(\omega L^S) - \gamma - v(L^S), \tag{46}$$

from which:

$$\frac{dV^S}{d\omega} = \left[1 - T_C' - C' \right] L^S. \tag{47}$$

Plugging (36), the condition reduces to (45). Because $d\tilde{V}^S/d\omega = dV^S/d\omega$, the first-order condition for incentive-compatibility (45) is equivalent to:

$$\frac{d\tilde{V}^S(\omega)}{d\omega} = \frac{v'(L^S(\omega))}{\omega} L^S(\omega). \tag{48}$$

For non-shifters, we obtain:

$$\frac{dV^{NS}(\omega)}{d\omega} = \frac{v'(L^{NS}(\omega))}{\omega} L^{NS}(\omega). \tag{49}$$

Incentive compatible allocations verify (48) and (49), in addition to the monotonicity constraints that gross-income be non-decreasing in skills within each set of agents (non-shifters and shifters). We below adopt the so-called "first-order approach" and do not formally account for the monotonicity constraints when writing the policy-maker's optimization problem. These constraints can be checked *ex post* in numerical simulations.

Feasibility

At the individual level, shifted earnings must not exceed the individual's total earnings. Hence, for all values of ω, we must have:

$$A(w) \leq \omega L^C(w). \tag{50}$$

In addition, optimal allocations must be budget balanced. The state's resource constraint can be written as follows:

$$\int_0^\infty \int_0^{\hat{\gamma}(\omega)} [\omega L^S - v(L^S) - \tilde{V}^S(\omega)] f(\omega, \gamma) \, d\gamma d\omega$$
$$+ \int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty [\omega L^{NS} - v(L^{NS}) - V^{NS}(\omega)] f(\omega, \gamma) \, d\gamma d\omega \geq R. \tag{51}$$

For simplicity, we set the exogenous revenue requirement to zero, i.e. $R = 0$. The following Lemma summarizes the policy-maker's optimization problem.

Problem 1. *Find $\tilde{V}^S(\omega)$, $V^{NS}(\omega)$, $L^S(\omega)$, $L^{NS}(\omega)$ and $A(\omega)$ which maximizes the social objective (43) subject to (i) the incentive compatibility conditions (48) and (49), (ii) the tax revenue constraint (51) with $R = 0$ and (iii) the inequality constraint (50).*

5.4 Optimal Tax Rules

When marginal tax rates are constant and shifting only involves a fixed cost (cf. Section 4), a rational agent either shifts nothing or her entire earnings. 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 as in the pure extensive model.

Proposition 3. *Assume $C'(A) > 0$. In the social optimum, agents choose $A = 0$ or $A = \omega L$.*

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

$$\mu C'(A) f_\omega(\omega) F_{\gamma|\omega}(\hat{\gamma}) = 0. \tag{52}$$

The shadow price of the resource constraint μ and the density $f_\omega(\omega)$ are strictly positive. Therefore, by (52), $C'(A) > 0$ if and only if $F_{\gamma|\omega}(\hat{\gamma}) = 0$. □

Proposition 3 shows that the sorting of agents between pure shifters and pure non-shifters, highlighted in Section 4, is robust to the introduction of a smooth shifting cost as well as to the relaxation of the linear tax assumptions. In addition, Proposition 1 can be seen as a subcase of Proposition 3. When $A = \omega L$ is not available, agents must choose $A = 0$. In that case, there is no shifting, corresponding to $\tau_P = \tau_C$. The next Proposition generalizes Proposition 2.

Proposition 4. *In the social optimum,*

$$\frac{T_P'(\omega L^{NS})}{1 - T_P'(\omega L^{NS})} = \left[1 + \frac{1}{e^{NS}(\omega)} \right]$$
$$\times \frac{\int_\omega^\infty \int_{\hat{\gamma}(\omega)}^\infty [1 - b(\omega, \gamma)] f(\omega, \gamma) \, d\gamma d\omega + \int_\omega^\infty \Delta T(\omega) f(\omega, \hat{\gamma}) d\omega}{\omega f_\omega(\omega)[1 - F_{\gamma|\omega}(\hat{\gamma})]} \quad (53)$$

$$\frac{T_C'(\omega L^S)}{1 - T_C'(\omega L^S)} = \left[1 + \frac{1}{e^S(L^S, \omega)} - \frac{C'(\omega L^S)}{1 - T_C'(\omega L^S)} \right]$$
$$\times \frac{\int_\omega^\infty \int_0^{\hat{\gamma}(\omega)} [1 - b(\omega, \gamma)] f(\omega, \gamma) \, d\gamma d\omega - \int_\omega^\infty \Delta T(\omega) f(\omega, \hat{\gamma}) d\omega}{\omega f_\omega(\omega) F_{\gamma|\omega}(\hat{\gamma})} \quad (54)$$

Proof. See Appendix C.   □

The optimal tax rules (53) and (54) have the same structure to those presented in Proposition 2. 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^{NS} and ωL^S. For example, increasing T_P' 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 (53), with $F_{\gamma|\omega}(\hat{\gamma}) = 0$ and $b(\omega, \gamma) = b(\omega)$.

A new element of (54) 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 (54) that this motivates a lower marginal tax rate than otherwise. However, if the total shifting cost (fixed cost + continuous cost) becomes sufficiently large it will not be optimal for the social planner to allow for shifting. Remember that (54) 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 5.

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, in this non-linear setting we to some extent account for this policy endogeneity as the government may affect the fixed shifting cost by varying $T_C(0)$ (undetermined sign), 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 also imagine other ways in which the government 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 government's maximization problem.

5.5 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 $\eta^{NS}(\omega)$ converge to $\eta^S \leq 0$ and $\eta^{NS} \geq 0$ respectively. Note that $\eta^{NS} = -\eta^S \frac{F_\gamma}{1-F_\gamma}$ and F_γ is constant.

Proposition 5. *Assume* $\frac{f_\omega(\omega)\omega}{1-F_\omega(\omega)} \to a$, $\frac{\int_\omega^\infty \frac{\Delta T(\omega)}{T_P(\omega L^P)} \eta^S(\omega) F_{\gamma|\omega} d\omega}{1-F_\omega(\omega)} \to \frac{\Delta T}{T}\eta^S F_\gamma$, $e^{NS}(\omega) \to e^{NS}$, $e^S(\omega) \to e^S$ 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:

$$\tau_P^* = \frac{1 - \frac{\Delta T}{T}\eta^{NS}}{1 + a\frac{e^{NS}}{1+e^{NS}} - \frac{\Delta T}{T}\eta^{NS}} \tag{55}$$

and

$$\tau_C^* = \frac{1 - \frac{\Delta T}{T}\eta^S - ca\frac{e^S}{1+e^S}}{1 + a\frac{e^S}{1+e^S} - \frac{\Delta T}{T}\eta^S}, \tag{56}$$

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 (53) 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, (53) yields:

$$\frac{T_P'}{1-T_P'} = \{1 + \frac{1}{1-F_\gamma} \lim_{\omega \to \infty} \frac{\int_\omega^\infty \Delta T(\omega,\hat\gamma)f(\omega,\hat\gamma)d\omega}{1-F_\omega(\omega)}\}\frac{1}{a}(1+\frac{1}{e^{NS}})$$
$$= \frac{1}{a}(1 - \frac{\Delta T}{T}\eta^{NS})(1+\frac{1}{e^{NS}}). \tag{57}$$

Solving for T_P', we obtain (55). Similarly, when ω tends to infinity, (54) yields:

$$\frac{T_C'}{1-T_C'} = \frac{1}{a}(1 - \frac{\Delta T}{T}\eta^S)(1+\frac{1}{e^S} - \frac{c}{1-T_C'}). \tag{58}$$

Solving for T_C', we obtain (56). □

It should be emphasized that the top marginal tax rates in Proposition 5 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 $a/(1 + e^{NS})$. 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^* negatively depends on the real labor supply elasticity of non-shifters e^{NS} 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 $\frac{\Delta T}{T}\eta^{NS}$. If the tax payment as a shifter is larger than the tax payment as a non-shifter, ΔT is positive. Because $\eta^{NS} \leq 0$, this implies $\frac{\Delta T}{T}\eta^{NS} \leq 0$. 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. This in particular 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^{NS} 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 (56). A positive c will reduce τ_C^* as the

29

total labor supply distortion of shifters is given by $\tau_C^* + c$. Of course, the revenue-maxizing top marginal tax rate (56) derived in a setting which 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 5 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 5 formalizes the potential importance of fiscal externalities in a new and novel way, which endogenizes income shifting.

6   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 carries 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 21 percentage points in Sweden (69% 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; individuals may e.g. 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 for 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

Jonas Agell and Mats Persson. Tax arbitrage and labor supply. *Journal of Public Economics*, 78(1-2):3–24, October 2000.

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

Annette Alstadsæter and Martin Jacob. Who participates in tax avoidance? evidence from swedish microdata. *Applied Economics*, forthcoming.

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

Raj Chetty. 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, August 2009.

Vidar Christiansen and Matti Tuomala. On taxing capital income with income shifting. *International Tax and Public Finance*, 15:527–545, 2008.

C.M. Cuadras. On the covariance between functions. *Journal of Multivariate Analysis*, 81 (1):19 – 27, 2002.

Peter A Diamond. Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates. *American Economic Review*, 88(1):83–95, March 1998.

Philipp Doerrenberg, Andreas Peichl, and Sebastian Siegloch. The elasticity of taxable income in the presence of deduction possibilities. *Journal of Public Economics*, pages –, 2015.

Martin Feldstein. 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, June 1995.

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

Renato Gomes, Jean-Marie Lozachmeur, and Alessandro Pavan. Differential Taxation and Occupational Choice. CESifo Working Paper Series 5054, CESifo Group Munich, 2014.

Roger H. Gordon and Joel Slemrod. Are "Real" Responses to Taxes Simply Income Shifting Between Corporate and Personal Tax Bases? NBER Working Papers 6576, National Bureau of Economic Research, Inc, May 1998.

Jarkko Harju and Tuomas Matikka. The elasticity of taxable income and income-shifting: what is "real" and what is not? *International Tax and Public Finance*, pages 1–30, 2016.

Johannes Hermle and Andreas Peichl. Jointly optimal income taxes for different types of income. Technical report, 2015.

Laurence Jacquet and Etienne Lehmann. Optimal taxation with heterogeneous skills and elasticities: Structural and sufficient statistics approaches. Technical report, THEMA Working Paper n2016-04. Universit de Cergy-Pontoise, France, 2016.

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

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

Thomas Piketty and Emmanuel Saez. Optimal labor income taxation. In Martin Feldstein Alan J. Auerbach, Raj Chetty and Emmanuel Saez, editors, *Handbook of public economics, vol. 5*, volume 5 of *Handbook of Public Economics*, pages 391 – 474. Elsevier, 2013.

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

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

Casey Rothschild and Florian Scheuer. Redistributive taxation in the roy model*. *The Quarterly Journal of Economics*, 2012.

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

Emmanuel Saez and Stefanie Stantcheva. A simpler theory of optimal capital taxation. Working Paper 22664, National Bureau of Economic Research, September 2016.

Emmanuel Saez, Joel Slemrod, and Seth H. Giertz. The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature*, 50(1):3–50, March 2012.

Florian Scheuer. Entrepreneurial Taxation with Endogenous Entry. *American Economic Journal: Economic Policy*, 6(2):126–63, May 2014.

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

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

Joel Slemrod. Income Creation or Income Shifting? Behavioral Responses to the Tax Reform Act of 1986. *American Economic Review*, 85(2):175–80, May 1995.

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

Joel Slemrod. A General Model of the Behavioral Response to Taxation. *International Tax and Public Finance*, 8(2):119–128, March 2001.

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

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

Appendix A : Proof of Proposition 2

The social planner solves the following problem:

$$
\max_{\tau_P, \tau_C, G} \int_\kappa \int_0^{\hat{\gamma}(\kappa)} g(\kappa, \gamma) V^C(\kappa) f(\kappa, \gamma)\, d\gamma d\kappa + \int_\kappa \int_{\hat{\gamma}(\kappa)}^\infty g(\kappa, \gamma) V^P(\kappa, \gamma) f(\kappa, \gamma)\, d\gamma d\kappa, \quad (59)
$$

subject to:

$$
\tau_C \int_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega L^C(\kappa) f(\kappa, \gamma) d\gamma d\kappa + \tau_P \int_\kappa \int_{\hat{\gamma}(\omega, \varepsilon)}^\infty \omega L^P(\kappa) f(\kappa, \gamma) d\gamma d\kappa - R - G = 0. \quad (60)
$$

We let λ be the Lagrange multiplier of the budget constraint (60). The derivative of (59) with respect to τ_P is:

$$
\int_\kappa \int_{\hat{\gamma}(\kappa)}^\infty g(\kappa, \gamma) \omega L^P(\kappa) f(\kappa, \gamma) d\gamma d\kappa. \quad (61)
$$

We used the fact that $V^P(\kappa) = V^C(\kappa, \gamma)$ for $\gamma = \hat{\gamma}(\kappa)$. We now compute the derivative of the budget constraint (60) with respect to τ_P. We obtain:

$$
\int_\kappa \int_{\hat{\gamma}(\kappa)}^\infty \omega L^P(\kappa) f(\kappa, \gamma) d\gamma d\kappa + \tau_P \int_\kappa \int_{\hat{\gamma}(\kappa)}^\infty \omega \frac{\partial L^P(\kappa)}{\partial \tau_P} f(\kappa, \gamma) d\gamma d\kappa
$$
$$
+ \int_\kappa \left[\tau_C \omega L^C(\kappa) - \tau_P \omega L^P(\kappa) \right] \frac{\partial \hat{\gamma}}{\partial \tau_P} f(\kappa, \hat{\gamma}) d\kappa \quad (62)
$$

From Lemma 1, we know that $\partial\hat{\gamma}/\partial\tau_P = \omega L^P$. We now write $(61) - \lambda(62) = 0$, rearrange and use the definition of $e(\varepsilon)$ to obtain (26).

To obtain (28), we compute the derivative of the social objective with respect to τ_C. Using the indifference condition at $\hat{\gamma}$, we obtain:

$$\int_\kappa \int_0^{\hat{\gamma}(\kappa)} g(\kappa,\gamma)\omega L^C(\kappa) f(\kappa,\gamma) d\gamma d\kappa. \tag{63}$$

We now compute the derivative of the budget constraint (60) with respect to τ_C:

$$\int_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega L^C(\kappa) f(\kappa,\gamma) d\gamma d\kappa + \tau_C \int_\kappa \int_0^{\hat{\gamma}(\kappa)} \omega \frac{\partial L^C(\kappa)}{\partial \tau_C} f(\kappa,\gamma) d\gamma d\kappa$$
$$+ \int_\kappa \left[\tau_C \omega L^C(\kappa) - \tau_P \omega L^P(\kappa) \right] \frac{\partial \hat{\gamma}}{\partial \tau_C} f(\kappa,\hat{\gamma}) d\kappa \tag{64}$$

From (20) we know that $\frac{\partial \hat{\gamma}}{\partial \tau_C} = -\omega L^C$. We now write $(63) - \lambda(64) = 0$, rearrange and use the definition of $e(\varepsilon)$ to obtain (28).

Appendix B: Calibration of the fixed cost model

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 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.[14] Ultimately, we therefore solely calibrate μ_γ.

---

[14]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 $\rho = 1$.

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.

We set our target, i.e. the actual fraction of shifters, to be 5 %. Alstadsæter and Jacob (forthcoming) 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.[15] 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 $\tau_P = 0.505$ and $\tau_C = 0.410$ 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.

Appendix C: Proof of Proposition 4

Lagrangian and FOC

We form the Lagrangian from the objective (43) and the two sums of incentive compatibility constraints defined by (48) and (49) and the resource constraint (51). At a given skill

---

[15]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.

level ω, we denote the Lagrange multiplier associated with the incentive compatibility constraints (48) and (49) by $\lambda(\omega)$ and $\lambda^{NS}(\omega)$ respectively. μ refers to the Lagrange multiplier of the resource constraint (51) and $\lambda^A(\omega)$ denotes the Lagrange multiplier of the last constraint (50).

By integration by parts, we obtain:

$$\int_0^\infty \lambda(\omega)[\frac{d\tilde{V}^S(\omega)}{d\omega} - \frac{v'(L^S)}{\omega}L^S]d\omega = \lim_{\omega\to\infty} \lambda(\omega)\tilde{V}^S(\omega) - \lambda(0)\tilde{V}(0)$$
$$- \int_0^\infty \lambda'(\omega)\tilde{V}^S(\omega) - \int_0^\infty \lambda(\omega)[\frac{v'(L^S)}{\omega}L^S]d\omega, \quad (65)$$

$$\int_0^\infty \lambda^{NS}(\omega)[\frac{dV^{NS}(\omega)}{d\omega} - \frac{v'(L^{NS})}{\omega}L^{NS}]d\omega = \lim_{\omega\to\infty} \lambda^{NS}(\omega)V^{NS}(\omega) - \lambda^{NS}(0)V^{NS}(0)$$
$$- \int_0^\infty \lambda^{NS}(\omega)V^{NS}(\omega) - \int_0^\infty \lambda^{NS}(\omega)[\frac{v'(L^{NS})}{\omega}L^{NS}]d\omega, \quad (66)$$

where $\lim_{\omega\to\infty} \lambda(\omega)\tilde{V}^S(\omega) - \lambda(0)\tilde{V}^S(0) = \lim_{\omega\to\infty} \lambda^{NS}(\omega)V^{NS}(\omega) - \lambda^{NS}(0)V^{NS}(0) = 0$ due to transversality conditions $\lim_{\omega\to\infty} \lambda(\omega) = \lambda(0) = \lim_{\omega\to\infty} \lambda^{NS}(\omega) = \lambda^{NS}(0) = 0$. Combining the social objective (43), the reformulated conditions for incentive compatibility (65) and (66), together with the resource constraint (51), the Lagrangian may be rewritten as:

$$\mathcal{L} = \int_0^\infty \int_0^{\hat{\gamma}(\omega)} g(\omega,\gamma)[\tilde{V}^S(\omega) - \gamma]f(\omega,\gamma)\,d\gamma d\omega + \int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty g(\omega,\gamma)V^{NS}(\omega)f(\omega,\gamma)\,d\gamma d\omega$$
$$- \int_0^\infty \lambda'(\omega)\tilde{V}^S(\omega) - \int_0^\infty \lambda(\omega)\frac{v'(L^S)}{\omega}L^S d\omega - \int_0^\infty \lambda^{NS}(\omega)V^{NS}(\omega) - \int_0^\infty \lambda^{NS}(\omega)\frac{v'(L^{NS})}{\omega}L^{NS}d\omega$$
$$+ \mu\int_0^\infty \int_0^{\hat{\gamma}(\omega)} [\omega L^S - v(L^S) - \tilde{V}^S(\omega)]f(\omega,\gamma)\,d\gamma d\omega$$
$$+ \mu\int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty [\omega L^{NS} - v(L^{NS}) - V^{NS}(\omega)]f(\omega,\gamma)\,d\gamma d\omega + \lambda^A(\omega)[A - \omega L^S]. \quad (67)$$

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 $\tilde{V}^S(\omega)$, $V^{NS}(\omega)$, $A(\omega)$, $L^S(\omega)$ and $L^{NS}(\omega)$ to arrive at necessary conditions that

hold at given levels of ω. The first-order condition with respect to $\tilde{V}^S(\omega)$ is:

$$\int_0^{\hat{\gamma}(\omega)} [g(\omega, \gamma) - \mu] f(\omega, \gamma) \, d\gamma - \lambda'(\omega) + \mu \Delta T(\omega) f(\omega, \hat{\gamma}) = 0, \tag{68}$$

where $\Delta T(\omega) = [\omega L^S - v(L^S) - \tilde{V}^S(\omega)] - [\omega L^{NS} - v(L^{NS}) - V^{NS}(\omega)]$ is *the extra tax paid by the marginal shifter*. When writing down (68), we have used the fact that $\hat{\gamma}(\omega) = \tilde{V}^S(\omega) - V^{NS}(\omega)$, which in turn implies $\partial \hat{\gamma}(\omega)/\partial \tilde{V}(w) = 1$. In a similar way, the first-order condition with respect to $V^{NS}(\omega)$ reads:

$$\int_{\hat{\gamma}(\omega)}^{\infty} [g(\omega, \gamma) - \mu] f(\omega, \gamma) \, d\gamma - \lambda^{NS'}(\omega) - \mu \Delta T(\omega) f(\omega, \hat{\gamma}) = 0. \tag{69}$$

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

$$\lambda(\omega)\left[-\frac{v''(L^S)}{\omega}L^S - \frac{v'(L^S)}{\omega}\right] + \mu \int_0^{\hat{\gamma}} (\omega - v'(L^S))f(\omega, \gamma)d\gamma - \lambda^A(\omega)\omega = 0. \tag{70}$$

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

$$-\mu C_A' \int_0^{\hat{\gamma}} f(\omega, \gamma)d\gamma + \lambda^A(\omega) = 0. \tag{71}$$

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

Combining (70) and (71), we obtain:

$$\lambda(\omega)\left[-\frac{v''(L^S)}{\omega}L^S - \frac{v'(L^S)}{\omega}\right] + \mu \int_0^{\hat{\gamma}} [\omega - v'(L^S) - \omega C'(\omega L^S)]f(\omega, \gamma)d\gamma = 0. \tag{72}$$

From (36), $v'(L^S) + \omega C' = (1 - T_C')\omega$. Using this relationship and dividing (72) by $v'(L^S)$ and rearranging, we obtain:

$$\frac{T_C'(\omega L^S)}{1 - T_C'(\omega L^S) - C'(\omega L^S)} = \frac{\lambda(\omega)}{\omega \mu \int_0^{\hat{\gamma}} f(\omega, \gamma)d\gamma}\left[1 + \frac{1}{e^S(\omega)}\right]. \tag{73}$$

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

$$\frac{T_P'(\omega L^{NS})}{1 - T_P'(\omega L^{NS})} = \frac{\lambda^{NS}(\omega)}{\omega \mu \int_{\hat{\gamma}}^{\infty} f(\omega, \gamma)d\gamma}\left[1 + \frac{1}{\epsilon^{NS}(\omega)}\right]. \tag{74}$$

38

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

It follows from (71) that:

$$\mu C_A' \int_0^{\hat{\gamma}} f(\omega, \gamma) d\gamma = \mu C'(A) f_\omega(\omega) F_{\gamma|\omega}(\hat{\gamma}) = 0. \tag{75}$$

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 (35) that the two marginal tax rates should be equalized; i.e., $T_P' = T_C'$.

Finding expressions for λ and λ^{NS}

Following Scheuer (2014), Appendix A.3, we integrate equations (68) and (69) 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 $\lim_{\omega \to \infty} \lambda(\omega) = \lambda(0) = \lim_{\omega \to \infty} \lambda^{NS}(\omega) = \lambda^{NS}(0) = 0$, we get:

$$\int_0^\infty \int_0^{\hat{\gamma}(\omega)} [g(\omega, \gamma) - \mu] f(\omega, \gamma) \, d\gamma d\omega - \int_0^\infty \lambda'(\omega) d\omega + \mu \int_0^\infty \Delta T(\omega) f(\omega, \hat{\gamma}) d\omega$$
$$+ \int_0^\infty \int_{\hat{\gamma}(\omega)}^\infty [g(\omega, \gamma) - \mu] f(\omega, \gamma) \, d\gamma d\omega - \int_0^\infty \lambda^{NS\prime}(\omega) d\omega - \mu \int_0^\infty \Delta T(\omega) f(\omega, \hat{\gamma}) d\omega$$
$$= \int_0^\infty \int_0^\infty [g(\omega, \gamma) - \mu] f(\omega, \gamma) \, d\gamma d\omega = \bar{g} - \mu = 0, \quad (76)$$

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 (68) and (69) between 0 and ω, using the relationship given by (76) and the fact that $\lambda(\omega) = \int_0^\omega \lambda'(\omega) d\omega$ and $\lambda^{NS}(\omega) = \int_0^\omega \lambda^{NS\prime}(\omega) d\omega$, we obtain:

$$\lambda(\omega) = \int_0^\omega \int_0^{\hat{\gamma}(\omega)} [g(\omega, \gamma) - \bar{g}] f(\omega, \gamma) \, d\gamma d\omega + \bar{g} \int_0^\omega \Delta T(\omega) f(\omega, \hat{\gamma}) d\omega = 0, \tag{77}$$

$$\lambda^{NS}(\omega) = \int_0^\omega \int_{\hat{\gamma}}^\infty [g(\omega, \gamma) - \bar{g}] f(\omega, \gamma) \, d\gamma d\omega - \bar{g} \int_0^\omega \Delta T(\omega) f(\omega, \hat{\gamma}) d\omega = 0. \tag{78}$$

Because $\lim_{\omega \to \infty} \lambda(\omega) = \lim_{\omega \to \infty} \lambda^{NS}(\omega) = 0$, we can rewrite (77) and (78) as:

$$\lambda(\omega) = \int_{\omega}^{\infty} \int_{0}^{\hat{\gamma}(\omega)} [\bar{g} - g(\omega, \gamma)] f(\omega, \gamma) \, d\gamma d\omega - \bar{g} \int_{\omega}^{\infty} \Delta T(\omega) f(\omega, \hat{\gamma}) d\omega = 0 \qquad (79)$$

$$\lambda^{NS}(\omega) = \int_{\omega}^{\infty} \int_{\hat{\gamma}}^{\infty} [\bar{g} - g(\omega, \gamma)] f(\omega, \gamma) \, d\gamma d\omega + \bar{g} \int_{\omega}^{\infty} \Delta T(\omega) f(\omega, \hat{\gamma}) d\omega = 0 \qquad (80)$$

Optimal tax rules in Proposition 4

Combining (74), (76), (80) while using the definition $b(\omega, \gamma) = g(\omega, \gamma)/\mu$ gives (53). Similarly, combining (73), (76), (79) while using the definition $b(\omega, \gamma) = g(\omega, \gamma)/\mu$ gives (54).