Income Shifting along the Intensive vs. the Extensive Margin: Implications for Optimal Tax Design

Håkan Selin† and Laurent Simula‡

May 31, 2016

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 optimal policy implications. We consider a population of agents differing both in terms of productivity and ability to shift income. We first investigate a model in which income shifting may only occur along an intensive margin. We demonstrate that the social planner should stop shifting when there is a negative dependency between skills and shifting costs, or if the two distributions are independent. In the extensive margin model the social planner should not in general combat shifting. In particular, numerical simulations suggest that the social planner should allow for income shifting if elasticities are heterogeneous in the population.

Keywords: Income Shifting, Optimal Taxation, Labor Income Tax, Capital Income Tax, Dual Taxation

JEL Classification: H21; H24

∗We are grateful to Spencer Bastani, Sören Blomquist, Pierre Boyer, Roger Gordon, Wojciech Kopczuk, Claus Thastrup Kreiner, Joel Slemrod and participants at the workshop on Individual and Small Business Taxation in Warsaw, the Workshop on Behavioral Responses to Income Taxation in Mannheim and seminar participants at Umeå University for comments and suggestions.

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

‡University Grenoble Alpes, UMR GAEL, CS 40700, 38058 Grenoble Cedex 09, France. Email: Laurent.Simula@univ-grenoble-alpes.fr
1 Introduction

Income shifting for tax purposes may appear in many forms. In comprehensive tax systems, like in the United States, people may shift income between the personal and the corporate income tax bases (Gordon and Slemrod, 1998). In dual income tax systems, with separate taxation of labor and capital incomes, people may start up closely held corporations and shift income between the labor and the capital income tax bases (Pirttilä and Selin, 2011, Alstadsæter and Jacob, 2014, and Harju and Matikka, 2016). In the optimal tax literature, it has become standard 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) and Hermle and Peichl (2015). More precisely, models typically assume that the cost of income shifting is smooth and convex, and consider individuals simultaneously choosing how much labor to supply and how much labor income to shift.¹

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, and the novelty of our article is to consider income shifting not only along the intensive, but also along the extensive margin. In particular, we show that the introduction of an extensive margin of income shifting radically modifies implications for policy making.

In this article, we chose to neither model capital accumulation nor tax competition and, accordingly, abstract from standard motives in favor of taxing capital income lower than labor income. Following Piketty and Saez (2013, Section 4), we rather focus on the income shifting mechanism per se and its interaction with labor supply.² In this setting, we examine whether it may be optimal for a benevolent social planner to allow for income shifting, even in the hypothetical situation where all income stems from labor effort.

From a formal viewpoint, we study a static income tax problem in the spirit of Atkinson and Stiglitz (1980) and Slemrod (1994).³ To focus on the income shifting

¹Convex cost functions are also widely used to analyze the normative implications of tax avoidance, see e.g. Slemrod and Kopczuk (2002), Kopczuk (2001) and Chetty (2009).
²Hence, we do not consider income splitting rules, which determine the business income shares to be taxed as capital and labor income and are present in the Nordic countries. The latter often depend on a ‘presumed rate of return’, corresponding to an imputed ‘normal’ rate of return to business assets, see e.g. Sørensen (2005).
³Our paper also relates to the literature on the taxation of entrepreneurial income, but has a different focus. Recently, Scheuer (2014) analyzes a model in which the production side is managed by entrepreneurs whilst both wages and the decision to become either a worker or an entrepreneur are endogenous. Because Scheuer (2014) derives formulas for both the optimal marginal profit and income tax
mechanism from labor earnings into capital earnings, we let each individual’s stock of capital be given exogenously and set it equal to zero. In this setting, a benevolent social planner implements a linear tax on labor incomes and a proportional tax on capital income, with a view of maximizing a weighted sum of individual utilities. Because the planner has two tax instruments at its disposal, it can always eliminate income shifting at no direct cost. Every agent has the possibility to engage in shifting activities between the two tax bases. She therefore faces a trade-off between reducing her total tax liability and paying the costs of shifting. Following Kopczuk (2001), we consider that agents differ with respect to two dimensions: the ability to transform effort into earnings $\omega$ as well as the capacity to shift income, that we capture through a cost parameter $\gamma$. The smaller the value of $\gamma$, the easier it is for an individual to shift income. This might be due to some intrinsic preferences or to access to specific shifting technologies. We allow for a joint distribution of the $\omega$ and $\gamma$, without making any restriction on their potential correlation. Individual utilities depend on both parameters, while social weights only depend on the skill level $\omega$.

Considering a population differing with respect to two dimensions is both empirically relevant and conceptually important. First, there is a large heterogeneity in income shifting opportunities across individuals. Employees, on the one hand, are subject to third-party reporting of wage income and, hence, typically cannot convert labor income into capital income. Business owners, on the other hand, have larger opportunities to shift. When the gap between the two tax rates increases, one would expect some people to transit from employment to self-employment for tax purposes. Second, from a conceptual viewpoint, there would be a concern that our results are contingent on the restriction to linear tax instruments if we only allowed for heterogeneity along the skill dimension.

Our key contribution in light of the earlier literature is to pay particular attention to the cost structure of income shifting. We first re-investigate the case in which income shifting operates along a pure intensive margin; specifically, we consider that for a given $\gamma$ the cost of shifting is increasing and convex in the amount of shifted income. This is a generalization of the assumption considered in most of the previous literature. When $\omega$ and $\gamma$ are either independent or negative quadrant dependent, we find that the rates, it is conceptually related to our article. In Scheuer’s article, agents differ in two dimensions: with respect to their skill level and to their fixed cost of becoming an entrepreneur. In contrast, we examine the role of the cost structure of shifting income into capital income, and we abstract from many of the issues that Scheuer analyzes, among which endogenous wages and non-linear tax functions.

The tax structure is also an important determinant of the choice of organizational form within the population of self-employed, as shown by Edmark and Gordon (2013).
government should equalize the marginal tax rates on labor income and capital income, therefore eliminating income shifting. Negative quadrant dependence, though not in the basic economic toolkit, is a widely used measure of dependence. It means that the joint probability that both \( \omega \) and \( \gamma \) are larger than a given threshold is always smaller as compared to the situation when these variables are independent. It in particular implies negative correlation between \( \omega \) and \( \gamma \). Under positive quadrant dependence instead, the government should set the capital income tax rate at a lower value than the labor income tax rate. In this case, income shifting may be regarded as socially optimal.\(^5\) We therefore see that the result of tax rate equalization, as formulated by Piketty and Saez (2013) in a model with a single dimension of heterogeneity (skills), does only generalize to a more realistic two-dimensional setting if the joint density exhibits certain properties.

We then investigate the case in which income shifting operates along a pure extensive margin, and thus involves a fixed cost. We show that it is usually not socially optimal to equalize the marginal tax rates for labor and capital earnings. In the social optimum, the population is typically partitioned into “shifters” and “non-shifters”. In the shifting sub-population, the marginal incentives to supply labor is determined by the capital income tax rate whilst, in the non-shifting group, by the labor income tax rate. Our optimal income tax formulae do not however inform us about the magnitude of the gap between the two marginal tax rates. We provide numerical simulations to investigate this issue. Regarding preferences, we focus on the case in which the social planner aims at maximizing the well-being of the worst-off individuals (maximin) and abstract from income effects on taxable income. Using Swedish data for skill levels, we calibrate the joint distribution of skills and shifting costs so that, for the actual average values of the labour and capital tax rates, the amount of individuals deciding to shift income corresponds to the empirical estimate. In our benchmark scenario, we find that it is socially optimal to set the labor income tax rate 9 to 15 percentage points higher than the capital income tax, depending on the correlation structure of \( \omega \) and \( \gamma \). Our main conclusions extend to a mixed model, in which both intensive and extensive margins are simultaneously accounted for.

The article is organized as follows. Section 2 reviews the related literature. Section 3 sets up a general model of labor supply and income shifting. Section 4 examines the optimal tax structure when income shifting operates along the intensive margin. Section 5 investigates the consequences of modelling income shifting along the extensive margin. Section 6 concludes.

\(^5\)This result is related to the point made by Kopczuk (2001) in the context of tax avoidance.
2 Related Literature

There is a substantial empirical literature on income shifting. However, in this literature review we focus on theoretical optimal tax papers on income shifting. As already mentioned, our work is closely related to the textbook model presented by Piketty and Saez (2013, Section 4), who model the cost of income shifting as a convex cost in a linear income tax setting. They show in a standard model with heterogeneity in skills only that 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 difference arising under the inverse elasticity rule).

Christiansen and Tuomala (2008) and Fuest and Huber (2001) focus on the role of income shifting in two-type models along the lines of Stiglitz (1982). The informational assumption in the two papers is 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. Christiansen and Tuomala (2008) show in a two-period model, with heterogeneity in the skill dimension and additively separable preferences, that a positive proportional capital income tax is 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 tax when people differ in the skill dimension only.

In the atemporal model of Fuest and Huber (2001), agents instead differ with respect to their wealth endowments. The government imposes non-linear income tax schedules for labor and capital incomes. In addition, the government levies a proportional source tax on capital used in production. It turns out that the optimal source tax is negative. For agents with large endowments, the optimal marginal tax rate is positive and equal to the marginal tax rate for capital income. For agents with small endowments, the marginal tax on capital income is larger than the marginal tax on labor income.

Finally, Hermle and Peichl (2015) derived optimal rules in a model with a 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.

To our knowledge, no one has earlier made the point that income shifting done at a fixed cost potentially can increase social welfare. This does not mean, however,
that related points have not been made. Alesina and Weil (1994) show that Pareto improvements can be achieved by offering individuals a menu of linear tax schedules, where individuals of different skill levels self-select into different tax schemes.\footnote{This idea of “tax buyouts” has recently been carried over to a dynamic overlapping generations economy by Del Negro et al. (2010). In an calibration exercise for the U.S. economy, they find that the introduction of the buyout benefits a significant fraction of the population.} While our “extensive margin model” also builds on the idea that individuals self-select into different tax schedules it nevertheless differs from the literature on “tax buyouts” in two important respects. First, in our extensive margin model the cost of choosing a lower tax rate is a pure waste from the society’s point of view (it does not appear in the government’s revenues). Second, we allow for heterogeneity in two dimensions.

3 The Model

We investigate the situation in which a benevolent policy-maker would like to redistribute income within its population. On the one hand, labor income is taxed in a linear manner, with lump-sum income $G$ and marginal rate $\tau_P$. On the other hand, capital income is taxed in a proportional manner, at the rate $\tau_C$.

The population consists of individuals differing with respect to two dimensions: the ability to transform effort into earnings $\omega$ and the difficulty for them to shift labor earnings into capital earnings, captured by the parameter $\gamma$. The joint distribution of $\omega$ and $\gamma$ is given by the probability density function $f(\omega, \gamma)$. We assume that its support is $\mathbb{R}^+ \times \mathbb{R}^+$ and do not make any restriction on the possible correlation between $\omega$ and $\gamma$. 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.

Individual Choices

To model individual choices, we use the canonical labor-leisure model, augmented with a possibility of income shifting. We denote individual consumption (or net income) by $Y$ and labor supplied by $L$. An individual of skill $\omega$ supplying $L$ units of effort receives a gross income equal to $\omega L$ but incurs a utility loss $v(L; \omega)$, with $v'_L > 0$, $v'_\omega < 0$, $v''_{LL} > 0$ and $v''_{L\omega} \leq 0$. The individual utility function is given by:

$$U(Y, L) = Y - v(L; \omega). \quad (1)$$
Given this specification, there is no income effect on supplied labor. We allow for the possibility that the disutility of labor depends on the skill level. Every individual has the possibility to reduce her income subject to the income tax, from \(\omega L\) to \(Z = \omega L - A\) 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 \(\Gamma(A, \gamma) = k(\gamma) + C(A; \gamma)\). Most of the previous literature has focused on the case where \(C(A; \gamma) = C(A)\). In the following, we investigate the implications of a more general cost structure. To focus on the main mechanisms at stake, we look into the two opposite cases when \(\Gamma(A, \gamma) = C(A; \gamma)\) and \(\Gamma(A, \gamma) = k(\gamma)\). The first specification is a generalization of the smooth and convex cost function traditionally used to analyze income shifting. The second one models income shifting as a fixed cost. All of the main conclusions we draw below extend to the situation where \(\Gamma(A, \gamma) = k(\gamma) + C(A; \gamma)\). The formulas look however more complicated and we therefore refrained from presenting results where the intensive and extensive margins are taken into account simultaneously.

Shifted income \(A\) is subject to the capital income tax. Overall, an individual pays a tax liability of \(\tau_p Z + G - \tau C A\) and thus receives a net income of:

\[
Y = \omega L - (\tau_p Z + G - \tau C A) - \Gamma(A, \gamma). \tag{2}
\]

Individual choices proceed from the maximization of the utility \(U(Y, L)\) subject to the budget constraint (2). The indirect utility is therefore defined as:

\[
V(\omega, \gamma) = \max_{L,A} \{\omega L - \tau_p(\omega L - A) - \tau C A - \Gamma(A, \gamma) + G - v(L; \omega)\}. \tag{3}
\]

Given the additive separability of this specification, the optimal value of \(L\) is independent of \(\gamma\). We call \(L(\omega)\) the optimal supply of effort and \(A(\omega, \gamma)\) the optimal amount of shifting for an individual of type \((\omega, \gamma)\). For later use, we also define:

\[
V^P(\omega) = \max_L \{(1 - \tau_p) \omega L + G - v(L; \omega)\}, \tag{4}
\]

\[
V^C(\omega, \gamma) = \max_L \{(1 - \tau C) \omega L + G - \Gamma(\omega L, \gamma) - v(L; \omega)\}. \tag{5}
\]

For an individual of type \((\omega, \gamma)\), the first optimization program provides the maximum utility \(V^P(\omega, \gamma)\) that can be obtained in the absence of any income shifting \((A = 0)\). We denote by \(L^P(\omega)\) its solution in \(L\). The second optimization program provides the maximum utility \(V^C(\omega, \gamma)\) when the entire labor earnings are shifted \((A = \omega L)\). We
denote by $L^C(\omega, \gamma)$ its solution in $L$.

Policy-Maker’s Choices

The policy-maker chooses the tax rates, $\tau_P$ and $\tau_C$, and the lump-sum income $G$ to maximize a weighted sum of individual utilities:

\[
\int \int g(\omega) V(\omega, \gamma) f(\omega, \gamma) d\gamma d\omega,
\]

subject to:

\[
G + R = \tau_P \int \int [\omega L(\omega) - A(\omega, \gamma)] f(\omega, \gamma) d\gamma d\omega + \tau_C \int \int A(\omega, \gamma) f(\omega, \gamma) d\gamma d\omega.
\]

We consider that the social weights $g(\cdot)$ only depend on productivity levels. A more general specification would be to allow social weights to both depend on $\omega$ and $\gamma$. However, the reasons for which the policy-maker would like to redistribute incomes based on the second dimension of heterogeneity $\gamma$ is not completely transparent to us. Moreover, the results obtained below for income shifting along the pure extensive margin (Section 5) are robust to replacing $g(\omega)$ with $g(\omega, \gamma)$. $R$ 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.

4 The Intensive Margin

In this section, we consider that the cost of income shifting is given by $\Gamma(A, \gamma) = C(A, \gamma)$ where $C > 0$, $C'_A > 0$ and $C''_{AA} > 0$. For a given $\gamma$, the cost is thus positive and convexly increasing in $A$. To avoid corner solutions, we further assume that an individual would face an arbitrarily large cost when shifting her entire earnings. In addition, given our interpretation of $\gamma$ as an underlying difficulty to shift income (e.g., the difficulty to access shifting technologies), we consider that the marginal cost to shift income $C'_A$ is increasing in $\gamma$, i.e., $C''_{A\gamma} > 0$.
4.1 Behavioral Responses

Differentiating (3), we obtain the following first-order conditions of the individual utility maximisation program:

\[ v'(L; \omega) = \omega(1 - \tau_P) \]  \hspace{1cm} (8)
\[ C'_A(A, \gamma) = \tau_P - \tau_C \]  \hspace{1cm} (9)

The optimal effort level depends on the productivity \( \omega \) and on the retention rate \( 1 - \tau_P \). It is in particular increasing with respect to productivity \( \omega \). The optimal amount of shifting, from the individual perspective, is therefore independent of \( \omega \). It only depends on \( \gamma \) and on the tax differential \( \tau_P - \tau_C \). By the implicit function theorem, we see that \( \partial A / \partial \gamma = -C''_A / C''_A < 0 \). On this basis, it is useful to summarize behavioral responses to taxation in terms of elasticities. We call \( \varepsilon_Z \) the elasticity of \( Z \) with respect to the retention rate \( 1 - \tau_P \), whilst keeping \( \tau_C \) constant:

\[ \varepsilon_Z(\omega, \gamma) = -\frac{1 - \tau_P}{Z} \frac{\partial Z}{\partial \tau_P} = \frac{1 - \tau_P}{Z} \left[ \frac{\omega^2}{v''(L)} - \frac{1}{C''_A} \right]. \]  \hspace{1cm} (10)

As shown in the square bracket, a small increase in \( \tau_P \) has two effects on the optimal individual choices. First, the usual substitution effect induces her to work less, which reduces labor earnings. Second, the incentive to shift income from the labor income tax to the capital income tax becomes larger. The part played by the second response in the total behavioral response is therefore:

\[ \sigma(\omega, \gamma) = \frac{\partial A / \partial (\tau_P - \tau_C)}{\partial A / \partial (\tau_P - \tau_C) - \omega \partial L / \partial \tau_P} = \frac{\partial A / \partial (\tau_P - \tau_C)}{\partial Z / \partial (1 - \tau_P)}. \]  \hspace{1cm} (11)

Following Piketty et al. (2014), the shifting elasticity component corresponds to the share of the behavioral response due to income shifting. It is formally defined as:

\[ \varepsilon_A(\omega, \gamma) = \sigma(\omega, \gamma) \varepsilon_Z(\omega, \gamma) = \frac{1 - \tau_P}{Z} \frac{\partial A}{\partial (\tau_P - \tau_C)}. \]  \hspace{1cm} (12)

4.2 Optimal Tax Policy

The social planner chooses the tax rates \( \tau_P, \tau_C \) and the lump-sum income \( G \) so as to maximize the welfare functional (6) subject to the tax revenue constraint (7). Denoting the shadow price of public funds by \( \lambda \), the Lagrangian of the optimization problem is
given by:

\[ \int_\omega \int_\gamma \{ g(\omega)V(\omega, \gamma) + \lambda [\tau_P \omega L(\omega) - (\tau_P - \tau_C)A(\omega, \gamma) - G - R] \} f(\omega, \gamma) d\gamma d\omega. \]  \hspace{0.5cm} (13)

To simplify notations, we omit the arguments of the different functions. The first-order conditions with respect to \( \tau_P \), \( \tau_C \) and \( G \) are respectively:

\[ \int_\omega \int_\gamma \left( b \frac{\partial V}{\partial \tau_P} + \omega L - A + \tau_P \frac{\partial A}{\partial \tau_P} - (\tau_P - \tau_C) \frac{\partial A}{\partial \tau_C} \right) f(\omega, \gamma) d\gamma d\omega = 0, \] \hspace{0.5cm} (14)

\[ \int_\omega \int_\gamma \left( b \frac{\partial V}{\partial \tau_C} + A - (\tau_P - \tau_C) \frac{\partial A}{\partial \tau_C} \right) f(\omega, \gamma) d\gamma d\omega = 0, \] \hspace{0.5cm} (15)

\[ \int_\omega \int_\gamma \left( b - 1 \right) f(\omega, \gamma) d\gamma d\omega = 0. \] \hspace{0.5cm} (16)

The quantity \( b(\omega) = g(\omega)/\lambda \) corresponds to the net social marginal valuation of income of an individual of skill \( \omega \). We call \( \bar{b} \) its average value in the population. Using these weights and the behavioral elasticities \( \varepsilon_A \) and \( \varepsilon_Z \) introduced above, we can rearrange the optimality conditions in a more transparent manner:

\[ \int_\omega \int_\gamma \left( -b \cdot Z + Z - \frac{\tau_P \varepsilon_Z Z}{1 - \tau_P} + \frac{\tau_C}{1 - \tau_P} \varepsilon_A Z \right) f(\omega, \gamma) d\gamma d\omega = 0, \] \hspace{0.5cm} (17)

\[ \int_\omega \int_\gamma \left( -b \cdot A + A + \frac{\tau_P - \tau_C}{1 - \tau_P} \varepsilon_A Z \right) f(\omega, \gamma) d\gamma d\omega = 0, \] \hspace{0.5cm} (18)

\[ \int_\omega \int_\gamma (b - 1) f(\omega, \gamma) d\gamma d\omega = 0 \hspace{0.5cm} \text{i.e.,} \hspace{0.5cm} \bar{b} = 1. \] \hspace{0.5cm} (19)

Combining the first-order conditions with respect to \( \tau_P \) and \( G \), we obtain:

\[ \int_\omega \int_\gamma Z \left( b - \bar{b} + \frac{\tau_P}{1 - \tau_P} \varepsilon_Z Z \right) f(\omega, \gamma) d\gamma d\omega - \int_\omega \int_\gamma \frac{\tau_C}{1 - \tau_P} \varepsilon_A Z f(\omega, \gamma) d\gamma d\omega = 0. \] \hspace{0.5cm} (20)

Introducing the covariance of \( Z \) and \( b \), and rearranging, we can write:

\[ \frac{\tau_P}{1 - \tau_P} = \frac{\text{cov}(Z, b)}{\int_\omega \int_\gamma \varepsilon_Z Z f(\omega, \gamma) d\gamma d\omega} + \frac{\tau_C}{1 - \tau_P} \frac{\int_\omega \int_\gamma \sigma Z \varepsilon_Z Z f(\omega, \gamma) d\gamma d\omega}{\int_\omega \int_\gamma \varepsilon_Z Z f(\omega, \gamma) d\gamma d\omega}. \] \hspace{0.5cm} (21)

We see that the marginal tax rate on labor earnings depends on two terms. The first one looks the same as in the usual optimal linear income tax, with \( Z \) replacing \( \omega L \). It clearly reflects the trade-off between equity, on the numerator, and efficiency, on the denominator. The second term is new. It involves a weighted average of the amount
of income shifted by the individuals, the weights being given by the elasticities $\varepsilon_Z$ and $\varepsilon_A = \sigma\varepsilon_Z$. Rearranging the first-order condition with respect to $\tau_C$, we obtain:

$$\frac{\tau_P - \tau_C}{1 - \tau_P} = \frac{\text{cov}(A, b)}{\int_\omega \int_\gamma \varepsilon_A Z f(\omega, \gamma)d\gamma d\omega}. \tag{22}$$

By definition, the denominator of the right-hand side is positive. This implies that $\tau_P$ is strictly larger than $\tau_C$ if and only if $\text{cov}(A, b) > 0$. Moreover, we know that the function $A$ is increasing in $\gamma$ and independent of $\omega$. Conversely, the social weights $b$ are decreasing in $\omega$ and independent of $\gamma$. Therefore, $\text{cov}(A, b)$ can be rewritten as:

$$\text{cov}(A, b) = \int_\omega \int_\gamma [F(\omega, \gamma) - f(\omega)g(\gamma)]db(\omega)dA(\gamma) \tag{23}$$

where $F(\omega, \gamma)$ is the cumulative density function of the joint distribution of $(\omega, \gamma)$ whilst $f$ and $g$ are the conditional probability density functions. Given that $db(\omega) < 0$ and $dA(\gamma) < 0$, a sufficient condition for $\text{cov}(A, b)$ to be positive is that the square bracket inside the double integral be positive, see Cuadras (2002, Theorem 1). Conversely, a sufficient condition for $\text{cov}(A, b)$ 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 $(\omega, \gamma)$ correspond to a generalization of correlation, called quadratic dependence. In words, positive quadrant dependence means that the joint probability that both $\omega$ and $\gamma$ are larger than a pair $(\hat{\omega}, \hat{\gamma})$ is larger than the product of the two independent probabilities for all possible pairs $(\hat{\omega}, \hat{\gamma})$. Negative quadrant dependence means that this joint probability is smaller. Loosely speaking, quadrant dependence is a stronger version of the more often used concept of correlation. Positive quadrant dependence implies positive correlation, whilst negative quadrant dependence implies negative correlation. The converse statements are not necessarily true. For the bivariate normal distribution however, the sign of correlation coefficient and the sign of the quadratic dependence are always the same.

Putting things together, we can summarize the results as follows.

**Proposition 1** When $\Gamma(A, \gamma) = C(A, \gamma)$, it is socially optimal to set:

1. $\tau_P = \tau_C$ when $\omega$ and $\gamma$ are independent or negative quadrant dependent.
2. $\tau_P > \tau_C$ when $\omega$ and $\gamma$ are positive quadrant dependent;

We therefore see that the result of tax rate equalization, as formulated by Piketty et al.
In a model with a single dimension of heterogeneity (skills), does only generalize to a more realistic two-dimensional setting if the joint density exhibits certain properties. In particular, under positive quadrant dependence, the government should set the capital income tax rate at a lower value than the labor income tax rate. In this case, income shifting may be regarded as socially optimal.

5 The Extensive Margin

We now investigate the case in which income shifting operates along the extensive margin. Without any loss of generality, we directly interpret the parameter of heterogeneity $\gamma$ as a fixed cost.

5.1 Behavioral Responses

When income shifting is done against a fixed cost, a rational individual either shifts nothing ($A = 0$) or her entire labor earnings ($A = \omega L$). In the first case, her utility amounts to $V^P(\omega)$ and in the latter to $V^C(\omega, \gamma)$. Consequently, she chooses $A = 0$ when $V^P(\omega) \geq V^C(\omega, \gamma)$ and $A = \omega L$ when $V^C(\omega, \gamma) > V^P(\omega)$.

Using (4) and (5), we see that

$$(1 - \tau_P)\omega L^P + G - v(L^P; \omega) < (1 - \tau_C)\omega L^C + G - \gamma - v(L^C; \omega), \quad (24)$$

which is equivalent to:

$$\gamma < [(1 - \tau_C)\omega L^C - (1 - \tau_P)\omega L^P] + [v(L^P; \omega) - v(L^C; \omega)]. \quad (25)$$

Given that the fixed costs enters the utility in an additively separable way, $L^C(\omega, \gamma)$ is equal to $L^C(\omega)$. Because $\tau_P \geq \tau_C$, the first square bracket on the right-hand side of (25) is increasing in $\omega$. Moreover, $v''_{L\omega} \leq 0$ implies that the second square bracket is non-decreasing. Consequently, the right-hand side of (25) is monotonically increasing in $\omega$. This implies the following:

**Lemma 2** There is a cut-off level $\hat{\gamma}(\omega)$, non-decreasing in $\omega$, solution in $\gamma$ to Equation (25) written with equality instead of $<$, such that:

1. for $\gamma < \hat{\gamma}(\omega)$, $A(\omega, \gamma) = \omega L^C(\omega)$ and $L(\omega, \gamma) = L^C(\omega)$;
2. for $\gamma \geq \hat{\gamma}(\omega)$, $A(\omega, \gamma) = 0$ and $L(\omega, \gamma) = L^P(\omega)$. 

12
We can alternatively invert the \( \hat{\gamma} \) function and formulate the above lemma in terms of \( \omega \). This will prove particularly useful when formulating the social planner’s optimization problem.

**Corollary 3** There is a cut-off level \( \hat{\omega}(\gamma) \), non-decreasing in \( \gamma \) such that:

1. for \( \omega < \hat{\omega}(\gamma) \), \( A(\omega, \gamma) = 0 \) and \( L(\omega, \gamma) = L^P(\omega) \);
2. for \( \omega \geq \hat{\omega}(\gamma) \), \( A(\omega, \gamma) = \omega L^C(\omega) \) and \( L(\omega, \gamma) = L^C(\omega) \).

As shown in Figure ??, Lemma 2 and Corollary 3 imply that there is a partition of the population into two subgroups. All agents with \( (\omega, \gamma) \) below the (plain blue) curve \( \hat{\omega}(\gamma) \) have no incentive to shift income and, thus, face the labor income tax for all of their earnings. In contrast, the rest of the population – with \( (\omega, \gamma) \) above the graph of \( \hat{\omega}(\gamma) \) – shift their entire earnings, which are then taxed at the rate \( \tau_C \). When the labor marginal income tax rate \( \tau_P \) increases, everything else being constant, a fraction of the population moves from the non-shifting population into the shifting one. This corresponds to a downward shift in \( \hat{\omega}(\gamma) \). Conversely, when the capital tax rate \( \tau_C \) increases, a fraction of agents who previously were shifting their entire earnings join the non-shifting population. This implies an upward shift in the curve of \( \hat{\omega}(\gamma) \).

It is clear from Figure ?? that there is a correspondance between the effect of a raise in \( \tau_P \) and that of a raise in \( \tau_C \) on \( \hat{\omega}(\gamma) \). Applying the implicit function theorem to the definition of \( \hat{\omega}(\gamma) \), it is straightforward to obtain expressions for \( \partial \hat{\omega}/\partial \tau_P \) and \( \partial \hat{\omega}/\partial \tau_C \). They are closely related, with:

\[
\frac{\partial \hat{\omega}}{\partial \tau_P} = \frac{\partial \tau_P}{\partial \tau_C} \frac{\partial \hat{\omega}}{\partial \tau_C} \quad \text{where} \quad \frac{\partial \tau_C}{\partial \tau_P} \leq 0.
\]

(26)

This relationship between \( \partial \hat{\omega}/\partial \tau_P \) and \( \partial \hat{\omega}/\partial \tau_C \) will help us better understand how the shifting margin affects the marginal tax rates \( \tau_P \) and \( \tau_C \) in the social optimum.

### 5.2 Optimal Tax Policy

Given the partition of the population explained above, we can reformulate the social planner’s optimization problem as:

\[
\max_{\tau_P, \tau_C, G} \int_0^\infty \int_0^{\hat{\omega}(\gamma)} g(\omega)V^P(\omega) f(\omega, \gamma) d\omega d\gamma + \int_0^\infty \int_0^{\hat{\omega}(\gamma)} g(\omega)V^C(\omega, \gamma) f(\omega, \gamma) d\omega d\gamma,
\]

(27)
subject to the tax revenue constraint:
\[
\tau_P \int_0^\infty \int_0^{\hat{\omega}(\gamma)} \omega L^P(\omega) f(\omega, \gamma) d\omega d\gamma + \tau_C \int_0^\infty \omega L^C(\omega, \gamma) f(\omega, \gamma) d\omega d\gamma - R - G = 0. \tag{28}
\]

Rearranging the first-order conditions as explained in the Appendix, we obtain the following expressions for the optimal marginal tax rates \(\tau_P\) and \(\tau_C\).

**Proposition 4** In the social optimum, the marginal tax rates \(\tau_P\) and \(\tau_C\) are given by:

\[
\tau_P = \frac{\int_0^\infty \int_0^{\hat{\omega}(\gamma)} \omega L^P(\omega)(1 - b(\omega)) f(\omega, \gamma) d\omega d\gamma + \int_0^\infty (\tau_P \hat{\omega} L^P(\hat{\omega}) - \tau_C \hat{\omega} L^C(\hat{\omega})) \frac{\partial}{\partial \tau_P} f(\hat{\omega}, \gamma) d\gamma}{\int_0^\infty \int_0^{\hat{\omega}(\gamma)} \omega L^P(\omega) \varepsilon L f(\omega, \gamma) d\omega d\gamma},
\]

and

\[
\tau_C = \frac{\int_0^\infty \int_0^{\hat{\omega}(\gamma)} \omega L^C(\omega)(1 - b(\omega)) f(\omega, \gamma) d\omega d\gamma + \int_0^\infty (\tau_P \hat{\omega} L^P(\hat{\omega}) - \tau_C \hat{\omega} L^C(\hat{\omega})) \frac{\partial}{\partial \tau_C} f(\hat{\omega}, \gamma) d\gamma}{\int_0^\infty \int_0^{\hat{\omega}(\gamma)} \omega L^C(\omega) \varepsilon L f(\omega, \gamma) d\omega d\gamma}. \tag{29}
\]

As explained above, the population is usually divided into two fractions in the social optimum. In the sub-population of individuals shifting income, the marginal incentive to supply labor is determined by the capital income tax rate \(\tau_C\). In the sub-population of agents who do not shift income at all, this marginal incentive depends on the labor income tax rate \(\tau_P\). However, we cannot rule out situations in which there would be no shifting in the optimum. In that case, the cut-off level \(\hat{\omega}(\gamma)\) tends to +\(\infty\) and the formulae of Proposition 4 collapse into the "usual" optimal income tax rules, with \(\tau_P = \tau_C\).

In order to gain further insights into the optimal tax rules shown in Proposition 4, it is useful to look at the effects of a small tax reform perturbation around the optimum. More precisely, let us investigate the effects of increasing \(\tau_P\) at the margin (say by an infinitesimal quantity \(\partial \tau_P\)). This tax reform has several effects:

- **Net mechanical effect in the non-shifting population:** The rise \(\partial \tau_P\) in \(\tau_P\) mechanically increases taxes collected from each agent in the non-shifting population, by an amount \(\omega L^P(\omega) \partial \tau_P\). Given quasilinear-in-income preferences, it also reduces each agent’s utility by \(\omega L^P(\omega) \partial \tau_P\). This decreases social welfare by \(g(\omega)\omega L^P(\omega) \partial \tau_P\). Because \(\lambda\) is the unit of count in welfare, the corresponding effect on tax revenue social welfare is \(b(\omega)\omega L^P(\omega) \partial \tau_P\) with \(b(\omega) = g(\omega)/\lambda\). The net mechanical effect of the considered tax reform is therefore \((1 - b(\omega))\omega L^P(\omega) \partial \tau_P\). Summing
over the non-shifting population, we have:

\[ E_1 = + \int_0^\infty \int_0^\infty \omega L^P(\omega) \partial \tau_P f(\omega, \gamma) d\omega d\gamma. \]  

(31)

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

\[-\frac{\omega L^P(\omega) \cdot \varepsilon_l}{1 - \tau_P} \times \partial \tau_P.\]  

(32)

As a result, taxes collected from this agent diminish by \( \tau_P \times (32) \). Summing over the non-shifting population, we obtain:

\[ E_2 = - \int_0^\infty \int_0^\infty \tau_P \omega L^P(\omega) \varepsilon_l \partial \tau_P f(\omega, \gamma) d\omega d\gamma. \]  

(33)

- **Extensive responses:** Because of the increase \( \partial \tau_P \) in \( \tau_P \), \( \hat{\omega}(\gamma) \) goes down at each \( \gamma \) (cf. Figure ??). This induces \( (\partial \hat{\omega}(\gamma) / \partial \tau_P) \partial \tau_P \times f(\hat{\omega}(\gamma), \gamma) \) agents with shifting cost \( \gamma \) to move from the non-shifting to the shifting population. For each of them, the loss in collected taxes amounts to \( \tau_P \hat{\omega}(\gamma) L^P(\hat{\omega}(\gamma) - \tau_C \hat{\omega}(\gamma) L^C(\hat{\omega}(\gamma)) \). Therefore, the overall change in collected taxes due to the extensive responses amounts to:

\[ E_3 = - \int_0^\infty [\tau_P \hat{\omega}(\gamma) L^P(\hat{\omega}(\gamma) - \tau_C \hat{\omega}(\gamma) L^C(\hat{\omega}(\gamma))] \frac{\partial \hat{\omega}(\gamma)}{\partial \tau_P} \partial \tau_P f(\hat{\omega}(\gamma), \gamma) d\gamma. \]  

(34)

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 (29).

In order to better understand (30), we may consider an increase \( \partial \tau_C \) in \( \tau_C \). The first two effects are given by \( E_1 \) and \( E_2 \), with \( \tau_P \) replaced by \( \tau_C \) and \( L^P(\omega) \) replaced by \( L^C(\omega) \). The third effect is the extensive response. Substituting (26) into (34),

\[ E_3 = - \int_0^\infty [\tau_P \hat{\omega}(\gamma) L^P(\hat{\omega}(\gamma) - \tau_C \hat{\omega}(\gamma) L^C(\hat{\omega}(\gamma))] \frac{\partial \hat{\omega}(\gamma)}{\partial \tau_P} \partial \tau_P f(\hat{\omega}(\gamma), \gamma) d\gamma, \]

\[ = - \int_0^\infty [\tau_P \hat{\omega}(\gamma) L^P(\hat{\omega}(\gamma) - \tau_C \hat{\omega}(\gamma) L^C(\hat{\omega}(\gamma))] \frac{\partial \hat{\omega}(\gamma)}{\partial \tau_C} \partial \tau_C f(\hat{\omega}(\gamma), \gamma) d\gamma. \]  

(35)

It simply remains to equalize the sum of the three effects to obtain (30).
5.3 Numerical Simulations

The analysis of a small tax reform perturbation around the social optimum illuminated the mechanisms behind the optimal tax rates formulas of Proposition 4. 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 is the difference between the optimal marginal tax rates $\tau_P$ and $\tau_C$.

First of all, we consider that the social objective is the maximin. In this case, the social planner chooses $\tau_P$ and $\tau_C$ such that tax revenues are maximized. It follows that the social planner would set $\tau_C$ lower than $\tau_P$ 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).

We then need to calibrate 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 $\omega$ and $\gamma$, it is convenient for us to consider that these two parameters are described by a bivariate log-normal distribution. We use Swedish data to calibrate 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 (2014). We provide a more detailed discussion in the Appendix.

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

$$U(Y, L) = Y - \alpha \frac{L^{1+1/g(\epsilon, \omega)}}{1 + 1/g(\epsilon, \omega)},$$

(36)

where $\alpha$ and $\epsilon$ are taste parameters, constant across the population. Importantly, we allow for the possibility that the labor supply elasticity depends positively on the skill level. Alternatively, we could introduce a third dimension of heterogeneity. We have chosen a deterministic relationship between elasticities and skills to ease exposition.

A key issue in empirical public finance is how the earnings elasticity depends on the
skill level. It should be emphasized that \( \varepsilon_{\omega L} \) reflects the real labor supply response, and therefore does not include tax avoidance responses, as opposed to the commonly estimated ‘taxable income elasticity’. As emphasized by Saez et al. (2012, p. 35), “there is compelling evidence of substantial responses of upper income taxpayers to changes in tax rates, at least in the short run. However, in all cases, the response is either due to short-term retiming or income shifting. There is no compelling evidence to date of real responses of upper income taxpayers to changes in tax rates.” While this discussion supports the view that income shifting technologies are concentrated to high-skilled people, it also suggests that we actually know quite little on real labor supply behavior of highly skilled people. If we interpret the real labor supply elasticity more broadly to also accommodate for the migration margin, there is some recent evidence that migration decisions of top income earners may be very sensitive to taxes, as suggested by Kleven et al. (2013) and Kleven et al. (2014). In the baseline simulations, we therefore assume an increasing elasticity, from 0.1 at the bottom of the income distribution to 0.5 for the upper tail.

In Figure 2, the blue curve shows the gap – in percentage points – between \( \tau_P \) and

![Figure 1: Features of the Optimal Allocation (Benchmark Case)](image-url)
\( \tau_C \) 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 \( \tau_P \) and \( \tau_C \), 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-maximizing 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 \( \tau_C \) is considerably more sensitive to changes in the correlation coefficient than \( \tau_P \).

We now investigate to which extent our results are sensitive to the elasticity range. For three different values of the correlation coefficient \( \rho \) (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 labor earnings and one capital earnings. 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. This is however a very particular case, which in practice is often looked at because it strongly eases the calibration. 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.

### 6 Concluding Discussion

This study highlights the distinction between the intensive and extensive margins of income shifting – a distinction neglected in the previous optimal tax literature which focused on the intensive margin. We show that the distinction is an important one from
the point of view of optimal tax policy.

As already emphasized, our model framework is very stylized, and the model can be developed further in several directions. One such obvious simplification is that the shifting costs are exogenous from the government’s point of view in our model. In reality, shifting costs are to some extent endogenous to policy. The government may e.g. affect the costs of starting up a closely held corporation 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.

Our analysis also highlights the need for more empirical work in the area. First, systematic evidence on how real labor earnings elasticities differ across different skill groups and across ‘shifters’ and ‘non-shifters’ would be of great value. Second, empirical evidence on the correlation between earnings abilities and shifting costs would also be important. This is challenging, however, as shifting costs 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. In accordance with earlier optimal tax literature on income shifting we have abstracted from such issues. However, in empirical work they have to be addressed.

Finally, we do not believe that the fact that we model two linear tax schedules is very important for the qualitative conclusions we make. With one skill dimension only, the social planner would always be able to increase social welfare by introducing a non-linear tax on labor income. But with two dimensions of heterogeneity, however, two individuals with the same earnings in the absence of income shifting possibilities, may supply different amounts of earnings in the presence of income shifting if they face different fixed costs. Hence, introducing a non-linear labor income tax would is unlikely to alter the basic mechanism at play.

References


Appendix A: Proof of Proposition 4

We call $\lambda$ the Lagrange multiplier of the budget constraint (28). The derivative of (27) with respect to $\tau_P$ is:

$$\int_0^\infty \int_0^{\hat{\omega}(\gamma)} g(\omega)\omega L^P(\omega)f(\omega,\gamma)d\omega d\gamma. \quad (37)$$

We used the fact that $V^P(\omega) = V^C(\omega,\gamma)$ for $\omega = \hat{\omega}(\gamma)$. We now compute the derivative of the budget constraint (28) with respect to $\tau_P$. We obtain:

$$\int_0^\infty \int_0^{\hat{\omega}(\gamma)} \omega L^P(\omega)f(\omega,\gamma)d\omega d\gamma + \tau_P \int_0^\infty \int_0^{\hat{\omega}(\gamma)} \omega \partial_{\tau_P} L^P(\omega)f(\omega,\gamma)d\omega d\gamma + \int_0^\infty \left( \tau_P \hat{\omega} L^P(\hat{\omega}) - \tau_C \hat{\omega} L^C(\hat{\omega}) \right) \frac{\partial \hat{\omega}}{\partial \tau_C} f(\hat{\omega},\gamma)d\gamma \quad (38)$$

We know write (37) $- \lambda(38) = 0$, rearrange and use the definition of $\varepsilon_{\omega L}$ to obtain (29).

To obtain (30), we compute the derivative of the social objective with respect to $\tau_C$. Using the indifference condition at $\hat{\omega}$, we obtain:

$$\int_0^\infty \int_0^{\hat{\omega}(\gamma)} g(\omega)\omega L^C(\omega)f(\omega,\gamma)d\omega d\gamma. \quad (39)$$

We now compute the derivative of the budget constraint (28) with respect to $\tau_C$:

$$\tau_P \int_0^\infty \hat{\omega} L^P(\hat{\omega}) \frac{\partial \hat{\omega}}{\partial \tau_C} f(\hat{\omega},\gamma)d\omega d\gamma + \int_0^\infty \int_0^{\hat{\omega}} \omega L^C(\omega,\gamma)d\omega d\gamma + \tau_C \left( - \int_0^\infty \hat{\omega} L^C(\hat{\omega}) \frac{\partial \hat{\omega}}{\partial \tau_C} f(\hat{\omega},\gamma)d\omega d\gamma + \int_0^\infty \omega \frac{\partial L^C}{\partial \tau_C} f(\omega,\gamma)d\omega d\gamma \right). \quad (40)$$

We know write (39) $- \lambda(40) = 0$, rearrange and use the definition of $\varepsilon_{\omega L}$ to obtain (30).

Appendix B: Calibration of the fixed cost model

As stated in Section 5.3, we assume that skills, $\omega$, and shifting costs, $\gamma$, follow a bivariate log normal distribution, i.e. $(\omega,\gamma) \sim \ln N(\mu_x, \mu_\gamma, \sigma^2_x, \sigma^2_\gamma, \rho)$, where $\mu_x$ and $\sigma_x$ are the mean and standard deviation of $\ln(x)$. $\rho$ is the correlation coefficient for the bivariate normal distribution of $\ln(\omega)$ and $\ln(\gamma)$. The skill distribution is typically approximated by the distribution of wage rates. We do so as well. We observe the mean and standard deviation on micro data (the LINDA data source) 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.

The basic strategy is to calibrate the shifting cost distribution by choosing $\mu_\gamma$ and $\sigma_\gamma$ 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 particular 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. Ultimately, we therefore solely calibrate $\mu_\gamma$.

We set our target, i.e. the actual fraction of shifters, to be 5%. Alstadsæter and Jacob (2013) 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. 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 $\omega$. Then we find that the fraction of shifters is 5% when $\mu_\gamma = 11.795$. Since our model is very stylized we want to emphasize that we in no way consider this to be a valid 'estimate' of the average shifting costs. The purpose of the calibration exercise is to get a reasonable figure that facilitates qualitative insights. The parameters used in the simulations are summarized in Table 2.

---

8 The correlation coefficient for the transformed distributions is given by $\frac{e^{\rho \sigma_\omega \sigma_\gamma} - 1}{\sqrt{(e^{\sigma_\omega^2 - 1})(e^{\sigma_\gamma^2 - 1})}}$, where the natural exponential function is denoted by $e$. When $\sigma_\omega = \sigma_\gamma$ the correlation coefficient for the transformed distributions is always relatively close to $\rho$, and identical for $\rho = 0$ and $\rho = 1$.

9 If an owner of a closely held corporation distributes profits as wage income her marginal tax rate is $\tau_{personal} + \tau_{payroll}$. 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.
Table 2: Parameter values used in the simulations

<table>
<thead>
<tr>
<th></th>
<th>log((\omega))</th>
<th>log((\gamma))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu)</td>
<td>10.194</td>
<td>11.795</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>0.302</td>
<td>0.302</td>
</tr>
</tbody>
</table>

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