Optimal nonlinear income taxation with endogenous wages and income misreporting

Firouz Gahvari  
Department of Economics  
University of Illinois at Urbana-Champaign, USA

Luca Micheletto  
Department of Law, University of Milan, Italy,  
UCFS Uppsala, Sweden, and  
CESifo Germany

May 2016
Abstract

In a two-type model Stiglitz (1982, 1987) has shown that Pareto efficient taxation requires that the marginal tax rate on the high-skilled agents should be negative, except in the limiting case where the two types of labor are perfect substitutes, in which case it is zero. He has also shown that, while the marginal tax rate on the low-skilled is always positive, its magnitude depends on the elasticity of substitution between the two types of labor; the smaller the elasticity of substitution, the larger the marginal tax rate. In this paper we revisit the aforementioned results by introducing the possibility for agents to misreport, at some cost, their earned income to the tax authority. We show that the extent to which the government uses the marginal income tax rates as an instrument to affect the before-tax relative wages depends on the curvature of the agents’ cost of income-misreporting function. Allowing the government to spend resources to make income-misreporting more difficult, we also characterize the optimal amount of resources devoted to counteract income-misreporting and compare it to the case when different types of labor are perfect substitutes.

*JEL classification:* H21.

*Keywords:* Nonlinear labor income taxation, endogenous wages, income-misreporting, redistribution.
1 Introduction

In a two-type model Stiglitz (1982, 1987) has shown that Pareto efficient taxation requires that (in the standard case where it is the high-skilled self-selection constraint which is binding) the marginal tax rate on the high-skilled agents should be negative, except in the limiting case where the two types of labor are perfect substitutes, in which case it is zero. He has also shown that, while the marginal tax rate on the low-skilled is always positive, its magnitude depends on the elasticity of substitution between the two types of labor; the smaller the elasticity of substitution, the larger the marginal tax rate: the government increasingly relies on the general equilibrium incidence of the tax, the change in the before-tax relative wages, to redistribute income.

In this paper we revisit the aforementioned results by allowing for the possibility that agents misreport their earned income to the tax authority.\(^1\) To keep things as simple as possible, we introduce the possibility of income misreporting by adopting the riskless approach introduced by Usher (1986) and since then used in a number of subsequent contributions.\(^2\) Specifically, once agents have incurred some cost that depends on the amount they misreport, they face no risk of detection.

Modelling the cost-of-misreporting as a smooth convex function of the amount of income misreporting, in the first part of the paper we highlight that the extent to which the government uses the marginal income tax rates as an instrument to affect redistribution via general equilibrium effects will crucially depend on the curvature of the cost-of-misreporting function. More precisely, a lower curvature makes less appealing for the government to manipulate the marginal income tax rates to reduce the wage gap between high- and low-skilled agents. This implies that allowing for the possibility

---

\(^1\)In a standard setting without income-misreporting, the consequences of imperfect substitutability between different skill types for the design of other dimensions of an optimal redistributive policy have been addressed by Naito (1999), who considered the structure of optimal commodity taxes, and by Pirttilä and Tuomala (2002), and Micheletto (2004), who both considered the implications for the desirability of using in-kind transfers (public provision of private goods) as a supplementary policy tool.

of income misreporting mitigates the general equilibrium effects of taxation found in Stiglitz (1982, 1987).³

In the second part of the paper we extend our model by assuming that the cost-of-misreporting function can be affected by the amount of resources spent by the government to counteract income misreporting. This will allow us to characterize the optimal amount of resources spent on fighting income misreporting, and to evaluate how this optimal amount relates to the one that would prevail in a setting without general equilibrium effects on wages. In particular, we will highlight that, compared to a model with exogenous wages, imperfect substitutability between high- and low-skilled workers generates an additional source of gains from raising, at the margin, the amount of resources spent on counteracting income misreporting. However, whether this implies that wage endogeneity strengthens the importance of spending resources to counteract income-misreporting will ultimately depend on the extent to which redistribution is made easier for the government when it can operate both on an ex-ante (wage channel) and on an ex-post basis (tax-transfer channel).

The paper is organized as follows. In Section 2 we set up our baseline model and provide expressions for the optimal marginal income tax rates characterizing a Pareto-efficient tax structure. Section 3 extends our baseline model to allow for the possibility that the government spends resources to affect the private cost of income-misreporting. In Section 4 we provide a numerical example illustrating the effects of the mechanisms investigated in the theoretical part of the paper. Section 5 offers concluding remarks.

³In a recent related paper, Rothschild and Scheuer (2013) have analyzed the problem of optimal redistributive taxation in the context of a Roy model of occupational choices. In that setting, they provide another example of mechanisms that weaken the general equilibrium effects found in Stiglitz (1982, 1987), namely the fact that the tax-induced change in wages can lead some individuals to shift out of the high-wage into the low-wage sector.
2 The baseline model

Consider an economy consisting of two agent types, low-skilled workers (denoted by superscript $\ell$) and high-skilled workers (denoted by superscript $h$). The number of workers of type $j$, with $j = \ell, h$, is denoted by $N^j$. All workers have the same preferences that are defined over consumption and labor and represented by the utility function $U = u(c, L)$, where $c$ denotes consumption of a normal good and $L$ denotes labor supply.

A single consumption good is produced according to a production function that depends on the hours of work supplied by each of the two types of agents:

$$Q = P \left( N^\ell L^\ell, N^h L^h \right) = N^\ell L^\ell P \left( \frac{N^h L^h}{N^\ell L^\ell} \right),$$

where $P$ exhibits constant returns to scale and $N^j$ denotes the number of workers of type $j$ (with $j = \ell, h$). Denoting by $w^j$ the wage rate for an agent of skill type $j$, with each factor receiving its marginal product, we have:

$$w^h = \frac{\partial P \left( N^\ell L^\ell, N^h L^h \right)}{\partial N^h L^h} \equiv P'_2,$$

$$w^\ell = \frac{\partial P \left( N^\ell L^\ell, N^h L^h \right)}{\partial N^\ell L^\ell} \equiv P'_1.$$

Thus, unless the elasticity of substitution between different types of labor is infinite, equilibrium wages are endogenous.

Defining $n$ as $n = \frac{N^h L^h}{N^\ell L^\ell} = \left( N^h \frac{M^h + d^h}{w^h} \right) / \left( N^\ell \frac{M^\ell + d^\ell}{w^\ell} \right)$, we can thus also write:

$$w^h = p'(n),$$

$$w^\ell = p(n) - np'(n).$$

The government’s problem is to design a Pareto-efficient tax policy that allows it to achieve some redistributive goals. The informational structure of the problem is the
following. The government knows the distribution of types in the population, but it does not know “who is who”. This rules out the possibility of using first-best type-specific lump-sum taxes/subsidies. In contrast to what is usually assumed in optimal taxation models, we will assume that, due to the possibility of income-misreporting by agents, earned incomes are not publicly observable either. Instead, what the fiscal authority can rely on is taxing income reported by agents, which will be denoted by $M$, via a general nonlinear income tax $T(M)$.

To model income-misreporting in the simplest possible way, we follow the riskless approach introduced by Usher (1986); specifically, once agents have incurred a cost that is increasing in the amount they misreport, they face no risk of detection. With the true income earned by an agent being equal to $wL$, we denote by $a = wL - M$ the amount of income misreporting. In principle, income-misreporting can either take the form of under-reporting $(a > 0)$ or over-reporting $(a < 0)$. The cost of misreporting is expressed by means of the non-negative function $f(a)$. Furthermore, we assume that $f(\cdot)$ is increasing in the absolute value of $a$, strictly convex, and that $f(0) = f'(0) = 0$.

Since there are only two types in our model, only two points on the optimal tax schedule are of interest. Denoting net-of-tax reported income by $B$ (with $B = M - T(M)$), the problem of choosing the income tax schedule can then be equivalently stated as the problem of selecting two bundles in the $(M, B)$-space, one intended for the low-skilled agents and another intended for the high-skilled ones. Since individuals are free to choose the bundle they prefer, the bundles must be designed by the policy maker subject to (a public budget constraint and) a set of self-selection constraints requiring each ability type to (weakly) prefer the $(M, B)$-bundle intended for him to that intended for the other type. An agent that misrepresents his type is called a mimicker.

Before formalizing the government’s problem, it is useful to consider the optimization problem solved by the agents. It is helpful to divide this problem into two stages. In the first stage, for given values of $M$ and $B$, the agent optimally chooses his labor supply. This gives a conditional labor supply function. The budget available for consumption
is the true labour earnings \( (wL) \) minus the tax based on reported income \( (T = M - B) \) and minus the concealment cost \( (f(a)) \). Formally, the individual optimization problem (suppressing type superscript) in the first stage is:

\[
\max_{c,L} u(c, L) \\
\text{s.t.:} \quad c = wL - M + B - f(wL - M)
\]

Denoting by \( \chi \) the private marginal utility of income, the first order conditions of the above problem are:

\[
\frac{\partial u}{\partial c} = \chi; \\
\frac{\partial u}{\partial L} = -\chi w \left[ 1 - f'(wL - M) \right]; \\
wL - M + B - f(wL - M) - c \geq 0.
\]

Denoting by \( V(M; B; w) \) the maximum value function of the problem above, in the second stage the agent determines how much income to report, i.e. he chooses his preferred \( (M, B) \)-bundle, subject to the link between \( M \) and \( B \) implied by the income tax schedule. This allows us to implicitly define the marginal income tax faced by an agent as:

\[
T'(M) = 1 + \frac{\partial V/\partial M}{\partial V/\partial B} = 1 - MRS_{MB},
\]

where \( MRS_{MB} \) denotes the marginal rate of substitution between \( M \) and \( B \).

Notice also that, by invoking the envelope theorem, we have \( \partial V/\partial M = -[1 - f'(a)] \chi \) and \( \partial V/\partial B = \chi \). Therefore, we also have:
\[ 1 + \frac{\partial V}{\partial M} = f'(a), \]  

implying:

\[ T'(M) = f'(a). \]  

For later purposes, consider the effect on \( L \) of an exogenous marginal variation in either \( B \) or \( M \). Totally differentiating the 3-equations system (1)-(3) with respect to \( B \), we obtain, in matrix form:

\[
\begin{pmatrix}
\frac{\partial^2 u}{\partial c \partial c} & \frac{\partial^2 u}{\partial c \partial L} - \chi w^2 f'' & -1 \\
\frac{\partial^2 u}{\partial L \partial L} & w[1 - f'] & 0
\end{pmatrix}
\begin{pmatrix}
\frac{dc}{dB} \\
\frac{dL}{dB}
\end{pmatrix}
= \begin{pmatrix}
0 \\
-1
\end{pmatrix}
\]

Denoting by \( \Psi \) the determinant of the square matrix above, i.e.

\[ \Psi \equiv - \frac{\partial^2 u}{\partial c \partial c} \left[w(1 - f')\right]^2 - 2 \frac{\partial^2 u}{\partial c \partial L} w[1 - f'] - \frac{\partial^2 u}{\partial L \partial L} + \chi w^2 f'’, \]  

and observing that \( \Psi > 0 \) (from the second-order conditions of the individual maximization problem), applying Cramer’s rule, we get:

\[ \frac{dL}{dB} = \frac{\frac{\partial^2 u}{\partial c \partial c} w[1 - f'] + \frac{\partial^2 u}{\partial c \partial L}}{\Psi} < 0, \]

under the assumption that leisure is a normal good.

To evaluate the effect on \( L \) of an exogenous marginal increase in \( M \), totally differentiate the 3-equations system (1)-(3) with respect to \( M \). We obtain, in matrix form:

\[
\begin{pmatrix}
\frac{\partial^2 u}{\partial c \partial c} & \frac{\partial^2 u}{\partial c \partial L} - \chi w^2 f'' & -1 \\
\frac{\partial^2 u}{\partial L \partial L} & w[1 - f'] & 0
\end{pmatrix}
\begin{pmatrix}
\frac{dc}{dM} \\
\frac{dL}{dM}
\end{pmatrix}
= \begin{pmatrix}
0 \\
-\chi w f''
\end{pmatrix},
\]
and applying Cramer’s rule gives:

\[
\frac{dL}{dM} = -\frac{\partial^2 u}{\partial c \partial c} w \left[1 - f'\right]^2 - \frac{\partial^2 u}{\partial c \partial L} (1 - f') + \chi w f''
\]

\[
= -\left[ \frac{\partial^2 u}{\partial c \partial c} w (1 - f') + \frac{\partial^2 u}{\partial c \partial L} \left(1 - f'\right) + \chi w f'' \right],
\]

where we know that \(-\left[ \frac{\partial^2 u}{\partial c \partial c} w (1 - f') + \frac{\partial^2 u}{\partial c \partial L} \left(1 - f'\right) + \chi w f'' \right] > 0\) and \(\chi w f'' > 0\). Thus, we can conclude that \(dL/dM > 0\).

We are now ready to formalize the government’s problem. A Pareto-efficient tax structure can be characterized as the solution to the following problem:

\[
\max_{M^h, B^h, M^\ell, B^\ell} V \left( M^h, B^h, w^h \right)
\]

subject to:

\[
V \left( M^\ell, B^\ell, w^\ell \right) \geq \bar{U}^\ell,
\]

\[
V \left( M^h, B^h, w^h \right) \geq V \left( M^\ell, B^\ell, w^\ell \right),
\]

\[
P \left( N^\ell \frac{M^\ell + a^\ell}{w^\ell}, N^h \frac{M^h + a^h}{w^h} \right) \geq \left( B^\ell + a^\ell \right) N^\ell + \left( B^h + a^h \right) N^h.
\]

The first constraint prescribes a minimum utility for the low-skilled agents; the second constraint is the binding self-selection constraint requiring high-skilled agents not to behave as mimickers, and the last constraint is the resource constraint of the economy.\(^5\)

Introducing the notation \(g(n) \equiv p(n) - np'(n)\), and treating \(w^\ell\) as an artificial control variable for the government, we can formulate the Lagrangian of the government’s problem above as:\(^6\)

\(^5\)Notice that \((M^j + a^j)/w^j\) denotes the labor supply of a \(j\)-type agent.

\(^6\)The formulation that we use parallels the one adopted by Stiglitz (1987, p.1020) in a setting without income misreporting.
\[ \mathcal{L} = V \left( M^h, B^h, w^h \right) + \mu V \left( M^\ell, B^\ell, w^\ell \right) \\
+ \gamma \left\{ P \left( \frac{N^\ell M^\ell + a^\ell}{w^\ell}, \frac{N^h M^h + a^h}{w^h} \right) - \left( B^\ell + a^\ell \right) N^\ell - \left( B^h + a^h \right) N^h \right\} \\
+ \lambda \left[ V \left( M^h, B^h, w^h \right) - V \left( M^\ell, B^\ell, w^\ell \right) \right] \\
+ \phi \left[ w^\ell - g \left( n \right) \right], \]

where \( \mu \), \( \gamma \), \( \lambda \), and \( \phi \) are the Lagrange multipliers associated with, respectively, the constraint requiring a given level of utility for the low-skilled, the economy’s resource balance constraint, the self-selection constraint, and the constraint capturing the equilibrium value of the wage rate of low-skilled agents.

Writing \( V^h \) and \( V^\ell \) for, respectively, \( V \left( M^h, B^h; w^h \right) \) and \( V \left( M^\ell, B^\ell; w^\ell \right) \), and using a superscript \( h^\ell \) to denote a variable pertaining to a high-skilled mimicker (so that, for instance, \( V^{h\ell} \) stands for \( V \left( M^\ell, B^\ell; w^h \right) \)), the next Proposition characterizes the optimal marginal income tax rates faced by high- and low-skilled agents, as well as the value of the shadow price \( \phi \).

**Proposition 1** At the solution to the government’s Pareto efficient tax problem with endogenous wages and the possibility of income-misreporting, the optimal marginal income tax rates faced by the high- and the low-skilled agents are, respectively, given by:

\[ T^\ell \left( M^h \right) = \frac{\partial V^h}{\partial B^h} \frac{w^h f'' \left( a^h \right)}{\gamma N^h \Psi^h N^h M^h L^h} \phi g' \]  

(10)

\[ T^\ell \left( M^\ell \right) = \frac{\lambda}{\gamma N^\ell} \frac{\partial V^{h\ell}}{\partial B^\ell} \left[ M R S^\ell_{MB} - M R S^{h\ell}_{MB} \right] - \frac{\partial V^\ell}{\partial B^\ell} \frac{w^\ell f'' \left( a^\ell \right)}{\gamma N^\ell L^\ell} \phi g'. \]  

(11)

Moreover, the shadow price \( \phi \) takes a negative value and is given by:

\[ \phi = \left( L^{h\ell} \frac{d w^h}{d w^\ell} - L^\ell \right) \lambda \frac{\partial V^{h\ell}}{\partial B^\ell} M R S^{h\ell}_{MB} < 0. \]  

(12)
Proof. See Appendix A. ■

Starting with (10), notice that, unless the elasticity of substitution between different types of labor is infinity, we have that \( g' = p' - p' - np'' = -np'' > 0 \). Taking into account that, as stated by (12), \( \phi < 0 \), and that both \( \Psi^h \) and \( f'' (a^h) \) are positive, we can conclude that \( T' (M^h) < 0 \). Albeit this result confirms previous findings in the literature (see, e.g., Stiglitz 1982 and 1987), it also highlights that the optimal distortion imposed on the high-skilled agents crucially depends on the curvature of the cost-of-misreporting function. In particular, the less convex is the \( f \)-function and the closer we get to a no-distortion at the top result. Loosely speaking, this means that the easier it is for agents to misreport their true income to the tax authority, and the less effective becomes the marginal income tax rate as an instrument to affect their labor supply and, through this channel, the equilibrium wage rates. For the high-skilled agents this means that, rather than just stimulating their labor supply and lowering their equilibrium wage rate, an increase in the absolute value of their (negative) marginal tax rate will also translate, to an extent that is inversely related to the curvature of the \( f \)-function, in a higher amount of income misreporting.

A similar mechanism applies for the low-skilled agents, whose optimal marginal tax rate is provided by (11). The first term on the right-hand side of (11) is positive due to the agent-monotonicity (single-crossing) condition that implies \( MRS_{MB}^h - MRS_{MB}^{h^l} > 0 \). This term is standard and characterizes the optimal marginal tax rate faced by

---

7 It is worth noticing that in deriving our results we rely on condition (6), which in turn hinges on the properties that we have assumed for the \( f \)-function, and in particular the fact that it is strictly convex, and that \( f (0) = f' (0) = 0 \). These assumptions about the \( f \)-function are standard in the literature, to which we refer in footnote 1, that models income-misreporting by adopting the riskless approach. However, this also implies that eqs. (10)-(11) cannot be applied for dealing with the case when \( f'' = 0 \).

8 Notice that the fact that high-skilled agents face a negative marginal income tax rate implies that for them income misreporting takes the form of over-reporting, i.e. over-stating their true income for tax purposes. For them, an increase in income misreporting means an increase in income over-reporting.

9 According to the agent-monotonicity condition, the indifference curves of low-skilled agents are steeper than those of high-skilled agents at any point in the \( (M,B) \)-space. A formal proof of this result is presented at the beginning of Appendix C. Even though Appendix C refers to the extended model that we consider in Section 3 (where the cost-of-misreporting function also depends on the amount of resources spent by the government to counteract income-misreporting), the result also applies (by
low-skilled agents in a setting where wages are fixed. The second term, which is positive as well (given that \( \phi < 0 \), and that both \( \Psi^f \) and \( f''(a^f) \) are positive) provides an additional reason, related to the exploitation of general equilibrium effects on wages, for distorting downward the labor supply of low-skilled agents. However, as it happened for the corresponding term in (10), also in this case the less convex is the \( f \)-function and the weaker is the incentive for the government to manipulate the marginal tax rate faced by low-skilled agents for the purpose of exploiting general equilibrium effects on wages.\(^{10}\)

3 The extended model

In this section we extend our baseline model to consider a setting where the private cost of income-misreporting can be affected by the amount of resources, denoted by \( R \), devoted by the government to fight tax avoidance/evasion. Specifically, we will assume that the cost of income misreporting is now given by the non-negative function \( f(a, R) \) with the following properties: \( f(a, R) \) is increasing in \( R \) and in the absolute value of \( a \), \( \frac{\partial^2 f(a, R)}{\partial a \partial a} > 0 \), \( \frac{\partial^2 f(a, R)}{\partial R \partial R} > 0 \), and \( f(0, R) = \partial f(0, R)/\partial a = 0 \); moreover, we will also assume that the function \( \partial f(a, R)/\partial R \) is increasing in the absolute value of \( a \).

To simplify notation, hereafter we will denote \( \frac{\partial f(a, R)}{\partial a}, \frac{\partial^2 f(a, R)}{\partial a \partial a}, \frac{\partial^2 f(a, R)}{\partial R \partial R}, \) and \( \frac{\partial^2 f(a, R)}{\partial a \partial R} \) by, respectively, \( f'_1, f'_2, f''_{11}, f''_{22}, \) and \( f''_{12} \).

With \( R \) as an additional policy tool, the Lagrangian of the government’s problem becomes:

\(^{10}\)Notice that the fact that low-skilled agents face a positive marginal income tax rate implies that for them income misreporting takes the form of under-reporting, i.e. under-stating their true income for tax purposes. For them, an increase in the marginal tax rate, rather than just lowering their labor supply and raising their equilibrium wage rate, will partly translate, to an extent that is inversely related to the curvature of the \( f \)-function, in a higher amount of income concealed from the tax authority.
\[ \mathcal{L} = V \left( M^h, B^h, R \right) + \mu V \left( M^\ell, B^\ell, R \right) \\
+ \gamma \left\{ P \left( N^\ell \frac{M^\ell + a^\ell}{w^\ell}, N^h \frac{M^h + a^h}{w^h} \right) - \left( B^\ell + a^\ell \right) N^\ell - \left( B^h + a^h \right) N^h - R \right\} \\
+ \lambda \left[ V \left( M^h, B^h, R \right) - V \left( M^\ell, B^\ell, R \right) \right] \\
+ \phi \left[ w^\ell - g(n) \right], \]

where the differences with the corresponding Lagrangian of the previous section are represented by the fact that the term \(-R\) has been added in the \(\gamma\)-constraint and \(V\) is now also a function of \(R\).

The first thing to notice is that the characterization of the optimal marginal income tax rates provided by (10)-(11) remains valid even after expanding the armory of government’s policy tool to include \(R\) as an additional instrument. The obvious reason for this result is that the optimal marginal income tax rates are obtained by properly combining the first order condition to the government’s problem with respect to \(M^h, B^h, M^\ell, \) and \(B^\ell\); for a given value for \(R\), these first order conditions take exactly the same form as in the problem considered in the previous section.\(^{11}\)

This leaves us with the task of characterizing the optimal policy with respect to the choice of \(R\). The first order condition of the government’s problem with respect to this policy variable is given by:

\(^{11}\)Of course, this does not mean that the numerical value of the marginal income tax rates characterized by (10)-(11) will be identical in the two cases (unless the optimal policy for the government is to set \(R = 0\)). Moreover, with the \(f\)-function now depending both on \(a\) and \(R\), one needs to replace \(f''(a^h)\) in (10) and \(f''(a^\ell)\) in (11) with, respectively, \(f''_h(a^h, R)\) and \(f''_\ell(a^\ell, R)\), and to properly redefine \(\Psi^j\) (for \(j = h, \ell\)) as follows:

\[ \Psi^j = -\frac{\partial^2 u^j}{\partial c \partial L} \left[ w^j \left( 1 - f'(a^j, R) \right) \right] + 2 \frac{\partial^2 u^j}{\partial c^2} w^j \left[ 1 - f'(a^j, R) \right] - \frac{\partial^2 u^j}{\partial L \partial L} + \chi^j \left( w^j \right)^2 f''_j(a^j, R). \]
\[
\frac{\partial L}{\partial R} = (1 + \lambda) \frac{\partial V^h}{\partial R} + \mu \frac{\partial V^\ell}{\partial R} - \lambda \frac{\partial V^{h\ell}}{\partial R} + \gamma \left[ P'_1 N^t \frac{\partial a^\ell}{\partial R} + P'_2 N^h \frac{\partial a^h}{\partial R} - N^t \frac{\partial a^\ell}{\partial R} - N^h \frac{\partial a^h}{\partial R} - 1 \right] - \phi' \frac{dn}{dR} \\
\leq 0,
\] (13)

with associated the Kuhn-Tucker condition:

\[
R \frac{\partial L}{\partial R} = 0,
\]

which allows for the possibility that the optimal value for \( R \) is zero.

Focusing attention on the case of an interior solution (i.e. \( R > 0 \)), Proposition 2 characterizes the optimal value for \( R \).

**Proposition 2** Assuming an interior solution (i.e. \( R > 0 \)), the optimal amount of resources devoted to fight income-misreporting is implicitly characterized by the following condition:

\[
\lambda \frac{\partial V^{h\ell}}{\partial B^\ell} \left[ f'_2 \left( a^{h\ell}, R \right) - f'_2 \left( a^\ell, R \right) \right] + \gamma \left[ \frac{\partial V^h}{\partial B^h} \frac{\partial f''_{12}}{\partial a^h} \left( a^h, R \right) - \frac{\partial V^\ell}{\partial B^\ell} \frac{\partial f''_{12}}{\partial a^\ell} \left( a^\ell, R \right) \right] - \phi' \frac{dn}{dR} \geq 0 \\
\] (14)

**Proof.** See Appendix C. □

To interpret condition (14) it is useful to think at the effects of a policy reform that marginally raise \( R \) while at the same time adjusting \( B^\ell \) and \( B^h \) in such a way to leave unaffected the utility of all non-mimicking agents. This requires accompanying the marginal increase in \( R \) with an upward adjustment in \( B^\ell \) and \( B^h \) given by, respectively, \( dB^\ell = f'_2 \left( a^\ell, R \right) \) and \( dB^h = f'_2 \left( a^h, R \right) \).
The overall revenue cost of the reform is then given by the right-hand side of condition (14).

The left hand side of (14) captures instead the gains that can be achieved by the proposed reform.

On one hand there is a gain associated with relaxing the binding self-selection constraint. This is given by the first term on the left-hand side of condition (14), the term labelled $\Gamma$. As we show in Appendix D, $a^{hl} > a^l > 0$, which in turn implies that $f'_2(a^{hl}, R) > f'_2(a^l, R)$; this means that, while leaving unchanged the utility of the low-skilled agents, the reform lowers the utility of a high-skilled mimicker.

On the other hand, there is a gain, working via general equilibrium effects, associated with a reduction in the wage gap between high- and low-skilled workers. This effect is captured by the second term on the left-hand side of condition (14), the term labelled $\Delta$. As we show in Appendix B, the proposed reform induces both an increase in the labor supply of high-skilled workers and a reduction in the labor supply of low-skilled workers; this implies an increase in the relative wage rate of the latter and a gain through its impact on the $\phi$-constraint.

Finally, notice that in a setting with exogenous wages the counterpart of (14) would be the following simplified condition:

$$\lambda \frac{\partial V^{hl}}{\partial B} \left[ f'_2(a^{hl}, R) - f'_2(a^l, R) \right] = \gamma \left[ 1 + N^l f'_2(a^l, R) + N^h f'_2(a^h, R) \right],$$

i.e. the term labelled $\Delta$ in (14), representing one of the two sources of gains from raising $R$, would vanish. Given that $\Delta > 0$, one might then be tempted to infer that spending resources to fight income-misreporting is relatively more valuable in a setting with endogenous wages than in a setting where wages are exogenous. Intuitive as it

\footnote{It is worth pointing out that also the gain represented by the term $\Delta$ in (14) is, ultimately, a gain that is due to mimicking-deterring effects. This is apparent once one notices that, as implied by (12), the value of the shadow price $\phi$ is related to the value of the Lagrange multiplier $\lambda$, and that $\phi = 0$ when $\lambda = 0$.}

12
may seem, this conclusion is however in general unwarranted since the common terms appearing in (14) and (15) will take a different value in the two settings. In particular, one would also need to take into account that, when different types of labor are imperfect substitutes, the government can rely on an additional channel, i.e. general equilibrium effects on wages, to achieve the desired equity goals. The fact that redistribution can operate both on an ex-ante (wage channel) and on an ex-post basis (tax-transfer channel) tends to weaken the tightness of the self-selection constraint and to lower the equilibrium value of the Lagrange multiplier $\lambda$.

4 A numerical example

For the purpose of illustrating the results of the previous sections, we will present a numerical example that considers a slightly more general framework. In particular, instead of restricting attention to a setting with only two groups of agents, we will assume that there are four different groups. More precisely, while we will maintain the assumption that there is imperfect substitutability between low-skilled and high-skilled agents, we will assume that within each skill type there are two different types of workers who are perfect substitutes even though they have different productivities. In particular, we will assume that the production function is given by:

$$Q = \left[ \frac{1}{10} (N^{\ell1} L^{\ell1} + 1.1 N^{\ell2} L^{\ell2})^\delta + \frac{9}{10} (N^{h1} L^{h1} + 1.4 N^{h2} L^{h2})^\delta \right]^{1/\delta}, \quad (16)$$

where we use the superscripts $\ell1$ and $\ell2$ to denote, respectively, the less productive and the more productive workers among the low-skilled agents, and the superscripts $h1$ and $h2$ to denote, respectively, the less productive and the more productive workers among the high-skilled agents. We assume that there is an equal proportion of agents of each skill type and set $N^{\ell1} = N^{\ell2} = N^{h1} = N^{h2} = 1$. The parameter $\delta$ controls the degree
of substitutability between different skill types.\footnote{In particular, the case of perfect substitutability can be obtained by setting $\delta = 1$, whereas the Leontief case can be approached setting $\delta = -\infty$.}

All agents have identical preferences represented by:

$$ U = \ln c - \frac{L^2}{2}. \quad (17) $$

Regarding the government’s objective function we will consider both the case of a purely utilitarian objective and the case of a max-min social welfare function.

Finally, the cost of income-misreporting function will be given by the following:

$$ f(a, R) = \beta a^2 + \frac{25}{4} a^2 R, \quad (18) $$

where $\beta$ is a parameter affecting the curvature of $f(a, R)$.

4.1 Utilitarian objective

We begin by solving for the case of a purely utilitarian social welfare function, i.e. the case when the government maximizes $W = \sum_{j=L, h} \sum_{i=1, 2} N^{ji} U^{ji}$. We will also initially assume that $R$ is not part of the armory of policy instruments at the government’s disposal (i.e. we set $R = 0$). Setting $\delta = -2$ in (16) and $\beta = 1/4$ in (18), the solution to the government’s problem yields:

<table>
<thead>
<tr>
<th>$M^{f1}$</th>
<th>$B^{f1}$</th>
<th>$a^{f1}$</th>
<th>$L^{f1}$</th>
<th>$w^{f1}$</th>
<th>$T^f(M^{f1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.074</td>
<td>0.210</td>
<td>0.829</td>
<td>0.253</td>
<td>10.49%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$M^{f2}$</th>
<th>$B^{f2}$</th>
<th>$a^{f2}$</th>
<th>$L^{f2}$</th>
<th>$w^{f2}$</th>
<th>$T^f(M^{f2})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.074</td>
<td>0.233</td>
<td>0.837</td>
<td>0.278</td>
<td>11.65%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$M^{h1}$</th>
<th>$B^{h1}$</th>
<th>$a^{h1}$</th>
<th>$L^{h1}$</th>
<th>$w^{h1}$</th>
<th>$T^f(M^{h1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.661</td>
<td>0.606</td>
<td>0.120</td>
<td>1.007</td>
<td>0.774</td>
<td>5.99%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$M^{h2}$</th>
<th>$B^{h2}$</th>
<th>$a^{h2}$</th>
<th>$L^{h2}$</th>
<th>$w^{h2}$</th>
<th>$T^f(M^{h2})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.266</td>
<td>1.171</td>
<td>-0.104</td>
<td>1.071</td>
<td>1.084</td>
<td>-5.19%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$U^{f1}$</th>
<th>$U^{f2}$</th>
<th>$U^{h1}$</th>
<th>$U^{h2}$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.642</td>
<td>-1.575</td>
<td>-0.832</td>
<td>-0.511</td>
<td>-1.1403</td>
</tr>
</tbody>
</table>
Increasing the curvature of (18) by setting $\beta = 2$ would determine the following solution:

Table 2

<table>
<thead>
<tr>
<th></th>
<th>$M^f_1$</th>
<th>$B^f_1$</th>
<th>$a^f_1$</th>
<th>$L^f_1$</th>
<th>$w^f_1$</th>
<th>$T'(M^f_1)$</th>
<th>$M^f_2$</th>
<th>$B^f_2$</th>
<th>$a^f_2$</th>
<th>$L^f_2$</th>
<th>$w^f_2$</th>
<th>$T'(M^f_2)$</th>
<th>$M^h_1$</th>
<th>$B^h_1$</th>
<th>$a^h_1$</th>
<th>$L^h_1$</th>
<th>$w^h_1$</th>
<th>$T'(M^h_1)$</th>
<th>$M^h_2$</th>
<th>$B^h_2$</th>
<th>$a^h_2$</th>
<th>$L^h_2$</th>
<th>$w^h_2$</th>
<th>$T'(M^h_2)$</th>
<th>$U^f_1$</th>
<th>$U^f_2$</th>
<th>$U^h_1$</th>
<th>$U^h_2$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.208</td>
<td>0.323</td>
<td>0.072</td>
<td>0.719</td>
<td>0.390</td>
<td>29.00%</td>
<td>0.208</td>
<td>0.323</td>
<td>0.089</td>
<td>0.694</td>
<td>0.429</td>
<td>35.86%</td>
<td>0.647</td>
<td>0.580</td>
<td>0.031</td>
<td>0.987</td>
<td>0.687</td>
<td>12.47%</td>
<td>1.094</td>
<td>0.931</td>
<td>-0.016</td>
<td>1.119</td>
<td>0.962</td>
<td>-6.47%</td>
<td>-1.212</td>
<td>-1.165</td>
<td>-0.982</td>
<td>-0.716</td>
<td>-1.019</td>
</tr>
</tbody>
</table>

As one can see comparing the results in Tables 1 and 2, the increase in the curvature of the cost-of-misreporting function produces an increase in the absolute values of the marginal tax rates faced by the various agents. Moreover, the increase appears to be especially large for the low-skilled agents. This outcome confirms the predictions of our theoretical analysis of Section 2. In particular, coherently with the results provided by Proposition 1, the less convex is the $f$-function and the weaker is the incentive for the government to manipulate the marginal tax rates for the purpose of exploiting general equilibrium effects on wages.

It is also useful to compare the results obtained when income-misreporting is an available option for the agents with those that would be obtained if income-misreporting were not feasible. In that case, we would get the following solution:
As we can see comparing the results of Table 3 with those in Tables 1 and 2, the optimal marginal tax rates are much more compressed when agents have the possibility to misreport their true income for tax purposes.

Assume now that \( R \) is part of the armory of policy instruments at the government’s disposal. For the case when \( \beta = 1/4 \) in (18), the solution to the government’s problem yields:

### Table 4

\[
\begin{align*}
M^{f1} &= 0.067, \quad B^{f1} = 0.168, \quad a^{f1} = 0.193, \quad L^{f1} = 0.755, \quad w^{f1} = 0.346, \quad T'\left(M^{f1}\right) = 26.39\%; \\
M^{f2} &= 0.067, \quad B^{f2} = 0.168, \quad a^{f2} = 0.218, \quad L^{f2} = 0.753, \quad w^{f2} = 0.380, \quad T'\left(M^{f2}\right) = 29.79\%; \\
M^{h1} &= 0.661, \quad B^{h1} = 0.560, \quad a^{h1} = 0.071, \quad L^{h1} = 1.026, \quad w^{h1} = 0.714, \quad T'\left(M^{h1}\right) = 9.72\%; \\
M^{h2} &= 1.151, \quad B^{h2} = 0.981, \quad a^{h2} = -0.037, \quad L^{h2} = 1.114, \quad w^{h2} = 0.999, \quad T'\left(M^{h2}\right) = -6.13\%; \\
U^{f1} &= -1.374, \quad U^{f2} = -1.319, \quad U^{h1} = -0.991, \quad U^{h2} = -0.679; \\
R/Q &= 2.88\%, \quad W = -1.091.
\end{align*}
\]

As we can see comparing the results in Table 4 with those of Table 1, by spending resources to fight income-misreporting the government is able to raise social welfare by further redistributing towards the group of low-skilled workers. Moreover, by making more costly for agents to misreport their true earned income, the government finds it
optimal to raise the marginal tax rates (in absolute value) in order to exploit to a larger extent the tax-induced general-equilibrium effects on wages.

Let’s now consider the effect of changing $\delta$, i.e. the parameter controlling the degree of substitutability between high- and low-skilled workers, on the optimal amount of resources, as a fraction of total output, spent on counteracting income-misreporting. With $\beta = 1/4$ in (18), we get:

$$\begin{align*}
\delta &= -4 \implies R/Q = 1.76\%; \quad W = -1.081; \\
\delta &= -2 \implies R/Q = 2.88\%; \quad W = -1.091; \\
\delta &= -1 \implies R/Q = 3.67\%; \quad W = -1.088; \\
\delta &= 0.5 \implies R/Q = 6.53\%; \quad W = -1.067.
\end{align*}$$

Table 5

As we can see from the table above, the optimal amount of resources, as a fraction of total output, spent by the government to fight income-misreporting increases as the degree of substitutability between different skill types increases. This result suggests that, by making it easier for the government to redistribute in favor of the low-skilled group, a lower substitutability between different skill types lessens the importance of spending resources to curtail income-misreporting.

4.2 Max-min objective

Let’s now consider the case of a max-min social welfare function, i.e. the case when the government maximizes $W = U^{\alpha_1}$. Starting with the case when $R$ is not part of the armory of policy instruments at the government’s disposal, and setting $\delta = -2$ in (16) and $\beta = 1/4$ in (18), the solution to the government’s problem yields:
As one can see comparing the results in Tables 7 and 8, also in the case of a max-min social welfare function the increase in the curvature of the cost-of-misreporting function produces an increase in the absolute values of the marginal tax rates faced by the various agents. Moreover, the increase appears in this case to be large also for the high-skilled agents.

If income-misreporting were not feasible, we would instead have gotten the following solution:

\begin{table}
\centering
\begin{tabular}{l}
\hline
$M^{f_1} = 0, \quad B^{f_1} = 0.147, \quad a^{f_1} = 0.312, \quad L^{f_1} = 0.569, \quad w^{f_1} = 0.548, \quad T'(M^{f_1}) = 62.41\%$; \\
$M^{f_2} = 0.123, \quad B^{f_2} = 0.195, \quad a^{f_2} = 0.279, \quad L^{f_2} = 0.669, \quad w^{f_2} = 0.603, \quad T'(M^{f_2}) = 55.99\%$; \\
$M^{h_1} = 0.489, \quad B^{h_1} = 0.421, \quad a^{h_1} = 0.094, \quad L^{h_1} = 0.967, \quad w^{h_1} = 0.603, \quad T'(M^{h_1}) = 18.79\%$; \\
$M^{h_2} = 1.095, \quad B^{h_2} = 0.945, \quad a^{h_2} = -0.093, \quad L^{h_2} = 1.188, \quad w^{h_2} = 0.844, \quad T'(M^{h_2}) = -18.63\%$; \\
$U^{f_1} = -1.178, \quad U^{f_2} = -1.148, \quad U^{h_1} = -1.148, \quad U^{h_2} = -0.876; \quad W = -1.178.$
\hline
\end{tabular}
\caption{Table 7}
\end{table}
Table 8

\[ M^{f1} = 0.239, \ B^{f1} = 0.381, \ L^{f1} = 0.435, \ w^{f1} = 0.548, \ T'(M^{f1}) = 69.72\% ; \]
\[ M^{f2} = 0.437, \ B^{f2} = 0.458, \ L^{f2} = 0.724, \ w^{f2} = 0.603, \ T'(M^{f2}) = 45.00\% ; \]
\[ M^{h1} = 0.485, \ B^{h1} = 0.487, \ L^{h1} = 0.804, \ w^{h1} = 0.603, \ T'(M^{h1}) = 35.04\% ; \]
\[ M^{h2} = 1.012, \ B^{h2} = 0.847, \ L^{h2} = 1.20, \ w^{h2} = 0.844, \ T'(M^{h2}) = -20.30\% ; \]
\[ U^{f1} = -1.060, \ U^{f2} = -1.043, \ U^{h1} = -1.043, \ U^{h2} = -0.885; \ W = -1.060. \]

As for the case of a utilitarian social welfare function, the results in Table 8 show that, absent the possibility of misreporting income for tax purposes, the regressivity of the optimal income tax schedule would be magnified.

Assume now that \( R \) is part of the armory of policy instruments at the government’s disposal. For the case when \( \beta = 1/4 \) in (18), the solution to the government’s problem yields:

Table 9

\[ M^{f1} = 0, \ B^{f1} = 0.148, \ a^{f1} = 0.279, \ L^{f1} = 0.509, \ w^{f1} = 0.548, \ T'(M^{f1}) = 69.31\% ; \]
\[ M^{f2} = 0.244, \ B^{f2} = 0.239, \ a^{f2} = 0.206, \ L^{f2} = 0.747, \ w^{f2} = 0.603, \ T'(M^{f2}) = 51.30\% ; \]
\[ M^{h1} = 0.499, \ B^{h1} = 0.399, \ a^{h1} = 0.090, \ L^{h1} = 0.977, \ w^{h1} = 0.603, \ T'(M^{h1}) = 22.36\% ; \]
\[ M^{h2} = 1.100, \ B^{h2} = 0.898, \ a^{h2} = -0.072, \ L^{h2} = 1.217, \ w^{h2} = 0.844, \ T'(M^{h2}) = -18.08\% ; \]
\[ U^{f1} = -1.236, \ U^{f2} = -1.213, \ U^{h1} = -0.213, \ U^{h2} = -0.941; \]
\[ R/Q = 6.76\%, \ W = -1.236. \]

As we can see comparing the results in Table 9 with those of Table 6, by spending resources to fight income-misreporting the government is able to raise the equilibrium utility of the least well off agents. Once again, by making more costly for agents to
misreport their true earned income, the government finds it optimal to raise the marginal
tax rates (in absolute value) in order to exploit to a larger extent the tax-induced
general-equilibrium effects on wages.

Finally, as we did for the utilitarian case, consider now the effect of changing $\delta$, i.e.
the parameter controlling the degree of substitutability between high- and low-skilled
workers, on the optimal amount of resources, as a fraction of total output, spent on
counteracting income-misreporting. With $\beta = 1/4$ in (18), we get:

Table 10

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$R/Q$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-4$</td>
<td>3.91%</td>
<td>$-1.206$</td>
</tr>
<tr>
<td>$-2$</td>
<td>6.76%</td>
<td>$-1.236$</td>
</tr>
<tr>
<td>$-1$</td>
<td>7.28%</td>
<td>$-1.270$</td>
</tr>
<tr>
<td>$0.5$</td>
<td>12.94%</td>
<td>$-1.276$</td>
</tr>
</tbody>
</table>

Table 10 shows that also for the case of a max-min objective it is true that the
optimal amount of resources, as a fraction of total output, spent by the government
to fight income-misreporting increases as the degree of substitutability between differ-
ent skill types increases. However, one can also see that, compared with the case of
a utilitarian objective (see Table 5), the fraction of resources spent on counteracting
income-misreporting is higher, for a given degree of substitutability between skill types,
in the max-min case. This result highlights the fact that, when all agents have access
to the same concealment technology, income-misreporting is relatively less beneficial
for those agents towards whom the government wishes to redistribute. Thus, as the
redistributive tastes of the government becomes stronger, as in the case when moving
from a utilitarian to a max-min social welfare function, the government finds it optimal
to exert more effort in curtailing income-misreporting.
5 Concluding remarks

In a two-type model with high- and low-skilled workers and imperfect substitutability between skill types, Stiglitz (1982, 1987) has shown that Pareto efficient taxation requires that the marginal tax rate on the high-skilled agents should be negative. He has also shown that, while the marginal tax rate on the low-skilled is always positive, its magnitude depends on the elasticity of substitution between the two types of labor; the smaller the elasticity of substitution, the larger the marginal tax rate: the government increasingly relies on the general equilibrium incidence of the tax, the change in the before-tax relative wages, to redistribute income.

In this paper we have revisited the aforementioned results by introducing the possibility for agents to misreport, at some cost, their earned income to the tax authority.

Modelling the concealment cost as a smooth convex function of the amount of income misreporting, in the first part of the paper we have highlighted that the extent to which the government uses the marginal income tax rates as an instrument to affect redistribution, via general equilibrium effects on wages, will crucially depend on the curvature of the cost-of-misreporting function. More precisely, a lower curvature makes less appealing for the government to manipulate the marginal income tax rates for the purpose of reducing the wage gap between high- and low-skilled agents. This implies that allowing for the possibility of income misreporting mitigates the general equilibrium effects of taxation found in Stiglitz (1982, 1987). Moreover, as we illustrate with a numerical example, the magnitude of this dampening effect can be substantial.

In the second part of the paper we have extended our analysis by assuming that the cost-of-misreporting function can be affected by the amount of resources spent by the government to counteract income misreporting. This has allowed us to characterize the optimal amount of resources spent on fighting income misreporting, and to evaluate how this optimal amount relates to the one that would prevail in a setting without general equilibrium effects on wages. In particular, we have shown that, compared to a model
with exogenous wages, imperfect substitutability between high- and low-skilled workers creates an additional source of gains from raising, at the margin, the amount of resources spent on counteracting income misreporting. On purely theoretical grounds, whether this implies that wage endogeneity strengthens the importance of spending resources to counteract income-misreporting by taxpayers will ultimately depend on the extent to which redistribution is easier for the government when it can operate both on an ex-ante (wage channel) and on an ex-post basis (tax-transfer channel). However, for the numerical example that we have considered, the results indicate that the optimal amount of resources, as a fraction of total output, spent by the government to fight income-misreporting increases as the degree of substitutability between different skill types increases.
Appendix

Proof of Proposition 1. Consider first (10). The first order conditions for $M^h$ and $B^h$. They are respectively given by:

$$\frac{\partial V^h}{\partial M^h} + \gamma \left[ P^h_2 \frac{N^h}{w^h} \left( 1 + \frac{N^h}{dM^h} \right) - N^h \frac{dM^h}{dM^h} \right] + \lambda \frac{\partial V^h}{\partial M^h} - \phi g' \frac{dn}{dM^h} = 0, \quad (A1)$$

$$\frac{\partial V^h}{\partial B^h} + \gamma \left[ P^h_2 \frac{N^h}{w^h} \frac{dM^h}{dB^h} - N^h - N^h \frac{dM^h}{dB^h} \right] + \lambda \frac{\partial V^h}{\partial B^h} - \phi g' \frac{dn}{dB^h} = 0. \quad (A2)$$

With $P^h_2 = w^h$, we can rewrite (A1)-(A2) as:

$$\frac{\partial V^h}{\partial M^h} + \gamma \left[ 1 + \frac{N^h}{dM^h} \right] N^h + \lambda \frac{\partial V^h}{\partial M^h} - \phi g' \frac{dn}{dM^h} = 0,$$

$$\frac{\partial V^h}{\partial B^h} + \gamma \left[ \frac{dM^h}{dB^h} - 1 \right] N^h + \lambda \frac{\partial V^h}{\partial B^h} - \phi g' \frac{dn}{dB^h} = 0,$$

or, simplifying terms as:

$$\frac{\partial V^h}{\partial M^h} + \gamma N^h + \lambda \frac{\partial V^h}{\partial M^h} - \phi g' \frac{dn}{dM^h} = 0, \quad (A3)$$

$$\frac{\partial V^h}{\partial B^h} - \gamma N^h + \lambda \frac{\partial V^h}{\partial B^h} - \phi g' \frac{dn}{dB^h} = 0. \quad (A4)$$

Rewrite (A3)-(A4) as:

$$(1 + \lambda) \frac{\partial V^h}{\partial M^h} = \phi g' \frac{dn}{dM^h} - \gamma N^h, \quad (A5)$$

$$(1 + \lambda) \frac{\partial V^h}{\partial B^h} = \phi g' \frac{dn}{dB^h} + \gamma N^h. \quad (A6)$$

Combining (A5)-(A6) gives $\frac{\partial V^h}{\partial M^h} [\phi g' \frac{dn}{dM^h} + \gamma N^h] = \phi g' \frac{dn}{dM^h} - \gamma N^h$ or, equivalently:
\[ 
\gamma N^h \left( 1 + \frac{\partial V^h}{\partial M^h} \right) = \phi g', \quad \frac{dN}{dM^h} = -\frac{\partial V^h}{\partial M^h} \phi g' \quad \frac{dN}{dB^h} = -\frac{\partial V^h}{\partial B^h} \phi g'. \tag{A7} 
\]

Given that the marginal income tax rate faced by a high-skilled agent can be implicitly defined as \( T'(M^h) = 1 - MRS^h_{MB} = 1 + \left( \frac{\partial V^h}{\partial M^h} / \frac{\partial V^h}{\partial B^h} \right) \), from (A7) we have:

\[ 
T'(M^h) = \frac{\phi g' \frac{dn}{dM^h}}{\gamma N^h} - \frac{1}{\gamma N^h} \frac{\partial V^h}{\partial M^h} \phi g' \frac{dn}{dB^h} = \frac{\frac{dn}{dM^h} - \frac{\partial V^h}{\partial M^h} \frac{dn}{dB^h}}{\gamma N^h} \phi g'. \tag{A8} 
\]

With \( \frac{dn}{dM^h} \) and \( \frac{dn}{dB^h} \) being given by:

\[ 
\frac{dn}{dM^h} = \left( 1 + \frac{da^h}{dM^h} \right) \frac{N^h}{w^h N^h M^h + a^h}, \tag{A9} 
\]

\[ 
\frac{dn}{dB^h} = \frac{da^h}{dB^h} \frac{N^h}{w^h N^h M^h + a^h}, \tag{A10} 
\]

we can substitute (A9)-(A10) in (A8) and obtain:

\[ 
T'(M^h) = \frac{\left( 1 + \frac{da^h}{dM^h} \right) N^h}{\gamma N^h w^h N^h M^h + a^h} - \frac{\frac{\partial V^h}{\partial M^h} \frac{da^h}{dM^h} N^h}{\frac{\partial V^h}{\partial M^h} N^h w^h N^h M^h + a^h} \phi g' 
\]

\[ 
= \frac{\left( 1 + \frac{da^h}{dM^h} \right) N^h}{\gamma N^h w^h N^h M^h + a^h} \phi g' 
\]

\[ 
= \frac{1 + \frac{da^h}{dM^h}}{\gamma N^h w^h N^h M^h + a^h} \phi g' 
\]

\[ 
= \frac{1 + \frac{da^h}{dM^h}}{\gamma w^h N^h M^h + a^h} \phi g' 
\]

\[ 
= \frac{1 + \frac{da^h}{dM^h}}{\gamma N^h (M^h + a^h)} \phi g'. \tag{A11} 
\]

25
Finally, notice that combining (8) and (9), one gets:

\[
\frac{dL}{dM} - \frac{dL}{dB} \frac{\partial V/\partial M}{\partial V/\partial B} = - \left[ \frac{\partial^2 u}{\partial \bar{c} \partial c} w (1 - f') + \frac{\partial^2 u}{\partial \bar{c} \partial L} (1 - f') + \chi w f'' \right] \\
\frac{\Psi}{\Psi} - \frac{\partial^2 u}{\partial \bar{c} \partial c} w [1 - f'] + \frac{\partial^2 u}{\partial \bar{c} \partial L} \frac{\partial V/\partial M}{\partial V/\partial B} \\
= - \left[ \frac{\partial^2 u}{\partial \bar{c} \partial c} w (1 - f') + \frac{\partial^2 u}{\partial \bar{c} \partial L} (1 - f') + \chi w f'' - \frac{\partial^2 u}{\partial \bar{c} \partial c} w [1 - f'] \frac{\partial V/\partial M}{\partial V/\partial B} - \frac{\partial^2 u}{\partial \bar{c} \partial L} \frac{\partial V/\partial M}{\partial V/\partial B} \right] \\
\frac{\Psi}{\Psi} \\
= \chi w f'' - \left[ \frac{\partial^2 u}{\partial \bar{c} \partial c} (1 - f') w + \frac{\partial^2 u}{\partial \bar{c} \partial L} \left( \frac{\partial V/\partial M}{\partial V/\partial B} + 1 - f' \right) \right] \\
\frac{\Psi}{\Psi},
\]

which can be simplified, using (5), to obtain:

\[
\frac{dL}{dM} - \frac{dL}{dB} \frac{\partial V/\partial M}{\partial V/\partial B} = \frac{\chi w f''}{\Psi} > 0. 
\]

(A13)

With \(1 + (\partial a^h/\partial M^h) = (dL^h/dM^h) w^h\) and \(\partial a^h/\partial B^h = (dL^h/dB^h) w^h\), we can then use (A13) to substitute in (A11) and thus rewrite the optimal marginal income tax rate faced by high-skilled agents as:

\[
T^h(M^h) = \frac{u^h dL^h}{dM^h} - \frac{\partial u^h}{\partial B^h} dL^h - \frac{\partial w^h}{\partial \bar{c}} M^h + a^h w^h}{\gamma w^h M^h + a^h w^h} \phi ng' \\
= \frac{\gamma w^h M^h + a^h w^h}{\phi ng'} \\
= \frac{\chi^h w^h f''(a^h)}{\Psi^h \gamma N^h M^h + a^h w^h} \phi ng' \\
= \frac{\partial V^h w^h f''(a^h)}{\partial B^h \gamma \Psi^h N^h L^h} \phi ng'.
\]

(A14)
Consider now (11). The first order conditions with respect to $M^t$ and $B^t$ are respectively given by:

\[
\mu \frac{\partial V^t}{\partial M^t} + \gamma \left[ P_i N_i \left(1 + d\right) - N_i \frac{d\alpha^t}{dM^t} \right] - \lambda \frac{\partial V^{ht}}{\partial M^t} - \phi \frac{d\nu}{dM^t} = 0, \tag{A15}
\]

\[
\mu \frac{\partial V^t}{\partial B^t} + \gamma \left[ P_i N_i \frac{d\alpha^t}{dB^t} - N_i \frac{d\alpha^t}{dB^t} \right] - \lambda \frac{\partial V^{ht}}{\partial B^t} - \phi \frac{d\nu}{dB^t} = 0. \tag{A16}
\]

With $P_i = w$, we can rewrite (A15)-(A16) as:

\[
\mu \frac{\partial V^t}{\partial M^t} = -\gamma \left[ N_i \left(1 + d\right) - N_i \frac{d\alpha^t}{dM^t} \right] + \lambda \frac{\partial V^{ht}}{\partial M^t} + \phi \frac{d\nu}{dM^t}, \tag{A17}
\]

\[
\mu \frac{\partial V^t}{\partial B^t} = \gamma N_i + \lambda \frac{\partial V^{ht}}{\partial B^t} + \phi \frac{d\nu}{dB^t}. \tag{A18}
\]

Combining (A17)-(A18) gives:

\[
\frac{\partial V^t}{\partial M^t} \left\{ -\gamma \left[ N_i \frac{d\alpha^t}{dB^t} - N_i \frac{d\alpha^t}{dM^t} \right] + \lambda \frac{\partial V^{ht}}{\partial M^t} + \phi \frac{d\nu}{dM^t} \right\} = \gamma \left[ N_i \left(1 + d\right) - N_i \frac{d\alpha^t}{dM^t} \right] + \lambda \frac{\partial V^{ht}}{\partial M^t} + \phi \frac{d\nu}{dM^t}
\]

Rearranging terms gives:

\[
\gamma N_i + \gamma N_i \frac{\partial V^t}{\partial B^t} = \lambda \frac{\partial V^{ht}}{\partial M^t} + \phi \frac{d\nu}{dM^t} - \frac{\partial V^t}{\partial M^t} \left[ \lambda \frac{\partial V^{ht}}{\partial B^t} + \phi \frac{d\nu}{dB^t} \right],
\]

or, equivalently:

\[
1 + \frac{\partial V^t}{\partial B^t} = \gamma N_i \left\{ \lambda \frac{\partial V^{ht}}{\partial M^t} + \phi \frac{d\nu}{dM^t} - \lambda \frac{\partial V^t}{\partial M^t} \left[ \lambda \frac{\partial V^{ht}}{\partial B^t} + \phi \frac{d\nu}{dB^t} \right] \right\}.
\]
Given that the marginal income tax rate faced by the low-skilled agents is implicitly given by \(1 + \frac{\partial V}{\partial M} / \left(\frac{\partial V}{\partial B}\right)\), we have:

\[
T^t \left(M^t\right) = \frac{1}{\gamma N^t} \left\{ \lambda \frac{\partial V^{h_t}}{\partial B^t} \left[ \frac{\partial V^{h_t}}{\partial M^t} - \frac{\partial V^{h_t}}{\partial B^t} \right] + \phi g' \left[ \frac{dn}{dM^t} - \frac{\partial V^t}{\partial B^t} dM^t \right] \right\}
\]

\[
= \frac{1}{\gamma N^t} \left\{ \lambda \frac{\partial V^{h_t}}{\partial B^t} \left[ MRS_{MB}^t - MRS_{MB}^{bh} \right] + \phi \left[ \frac{dn}{dM^t} - \frac{\partial V^t}{\partial B^t} dM^t \right] g' \right\}.
\]

With \(dn/dM^t\) and \(dn/dB^t\) being given by:

\[
\frac{dn}{dM^t} = -\left(\frac{N h M^{h+a h}}{w^h}\right) \left(1 + \frac{da^t}{dM^t}\right) N^t \frac{w^t}{\left(\frac{M^t + a^t}{w^t}\right)^2}, \tag{A19}
\]

\[
\frac{dn}{dB^t} = -\left(\frac{N h M^{h+a h}}{w^h}\right) \frac{da^t}{dB^t} N^t \frac{w^t}{\left(\frac{M^t + a^t}{w^t}\right)^2}, \tag{A20}
\]

we can rewrite \(T^t \left(M^t\right)\) as:

\[
T^t \left(M^t\right) = \frac{\lambda}{\gamma N^t} \frac{\partial V^{h_t}}{\partial B^t} \left[ MRS_{MB}^t - MRS_{MB}^{bh} \right] + \frac{1}{\gamma N^t} \phi \left[ -\left(\frac{N h M^{h+a h}}{w^h}\right) \frac{w^t}{\left(\frac{M^t + a^t}{w^t}\right)^2} + \frac{\partial V^t}{\partial B^t} \left(\frac{N h M^{h+a h}}{w^h}\right) \frac{da^t}{dB^t} N^t \right] g'
\]

\[
= \frac{\lambda}{\gamma N^t} \frac{\partial V^{h_t}}{\partial B^t} \left[ MRS_{MB}^t - MRS_{MB}^{bh} \right] + \phi \frac{N h M^{h+a h} N^t}{w^t} \left[ \frac{w^t}{\left(\frac{M^t + a^t}{w^t}\right)^2} + \frac{\partial V^t}{\partial B^t} \frac{da^t}{dB^t} \right] g'
\]

\[
= \frac{\lambda}{\gamma N^t} \frac{\partial V^{h_t}}{\partial B^t} \left[ MRS_{MB}^t - MRS_{MB}^{bh} \right] - \frac{\phi}{\gamma N^t} \frac{N h M^{h+a h} N^t}{w^t} \frac{w^t}{\left(\frac{M^t + a^t}{w^t}\right)^2} \left[ \frac{dL^t}{dM^t} - \frac{\partial V^t}{\partial B^t} \frac{dL^t}{dB^t} \right] g'. \tag{A21}
\]

Adapting (A13) for the case of a low-skilled agent, and substituting in (A21) gives:
\[ T'(M^t) = \frac{1}{\gamma N^t} \left\{ \lambda \frac{\partial V^{ht}}{\partial B^t} \left[ MRS_{MB} - MRS_{MB}^{ht} \right] + \phi \frac{-N^h M^{h+a^t} N^t}{w^t \left( N^t M^{h+a^t} \right)^2} w^t \chi^t w^t f'' g' \right\} \]

\[ = \frac{1}{\gamma N^t} \left\{ \lambda \frac{\partial V^{ht}}{\partial B^t} \left[ MRS_{MB} - MRS_{MB}^{ht} \right] - \phi \frac{N^h M^{h+a^t} N^t}{\left( N^t M^{h+a^t} \right)^2} \chi^t w^t f'' g' \right\} \]

\[ = \frac{\lambda}{\gamma N^t} \frac{\partial V^{ht}}{\partial B^t} \left[ MRS_{MB} - MRS_{MB}^{ht} \right] - \frac{\phi}{\gamma N^t} \frac{N^h M^{h+a^t} N^t}{\left( N^t M^{h+a^t} \right)^2} \chi^t w^t f'' g' \]

\[ \]

Thus, we can express the derivative of the Lagrangian of the government’s problem with respect to \( w^t \) as:

\[ \frac{\partial \mathcal{L}}{\partial w^t} = \left( 1 + \lambda \right) L^h \frac{\partial V^h}{\partial B^t} \left[ 1 - f' \left( a^h \right) \right] \frac{dw^h}{dw^t} + \mu L^t \frac{\partial V^t}{\partial B^t} \left[ 1 - f' \left( a^t \right) \right] \]

\[- \lambda L^h \frac{\partial V^h}{\partial B^t} \left[ 1 - f' \left( a^h \right) \right] \frac{dw^h}{dw^t} \]

\[ + \gamma \left\{ w^t N^t \frac{dL^h}{dw^t} + w^h N^h \frac{dL^h}{dw^h} \frac{dw^h}{dw^t} - N^t \frac{da^t}{dw^t} - N^h \frac{da^h}{dw^h} \right\} + \left( 1 - g' \frac{dn}{dw^t} \right) \phi. \]

Multiplying (A6) and (A18) by, respectively, \(-L^h \left[ 1 - f' \left( a^h \right) \right] \frac{dw^h}{dw^t} \) and \(-L^t \left[ 1 - f' \left( a^t \right) \right],\)

we get:
\[-(1 + \lambda) \frac{\partial V^h}{\partial B^h} L^h \left[ 1 - f' \left( a^h \right) \right] \frac{dw^h}{dw^\ell} + \phi g \frac{dn}{dB^h} L^h \left[ 1 - f' \left( a^h \right) \right] \frac{dw^h}{dw^\ell} \]
\[+ \gamma L^h \left[ 1 - f' \left( a^h \right) \right] \frac{dw^h}{dw^\ell} = 0; \quad (A24)\]
\[-\mu \frac{\partial V^\ell}{\partial B^\ell} L^\ell \left[ 1 - f' \left( a^\ell \right) \right] + \gamma N^\ell L^\ell \left[ 1 - f' \left( a^\ell \right) \right] + \lambda \frac{\partial V^h \ell}{\partial B^h} L^\ell \left[ 1 - f' \left( a^\ell \right) \right]
+ \phi g \frac{dn}{dB^h} L^\ell \left[ 1 - f' \left( a^\ell \right) \right] = 0. \quad (A25)\]

Adding (A24)-(A25) to (A23), and simplifying terms, gives:

\[\frac{\partial L}{\partial w^\ell} = \lambda \frac{\partial V^h \ell}{\partial B^h} \left\{ L^\ell \left[ 1 - f' \left( a^\ell \right) \right] - L^h \left[ 1 - f' \left( a^h \right) \right] \right\} \frac{dw^h}{dw^\ell}
+ \gamma \left[ w^\ell N^\ell \frac{dL^\ell}{dw^\ell} + w^h N^h \frac{dL^h}{dw^\ell} - N^\ell \frac{da^\ell}{dw^\ell} - N^h \frac{da^h}{dw^\ell} \right]
+ \gamma \left\{ N^h L^h \left[ 1 - f' \left( a^h \right) \right] \frac{dw^h}{dw^\ell} + N^\ell L^\ell \left[ 1 - f' \left( a^\ell \right) \right] \right\}
+ \phi \left\{ 1 - g \frac{dn}{dw^\ell} + g \frac{dn}{dB^h} L^h \left[ 1 - f' \left( a^h \right) \right] \frac{dw^h}{dw^\ell} + g' \frac{dn}{dB^h} L^\ell \left[ 1 - f' \left( a^\ell \right) \right] \right\} \quad (A26)\]

Noticing that, for \( j = \ell, h \)

\[w^j N^j \frac{dL^j}{dw^j} = w^j N^j \frac{d}{dw^j} \left( \frac{M^j + a^j}{w^j} \right) = w^j N^j \frac{da^j}{dw^j} - \left( \frac{M^j + a^j}{w^j} \right) = N^j \frac{da^j}{dw^j} - N^j L^j, \]

we can further simplify (A26) and obtain:
\[
\frac{\partial L}{\partial w^t} = \lambda \frac{\partial V^h}{\partial B^t} \left[ L^t \left( 1 - f' \left( a^t \right) \right) - L^h \left( 1 - f' \left( a^h \right) \right) \frac{dw^h}{dw^t} \right] \\
- \gamma \left[ N^t L^t f' \left( a^t \right) + N^h L^h f' \left( a^h \right) \frac{dw^h}{dw^t} \right] \\
+ \phi \left\{ 1 - g' \left[ \frac{dn}{dw^t} - \frac{dn}{dB^h} L^h \left( 1 - f' \left( a^h \right) \right) \frac{dw^h}{dw^t} - \frac{dn}{dB^t} L^t \left( 1 - f' \left( a^t \right) \right) \right] \right\}.
\]

Requiring \( \frac{\partial L}{\partial w^t} = 0 \), we then end up with the following expression for \( \phi \):

\[
\phi \left\{ 1 - g' \left[ \frac{dn}{dw^t} - \frac{dn}{dB^h} L^h \left( 1 - f' \left( a^h \right) \right) \frac{dw^h}{dw^t} - \frac{dn}{dB^t} L^t \left( 1 - f' \left( a^t \right) \right) \right] \right\} =
\]

\[
\gamma \left[ N^t L^t f' \left( a^t \right) + N^h L^h f' \left( a^h \right) \frac{dw^h}{dw^t} \right] \\
+ \lambda \frac{\partial V^h}{\partial B^t} \left[ L^h \left( 1 - f' \left( a^h \right) \right) \frac{dw^h}{dw^t} - L^t \left( 1 - f' \left( a^t \right) \right) \right]. \tag{A27}
\]

Let’s now find an expression for \( \frac{dn}{dw^t} \). For this purpose, consider the decision problem faced by an agent for given values of \( M \) and \( B \):

\[
\max_{c,L} u \left( c, L \right) \\
\text{s.t.:} \quad c = wL - M + B - f \left( wL - M \right)
\]

Denoting by \( \chi \) the private marginal utility of income, the first order conditions of the above problem are:

\[
\frac{\partial u}{\partial c} = \chi; \\
\frac{\partial u}{\partial L} = -\chi w \left[ 1 - f' \left( wL - M \right) \right]; \\
wL - M + B - f \left( wL - M \right) - c \geq 0.
\]

To evaluate the effect on \( L \) of an exogenous increase in \( w \), totally differentiate the 3-equations system above with respect to \( w \). We obtain, in matrix form:

31
Applying Cramer’s rule, we get:

\[
\frac{dL}{dw} = \left( \frac{\partial^2 u_{dc}}{\partial c^2} wL \left[ 1 - f' \right]^2 + \frac{\partial^2 u_{dc}}{\partial c^2} (1 - f') L - \chi w L f'' \right) \Psi^{-1} \\
= \left( \frac{\partial^2 u_{dc}}{\partial c^2} w \left[ 1 - f' \right]^2 + \frac{\partial^2 u_{dc}}{\partial c^2} (1 - f') - \chi w f'' \right),
\]

where \( \Psi \) is defined as \( \Psi = -\frac{\partial^2 u_{dc}}{\partial c^2} [w (1 - f')]^2 - 2 \frac{\partial^2 u_{dc}}{\partial c^2} w [1 - f'] - \frac{\partial^2 u_{dc}}{\partial c^2} \chi w^2 f'' > 0. \)

With \( n \equiv N^{hL^h} \frac{\partial L}{\partial L^h} = \left( N^{hM^{h+a^h}} \frac{\partial L}{\partial (L^h)} \right) / \left( N^{\ell M^{\ell+a^\ell}} \frac{\partial L}{\partial (L^\ell)} \right) \), we have:

\[
\frac{dn}{dw^\ell} = \frac{N^{hL^h} dL^h / dB^h}{N^{\ell L^\ell} / dB^\ell} - \frac{N^{hL^h} dL^h / dw^\ell}{N^{\ell L^\ell} / dw^\ell} \\
= \frac{N^{hL^h} dL^h / dw^\ell}{N^{\ell L^\ell} / dw^\ell} \left( \frac{\partial^2 u_{dc}}{\partial c^2} w^h \left[ 1 - f' (a^h) \right]^2 + \frac{\partial^2 u_{dc}}{\partial c^2} (1 - f' (a^h)) - \chi w^h f'' (a^h) \right),
\]

and:

\[
\frac{dn}{dB^h} = \frac{N^{hL^h} dL^h / dB^h}{N^{\ell L^\ell}} = \frac{N^{hL^h} dL^h / dB^h}{N^{\ell L^\ell}} \left( \frac{\partial^2 u_{dc}}{\partial c^2} w^h \left[ 1 - f' (a^h) \right] + \frac{\partial^2 u_{dc}}{\partial c^2} \right),
\]

\[
\frac{dn}{dB^\ell} = -\frac{N^{hL^h} dL^h / dB^\ell}{N^{\ell L^\ell}} = -\frac{N^{hL^h} dL^h / dB^\ell}{N^{\ell L^\ell}} \left( \frac{\partial^2 u_{dc}}{\partial c^2} w^\ell \left[ 1 - f' (a^\ell) \right] + \frac{\partial^2 u_{dc}}{\partial c^2} \right),
\]

32
where in deriving the last two equations we have exploited (8).

Substituting (A28)-(A30) in (A27), and taking into account that $\chi^\ell = \partial V^\ell / \partial B^\ell$ and $\chi^h = \partial V^h / \partial B^h$, we get:

$$
\phi \left\{ 1 - g' \right\} \left[ \frac{\partial^2 u^h (1-f'(a^h))^2 + \frac{\partial^2 u}{\partial x^2} (1-f'(a^h)) - \frac{\partial V^h}{\partial x} u^h f''(a^h)}{\partial w^h} \right]

- \frac{\partial^2 u}{\partial x^2} w^h (1-f'(a^h))^2 + \frac{\partial^2 u}{\partial x^2} (1-f'(a^h)) - \frac{\partial V^h}{\partial x} w^h f''(a^h) + \frac{\partial V^h}{\partial x} w^h (1-f'(a^h)) \frac{dw^h}{dw^c} \right] \right)

= \gamma \left[ N^\ell f^\ell (a^\ell) + N^h f^h (a^h) \right] \frac{dw^h}{dw^c} + \lambda \frac{\partial V^h}{\partial B^c} \left[ L^\ell (1-f'(a^\ell)) \frac{dw^h}{dw^c} - L^h (1-f'(a^h)) \right],

or, equivalently:

$$
\phi \left\{ 1 - g' \right\} \left[ \frac{\partial^2 u^h (1-f'(a^h))^2 + \frac{\partial^2 u}{\partial x^2} (1-f'(a^h)) - \frac{\partial V^h}{\partial x} u^h f''(a^h)}{\partial w^h} \right]

- \frac{\partial^2 u}{\partial x^2} w^h (1-f'(a^h))^2 + \frac{\partial^2 u}{\partial x^2} (1-f'(a^h)) - \frac{\partial V^h}{\partial x} w^h f''(a^h) + \frac{\partial V^h}{\partial x} w^h (1-f'(a^h)) \frac{dw^h}{dw^c} \right] \right)

= \gamma \left[ N^\ell f^\ell (a^\ell) + N^h f^h (a^h) \right] \frac{dw^h}{dw^c} + \lambda \frac{\partial V^h}{\partial B^c} \left[ L^\ell (1-f'(a^\ell)) \frac{dw^h}{dw^c} - L^h (1-f'(a^h)) \right],

Simplifying terms in the equation above, we end up with:
\[
\phi \left\{ 1 - ng' \left[ \frac{\partial V^\ell \, w^\ell \, f^{''}(a^\ell)}{\Psi^\ell} - \frac{\partial V^h \, w^h \, f^{''}(a^h)}{\Psi^h} \frac{dw^h}{dw^\ell} \right] \right\} \\
= \gamma \left[ N^\ell L^\ell \, f'(a^\ell) + N^h L^h \, f'(a^h) \frac{dw^h}{dw^\ell} \right] + \lambda \frac{\partial V^h}{\partial B^\ell} \left[ L^h \left( 1 - f'(a^h) \right) \frac{dw^h}{dw^\ell} - L^\ell \left( 1 - f'(a^\ell) \right) \right], \quad (A31)
\]

Exploiting (6), we can replace \( f'(a^\ell) \) and \( f'(a^h) \) in the last term on the right-hand side of (A31) with, respectively, (11) and (10), and write:

\[
\phi \left\{ 1 - ng' \left[ \frac{\partial V^\ell \, w^\ell \, f^{''}(a^\ell)}{\Psi^\ell} - \frac{\partial V^h \, w^h \, f^{''}(a^h)}{\Psi^h} \frac{dw^h}{dw^\ell} \right] \right\} \\
= \gamma N^\ell L^\ell \left[ \frac{\lambda}{\gamma N^\ell} \frac{\partial V^h}{\partial B^\ell} \left( MR S^\ell_{MB} - MR S^h_{MB} \right) - \frac{\partial V^\ell \, w^\ell \, f^{''}(a^\ell)}{\Psi^\ell} \phi ng' \right] \\\n+ \gamma N^h L^h \frac{dw^h}{dw^\ell} \frac{\partial V^h \, w^h \, f^{''}(a^h)}{\partial B^\ell \, \gamma \Psi^h N^\ell L^\ell} + \lambda \frac{\partial V^h}{\partial B^\ell} \left[ L^h \left( 1 - f'(a^h) \right) \frac{dw^h}{dw^\ell} - L^\ell \left( 1 - f'(a^\ell) \right) \right],
\]

or, rearranging and simplifying terms:

\[
\phi \left\{ 1 - ng' \left[ \frac{\partial V^\ell \, w^\ell \, f^{''}(a^\ell)}{\Psi^\ell} - \frac{\partial V^h \, w^h \, f^{''}(a^h)}{\Psi^h} \frac{dw^h}{dw^\ell} \right] \right\} \\
+ \phi \frac{\partial V^\ell}{\partial B^\ell} \frac{f^{''}(a^\ell)}{\Psi^\ell} - \phi \frac{\partial V^h}{\partial B^\ell} \frac{w^h \, f^{''}(a^h)}{\Psi^h} ng' \\
= \lambda \frac{\partial V^h}{\partial B^\ell} \left( MR S^\ell_{MB} - MR S^h_{MB} \right) L^\ell \\\n+ \lambda \frac{\partial V^h}{\partial B^\ell} \left[ L^h \left( 1 - f'(a^h) \right) \frac{dw^h}{dw^\ell} - L^\ell \left( 1 - f'(a^\ell) \right) \right], \quad (A32)
\]

Notice that, by combining (4) and (6) we have that \( MR S^\ell_{MB} = 1 - T^\prime (M^\ell) = 1 - f'(a^\ell) \) and \( 1 - f'(a^h) = MR S^h_{MB} \). Thus, we can simplify terms in (A32) and
\( \phi = \lambda \frac{\partial V^{ht}}{\partial B^t} \left[ (MRS_{MB}^t - MRS_{MB}^{ht}) L^t + L^h MRS_{MB}^{ht} \frac{dw^h}{dw^t} - L^t MRS_{MB}^t \right] \). (A33)

Simplifying terms in (A33), and noticing that \( dw^h/dw^t < 0 \), gives the result stated in (12).

**Proof of Proposition 2.**

Taking into account that \( P_1 = w^t \) and \( P_2 = w^t \), (13) can be simplified and rewritten as:

\[
(1 + \lambda) \frac{\partial V^h}{\partial R} + \mu \frac{\partial V^t}{\partial R} - \lambda \frac{\partial V^{ht}}{\partial R} - \gamma - \phi g \frac{dn}{dR} \leq 0,
\]

where we also have:

\[
\frac{dn}{dR} = \left( \frac{N^t M^t + a^t}{w^t} \right) \frac{N^h dw^h}{dw^t} - \left( \frac{N^h M^h + a^h}{w^h} \right) \frac{N^t dw^t}{dw^h}.
\]

We will later derive expressions for \( dL^h/dR \) and \( dL^t/dR \) by looking at the optimization problem solved by a given agent. Before doing that, however, notice that, by invoking the envelope theorem, we have that

\[
\frac{\partial V^j}{\partial R} = -\frac{\partial V^j}{\partial B^j} f_2^t \left( a^j, R \right),
\]

\[
\frac{\partial V^{jk}}{\partial R} = -\frac{\partial V^{jk}}{\partial B^k} f_2^t \left( a^{jk}, R \right).
\]

Thus, we can rewrite (B1) as:

\[-(1 + \lambda) \frac{\partial V^h}{\partial B^h} f_2^t \left( a^h, R \right) + (1 + \lambda) \frac{\partial V^{ht}}{\partial B^{ht}} f_2^t \left( a^{ht}, R \right) - \gamma - \phi g \frac{dn}{dR} \leq 0. \]
Multiplying (A6) by \( f'_2 (a^h, R) \) and (A18) by \( f'_2 (a^\ell, R) \) gives the following two equations:

\[
(1 + \lambda) \frac{\partial V^h}{\partial B^h} f'_2 (a^h, R) - \phi g' \frac{dn}{dB^h} f'_2 (a^h, R) - \gamma N^h f'_2 (a^h, R) = 0,
\]

\[
\mu \frac{\partial V^\ell}{\partial B^\ell} f'_2 (a^\ell, R) - \gamma N^\ell f'_2 (a^\ell, R) - \lambda \frac{\partial V^{h\ell}}{\partial B^{h\ell}} f'_2 (a^{h\ell}, R) = 0.
\]

(B4)

Adding up (B4)-(B5) and (B3) gives:

\[
- (1 + \lambda) \frac{\partial V^h}{\partial B^h} f'_2 (a^h, R) - \mu \frac{\partial V^\ell}{\partial B^\ell} f'_2 (a^\ell, R) + \lambda \frac{\partial V^{h\ell}}{\partial B^{h\ell}} f'_2 (a^{h\ell}, R) - \gamma - \phi g' \frac{dn}{dR} \\
+ (1 + \lambda) \frac{\partial V^h}{\partial B^h} f'_2 (a^h, R) - \phi g' \frac{dn}{dR} f'_2 (a^h, R) - \gamma N^h f'_2 (a^h, R) \\
+ \mu \frac{\partial V^\ell}{\partial B^\ell} f'_2 (a^\ell, R) - \gamma N^\ell f'_2 (a^\ell, R) - \lambda \frac{\partial V^{h\ell}}{\partial B^{h\ell}} f'_2 (a^{h\ell}, R) - \phi g' \frac{dn}{dR} f'_2 (a^\ell, R) \leq 0.
\]

(B5)

Simplifying and rearranging terms, and focusing on the case of an interior optimum (i.e., the case where the optimal value of \( R \) is non-zero), we obtain:

\[
\lambda \frac{\partial V^{h\ell}}{\partial B^{h\ell}} f'_2 (a^{h\ell}, R) - f'_2 (a^\ell, R) - \phi g' \left[ \frac{dn}{dR} + \frac{dn}{dR} f'_2 (a^h, R) + \frac{dn}{dR} f'_2 (a^\ell, R) \right] \\
= \gamma \left[ 1 + N^\ell f'_2 (a^\ell, R) + N^h f'_2 (a^h, R) \right].
\]

(B6)

Consider now the sum \( \frac{dn}{dR} + \frac{dn}{dR} f'_2 (a^h, R) + \frac{dn}{dR} f'_2 (a^\ell, R) \) appearing on the left hand side of (B6). We have:

\[
\frac{dn}{dR} + \frac{dn}{dR} f'_2 (a^h, R) + \frac{dn}{dR} f'_2 (a^\ell, R) = \frac{(N^\ell L^\ell) N^h dL^h}{(N^\ell L^\ell)^2} - \frac{(N^h L^h) N^\ell dL^\ell}{(N^\ell L^\ell)^2} \\
+ \frac{N^h}{N^\ell L^\ell} \frac{dL^h}{dR} f'_2 (a^h, R) - \frac{N^h L^h N^\ell dL^\ell}{(N^\ell L^\ell)^2} f'_2 (a^\ell, R).
\]

(B7)
To evaluate \( dL^h / dR \), \( dL^f / dR \), \( dL^h / dB^h \), and \( dL^f / dB^f \), consider the decision problem faced by an agent for given values of \( M \), \( B \) and \( R \):

\[
\begin{align*}
\max_{c,L} u(c, L) \\
\text{s.t.:} & \quad c = wL - M + B - f(wL - M, R) \\
\end{align*}
\]

With \( \chi \) denoting the private marginal utility of income, the first order conditions are:

\[
\begin{align*}
\frac{\partial u}{\partial c} &= \chi; \\
\frac{\partial u}{\partial L} &= -\chi w [1 - f'_1(wL - M, R)]; \\
wL - M + B - f(wL - M, R) - c &\geq 0.
\end{align*}
\]

To evaluate the effect on \( L \) of an exogenous increase in \( R \), totally differentiate the 3-equations system above with respect to \( R \). We obtain, in matrix form:

\[
\begin{pmatrix}
\frac{\partial^2 u}{\partial c^2} & \frac{\partial^2 u}{\partial c \partial L} \\
\frac{\partial^2 u}{\partial L^2} - \chi w^2 f''_{11} & w [1 - f'_1] \\
-1 & 0
\end{pmatrix}
\begin{pmatrix}
\frac{dc}{dR} \\
\frac{dL}{dR} \\
\frac{dB}{dR}
\end{pmatrix}
= \begin{pmatrix}
0 \\
\chi w f'_{12} \\
0
\end{pmatrix},
\]

and applying Cramer’s rule, we get:

\[
\begin{align*}
\frac{dL}{dR} &= - \frac{\left[ \frac{\partial^2 u}{\partial c \partial L} (1 - f'_1) w + \frac{\partial^2 u}{\partial c^2} f''_{11} \right] f'_2 + \chi w f''_{12}}{\Psi}, \quad (B8) \\
\frac{dL}{dB} &= \frac{\frac{\partial^2 u}{\partial c \partial L} w [1 - f'_1] + \frac{\partial^2 u}{\partial L^2}}{\Psi} < 0, \quad (B9)
\end{align*}
\]

where \( \Psi \) is a positive term defined by \( \Psi \equiv -\frac{\partial^2 u}{\partial c \partial L} [w (1 - f'_1)]^2 - 2 \frac{\partial^2 u}{\partial c^2} w [1 - f'_1] - \frac{\partial^2 u}{\partial L^2} + \chi w^2 f''_{11} \).
Substituting (B8)-(B9) in (B7) gives:

\[
\frac{dn}{dR} + \frac{dn}{dB^h} f'_2(a^h, R) + \frac{dn}{dB^\ell} f'_2(a^\ell, R)
= \frac{(N^\ell L^\ell)N^h}{(N^\ell L^\ell)^2 \Psi^h} \left[ \frac{\partial^2 w_h}{\partial c \partial c} (1 - f_1'(a^h, R)) w^h + \frac{\partial^2 w_h}{\partial c \partial L} f'_2(a^h, R) + \chi^h w_h f''_{12}(a^h, R) \right]
+ \frac{(N^h L^h)N^\ell}{(N^\ell L^\ell)^2 \Psi^\ell} \left[ \frac{\partial^2 w_\ell}{\partial c \partial c} (1 - f_1'(a^\ell, R)) w^\ell + \frac{\partial^2 w_\ell}{\partial c \partial L} f'_2(a^\ell, R) + \chi^\ell w_\ell f''_{12}(a^\ell, R) \right]
+ \frac{N^h}{N^\ell L^\ell \Psi^h} \left[ \frac{\partial^2 w_h}{\partial c \partial c} w^h [1 - f_1'(a^h, R)] + \frac{\partial^2 w_h}{\partial c \partial L} \right] f'_2(a^h, R)
- \frac{N^h L^h N^\ell}{(N^\ell L^\ell)^2 \Psi^\ell} \left[ \frac{\partial^2 w_\ell}{\partial c \partial c} w^\ell [1 - f_1'(a^\ell, R)] + \frac{\partial^2 w_\ell}{\partial c \partial L} \right] f'_2(a^\ell, R).
\]

Therefore, simplifying terms we end up with:

\[
\frac{dn}{dR} + \frac{dn}{dB^h} f'_2(a^h, R) + \frac{dn}{dB^\ell} f'_2(a^\ell, R) = -\frac{(N^\ell L^\ell)N^h}{(N^\ell L^\ell)^2 \Psi^h} \chi^h w_h f''_{12}(a^h, R) + \frac{(N^h L^h)N^\ell}{(N^\ell L^\ell)^2 \Psi^\ell} \chi^\ell w_\ell f''_{12}(a^\ell, R),
\]

or, equivalently:

\[
\frac{dn}{dR} + \frac{dn}{dB^h} f'_2(a^h, R) + \frac{dn}{dB^\ell} f'_2(a^\ell, R) = \left[ -\frac{\chi^h w_h f''_{12}(a^h, R)}{\Psi^h L^h} + \frac{\chi^\ell w_\ell f''_{12}(a^\ell, R)}{\Psi^\ell L^\ell} \right] n.
\]

(B10)

Finally, notice that, since \( \chi^h = \partial V^h / \partial B^h \) and \( \chi^\ell = \partial V^\ell / \partial B^\ell \), we can rewrite (B10) as:

\[
\frac{dn}{dR} + \frac{dn}{dB^h} f'_2(a^h, R) + \frac{dn}{dB^\ell} f'_2(a^\ell, R) = \left[ -\frac{\partial V^h w_h f''_{12}(a^h, R)}{\Psi^h L^h} + \frac{\partial V^\ell w_\ell f''_{12}(a^\ell, R)}{\Psi^\ell L^\ell} \right] n.
\]

(B11)

Moreover, given our assumption that \( \partial f(a, R) / \partial R \) is increasing in the absolute value of \( a \), and given that \( a^h < 0 \) and \( a^\ell > 0 \), we have that \( f''_{12}(a^h, R) < 0 \) and \( f''_{12}(a^\ell, R) \). Thus, we can conclude that

\[
\frac{dn}{dR} + \frac{dn}{dB^h} f'_2(a^h, R) + \frac{dn}{dB^\ell} f'_2(a^\ell, R) > 0.
\]

38
Finally, using (B11) to substitute in (B6), we obtain (14).

**Proof of the result that** $a^{h\ell} > a^\ell$.

Consider the utility of a $j$-type individual at a given $(M, B)$-bundle and conditional on a given value for $a$ and a given value for $R$. We have:

$$U(M, B, w^j; a) \equiv u(B + a - f(a, R), (M + a)/w^j).$$

Observe that normality of $c$ ensures normality of $B$ in $U(\cdot)$. In turn, this ensures that the “single-crossing” property is satisfied for $U(\cdot)$.

Next define the marginal rate of substitution between $M$ and $B$ for an agent of type $j$ as,

$$MRS_{MB}(M, B, R, w^j; a) = -\frac{\partial U(B, M, w^j; a)}{\partial U(B, M, w^j; a)} = -\frac{1}{w^j u_c(B + a - f(a, R), (M + a)/w^j)}. \quad (C1)$$

Observe that the normality of $B$ also implies that $MRS_{MB}(M, B, R, w^j; a)$ is increasing in $a$. This happens both because an increase in $a$, for a given $M$ and $w^j$, implies a higher labor supply, and because it implies a larger disposable income, $B + a - f(a, R)$.

A low-skilled agent selecting the $(M^\ell, B^\ell)$-bundle intended for him by the planner chooses $a$ to satisfy,

$$MRS_{MB}(M^\ell, B^\ell, R, w^\ell; a) = 1 - f_1^\ell(a, R). \quad (C2)$$

On the other hand, a high-skilled mimicker would choose $a$ such that

$$MRS_{MB}(M^h, B^h, R, w^h; a) = 1 - f_1^h(a, R). \quad (C3)$$

---

14 The single-crossing or “agent monotonicity” condition requires that the marginal rate of substitution between consumption and income, $B$ and $M$ in this case, to be decreasing in wage so that at any $(M, B)$-bundle, the high-ability agent will have a flatter indifference curve than the low-ability agent. In this way, they can cross only once. See, e.g., Salanié (2011).
Denote the solution to (C2) by \( a^\ell \) and the solution to (C3) by \( a^{hl} \). It follows from (C2)–(C3) that

\[
MRS_{MB}\left(M^\ell, B^\ell, R, w^{\ell}; a^\ell\right) + f'_1\left(a^\ell, R\right) = MRS_{MB}\left(M^\ell, B^\ell, R, w^{h}; a^{hl}\right) + f'_1\left(a^{hl}, R\right).
\]

(C4)

At the same time, the single-crossing property implies that for the same value of \( a \),

\[
MRS_{MB}\left(M^\ell, B^\ell, w^{\ell}; a\right) > MRS_{MB}\left(M^\ell, B^\ell, w^{h}; a\right); \text{ or}
\]

\[
MRS_{MB}\left(M^\ell, B^\ell, R, w^{\ell}; a^\ell\right) + f'_1\left(a^\ell, R\right) > MRS_{MB}\left(M^\ell, B^\ell, R, w^{h}; a^{hl}\right) + f'_1\left(a^{hl}, R\right).
\]

(C5)

Substituting from (C4) for the left-hand side of (C5),

\[
MRS_{MB}\left(M^\ell, B^\ell, R, w^{h}; a^{hl}\right) + f'_1\left(a^{hl}, R\right) > MRS_{MB}\left(M^\ell, B^\ell, R, w^{h}; a^{hl}\right) + f'_1\left(a^{hl}, R\right).
\]

Now with \( MRS_{MB}\left(M, B, R, w^{\ell}; a\right) \) increasing in \( a \) as shown earlier and given our assumption that \( f'_1\left(a, R\right) \) is increasing in \( a \), it follows from the above inequality that \( a^{hl} > a^\ell \).
References


