Universität Trier

Optimal Nonlinear Taxation, Minimum Hours, and the Earned Income Tax Credit

Normann Lorenz
Dominik Sachs

Research Papers in Economics
No. 11/11

Optimal Nonlinear Taxation, Minimum Hours, and the Earned Income Tax Credit

Normann Lorenz* and Dominik Sachs†‡

December 7, 2011

Abstract

We characterize the solution to the optimal nonlinear income taxation problem if individuals face a minimum hours constraint that gives rise to labor supply responses along the extensive margin. We provide conditions for optimal marginal tax rates to be positive everywhere and derive a formula for the optimal participation taxes. This formula shows the additional forces in comparison to the pure extensive labor supply model, can easily be generalized to other contexts of extensive and intensive labor supply responses, and provides a new condition under which an Earned Income Tax Credit (EITC) can be ruled out. In addition, we develop a test for the second-best Pareto-efficiency of any income tax schedule. The test is expressed in reduced form and can be applied if the income distribution and empirical estimates of the extensive and intensive labor supply elasticities are known. Carefully parameterized simulations suggest that an EITC is optimal. An exogenous restriction that the welfare benefit cannot be set below a certain level causes the EITC to be less pronounced. On the other hand, exogenous government revenue requirements cause the EITC to be more pronounced in relative terms, because the welfare benefit decreases while the participation subsidy remains fairly constant. However, with the restriction of a fixed welfare benefit an increase in revenue requirements leads to a sharp decline of the participation subsidy.

JEL-**classification:** H 21, H 23.

Keywords: Optimal taxation, participation taxes, extensive margin, intensive margin.

---

*Universität Trier, Universitätsring 15, 54286 Trier, Germany; email: Normann.Lorenz@uni-trier.de; phone: 0049-651-2012624

†Universität Konstanz, Fach D 136, 78457 Konstanz, Germany; email: Dominik.Sachs@uni-konstanz.de; phone: 0049-7531-882189

‡We thank Friedrich Breyer, Sarah Brockhoff, Sebastian Findeisen, Laurence Jacquet, Mathias Kifmann, Etienne Lehmann, Nick Netzer, Florian Scheuer, Dirk Schindler and Iván Werning as well as seminar participants in Gerzensee and participants of the 4th RGS Doctoral Conference in Dortmund for helpful comments and suggestions. Sachs acknowledges the hospitality of Stanford University where part of this research was undertaken.

1 Introduction

Redistribution schemes that support the unemployed and individuals with low income exist in all developed countries. There is, however, a public debate on the appropriate design of such schemes: Should redistribution mainly be targeted to the unemployed with a Negative Income Tax (NIT) or mainly to low incomes with an Earned Income Tax Credit (EITC)?

Economists can contribute to that debate by analyzing the equity-efficiency trade-off inherent in such redistribution schemes and by deriving conditions under which one of these schemes may be better suited than the other to achieve the distributional goals of society.

A large part of the literature related to that debate, including the classic article by Mirrlees (1971), derives the properties of a tax transfer system that maximizes a social welfare function (SWF)[1] if individuals' productivity is unobservable and labor supply is continuous, i.e. individuals adjust their labor supply along the intensive margin.The main result of this literature is that in general marginal tax rates are positive,[2] which results in a Negative Income Tax to be optimal. Zero and low incomes receive a transfer, but face a positive marginal tax rate. This implies that the tax when working is always higher than the tax when being unemployed, so all individuals face a positive participation tax when entering the labor market: $T_{part}(Y) = T(Y) - T(0) > 0$, see Figure 1(a).

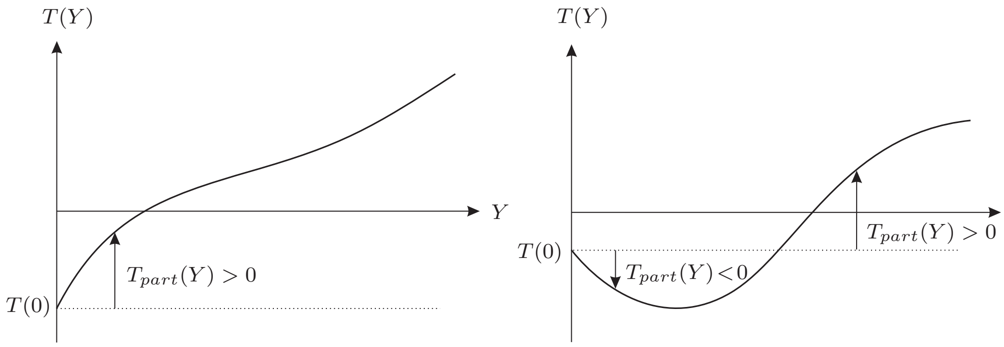


Figure 1: (a) Negative Income Tax (NIT) (b) Earned Income Tax Credit (EITC)

However, as pointed out by Diamond (1980) and more recently by Saez (2002) and Choné and Laroque (2011b), this result critically hinges on the underlying labor supply model. If individuals cannot choose the number of hours they work, but only, whether to work or not, (so there is only an extensive margin), results change drastically. Under the reasonable assumption that the social marginal utility of income of those with very low income is higher than the marginal value of public funds, participation taxes for this group are negative, i.e. individuals with low income receive a higher transfer than the unemployed. This makes an Earned Income Tax Credit (EITC) optimal, see Figure 1(b).[3]

These contradicting results raise the question which labor supply model applies. As the empirical literature on this topic points out, individuals adjust their labor supply along both, the extensive

[1] There is also a literature that characterizes the whole set of second-best Pareto-efficient nonlinear tax schedules; see Stiglitz (1982), Werning (2007) and Choné and Laroque (2011a) for the intensive and Laroque (2005) for the extensive labor supply model.

[2] See Mirrlees (1971), Seade (1982), Tuomala (1990), Diamond (1998), Werning (2000), Saez (2001) and Hellwig (2007).

[3] Choné and Laroque (2011b) consider the two different types of an EITC: either a tax schedule with negative marginal tax rates (as in Figure 1(b)) or one with a discontinuity (as in Figure 2).

and the intensive margin.[4] This implies that the properties of an optimal redistribution scheme should be derived within a framework that accounts for the decision of how many hours to work as well as the decision of whether to work at all.

Saez (2002) was first to consider the optimal income tax problem within such a framework. In his model, each individual can choose among two different occupations (intensive margin) and unemployment (extensive margin), where occupations differ in earnings and disutility of work. Besides deriving a formula for the optimal marginal tax rates[5] he calibrates his model for the US and shows that the EITC rather than the NIT is optimal if participation elasticities for low income earners compared to hours of work elasticities are sufficiently high.

Putting the focus on the interplay of these two elasticities, Saez (2002) did not refer to a particular underlying labor supply model and a specific reason for the extensive margin. These two aspects have subsequently been analyzed in greater detail by Jacquet, Lehmann, and Van der Linden (2010). They incorporate disutility of participation as a reason for the extensive margin in a model of continuous labor supply with income effects and derive the optimal tax schedule for the case of a continuous earnings distribution.[6] They show conditions under which optimal participation taxes are positive so that an EITC can be ruled out: The social marginal utility of income for the lowest income group has to be smaller than the marginal value of public funds, so that participation taxes are positive for this group. Adding the conditions for positive marginal tax rates then ensures that participation taxes are positive for all income levels. However, their simulation results show that usually only the second condition is met, so that an EITC is optimal that is characterized by positive marginal tax rates, but negative participation taxes for low incomes. This is due to a discontinuity in the tax schedule, see Figure 2.

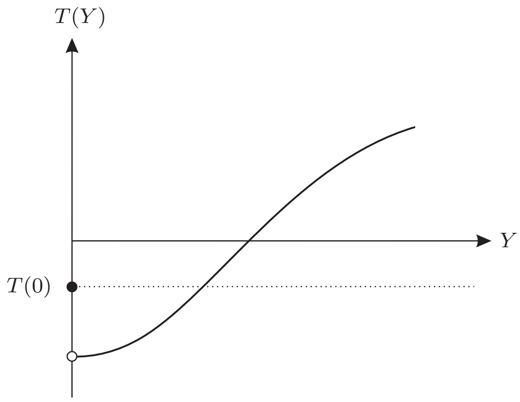


Figure 2: Negative participation tax for low incomes (i.e. Earned Income Tax Credit) without negative marginal tax rates due to a discontinuity in the tax schedule

Boone and Bovenberg (2004) also consider the optimal non-linear tax problem in the presence of both margins. In their model individuals have to search for a job and can either be unemployed voluntarily (no search) or involuntarily (search without finding a job). Unlike Saez (2002) and

[4]See Heckman (1993), Eissa and Liebman (1996), Meyer and Rosenbaum (2001) and Meghir and Phillips (2008).

[5]Since there is a finite number of occupations (and thus income levels) in this discrete setting, the marginal tax rate is the change in taxes relative to the change in income of two 'adjacent' occupations. However, in Saez (2000), an earlier version of the paper, he also derives the formula for the marginal tax rate for the continuous case.

[6]The concept of disutility of participation is closely related to the concept of fixed costs of work, that are considered as an important reason for extensive labor supply responses in the labor economics literature, see Hausman (1985) and Cogan (1981).

Jacquet, Lehmann, and Van der Linden (2010) they consider the case of one-dimensional heterogeneity where individuals only differ in productivity w; thus all individuals of a given w either search or do not search. Those who search are then divided into two groups: those who find a job and those who are involuntarily unemployed. They show that participation taxes can be negative and that in this case the marginal tax rate for the lowest productivity is negative, too.

We add to this literature on optimal nonlinear income taxation with both extensive and intensive margin in the following respects:

First, we analyze the consequences of a minimum hours constraint. Such a constraint can be due to several reasons: Some tasks require the worker to be present for a certain amount of time and some occupations need constant exercise (and thus a minimum amount of working time per week or month) to keep quality at the desired level. There may also be fixed costs on the side of the firm (e.g., for training or for providing equipment) on which the firm wants to economize. Numerous empirical papers provide strong evidence for a minimum hours constraint.[7] Such a constraint has two consequences. First, it brings about the extensive margin: Some of the individuals, who would like to work less than L_{min}, will prefer to be unemployed rather than to work L_{min}. Secondly – and in contrast to a model with fixed costs of work or disutility of participation – it qualifies the response along the intensive margin: Those individuals who are constrained by L_{min} do not respond to (small) changes of the marginal tax rate, so that for each productivity level, the intensive margin is only present for part of the workers.

Secondly, in the literature on optimal income taxation with both margins, the tax perturbation method has so far only been used to derive a formula for the marginal tax rates. We show, how a different tax perturbation can be used to also derive a formula for the participation taxes. With this formula we can extend the results of Jacquet, Lehmann, and Van der Linden (2010): If the social marginal utility for the lowest income group is smaller than the marginal value of public funds, participation taxes are positive for all income levels, regardless of whether the conditions for positive marginal tax rates are satisfied or not. We can therefore show that only part of the results from the models with only the extensive margin carry over to the case of both margins: If the social marginal utility for the lowest income group is smaller than the marginal value of public funds, the NIT remains optimal. The reverse, while true in a model with only the extensive margin, does not hold: If the social marginal utility of the lowest income group is larger than the marginal value of public funds, the EITC is not necessarily optimal.

Thirdly, we derive a sufficient-statistics test for the second-best Pareto-efficiency of any given tax schedule in the presence of both labor supply responses and given quasi-linear preferences. This test can also be applied when the extensive margin is due to another reason than a minimum hours constraint and only requires knowledge of the following observable variables: extensive margin elasticities, intensive margin elasticities and the income distribution.

Finally, in carefully parameterized simulations we confirm the results of Jacquet, Lehmann, and Van der Linden (2010) that an optimal tax schedule is characterized by negative participation taxes (i.e. participation subsidies) for low incomes and positive marginal tax rates. We then show that the relationship of the degree of redistribution desired by society and the optimal participation subsidy is inversely u-shaped. In a next step we explore the robustness of the optimality

[7] Moffitt (1982) and Chen (1991) explicitly test for a minimum hours constraint and find it to be statistically significant. Sachiko and Isamu (2011) show that higher fixed costs on the side of the firm lead to higher minimum hours. Euwals and Van Soest (1999) show that there are less part time jobs than desired by workers in the Netherlands. Ilmakunnas and Pudney (1990) find similar results for Finland. Van Soest, Woittiez, and Kapteyn (1990) and Tummers and Woittiez (1991) suggest hours constraints to be a reason that many female unemployed cannot find jobs with a low number of hours per week.

of participation subsidies with regard to exogenous restrictions on the level of the welfare benefit: If this restriction causes the welfare benefit to be above its optimal level, the participation subsidy should be decreased. Thus, if a government is restricted not to set the welfare benefit below a given level, e.g. the subsistence level prescribed by constitution or deemed necessary by moralities, an EITC becomes less pronounced. This may explain why the EITC as an element of social policy is more important in the US than in continental Europe with its tradition on high welfare benefits.[8] In a last step, we investigate the question whether an EITC should rather be in place in countries with low or high exogenous government revenue requirements (e.g. for public goods or interest on public debt). Interestingly, our results show that higher revenue requirements strengthen the case for an EITC: While the welfare benefit declines, the participation subsidy remains fairly constant and therefore increases in relative terms. A consequence of the current public debt crisis in Europe might therefore be a greater reliance on EITC-type tax transfer systems. However, if the welfare benefit is fixed, an increase in revenue requirements leads to a sharp decline of the participation subsidy, so that the EITC becomes less pronounced. A greater reliance on EITC-type tax transfer systems should therefore only be observed in those countries, in which a substantial reduction of the welfare benefit is conceivable.

The remainder of this paper is organized as follows: In Section 2 we present our model of labor supply. We first consider the case without the minimum hours constraint in Section 2.1 and then show how it must be modified due to this constraint in Section 2.2. We then state the government's problem and formulate it as a mechanism design problem in Section 3. We derive its solution and develop the test for the Pareto-efficiency of any given tax schedule in Section 4. Section 5 provides the simulation results and Section 6 concludes.

2 The Model

Individuals' preferences over consumption C and hours of work L are characterized by

$$U(C, L; \alpha) = C - v(\alpha L), \tag{1}$$

with $v(0) = 0$, $v' \geq 0$, $v'' \geq 0$. We assume quasi-linear preferences only to simplify the exposition; incorporating income effects with a utility function $U = u(C) - v(\alpha L)$ with $u' > 0$, $u'' < 0$ is straightforward.[9]

Individuals differ in the parameter α, which measures preferences for leisure and is assumed to enter the utility function in this way to render the two dimensional screening problem tractable. Individuals also differ in their productivity w. The parameters w and α are distributed according to a joint density function $k(w, \alpha)$, which we represent by the marginal density $f(w)$ and the conditional density $g(\alpha|w)$:

$$k(w, \alpha) = f(w)\, g(\alpha|w).$$

The density functions f and g have support $[w_0, w_1]$ and $[\alpha_0, \alpha_2]$ respectively, with $w_0, \alpha_0 > 0$. The corresponding distribution functions are $F(w)$ and $G(\alpha|w)$.

[8] We provide a political economy interpretation of this result based on Coughlin (1986) and Lindbeck and Weibull (1987) in Section 5.

[9] We state how the results change if we allow for income effects in Section 4.

When choosing their labor supply L, individuals have to take into account the minimum hours constraint $L \geq L_{min}$. We could allow for L_{min} to depend on w, and for this constraint to apply to only a share of the individuals, so that some individuals are entirely free when deciding on the number of hours they work; however, in order to focus on the main mechanisms we simply assume that the minimum hours constraint applies to all individuals, and that L_{min} is the same for all productivity levels w.

Given an income tax schedule $T(Y)$, the optimization problem of an individual of type (w, α) is

$$\max_{C,L} U(C, L; \alpha) = C - v(\alpha L) \quad \text{s.t.} \quad C \leq wL - T(wL) \tag{2}$$

$$L = 0 \ \vee \ L \geq L_{min}. \tag{3}$$

2.1 Labor Supply without the Minimum Hours Constraint

Let $\hat{L}$ be the optimal labor supply if there were no minimum hours constraint. It is the solution to (2) without (3) and given by the first order condition[10]

$$\frac{\partial U}{\partial L} = (1 - T'(w\hat{L}))w - \alpha v'(\alpha \hat{L}) = 0. \tag{4}$$

Denote by $\hat{Y} = w\hat{L}$ gross income associated with $\hat{L}$. Replacing $w\hat{L}$ in (4) yields

$$(1 - T'(\hat{Y})) - v'\left(\hat{Y}/\tfrac{w}{\alpha}\right)/\tfrac{w}{\alpha} = 0, \tag{5}$$

and shows that $\hat{Y}$ depends on w and α only through the one-dimensional aggregate $\frac{w}{\alpha}$. We denote this aggregate by β. If there is no minimum hours constraint, all individuals – or types (w, α) – with identical $\frac{w}{\alpha} = \beta$ earn the same income $\hat{Y}(\beta)$. They can be found in α-w-space along a straight line through the origin with slope β, see Figure 3.[11]

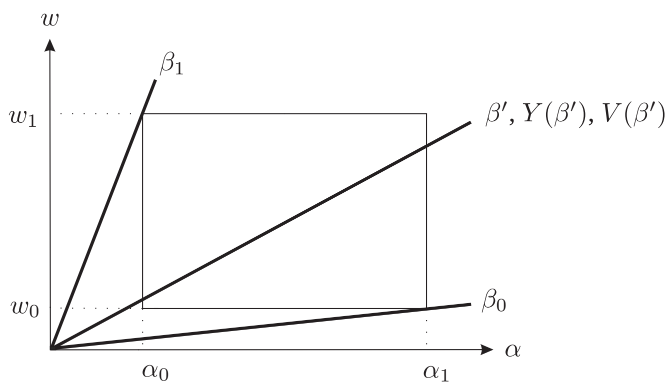


Figure 3: Identical income $Y(\beta)$ and identical maximum utility $V(\beta)$ for all combinations of α and w along the line with slope $\beta = \frac{w}{\alpha}$. β, $Y(\beta)$ and $V(\beta)$ increasing counter-clock-wise.

[10]The second order condition $SOC = -w^2T''(wL) - \alpha^2 v''(\alpha L) < 0$ is satisfied if the tax schedule T is not too concave, which we assume to be the case.

[11]Note that the smallest and largest value of β are $\beta_0 = w_0/\alpha_1$ and $\beta_1 = w_1/\alpha_0$ respectively.

They also receive the same utility, since

$$V(\beta) = Y(\beta) - T(Y(\beta)) - v\left(Y(\beta)/\beta\right). \tag{6}$$

While those individuals of type β with a higher productivity w have to work fewer hours to earn $Y(\beta)$, they suffer from a higher disutility of work α. Along a β-line, these two effects cancel out and utility is constant.

Without the minimum hours constraint, determining the optimal tax schedule in this setting just constitutes a one-dimensional screening problem, even though individuals originally differ among two characteristics.[12] With the minimum hours constraint this is no longer the case.

2.2 Labor Supply with the Minimum Hours Constraint

As can be seen from Figure 3, for a given productivity w, income is decreasing in α, and thus optimal labor supply $\hat{L}$ as well. Therefore individuals with a large enough α want to work less than L_{min}. With the minimum hours constraint this is not feasible: If they decide to work at all, they have to work longer hours than is optimal for them, i.e., L_{min} instead of $\hat{L}$. Denote by $\alpha^m(w)$ the threshold that separates those working L_{min} from those working more than L_{min}, see Figure (4). It is implicitly defined by the FOC (4) evaluated at L_{min}:

$$(1 - T'(wL_{min}))w - v'(\alpha^m(w)L_{min})\alpha^m(w) = 0. \tag{7}$$

Typically $\alpha^m(w)$ is increasing in w since for a given value of α, individuals with a higher productivity w work more. However, if the tax schedule is quite convex, $\alpha^m(w)$ could also be decreasing in w.

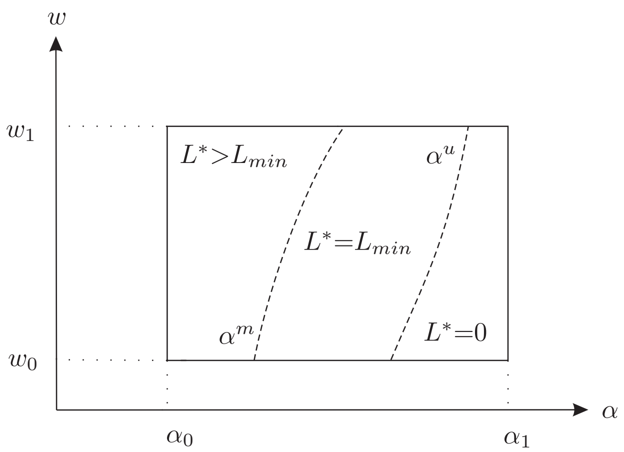


Figure 4: Partition of w-α-space by $\alpha^m(w)$ and $\alpha^u(w)$

In order to determine overall optimal labor supply L^*, we now turn to the individual's decision of whether to work at all. Denote the welfare benefit of an unemployed by $b = -T(0)$, which gives utility $U(b, 0, \alpha) = b$. Utility, when working, decreases in α, so that the individual prefers

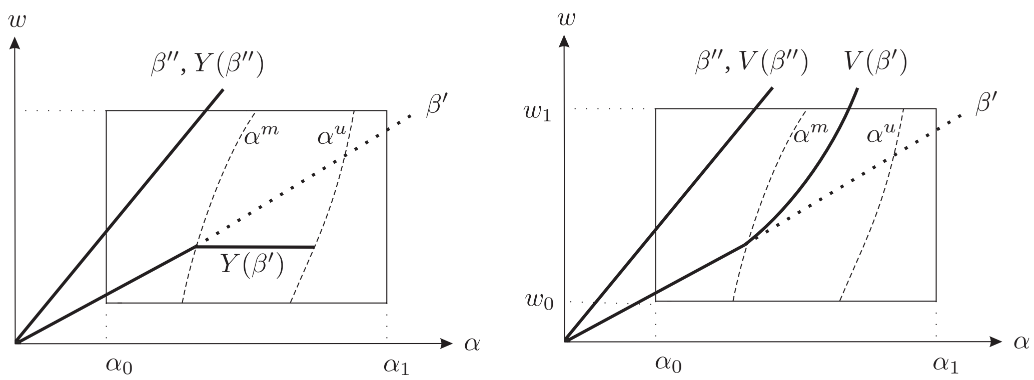
[12] Choné and Laroque (2011a) consider a similar model; however, they allow for a more general aggregation function than $\beta = w/\alpha$. Also, Brett and Weymark (2003) consider a similar 'type aggregator' in a model with endogenous education.

to be unemployed if α is large enough. Denote by $\alpha^u(w)$ the threshold for which the individual is indifferent, (see Figure 4 again). It is implicitly defined by

$$w \cdot \max[\hat{L}, L_{min}] - T(w \cdot \max[\hat{L}, L_{min}]) - v(\alpha^u(w) \cdot \max[\hat{L}, L_{min}]) = b. \tag{8}$$

As can easily be shown, $\alpha^u(w)$ is increasing in w.

Figure 4 shows functional forms of $\alpha^u(w)$ and $\alpha^m(w)$, so that for each productivity w all three groups, those working more than L_{min}, those working L_{min} and those not working, exist. This, however, need not always be the case. First, for some productivity levels the functions could be outside the interval $[\alpha_0, \alpha_1]$. This case is captured in the formulas for the optimal tax schedule we derive and does not have to be considered separately. Secondly, with positive participation taxes, the two functions could cross if L_{min} is small. In this case, there are some productivity levels for which no individuals work L_{min}. However, since we want to analyze the impact of the minimum hours constraint, we assume L_{min} to be large enough, so that for each w some individuals are affected by L_{min}.[13]



(a) $Y(\beta)$-curves

(b) $V(\beta)$-curves

Figure 5: Shape of β-, $Y(\beta)$- and $V(\beta)$-curve: identical for β'', different for β'

Not all income levels are affected by this constraint, as can be seen in Figure 5. While none of the individuals earning $Y(\beta'')$ is constrained by L_{min}, some of the individuals earning $Y(\beta')$ are. Because part of the β'-line is in the area where individuals would like to work fewer hours than L_{min}, but cannot, Y is increasing along the β-line between α^m and α^u. In fact, since labor supply for a given productivity is constant in this area, income equals $Y(\beta')$ along the horizontal line between α^m and α^u, see Figure 5(a).

Along this horizontal line, utility is decreasing, since all individuals work L_{min}, get the same income $Y(\beta')$, but have a higher and higher disutility of labor. In fact, utility is even decreasing along the β'-line between α^m and α^u; it is constant along this line only without the minimum hours constraint. Because individuals cannot optimally choose L, utility equals $V(\beta')$ along a curve that lies above the β'-line in the area between α^m and α^u.

The iso-income line does not coincide with the iso-indirect utility curve, so the type aggregator $\beta = \frac{w}{\alpha}$ does not apply in the area between α^m and α^u. Therefore, with the minimum hours constraint the optimal tax schedule is not simply the solution to a one-dimensional screening problem.

[13]The formulas we derive can easily be adapted to the case that α^m and α^u cross; however, for notational simplicity, we refrain from doing that.

3 The Government's Problem

The government's objective is to maximize the social welfare function

$$W = \int_{w_0}^{w_1} \int_{\alpha_0}^{\alpha_1} \Psi(V(w,\alpha)) dG(\alpha|w) dF(w), \tag{9}$$

where $V(w,\alpha)$ is the indirect utility function of an individual of type (w,α) and $\Psi(\cdot)$ is increasing and concave. $\Psi(\cdot)$ may either represent redistributive preferences of the government or a concave transformation of individual utilities that does not change preferences over leisure and consumption.

The government can only observe income Y, but neither labor supply L nor w or α. However, it knows the distribution functions $F(w)$ and $G(\alpha|w)$. The obvious strategy to solve this problem is to formulate it as a mechanism where the government determines the optimal income $Y(w,\alpha)$ and consumption $C(w,\alpha)$ for each type (w,α). The government then maximizes (9) subject to the budget constraint, the minimum hours constraint

$$Y(w,\alpha) \geq wL_{min} \;\vee\; Y(w,\alpha) = 0 \tag{MHC}$$

and the incentive compatibility constraint

$$C(w,\alpha) - v\left(\alpha \frac{Y(w,\alpha)}{w}\right) \geq C(w',\alpha') - v\left(\alpha \frac{Y(w',\alpha')}{w}\right) \quad \forall \;\; w', \alpha', w, \alpha. \tag{IC}$$

So far this two-dimensional screening problem is difficult to solve since there is no obvious ordering of the incentive constraints and because of the minimum hours constraint. We first show how the problem can be solved if there were no minimum hours constraint, and then use the result to show that in the above problem the incentive compatibility constraints are only locally binding, which renders the problem tractable.

3.1 The Problem without the Minimum Hours Constraint

As shown in Section 2, without the minimum hours constraint all individuals along a β-line earn the same income $Y(\beta)$ and receive the same utility $V(\beta)$. Because the type aggregator then applies to all individuals, the government's problem can be stated entirely in terms of β. The incentive constraint in this case reads as

$$C(\beta) - v\left(\frac{Y(\beta)}{\beta}\right) \geq C(\beta') - v\left(\frac{Y(\beta')}{\beta}\right) \quad \forall \;\; \beta', \beta. \tag{IC_β}$$

Since preferences satisfy the Spence-Mirrlees condition, it can easily be shown that (IC_β) can be replaced by an envelope condition

$$V'(\beta) = v'\left(\frac{Y(\beta)}{\beta}\right) \frac{Y(\beta)}{\beta^2} \tag{EC_β}$$

and a monotonicity constraint

$$Y'(\beta) \geq 0. \tag{MC}$$

For those individuals who work more than L_{min} this incentive constraint also holds in the full problem. We will now argue why all the other incentive constraints also bind only locally.

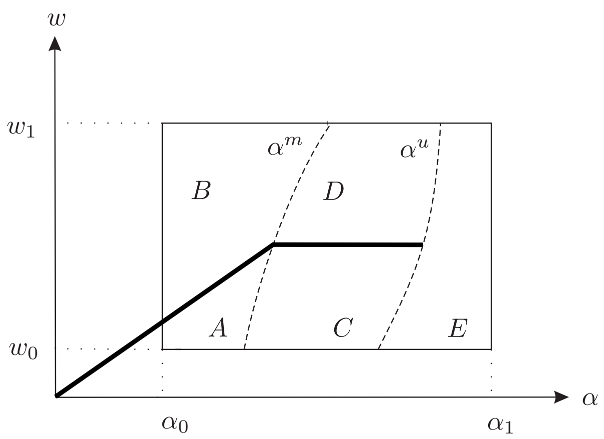


Figure 6: Incentive compatibility constraints

3.2 The Full Problem

Consider a representative iso-income-curve as shown in Figure 6. First note that along this curve, by definition, income is constant, so consumption must be equal, too. Secondly, all individuals to the right of the α^u-curve must get the same income-consumption-bundle because the government cannot distinguish between them. Now, three more steps are necessary to show that incentive constraints bind only locally with respect to this iso-income-curve; the argument then applies to all such curves:

1. For all individuals on the increasing part of the iso-income-curve, income only depends on $\frac{w}{\alpha}$, i.e. $Y = Y(\frac{w}{\alpha})$. Using the results of Section 3.1, we know that within the areas A and B incentive constraints bind only locally in the 'β-direction'. By construction of the iso-income-curves, all income-consumption-bundles in C and D can also be found in A and B. Hence individuals in A or B do not prefer any income-consumption-bundles in C and D.
2. By definition, the α^u-type is indifferent between working L_{min} and being unemployed. Thus on the α^u-curve incentive constraints bind locally. It immediately follows that all individuals to the right of this curve strictly prefer being unemployed to earning wL_{min}. Furthermore it follows that all individuals to the left of the α^u-curve strictly prefer their income-consumption-bundle to that of an unemployed individual since along the iso-income-curve income is constant while α is decreasing. This also implies that all individuals on the increasing part of the iso-income-curve prefer their income-consumption-bundle to that of the unemployed.
3. Because the α^m-type does not prefer any income-consumption bundle with higher income, all other individuals on the horizontal line also do not prefer such a bundle since they have a higher α and the same w. This reasoning also applies to individuals on the horizontal line to the right of the α^u-curve since they have the same w but higher α.

This shows that for all individuals to the left of the α^m-curve, incentive constraints bind only locally in the 'β-direction'. For all individuals on the α^u-curve, incentive constraints bind locally in the 'α-direction'. For all other individuals incentive constraints are not binding. Using these results, we can reformulate the full problem in a tractable way.

For each β, let $\underline{\alpha}(\beta)$ be the smallest and $\overline{\alpha}(\beta)$ be the largest value of α on the respective β-line. Also, let $g(\alpha|\beta)$ be the density of α given β, i.e. the density along the β-line, with support $[\underline{\alpha}(\beta), \overline{\alpha}(\beta)]$, and by $h(\beta)$ the density of β, with corresponding distribution functions $G(\alpha|\beta)$ and $H(\beta)$. Finally, denote by $\underline{\beta}$ the value of β associated with the lowest attainable income:[14]

$$Y(\underline{\beta}) = Y_{min} = w_0 L_{min}.$$

The government's objective then is

$$\begin{aligned} W \quad = \quad & \int_{\underline{\beta}}^{\beta_1} \int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)} \Psi(V(\beta))\, dG(\alpha|\beta)\, dH(\beta) \\ & + \int_{w_0}^{w_1} \left[\int_{\alpha^m(w)}^{\alpha^u(w)} \Psi(V(w,\alpha))\, dG(\alpha|w) + \int_{\alpha^u(w)}^{\alpha_1} \Psi(b)\, dG(\alpha|w) \right] dF(w), \end{aligned} \tag{10}$$

where the first term represents the individuals with $L^* > L_{min}$, the second those with $L^* = L_{min}$, and the third the unemployed.[15]

When maximizing (10) the government has to satisfy the balanced budget constraint:

$$\begin{aligned} \int_{w_0}^{w_1} \int_{\alpha^u(w)}^{\alpha_1} b\, dG(\alpha|w) dF(w) + R \quad = \quad & \int_{\underline{\beta}}^{\beta_1} (Y(\beta) - C(\beta))\, G(\alpha^m(\beta)|\beta) dH(\beta) \\ + \quad & \int_{w_0}^{w_1} \int_{\alpha^m(w)}^{\alpha^u(w)} (w L_{min} - C(w,\alpha))\, dG(\alpha|w) dF(w). \end{aligned} \tag{11}$$

Note that $Y(\beta) - C(\beta) = T(Y(\beta))$ and $wL_{min} - C(w,\alpha) = T(wL_{min})$, so the first term on the right hand side reflects the taxes collected from the individuals that are not constrained by the minimum hours requirement, the second the taxes from those that work L_{min}. On the left-hand side are the fiscal costs of welfare benefits and additional exogenous revenue requirements R.

Further, the government has to consider a 'no-discrimination-constraint'

$$C(\beta) = C\left(\frac{Y(\beta)}{L_{min}}, \alpha\right), \tag{NDC}$$

which states that individuals who earn the same income, must attain the same level of consumption.[16] The LHS reflects the consumption of an individual of type β that is not affected by the minimum hours constraint, and the RHS reflects the consumption of an individual that works L_{min} and earns $Y(\beta)$ since his wage is $w = \frac{Y(\beta)}{L_{min}}$.

Finally, the envelope condition has to hold:[17]

$$V'(\beta) = v'\left(\frac{Y(\beta)}{\beta}\right) \frac{Y(\beta)}{\beta^2}, \quad \forall\, \beta \in [\underline{\beta}, \beta_1]. \tag{EC'_β}$$

[14] Note that the β-line associated with $\underline{\beta}$ passes through the point $(w_0, \alpha^m(w_0))$.

[15] Note that in the first line the limit of integration is $min[\overline{\alpha}(\beta), \alpha^m(\beta)]$, but can be simplified to $\alpha^m(\beta)$, since $g(\alpha|\beta) = 0$ for $\overline{\alpha}(\beta) < \alpha < \alpha^m(\beta)$. This applies for all the following expressions where $\overline{\alpha}(\beta) < \alpha^m(\beta)$ is possible for some β.

[16] Scheuer (2011), considering optimal taxation in a framework with endogenous occupational choice, imposes a technically similar constraint when deriving properties of any Pareto-optimal tax schedule for the case that the government cannot treat entrepreneurs and employees differently, so that entrepreneurial profits and wage income are taxed according to the same tax schedule.

[17] Following common practice in optimal tax theory, we solve the problem without the monotonicity constraint and verify ex-post that it is fulfilled.

4 Properties of the Optimal Tax Schedule

4.1 Marginal Tax Rates

Proposition 1. *The solution to the government's problem in terms of marginal tax rates is*

$$\frac{T'(Y(\beta))}{1-T'(Y(\beta))} = \frac{\mathcal{A}(Y(\beta))}{\mathcal{B}(Y(\beta))}, \tag{12}$$

with

$$\begin{aligned} \mathcal{A}(Y(\beta)) \quad = \quad & \int_{\beta}^{\beta_1} \int_{\underline{\alpha}(\beta')}^{\alpha^m(\beta')} (\lambda - \Psi'(V(\beta')))\, dG(\alpha|\beta') dH(\beta') \\ & + \int_{\frac{Y(\beta)}{L_{min}}}^{w_1} \left[\int_{\alpha^m(w)}^{\alpha^u(w)} (\lambda - \Psi'(V(w,\alpha)))\, dG(\alpha|w) \right. \\ & \left. + \lambda g(\alpha^u(w)|w) \frac{\partial \alpha^u(w)}{\partial T(wL_{min})} (T(wL_{min}) + b) \right] dF(w) \end{aligned} \tag{13}$$

and

$$\mathcal{B}(Y(\beta)) = \lambda\beta \frac{\varepsilon_{Y,1-T'}}{\varepsilon_{Y,1-T'}+1} G(\alpha^m(\beta)|\beta) h(\beta).$$

The Lagrange multiplier λ, associated with the government's budget constraint (11), is equal to the average social marginal utility of income, i.e.

$$\lambda = \int_{w_0}^{w_1} \int_{\alpha_0}^{\alpha_1} \Psi'(V(w,\alpha)) dG(\alpha|w) dF(w). \tag{14}$$

Further, we have

$$\mathcal{A}(Y(\underline{\beta})) = 0. \tag{15}$$

Proof. See Appendix A.1 □

The term $\mathcal{B}(Y(\beta))$ captures the effect of marginal tax rates on labor supply along the intensive margin.[18] The higher the mass $G(\alpha^m(\beta)|\beta)h(\beta)$ of individuals whose marginal incentives are distorted,[19] the larger their income (reflected by β), and the stronger their response to an increase in marginal tax rates $\varepsilon_{Y,1-T'}/(\varepsilon_{Y,1-T'}+1)$, the lower marginal tax rates should be.

The first two lines of the term $\mathcal{A}(Y(\beta))$ represent the difference between the marginal value of public funds λ, and the social marginal utility Ψ', summed up over all individuals with income greater than Y. If this expression is greater than zero, welfare is raised if taxes for this group are increased. Such an increase can be achieved by higher marginal tax rates at Y. This is why T' should be the larger, the larger this expression is.

[18] Equation (12) can also be derived by the tax perturbation method; in its terminology $\mathcal{B}(Y(\beta))$ represents the elasticity effect (Saez 2001) or the substitution effect (Jacquet, Lehmann, and Van der Linden 2010).

[19] Recall that $h(\beta)$ denotes the density of β, and $G(\alpha^m(\beta)|\beta)$ the share of individuals to the left of the α^m-curve on this particular β-line, so that $G(\alpha^m(\beta)|\beta)h(\beta)$ represents the mass of individuals with income $Y(\beta)$ which are not restricted by L_{min}.

However, an increase in taxes for individuals with income above Y also increases their participation taxes. This gives rise to labor supply responses along the extensive margin, captured by the third line in (13). For a given productivity w, the mass of individuals responding is $g(\alpha^u(w)|w)\left(-\frac{\partial \alpha^u(w)}{\partial T(wL_{min})}\right)$, and an individual responding then receives the welfare benefit b instead of paying taxes $T(wL_{min})$; this decreases government revenues by $T(wL_{min})+b$. Again, this has to be summed up over all individuals with income greater than Y. If this participation effect is large, then marginal tax rates at Y should be small.[20]

We now turn to the condition for optimal marginal tax rates to be positive for all income levels Y. Denote the density of Y by $\tilde{h}(Y)$.[21] Since $\tilde{h}(Y(\beta_1)) = 0$, in our model the result of no distortion at the top does not hold; however, this is a purely technical issue, see Brett and Weymark (2003, p. 2565). Also, the result of no distortion at the bottom does not hold either, so we can have $T'(Y_{min}) \neq 0$.[22]

Proposition 2. *Let $\overline{\Psi}'(Y)$ be the average social marginal utility of income of individuals with income Y, and $\xi_{\tilde{h},b}$ the semi-elasticity of unemployment for income Y with respect to a marginal increase in b, i.e. the share of individuals with income Y that would leave the labor force if unemployment benefits were marginally increased.*

(i) For $Y \leq w_1 L_{min}$, marginal tax rates are positive if

$$\frac{\partial}{\partial Y}\left(\frac{\lambda - \overline{\Psi}'(Y)}{\xi_{\tilde{h},b}(Y)}\right) > 0. \tag{16}$$

(ii) For $Y > w_1 L_{min}$, marginal tax rates are positive if $\overline{\Psi}'(Y)$ is decreasing in income.

Proof. See Appendix A.2. □

This proposition states the conditions for the classical Mirrlees result of positive marginal tax rates for a situation, in which individuals are restricted by a minimum hours constraint.[23] For incomes greater than $w_1 L_{min}$, the extensive margin is absent, so the condition is as in the standard Mirrlees case: $\overline{\Psi}'$ has to be decreasing in income. If the government wants to redistribute from the top to the bottom, negative marginal tax rates cannot be optimal since they distort labor supply and redistribute in the 'wrong' direction.

[20] In terminology of the tax perturbation method, $\mathcal{A}(Y(\beta))$ represents the mechanical and the participation effect. If we allowed for income effects in the utility function (i.e. $U(C, L; \alpha) = u(C) - v(\alpha L)$ with $u(c)$ increasing and concave), $\mathcal{A}(Y(\beta))$ would have to be extended by

$$\lambda \int_{\beta}^{\beta_1} \int_{\underline{\alpha}(\beta')}^{\alpha^m(\beta')} \eta T'(Y(\beta')) dG(\alpha|\beta') dH(\beta'),$$

where $\eta = \dfrac{\partial Y}{\partial \tau}$, with τ being an additional lump-sum transfer. Individuals with income greater than $Y(\beta)$ have a loss of net income that can be interpreted as an additional lump sum tax τ. If leisure is a normal good, as it would be in that case, this will make them increase their labor supply. This "reinforce[s] the mechanical effect" (Saez 2001, p. 217) by increasing tax revenues of the government.

[21] Note that $\tilde{h}(Y(\beta)) = G(\alpha^m(\beta)|\beta)h(\beta)\frac{d\beta}{dY} + \left(G\left(\alpha^u(w_\beta)|w_\beta\right) - G\left(\alpha^u(w_\beta)|w_\beta\right)\right)\frac{f(w_\beta)}{L_{min}}$ with $w_\beta = \frac{Y(\beta)}{L_{min}}$.

[22] This is because in our model for $\beta \to \underline{\beta}$, both $\mathcal{A}(Y(\beta)) \to 0$ and $\mathcal{B}(Y(\beta)) \to 0$. The result of no distortion at the bottom, derived by Jacquet, Lehmann, and Van der Linden (2010), in their model holds because the substitution effect does not vanish at the bottom of the income distribution.

[23] Condition (16) is the same as in the model of Jacquet, Lehmann, and Van der Linden (2010), see their Proposition 2. This shows that their result can be extended to a situation with heterogeneity in labor supply conditional on w. In addition, note that condition (16) is necessary for marginal tax rates to be positive in the pure extensive model, see Jacquet, Lehmann, and Van der Linden (2010).

In the presence of extensive responses that condition alone does not guarantee that marginal tax rates are positive. They also depend on $\xi_{\tilde{h},b}$, the semi-elasticity of unemployment for income Y with respect to the marginal increase in b. As Jacquet, Lehmann, and Van der Linden (2010) point out, it is reasonable to assume that it is decreasing in income, i.e. the increase in the income-specific unemployment rate due to a 1$-increase in unemployment benefits is higher for a low-income than for a high-income group.

If $\xi_{\tilde{h},b}$ is decreasing in income, this seems to be an additional force for positive marginal tax rates: If higher incomes are less inclined to become unemployed, the distortion of higher participation taxes for higher incomes (due to $T' > 0$) is reduced. However, two cases have to be distinguished.

Case 1: $\overline{\Psi}' < \lambda$

For (high) income levels with $\overline{\Psi}' < \lambda$, condition (16) is fulfilled if both $\overline{\Psi}'$ and $\xi_{\tilde{h},b}$ are decreasing. For these income levels, marginal tax rates can only be negative if $\xi_{\tilde{h},b}$ is increasing, so that participation effects are increasing and particularly high for a certain income level, say $\widetilde{Y}$. In this case, even though negative marginal tax rates just below $\widetilde{Y}$ have an undesirable redistributive effect and lead to an upward distortion of labor supply along the intensive margin, they also reduce the participation tax for $\widetilde{Y}$ and – because of the high $\xi_{\tilde{h},b}$ at $\widetilde{Y}$ – induce a relatively large number of individuals to earn $\widetilde{Y}$ instead of being unemployed , which then increases government revenues.

Case 2: $\overline{\Psi}' > \lambda$

For (low) income levels with $\overline{\Psi}' > \lambda$, both $\overline{\Psi}'$ and $\xi_{\tilde{h},b}$ decreasing is not sufficient for optimal marginal tax rates to be positive. Consider the case that for an income interval $[Y_1, Y_2]$ participation taxes are negative, $\overline{\Psi}'$ decreases 'slowly' and $\xi_{\tilde{h},b}$ decreases 'rapidly', and $T(Y)$ is constant (or increasing). Could welfare be higher with a decreasing tax schedule instead, (i.e. higher taxes close to Y_1 and lower taxes close to Y_2)? With $\overline{\Psi}'$ almost constant, the direct effect on welfare due to this redistribution within the interval is negligible. Since $\xi_{\tilde{h},b}$ decreases rapidly, more individuals will become unemployed due to the tax increase close to Y_1 than will begin working due to the tax decrease close to Y_2.[24] This overall increase in unemployment actually increases government revenue, because with negative participation taxes the government saves this extra transfer (the participation subsidy) for every individual that becomes unemployed.[25]

4.2 Participation Taxes

Proposition 3. *Optimal participation taxes are given by*

$$T_{part}(Y(\beta))\,\lambda g(\alpha^u(w_\beta)|w_\beta)\left(-\frac{\partial\alpha^u(w_\beta)}{\partial T_{part}(Y(\beta))}\right)\frac{f(w_\beta)Y'(\beta)}{L_{min}}+\left(\overline{\Psi}'(Y(\beta))-\lambda\right)\tilde{h}(Y(\beta))$$
$$=\frac{\partial}{\partial\beta}\left[\lambda\beta\left(\frac{\varepsilon_{Y,1-T'}}{\varepsilon_{Y,1-T'}+1}\right)\frac{T'(Y(\beta))}{1-T'(Y(\beta))}G(\alpha^m(\beta)|\beta)h(\beta)\right] \qquad (17)$$

where $w_\beta = \frac{Y(\beta)}{L_{min}}$.

[24]Since these individuals are indifferent between working and not working, their change in utility is of second order only.

[25]This argument for the possibility of negative marginal tax rates does not apply for Case 1, because we cannot have $\overline{\Psi}'(Y) < \lambda$, $T_{part}(Y) < 0$ and $T'(Y) < 0$ at the same time, see the proof of Corollary 2.

The derivation of this formula can be found in Appendix A.3. Here, we instead derive it intuitively. Therefor, consider first equation (17) with the right hand side equal to zero. We then have the standard interpretation of a model with only an extensive margin: The sign of the optimal participation tax only depends on the social marginal utility of income compared to the marginal value of public funds:[26] For income levels with $\overline{\Psi}' < \lambda$, participation taxes are positive, for those with $\overline{\Psi}' > \lambda$, they are negative (Diamond 1980, Saez 2002, Choné and Laroque 2011b). This result can most easily be understood by considering an (infinitesimally) small perturbation of a tax schedule as shown in Figure 7, so that the tax at income Y is reduced by dT due to a small decrease of the marginal tax rate in the interval $[Y - dY, Y]$ and a small increase of the marginal tax rate in the interval $[Y, Y + dY]$.[27] Without intensive labor supply responses, this only has a mechanical effect (individuals with income Y pay lower taxes) and a participation effect (some of the unemployed start working as the participation tax is reduced). For an optimal tax schedule, these two effects on welfare which are captured by the LHS of (17) have to add up to zero and therefore the sign of the participation tax is equal to the sign of $\lambda - \overline{\Psi}'$.

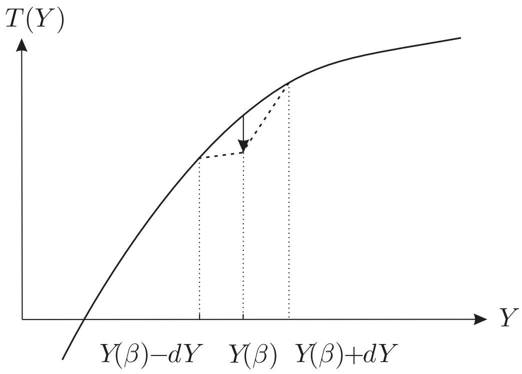


Figure 7: Tax perturbation

With labor supply responses along the intensive margin such a perturbation also has a substitution effect because of the change in marginal tax rates. Individuals with income in $[Y - dY, Y]$ will increase their labor supply, and those with income in $[Y, Y + dY]$ will decrease their labor supply. By the envelope theorem, these labor supply responses only change welfare by their impact on public funds. Whether government revenues increase or decrease due to the substitution effect, depends on the difference of these two effects, which in the limit, as $dT \to 0$, is captured by the derivative of the substitution effect, i.e. the RHS of (17). For a constant mass of individuals, a constant elasticity and a constant marginal tax rate, the substitution effect is increasing (so that the RHS of (17) is positive), which then makes negative participation taxes less likely compared to the pure extensive model. This shows that we can have $\overline{\Psi}' > \lambda$ and still $T_{part} > 0$. [28]

This raises the question if at least at the bottom of the income distribution, where there is no substitution effect, the result of the pure extensive model holds. Note that the right hand side of (17) can be decomposed into two terms:

[26]Note that $\lambda g(\alpha^u(w_\beta)|w_\beta)\left(-\frac{\partial \alpha^u(w_\beta)}{\partial T_{part}(Y(\beta))}\right)\frac{f(w_\beta)Y'(\beta)}{L_{min}} > 0$.

[27]Werning (2007) considers such a tax reform in a classical Mirrlees framework with intensive labor supply responses in order to test whether any given income tax schedule is Pareto-efficient. In the next section we extend his Pareto-efficiency test to models with intensive *and* extensive labor supply responses.

[28]The inclusion of income effects would make negative participation taxes less likely if leisure is a normal good. The reason is that an increase in the participation tax induces an increase in labor supply along the intensive margin due to income effects; this effect works against negative participation taxes.

$$\lambda\left[\frac{T'(Y(\beta))}{1-T'(Y(\beta))}\frac{\partial}{\partial\beta}\left(\beta\frac{\varepsilon_{Y,1-T'}}{\varepsilon_{Y,1-T'}+1}G(\alpha^m(\beta)|\beta)h(\beta)\right)\right.$$
$$\left.+\beta\frac{\varepsilon_{Y,1-T'}}{\varepsilon_{Y,1-T'}+1}G(\alpha^m(\beta)|\beta)h(\beta)\frac{\partial}{\partial\beta}\left(\frac{T'(Y(\beta))}{1-T'(Y(\beta))}\right)\right]. \tag{18}$$

Evaluating at $\beta \to \underline{\beta}$, the term in the second line vanishes because $G(\cdot) = 0$, while the derivative in the first line is unambiguously positive. In fact, this is just the situation depicted in Figure 7, when we let $Y(\beta)$ be equal to Y_{min}, i.e. the tax schedule starts at $Y(\beta)$ in Figure 7. Since there is no substitution effect to the left of Y_{min}, but there is one to the right of Y_{min}, the substitution effect is increasing. Although there is no substitution effect at the bottom of the income distribution, the result from the binary model that the sign of the participation tax only depends on $\overline{\Psi}'$ relative to λ does not carry over.

Corollary 1. *In a model with two margins, $\overline{\Psi}' > \lambda$ at the bottom of the income distribution is not sufficient for the participation tax $T_{part}(Y_{min})$ to be negative although there is no substitution effect at Y_{min}.*

However, one can show that the reverse holds:[29]

Corollary 2. *With $\overline{\Psi}' < \lambda$ at the bottom of the income distribution, $T_{part}(Y)$ is positive for all $Y \geq Y_{min}$.*

Proof. If $\overline{\Psi}' < \lambda$ at the bottom, $T_{part}(Y_{min})$ can only be negative if $\frac{T'(Y_{min})}{1-T'(Y_{min})}$ is negative. However, T_{part} has to be positive for some Y so that the government budget constraint is satisfied. This means that T' has to turn positive for some value of Y, say $\tilde{Y}$, where T_{part} is still negative. At $\tilde{Y}$, $T'(\tilde{Y}) = 0$ and $\frac{\partial}{\partial\beta}\left(\frac{T'(\tilde{Y})}{1-T'(\tilde{Y})}\right) > 0$, so the right hand side is unambiguously positive, a contradiction to $T_{part}(\tilde{Y})$ still being negative at that point. □

4.3 A Test for Pareto-Efficiency

So far we focused on characterizing that part of the Pareto-frontier that corresponds to concave social welfare functions. We now show that our analysis can be extended to test whether any given income tax schedule is second-best Pareto-efficient.[30] Rewriting the government's objective (9) as

$$W = \int_{w_0}^{w_1}\int_{\alpha_0}^{\alpha_1} V(w,\alpha)d\tilde{G}(\alpha|w)d\tilde{F}(w) \tag{19}$$

allows to derive a formula that characterizes the whole Pareto-frontier with $\tilde{g}(\alpha|w)$ and $\tilde{f}(w)$ being the Pareto-weights, and $\tilde{G}(\alpha|w)$ and $\tilde{F}(w)$ the cumulated Pareto-weights.[31] Replacing so-

[29] The result also applies to the model of Jacquet, Lehmann, and Van der Linden (2010): They also do not have a 'substitution effect' for the lowest income, because in their model $T'(Y_{min}) = 0$.

[30] Saez (2001) first proposed this method. Werning (2007) elaborates it for the classical Mirrlees model with intensive labor supply responses. Scheuer (2011) pursues this method for the case of differential tax treatment of labor income and profits; as Scheuer (2011), we focus on the integral form of the efficiency condition.

[31] It is well known that every Pareto-optimum can also be interpreted as a Utilitarian optimum with respective weights in the welfare function. Let the weight of an individual of type (w,α) be $\omega(w,\alpha)$. Then the Pareto-weight is $\omega(w,\alpha)g(\alpha|w)f(w) = \tilde{g}(\alpha|w)\tilde{f}(w)$.

cial marginal utility in (12) by the (averaged) Pareto-weights then leads to the following corollary that provides a condition for an income tax schedule to be Pareto-efficient.

Corollary 3. *A tax schedule $T(Y)$ is Pareto-efficient if and only if*

$$\frac{T'(Y(\beta))}{1-T'(Y(\beta))}\mathcal{B}(Y(\beta)) - \lambda\Bigg[\int_{\beta}^{\beta_1} G(\alpha^m(\beta')|\beta')dH(\beta') + \int_{\frac{Y(\beta)}{L_{min}}}^{w_1} (G(\alpha^m(w)|w) - G(\alpha^u(w)|w))dF(w) + \int_{\frac{Y(\beta)}{L_{min}}}^{w_1} g(\alpha^u(w)|w)\frac{\partial \alpha^u(w)}{\partial T(wL_{min})}(T(wL_{min})+b)dF(w)\Bigg] \tag{20}$$

is non-decreasing in β.

Rewriting formula (12) yields an equation with expression (20) on the LHS, and the negative of the cumulated Pareto-weights of all individuals with income above $Y(\beta)$ on the RHS. The RHS decreasing in β would imply a negative Pareto-weight at this point and we would have Pareto-inefficiency.

A problem with expression (20) is that neither the parameters α and w nor the distribution functions $f(w)$ and $g(\alpha|w)$ can be inferred from the income distribution, marginal tax rates and a given utility function as can be done with only one-dimensional heterogeneity (Saez 2001).

We now argue that one can rewrite (20) in reduced form making this information redundant:

Proposition 4. *Let $\hat{\varepsilon}_{Y,1-T'}$ be the empirically estimated elasticity of income along the intensive margin and $\hat{\xi}_{\tilde{h},b}$ the empirically estimated participation semi-elasticity at income level Y. Further, let $\tilde{H}(Y)$ be the cdf of the observed income distribution. Then, for quasi-linear preferences a tax schedule $T(Y)$ is Pareto-efficient if and only if*

$$\frac{T'(Y)}{1-T'(Y)}\hat{\varepsilon}_{Y,1-T'}\tilde{h}(Y)Y - (1-\tilde{H}(Y)) - \int_{Y}^{Y_{max}} \hat{\xi}_{\tilde{h},b}(T(Y)+b)d\tilde{H}(Y) \tag{21}$$

is non-decreasing in Y.

Proof. See Appendix A.4. □

As long as we can observe elasticities and the income distribution, we can test for the Pareto-efficiency of a tax schedule under the assumption of quasi-linear preferences. We thus extended the analysis of Werning (2007) for the case with intensive and extensive labor supply responses.[32]

Following Werning (2007) we now briefly discuss the idea of *simple* Pareto-improving reforms. Consider therefor again a tax reform as illustrated in Figure 7. This reform makes some individuals better off and no individual worse off. If, in addition, tax revenue does not decrease the reform unambiguously induces a Pareto-improvement. As in Werning (2007), this is rather likely

[32]Instead of testing for Pareto-efficiency, this analysis could also be used to check whether marginal reforms increase welfare for given welfare weights as proposed by Chetty (2009). Immervoll, Kleven, Kreiner, and Saez (2007) actually consider two kinds of marginal reforms for several European countries: increasing the welfare benefit and increasing in-work benefits.

if the income density, and therefore the substitution effect, is falling rapidly at $Y(\beta)$. In the presence of extensive labor supply responses the possibility of such a Pareto-improvement is further influenced by the participation elasticity and the sign of the participation tax; if $T_{part}(Y(\beta)) > 0$ a Pareto-improvement is more likely.[33]

4.4 Exogenous Welfare Benefit

So far we assumed that the government is only restricted by informational asymmetries. We now consider the case where the government cannot set the level of the welfare benefit b below an exogenous threshold $\bar{b}$. Such a threshold could exist for several reasons: On the one hand, it could be predetermined by constitution or set by the welfare court to cover a subsistence level that might be higher than the welfare maximizing one. On the other hand, there may simply be a tradition of high welfare benefits that is difficult to overcome by a government without being accused of lack of solidarity with the poorest poor.[34]

The following corollary summarizes:

Corollary 4. *If the government is restricted not to set the welfare benefit b below an exogenous threshold $\bar{b}$, then the formulas for the optimal marginal tax rates and the optimal participation taxes do not change. However, the marginal value of public funds is now described by*

$$\int_{w_0}^{w_1} \int_{\alpha_0}^{\alpha_1} \Psi'(V(w,\alpha)) dG(\alpha|w) dF(w) + \gamma = \lambda,$$

where γ is the Lagrange multiplier of the constraint $b \geq \bar{b}$.

Proof. See Appendix A.5 □

If the constraint is binding, then $\gamma > 0$, so that λ is then larger than the average social marginal utility of income. In the following section we numerically investigate the effect of such an increase in λ, i.e. how such a binding constraint influences the shape of the optimal tax system.

5 Simulations

In this section we numerically investigate our model. After parameterizing the model, we briefly document on the shape of the marginal tax rates as well as the optimality of a negative participation tax and compare these results to the literature. Then we investigate to what extent the optimality of a negative participation tax is robust to exogenous restrictions on the welfare benefit and to exogenous revenue requirements of the government.

[33] Note, however, that Pareto-inefficient income tax schedules exist for which such simple Pareto-improving reforms are not possible. To obtain Pareto-improvements in this case, more sophisticated tax reforms are necessary; nevertheless conditions (20) and (21) can identify the Pareto-inefficiency of such schedules.

[34] Boone and Bovenberg (2004, 2006) investigate a similar case where b is exogenous. They propose that social assistance and the tax system might be chosen by different governmental institutions. Further, they argue, that one can view such an analysis 'as exploring how the tax system can be employed to address the possibly suboptimal aspects of social assistance' (Boone and Bovenberg 2004, p. 2229).

5.1 Choice of Parameters

They key parameters of the model are the distribution of α, β and w and the functions U and Ψ. The latter is assumed to be

$$\Psi(V(\cdot)) = \frac{1}{1-\rho} V(\cdot)^{1-\rho}.$$

The higher ρ the higher the degree of inequality aversion. For our benchmark simulations we set $\rho = 1.5$.

For the utility function we assume $U(C, L; \alpha) = C - (\alpha L)^k$, leading to a constant elasticity of $\varepsilon_{Y,1-T'} = \frac{1}{k-1}$. In the following we choose $k = 4$, which implies an elasticity of 0.33. Due to the minimum hours restrictions, this parametric assumption actually leads to lower average elasticities for low income levels, whose exact size depends on the share of individuals affected by the constraint at the respective income level.

Since the work of Diamond (1998) and Saez (2001) it is known that the skill distribution plays a key role for the shape of the optimal tax schedule. We assume the parameter β to be distributed according to a lognormal distribution with parameters $(\mu, \sigma) = (2.757, 0.5611)$[35] in the interval $[0.06, 3500]$ and append a Pareto-tail with parameter $a = 2$ at the 98.5%-percentile. We choose the minimum possible value of the Pareto-distribution such that the resulting density is continuous.

The distribution function $H(\beta)$ is consistent with many distribution functions $G(\alpha|w)$ and $F(w)$ and many values of w_0 and w_1.[36] We specify these as follows: Along a β-line the mass of individuals is normally distributed with the mode at the center of the β-line; the variance of the normal distribution as well as w_0, w_1 and L_{min} are chosen so that the extensive margin is present up to the 75%-quantile of the income distribution, participation elasticities are about 0.25 on average, and average intensive elasticities are between 0.1 (very low incomes) and 0.33 (medium and higher incomes) consistent with empirical estimates[37] if the following tax schedule was in place: a constant marginal tax rate of 40% and a welfare benefit such that the government budget constraint is met.[38] The density in (w, α)-space is illustrated in Figure 8.

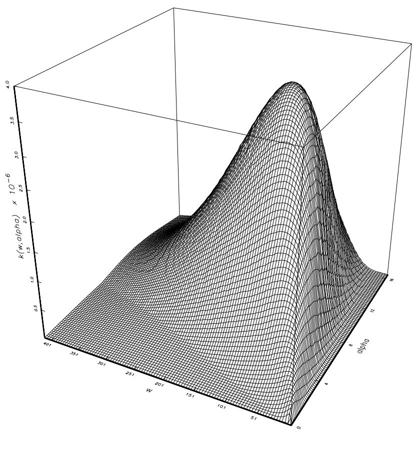


Figure 8: Density $k(w, \alpha)$

[35] See the online appendix of Mankiw, Weinzierl, and Yagan (2009).

[36] With β_0, β_1, w_0 and w_1 given, α_0 and α_1 can be inferred.

[37] See, e.g., Chetty, Guren, Manoli, and Weber (2011) and Blundell, Bozio, and Laroque (2011).

[38] Note that these restrictions do not pin down the values of the variance, w_0, w_1 and L_{min}, so the parameter combination we chose is one of many possible combinations. Results are, however, very similar for other combinations.

5.2 Results

Figure 9 illustrates optimal marginal tax rates as a function of β for the benchmark case. They first increase and then follow a U-shaped pattern. Further, participation subsidies are optimal ($T_{part}(Y_{min}) < 0$); we have $|T_{part}(Y_{min})/b| = 0.36$, implying that the transfer a worker with the lowest income receives is 36% higher than the welfare benefit b.

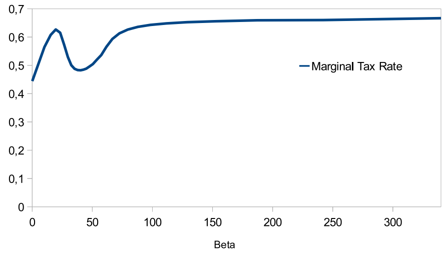
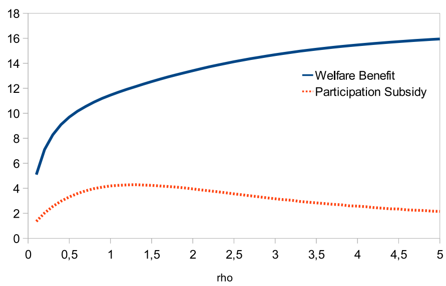
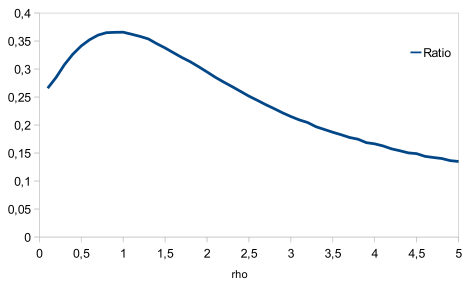


Figure 9: Marginal Tax Rates

We first test how the result of negative participation taxes depends on the assumed redistributive preferences of the government, see Figure 10(a). Not surprisingly the welfare benefit b increases in ρ. The participation subsidy $|T_{part}(Y_{min})|$ for the lowest income, and the ratio of this subsidy relative to the welfare benefit $|T_{part}(Y_{min})/b|$ first increases and then decreases. This implies that participation subsidies (relative to the welfare benefit) are more important for intermediate values of ρ, with a maximum at $\rho \approx 1.2$.



(a) b and $|T_{part}(Y_{min})|$



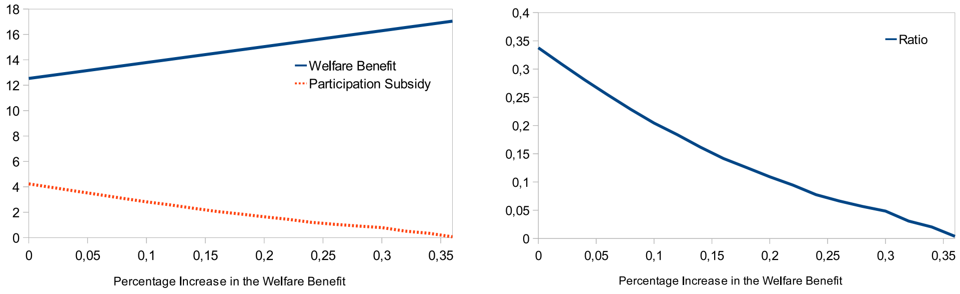
(b) $|T_{part}(Y_{min})/b|$

Figure 10: Optimality of EITC-type tax schedule as a function of ρ

Several unreported simulations support these results as long as the parameters are such that β is distributed log-normally with a Pareto-tail and the extensive and intensive elasticities are similar. We refrain from investigating to what extent the results are robust to other sizes of the elasticities, but refer to Jacquet, Lehmann, and Van der Linden (2010) and Saez (2002) who provide elaborations of that question. Instead we investigate to what extent the optimality of negative participation taxes for a given set of parameters is robust to exogenous restrictions of the government.

5.3 Optimality of Negative Participation Taxes with Restrictions on Welfare Benefits

As we discussed in Section 4.4, the government might be restricted not to set the welfare benefit b below a certain threshold $\bar{b}$. We now investigate whether negative participation taxes stay optimal if this restriction is binding. As shown in Figure 11 for the benchmark case of $\rho = 1.5$, an increase of the welfare benefit b above its optimal value leads to a decrease in the participation subsidy rendering the EITC less pronounced. We illustrate increases only until 35%, because then the Rawlsian, i.e. the maximum value of b is reached. Although participation subsidies do not vanish entirely, they become small relative to the welfare benefit, see Figure 11(b).[39]



(a) b and $|T_{part}(Y_{min})|$ (b) $|T_{part}(Y_{min})/b|$

Figure 11: Increasing b with $\rho = 1.5$

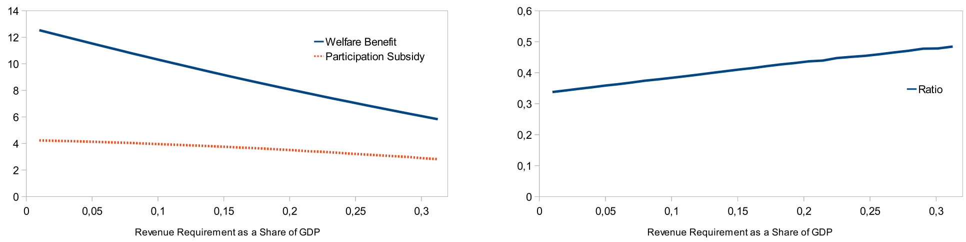
These results may describe one of the reasons why EITC-type tax schedules are a more important element of social policy in the US than in continental Europe. If there are constitutional constraints not to set the welfare benefit below the subsistence level, which may be higher in Europe than in the US, or there simply is a tradition of higher welfare benefits in Europe, which makes it difficult for a government to lower welfare benefits without being accused of having abandoned solidarity with those who are most in need, European governments may be confined to implement a 'third best' without (or with rather low) participation subsidies but high welfare benefits. In fact, if $\Psi(\cdot)$ is considered as concavity of individual utility, so that (9) represents a utilitarian objective, this 'third best' could be given a political economy interpretation: As Coughlin (1986) as well as Lindbeck and Weibull (1987) have shown, the political outcome in a probabilistic voting framework is described by the solution of the maximization of a utilitarian welfare function.

This is obviously not a closed theory since the restriction on the welfare benefit is exogenous to our model. However, it seems quite reasonable to assume that constitutional constraints or tradition regarding society's generosity with respect to the poorest poor may evolve rather slowly, so that the welfare benefit can not as easily be set as the tax schedule.

[39] The graphs are very similar for different values of ρ, however, all three curves get steeper as ρ increases.

5.4 Optimality of Negative Participation Taxes with Exogenous Government Spending

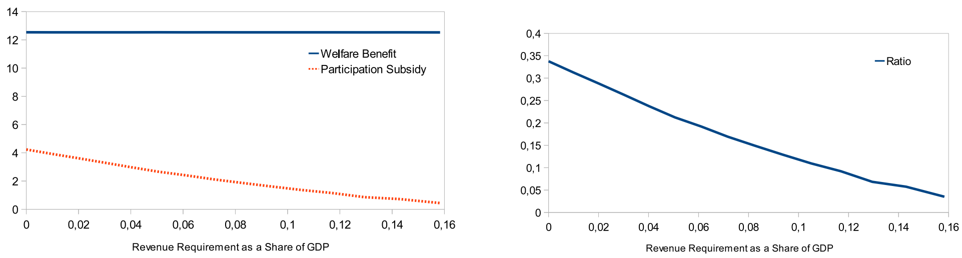
The current public debt crises in Europe will probably lead to tighter budgets in the future. We therefore analyze whether higher fiscal obligations will lead governments to rely on EITC-type tax-transfer system to a greater extent. Figure 12 shows comparative statics for b and $T_{part}(Y_{min})$ with respect to an increase in revenue requirements R. Interestingly, an increase in the revenue requirement leads to an almost parallel upwards shift of the entire tax schedule, so that the participation subsidy stays almost constant while the welfare benefit decreases strongly. With higher revenue requirements the EITC is more pronounced in relative terms.[40] One might therefore expect EITC-type tax transfer schedules to play a more important role in Europe as a consequence of the public debt crisis.



(a) b and $|T_{part}(Y_{min})|$ (b) $|T_{part}(Y_{min})/b|$

Figure 12: Increase in R

In contrast, and not surprisingly, if the welfare benefit is set exogenously, an increase in R clearly makes an EITC less likely. In Figure 13, the welfare benefit is set at its optimal level without additional revenue requirements. Keeping the welfare benefit at this level while increasing R shows that the participation tax decreases. The ratio $|(T_{part}(Y_{min})/b|$ then, of course, decreases as well.[41] A greater reliance on EITC-type tax transfer systems should therefore only be observed in those countries, in which a substantial reduction of the welfare benefit is conceivable.



(a) b and $|T_{part}(Y_{min})|$ (b) $|T_{part}(Y_{min})/b|$

Figure 13: Increase in R with fixed b

[40] This result is robust to other values of ρ: The welfare benefit always decreases, while the participation tax remains fairly constant or decreases only slowly.

[41] This result is robust to other exogenously set values of the welfare benefit (as well as to other values of ρ).

6 Conclusion

We characterized the solution to the optimal non-linear income tax problem when individuals cannot work less than a certain number of hours. We provided the conditions for optimal marginal tax rates to be positive and derived a formula for the optimal participation taxes. We showed that participation taxes need not be negative at the bottom of the income distribution if the social marginal utility of those with the lowest income is greater than the marginal value of public funds. This shows that only part of the results of the models with only an extensive margin carry over to the case of two margins.

In addition, we developed a test for the second-best Pareto-efficiency of any given income tax schedule in the presence of quasi-linear preferences and intensive and extensive labor supply responses. When stated in reduced form, the test only requires knowledge of labor supply elasticities and the income distribution. This test complements the work of Werning (2007) by incorporating extensive labor supply responses.

A numerical exploration of the model yielded several results. Reasonable parameterizations of the model confirmed the results of Jacquet, Lehmann, and Van der Linden (2010) that optimal non-linear tax schedules are characterized by a U-shaped pattern of marginal tax rates and participation subsidies for low incomes through a discontinuity in the tax function. Additionally, we contribute to the literature by elaborating the relationship between the degree of inequality aversion and the magnitude of participation subsidies for low incomes, where we find a hump-shaped relationship. For low inequality aversion, welfare benefits and participation subsidies should be low. Whereas the welfare benefit unambiguously rises in the degree of inequality aversion, the participation subsidy first increases and then decreases indicating that the optimality of an EITC-type tax transfer system is most pronounced for intermediate redistributive preferences.

Finally, we investigate whether the optimality of an EITC-type tax schedule is robust to exogenous restrictions on the welfare benefits, exogenous government spending and its interaction. First, if the government is restricted not to set the welfare benefit below a certain value and this restriction is binding, we find that the level of participation subsidies is strictly declining in this exogenous value; from a political economy perspective this result might explain why labor supply of individuals with low income is subsidized to a greater extend in the US than in continental Europe. Secondly, we find that an increase in exogenous government revenue requirements mainly leads to a decrease in the welfare benefit while the participation subsidy stays fairly constant; this result indicates that the importance of participation subsidies relative to welfare benefits should be higher in countries with higher debt. If, however, the constraint on the welfare benefit is already binding, then only small increases in exogenous government obligations lead to a strong decrease in the participation subsidies.

A Appendix

A.1 Proof of Proposition 1

The Lagrangian for the Problem as stated in Section 3.2 reads as:

$$
\begin{aligned}
\mathcal{L} \;=\; & \int_{\underline{\beta}}^{\beta_1} \int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)} \Psi(V(\beta))\, dG(\alpha|\beta)\, dH(\beta) \\
& + \int_{w_0}^{w_1} \int_{\alpha^m(w)}^{\alpha^u(w)} \Psi(V(w,\alpha))\, dG(\alpha|w) dF(w) + \int_{w_0}^{w_1} \int_{\alpha^u(w)}^{\alpha_1} \Psi(u(b))\, dG(\alpha|w) dF(w) \\
& + \lambda \Bigg[\int_{\underline{\beta}}^{\beta_1} \left(Y(\beta) - \left(V(\beta) + v\left(\frac{Y(\beta)}{\beta} \right) \right) \right) G(\alpha^m(\beta)|\beta) dH \\
& \qquad + \int_{w_0}^{w_1} \int_{\alpha^m(w)}^{\alpha^u(w)} \left(wL_{min} - \left(V(w,a) + v\left(\alpha L_{min}\right) \right) \right) dG(\alpha|w) dF(w) \\
& \qquad - \int_{w_0}^{w_1} \int_{\alpha^u(w)}^{\alpha_1} b\, dG(\alpha|w) dF(w) \Bigg] \\
& + \int_{\underline{\beta}}^{\beta_1} \int_{\alpha^u(Y(\beta)/L_{min})}^{\alpha^m(Y(\beta)/L_{min})} \eta(\beta,\alpha) \left[V(\beta) + v\left(\frac{Y(\beta)}{\beta} \right) - V(w,\alpha) - v\left(\alpha L_{min}\right) \right] d\alpha d\beta \\
& + \int_{\underline{\beta}}^{\beta_1} \left(\mu(\beta) V'(\beta) - \mu(\beta) v' \left[\frac{Y(\beta)}{\beta} \right] \frac{Y(\beta)}{\beta^2} \right) d\beta. \qquad (22)
\end{aligned}
$$

Partially integrating (EC'_β) and using $\mu(\underline{\beta}) = \mu(\beta_1) = 0$ yields $\int_{\underline{\beta}}^{\beta_1} \mu(\beta) V'(\beta) = -\int_{\underline{\beta}}^{\beta_1} \mu'(\beta) V(\beta)$, so that the last line of the Lagrangian can be replaced by

$$
+ \int_{\underline{\beta}}^{\beta_1} \left(-\mu'(\beta) V(\beta) - \mu(\beta) v' \left[\frac{Y(\beta)}{\beta} \right] \frac{Y(\beta)}{\beta^2} \right) d\beta. \qquad (23)
$$

The first order conditions are:

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial V(\beta)} \;=\; & \int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)} \left(\Psi'(V(\beta)) - \lambda \right) dG(\alpha|\beta)\, h(\beta) - \mu'(\beta) \qquad (24) \\
& + \int_{\alpha^u(Y(\beta)/L_{min})}^{\alpha^m(Y(\beta)/L_{min})} \eta(\beta,\alpha) d\alpha = 0
\end{aligned}
$$

$$
\left. \frac{\partial \mathcal{L}}{\partial V(w,\alpha)} \right|_{\alpha < \alpha^u} = \left(\Psi'(V(w,\alpha)) - \lambda \right) g(\alpha|w) f(w) - \eta(Y^{-1}((wL_{min}), \alpha) = 0 \qquad (25)
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial V(w,\alpha^u(w))} \;=\; & \left(\Psi'(V(w,\alpha)) - \lambda \right) g(\alpha|w) f(w) - \eta(Y^{-1}((wL_{min}), \alpha) \qquad (26) \\
& + \lambda g(\alpha^u(w)|w) \frac{\partial \alpha^u(w)}{\partial V} \underbrace{\left(b + wL_{min} - \left(V(w,a) + v\left(L_{min}\right) \right) \right)}_{T(wL_{min}) + b} = 0
\end{aligned}
$$

$$\frac{\partial \mathcal{L}}{\partial Y(\beta)} = \lambda\left(1 - v'\left(\frac{Y(\beta)}{\beta}\right)\frac{1}{\beta}\right)G(\alpha^m(\beta)|\beta)h(\beta) \tag{27}$$

$$-\mu(\beta)\frac{v'\left(\frac{Y(\beta)}{\beta}\right) + v''\left(\frac{Y(\beta)}{\beta}\right)\frac{Y(\beta)}{\beta}}{\beta^2}$$

$$+\int_{\alpha^u(Y(\beta)/L_{min})}^{\alpha^m(Y(\beta)/L_{min})} \eta(\beta,\alpha)\left[v'\left(\frac{Y(\beta)}{\beta}\right)\frac{1}{\beta} - \frac{\partial V}{\partial w}\frac{1}{L_{min}}\right]d\alpha = 0$$

$$\frac{\partial \mathcal{L}}{\partial b} = \int_{w_0}^{w_1}\int_{\alpha^u(w)}^{\alpha_1}\Psi'(b)dG(\alpha|w)dF(w) - \lambda\int_{w_0}^{w_1}\int_{\alpha^u(w)}^{\alpha_1}dG(\alpha|w)dF(w)$$

$$-\lambda\int_{w_0}^{w_1}\frac{\partial\alpha^u(w)}{\partial b}g(\alpha^u(w)|w)(T(wL_{min}) + b)dF(w). \tag{28}$$

Solving this set of equations for λ yields[42]

$$\lambda = \int_{w_0}^{w_1}\int_{\alpha_0}^{\alpha_1}\Psi'(V(w,\alpha))dG(\alpha|w)dF(w). \tag{29}$$

Integrating (24) yields

$$\mu(\beta) = \int_{\beta}^{\beta_1}\int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)}\left(\lambda - \Psi'(V(\beta))\right)dG(\alpha|\beta)dH(\beta) - \int_{\beta}^{\beta_1}\int_{\alpha^u(Y(\beta)/L_{min})}^{\alpha^m(Y(\beta)/L_{min})}\eta(\beta,\alpha)d\alpha. \tag{30}$$

Inserting (25) and (26) into (30) then results in

$$\mu(\beta) = \int_{\beta}^{\beta_1}\int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)}\left[\lambda - \Psi'(V(\beta))\right]dG(\alpha|\beta)dH(\beta) \tag{31}$$

$$+\int_{\beta}^{\beta_1}\left[\int_{\alpha^u(Y(\beta)/L_{min})}^{\alpha^m(Y(\beta)/L_{min})}\left[\lambda - \Psi'\left(V\left(\frac{Y(\beta)}{L_{min}},\alpha\right)\right)\right]d\alpha\right.$$

$$\left.-\lambda g(\alpha^u(w)|w)\frac{\partial\alpha^u(w)}{\partial V(w,\alpha^u(w))}\left(T(wL_{min}) + b\right)\right]dH(\beta).$$

Using $\frac{\partial V}{\partial w} = (1 - T'(Y(\beta)))L_{min}$ and $v'\left(\frac{Y(\beta)}{\beta}\right)\frac{1}{\beta} = 1 - T'(Y(\beta))$ to simplify (27) yields:

$$\lambda\left(1 - v'\left(\frac{Y(\beta)}{\beta}\right)\frac{1}{\beta}\right)G(\alpha^m(\beta)|\beta)h(\beta) - \mu(\beta)\frac{v'\left(\frac{Y(\beta)}{\beta}\right) + v''\left(\frac{Y(\beta)}{\beta}\right)\frac{Y(\beta)}{\beta}}{\beta^2} = 0. \tag{32}$$

[42] First integrate (25) over α^m to α^u and add (26) and (28), then integrate this whole expression over $\underline{\beta}$ to β_1, and add (24) integrated over $\underline{\beta}$ to β_1.

Inserting (30) into (32) and using $\varepsilon_{Y,1-T'} = \frac{\beta^2}{v''}\frac{1-T'}{Y(\beta)}$, (where $\frac{\partial Y}{\partial(1-T')} = \frac{\beta^2}{v''}$ can be derived by implicitly differentiating the FOC of the unconstrained individuals), we have

$$\begin{aligned}
&\frac{T'(Y(\beta))}{1-T'(Y(\beta))}\lambda\beta\left(\frac{\varepsilon_{Y,1-T'}}{\varepsilon_{Y,1-T'}+1}\right)G(\alpha^m(\beta)|\beta)h(\beta) \\
&\quad= \int_{\beta}^{\beta_1}\int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)}\left(\lambda - \Psi'(V(\beta'))\right)dG(\alpha|\beta')dH(\beta') \\
&\quad+ \int_{\frac{Y(\beta)}{L_{min}}}^{w_1}\left[\int_{\alpha^m(w)}^{\alpha^u(w)}\left(\lambda - \Psi'(V(w,\alpha))\right)dG(\alpha|w)\right. . \\
&\quad\left. +\lambda g(\alpha^u(w)|w)\frac{\partial\alpha^u(w)}{\partial T(wL_{min})}(T(wL_{min})+b)\right]dF(w).
\end{aligned} \tag{33}$$

Together with the first order condition with respect to b

$$\begin{aligned}
&\int_{w_0}^{w_1}\int_{\alpha^u(w)}^{\alpha_1}\Psi'(b)\,dG(\alpha|w)dF(w) \\
&\quad -\lambda\int_{w_0}^{w_1}\left[\int_{\alpha^u(w)}^{\alpha_1}dG(\alpha|w) + g(\alpha^u(w)|w)(T(wL_{min})+b)\right]dF(w) = 0
\end{aligned} \tag{34}$$

and the transversality condition $\mu(\underline{\beta}) = 0$, i.e.

$$\begin{aligned}
&\int_{\underline{\beta}}^{\beta_1}\int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)}\left(\lambda - \Psi'(V(\beta))\right)dG(\alpha|\beta)dH(\beta) + \int_{w_0}^{w_1}\left[\int_{\alpha^m(w)}^{\alpha^u(w)}\left(\lambda - \Psi'(V(w,\alpha))\right)dG(\alpha|w)\right. \\
&\quad\left. +\lambda g(\alpha^u(w)|w)\frac{\partial\alpha^u(w)}{\partial T(wL_{min})}(T(wL_{min})+b)\right]dF(w) = 0
\end{aligned} \tag{35}$$

this constitutes the solution.

A.2 Proof of Proposition 2

Negative marginal tax rates can only arise if $\mathcal{A}(Y(\beta)) < 0$. The transversality conditions imply $\mathcal{A}(Y(\beta_1)) = \mathcal{A}(Y(\underline{\beta})) = 0$, so that for $\mathcal{A}(Y(\beta)) < 0$ in an interval $]Y(\beta_2), Y(\beta_3)[$, we must have $\mathcal{A}'(Y(\beta_2)) \leq 0$, $\mathcal{A}'(Y(\beta_3)) \geq 0$ and $T(Y(\beta_2)) > T(Y(\beta_3))$.

In the following, we prove by contradiction that this cannot hold. We start with part (i) of Proposition 2, i.e. with the income interval where the intensive margin is present. Note that

$$\begin{aligned}
\mathcal{A}'(Y(\beta)) \quad = \quad & \int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)}[\Psi'(V(\beta)) - \lambda]dG(\alpha|\beta)h(\beta)\frac{\partial\beta}{\partial Y} \\
& +\left[\int_{\alpha^m(w_\beta)}^{\alpha^u(w_\beta)}[\Psi'(V(w,\alpha)) - \lambda]dG(\alpha|w_\beta)\right. \\
& \left. -\lambda g\left(\alpha^u(w_\beta)|w_\beta\right)\frac{\partial\alpha^u(w_\beta)}{\partial T(Y(\beta))}(T(Y(\beta))+b)\right]\frac{f(w_\beta)}{L_{min}},
\end{aligned} \tag{36}$$

with $w_\beta = \frac{Y(\beta)}{L_{min}}$.

Solving $\mathcal{A}'(Y(\beta_2)) \leq 0$ for $T(Y(\beta_2)) + b$ and $\mathcal{A}'(Y(\beta_3)) \geq 0$ for $T(Y(\beta_3)) + b$, and using $T(Y(\beta_2)) > T(Y(\beta_3))$ we have

$$\frac{\int_{\underline{\alpha}(\beta_2)}^{\alpha^m(\beta_2)} [\Psi'(V(\beta_2)) - \lambda] dG(\alpha|\beta_2) h(\beta_2) \frac{\partial \beta}{\partial Y} + \int_{\alpha^m(w_{\beta_2})}^{\alpha^u(w_{\beta_2})} [\Psi'(V(w_{\beta_2}, \alpha)) - \lambda] dG(\alpha|w_{\beta_2}) \frac{f(w_{\beta_2})}{L_{min}}}{-\frac{\partial \alpha^u(w_{\beta_2})}{\partial T} g(\alpha^u(w_{\beta_2})|w_{\beta_2}) \frac{f(w_{\beta_2})}{L_{min}}}$$
$$< \frac{\int_{\underline{\alpha}(\beta_3)}^{\alpha^m(\beta_3)} [\Psi'(V(\beta_3)) - \lambda] dG(\alpha|\beta_3) h(\beta_3) \frac{\partial \beta}{\partial Y} + \int_{\alpha^m(w_{\beta_3})}^{\alpha^u(w_{\beta_3})} [\Psi'(V(w_{\beta_3}, \alpha)) - \lambda] dG(\alpha|w_{\beta_3}) \frac{f(w_{\beta_3})}{L_{min}}}{-\frac{\partial \alpha^u(w_{\beta_3})}{\partial T} g(\alpha^u(w_{\beta_3})|w_{\beta_3}) \frac{f(w_{\beta_3})}{L_{min}}} \tag{37}$$

Using $\overline{\Psi}'(Y)$, the average marginal utility of income of all individuals earning income Y, which is given by

$$\overline{\Psi}'(Y(\beta)) = \frac{\int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)} \Psi'(V(\beta)) dG(\alpha|\beta) h(\beta) \frac{\partial \beta}{\partial Y} + \int_{\alpha^m(w_\beta)}^{\alpha^u(w_\beta)} \Psi'(V(w_\beta, \alpha)) dG(\alpha|w_\beta) \frac{f(w_\beta)}{L_{min}}}{\tilde{h}(Y(\beta))}$$

and the definition of $\tilde{h}(Y)$,[43] we can rewrite (37) as

$$\frac{[\overline{\Psi}'(Y(\beta_2)) - \lambda] \; \tilde{h}(Y(\beta_2))}{\left(-\frac{\partial \alpha^u(w_{\beta_2})}{\partial T} g(\alpha^u(w_{\beta_2})|w_{\beta_2}) \frac{f(w_{\beta_2})}{L_{min}}\right)} < \frac{[\overline{\Psi}'(Y(\beta_3)) - \lambda] \; \tilde{h}(Y(\beta_3))}{\left(-\frac{\partial \alpha^u(w_{\beta_3})}{\partial T} g(\alpha^u(w_{\beta_3})|w_{\beta_3}) \frac{f(w_{\beta_3})}{L_{min}}\right)} \tag{38}$$

Note that in the denominator we have the mass of individuals (that have earned $Y(\beta_i)$), who decide to become unemployed due to an increase of the welfare benefit. Therefore, the expression without the bracket $[\overline{\Psi}'(Y(\beta_i)) - \lambda]$ is just the inverse of the relative increase of the unemployed among the group earning Y due to an absolute increase in b (or T), i.e. the semi-elasticity $\xi_{\tilde{h},b}$ for unemployment with respect to b (or T).

So we have

$$\frac{[\overline{\Psi}'(Y_2) - \lambda]}{\xi_{\tilde{h},b}(Y_2)} < \frac{[\overline{\Psi}'(Y_3) - \lambda]}{\xi_{\tilde{h},b}(Y_3)}.$$

Since we assumed

$$\frac{\partial}{\partial Y} \left(\frac{\lambda - \overline{\Psi}'(Y)}{\xi_{\tilde{h},b}(Y)} \right) > 0, \tag{39}$$

this is a contradiction and we get part (i) of Proposition 2.

For part (ii) of Proposition 2, i.e. for $Y > w_1 L_{min}$, the extensive margin is absent. In that case $\mathcal{A}'(Y(\beta))$ simplifies to

$$\int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)} [\Psi'(V(\beta)) - \lambda] dG(\alpha|\beta) h(\beta) \frac{\partial \beta}{\partial Y}. \tag{40}$$

With the same reasoning as for part (i), part (ii) immediately follows.

[43] Recall that $\tilde{h}(Y(\beta)) = G(\alpha^m(\beta)|\beta) h(\beta) \frac{d\beta}{dY} + \left(G\left(\alpha^u(w_\beta)|w_\beta\right) - G\left(\alpha^u(w_\beta)|w_\beta\right)\right) \frac{f(w_\beta)}{L_{min}}$.

A.3 Proof of Proposition 3

Since equation (12) holds for all values of β, one can take the derivative with respect to β:

$$\int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)} (\Psi'(V(\beta)) - \lambda) dG(\alpha|\beta) h(\beta) \tag{41}$$

$$+ \left[\int_{\alpha^m(w)}^{\alpha^u(w)} (\Psi'(V(w,\alpha)) - \lambda) \; dG(\alpha|w) \right.$$

$$\left. -\lambda g(\alpha^u(w)|w) \frac{\partial \alpha^u(w)}{\partial T(wL_{min})} (T(wL_{min}) + b) \right] f\left(\frac{Y(\beta)}{L_{min}}\right) \frac{Y'(\beta)}{L_{min}}$$

$$-\frac{\partial}{\partial \beta}\left[\lambda\beta\left(\frac{\varepsilon_{Y,1-T'}}{\varepsilon_{Y,1-T'}+1}\right)\frac{T'(Y(\beta))}{1-T'(Y(\beta))} G(\alpha^m(\beta)|\beta) h(\beta)\right] = 0.$$

Rearranging terms yields Proposition 3.

This result can also be derived by considering the effects of the tax perturbation described in Section 4.2. The first one is the mechanical effect:

$$dW^M = \left[\int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)} (\Psi'(V(\beta)) - \lambda) \; dG(\alpha|\beta) h(\beta) \frac{\partial \beta}{\partial Y} \right. \tag{42}$$

$$\left. + \int_{\alpha^m(w_\beta)}^{\alpha^u(w_\beta)} (\Psi'(V(w_\beta,\alpha)) - \lambda) \; dG(\alpha|w_\beta) \frac{f(w_\beta)}{L_{min}} \right] dT' dY dY,$$

where $dT'dY = dT$ and $w_\beta = Y(\beta)/L_{min}$.[44] The mechanical effect has to be integrated over the interval $[Y(\beta') - dY, Y(\beta') + dY]$. This triangular area can be approximated by a rectangular area with length $2dY$ and height $\frac{1}{2}dT = \frac{1}{2}dT'dY$. The difference between the integration over the triangular area and the rectangular area will be of second order as $dY \to 0$. The term in brackets therefore has to be weighted by $dT'dYdY$. This reasoning also applies for the participation effect:

$$dW^P = \lambda g(\alpha^u(w_\beta)|w_\beta) \frac{\partial \alpha^u(w_\beta)}{\partial T(Y(\beta))} (T(Y(\beta)) + b) \frac{f(w_\beta)}{L_{min}} dT' dY dY. \tag{43}$$

The substitution effect consists of two parts: the first one to the left, the second, to the right of $Y(\beta)$.[45] Again taking limits ($dY \to 0$), the sum of these two effects can be replaced by the derivative:

$$dW^S = -\frac{\partial}{\partial \beta'}\left[\lambda \frac{\beta'}{\tilde{\varepsilon}_{Y,\beta'}} G(\alpha^m(\beta')|\beta') h(\beta') \left(-\frac{\tilde{\varepsilon}_{Y,1-T'}}{1-T'} T'\right)\right] \frac{\partial \beta}{\partial Y} dT' dY dY. \tag{44}$$

[44] The first term is expressed in terms of β, and should therefore be weighted by $d\beta$, which can be replaced by $d\beta = \frac{\partial \beta}{\partial Y} dY$; in the second line the same applies for dw.

[45] To derive the formula for the substitution effect, one first needs the mass of individuals for whom marginal incentives change. It is $G(\alpha^m(\beta)|\beta)h(\beta)d\beta$, where $h(\beta)d\beta$ is the mass of individuals in the interval $[\beta - d\beta, \beta]$ and $G(\alpha^m(\beta)|\beta)$ is the share of all individuals on the β-line with $\alpha < \alpha^m$, i.e., $L^* > L_{min}$. Denote by $\tilde{\varepsilon}_{Y,\beta} = \frac{dY}{d\beta}\frac{\beta}{Y}$ the elasticity of income with respect to β along the nonlinear tax schedule as defined by Jacquet, Lehmann, and Van der Linden (2010). $d\beta$ can then be replaced by $(\beta dY)/(\tilde{\varepsilon}_{Y,\beta} Y)$. Individuals affected by this change in marginal tax rates adjust their income according to $-\frac{\tilde{\varepsilon}_{Y,1-T'}}{1-T'} Y dT'$, where $\tilde{\varepsilon}_{Y,1-T'}$ is also defined along a nonlinear tax schedule. Multiplying this income change by T' then yields the effect on tax revenues. For the effect on welfare, this change in tax revenues has to be multiplied by λ.

Taking the sum of these three effects (and dividing by $\frac{\partial \beta}{\partial Y} dT' dY dY$) yields

$$\begin{aligned} d\widetilde{W} \quad = \quad & \int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)} (\Psi'(V(\beta)) - \lambda) dG(\alpha|\beta) h(\beta) \\ & + \left[\int_{\alpha^m(w)}^{\alpha^u(w)} (\Psi'(V(w,\alpha)) - \lambda) \ dG(\alpha|w) \right. \\ & \left. \qquad - \lambda g(\alpha^u(w')|w) \frac{\partial \alpha^u(w')}{\partial T(wL_{min})} (T(wL_{min}) + b) \right] f\left(\frac{Y(\beta)}{L_{min}}\right) \frac{Y'(\beta)}{L_{min}} \\ & - \frac{\partial}{\partial \beta} \left[\lambda \beta \frac{\tilde{\varepsilon}_{Y,1-T'}}{\tilde{\varepsilon}_{Y,\beta'}} \frac{T'}{1-T'} G(\alpha^m(\beta)|\beta) h(\beta) \right] = 0. \end{aligned} \tag{45}$$

Using the definitions of $\tilde{\varepsilon}_{Y,\beta'}$ and $\tilde{\varepsilon}_{Y,1-T'}$ and the FOC $(1-T')\beta = v'\left(\frac{Y(\beta)}{\beta}\right)$ of those individuals that are not constraint by the minimum hours requirement, we have

$$\frac{\tilde{\varepsilon}_{Y,\beta}}{\tilde{\varepsilon}_{Y,1-T'}} = \frac{v''\left(\frac{Y(\beta)}{\beta}\right)\frac{Y(\beta)}{\beta^2} + v'\left(\frac{Y(\beta)}{\beta}\right)\frac{1}{\beta}}{(1-T')} = 1 + \frac{v''\left(\frac{Y(\beta)}{\beta}\right)\frac{Y(\beta)}{\beta^2}}{(1-T')} = 1 + \frac{1}{\varepsilon_{Y,1-T'}},$$

which then yields Proposition 3.

A.4 Proof of Proposition 4

To derive the reduced form equation first note that

$$\int_{\beta}^{\beta_1} G(\alpha^m(\beta')|\beta') dH(\beta') + \int_{\frac{Y(\beta)}{L_{min}}}^{w_1} (G(\alpha^m(w)|w) - G(\alpha^u(w)|w)) dF(w) = 1 - \tilde{H}(Y(\beta)). \tag{46}$$

The observable elasticity $\hat{\varepsilon}_{Y,1-T'}$ for income Y and the elasticity $\varepsilon_{Y,1-T'}$ corresponding to the assumed preferences are linked by

$$\frac{\hat{\varepsilon}_{Y,1-T'}}{\varepsilon_{Y,1-T'}} = \frac{G(\alpha^m(\beta)|\beta)h(\beta)}{\tilde{h}(Y)\frac{\partial Y}{\partial \beta}}. \tag{47}$$

Finally, applying integration by substitution, the term

$$\int_{\frac{Y(\beta)}{L_{min}}}^{w_1} g(\alpha^u(w)|w) \frac{\partial \alpha^u(w)}{\partial T(wL_{min})} (T(wL_{min}) + b) f(w) dw$$

can be rewritten as

$$\int_{Y(\beta)}^{w_1 L_{min}} \underbrace{g\left(\alpha^u\left(\frac{Y}{L_{min}}\right) \middle| \frac{Y}{L_{min}}\right) \frac{\partial \alpha^u\left(\frac{Y}{L_{min}}\right)}{\partial T(Y)} \frac{1}{\tilde{h}(Y)} \frac{f\left(\frac{Y}{L_{min}}\right)}{L_{min}}}_{\xi_{\tilde{h},b}} (T(Y) + b) d\tilde{H}(Y). \tag{48}$$

Inserting (46), (47) and (48) into (20) and using the definition $\varepsilon_{Y,\beta} = \frac{\partial Y}{\partial \beta}\frac{\beta}{Y}$ then yields Proposition 4.

A.5 Proof of Corollary 4

If we have the additional constraint that $b \geq \bar{b}$, the Lagrangian is extended by $\gamma(b-\bar{b})$. This does not change the first order conditions with respect to $V(\beta), V(w,\alpha), Y(\beta)$. Also, the formula for the optimal marginal tax rates and for the participation taxes do not change. There is however a change in the value of λ,

$$\int_{w_0}^{w_1}\int_{\alpha_0}^{\alpha_1} \Psi'(V(w,\alpha))dG(\alpha|w)dF(w) + \gamma = \lambda, \tag{49}$$

and in the first order condition with respect to b:

$$\int_{w_0}^{w_1}\int_{\alpha^u(w)}^{\alpha_1} \Psi'(u(b))\, dG(\alpha|w)dF(w) + \gamma - \lambda \int_{w_0}^{w_1}\int_{\alpha^u(w)}^{\alpha_1} dGdF$$
$$-\lambda \int_{w_0}^{w_1} g(\alpha^u(w)|w)(T(wL_{min}) + b)dF(w) = 0. \tag{50}$$

This can be solved for γ and inserted into (4):

$$\int_{\underline{\beta}}^{\beta_1}\int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)} \Psi'(V(\beta))\, dG(\alpha|\beta)\, dH(\beta) + \int_{w_0}^{w_1}\int_{\alpha^m(w)}^{\alpha^u(w)} \Psi(V(w,\alpha))\, dG(\alpha|w) =$$
$$\lambda\left[1 - \int_{w_0}^{w_1}\int_{\alpha^u(w)}^{\alpha_1} dGdF - \int_{w_0}^{w_1} g(\alpha^u(w)|w)(T(wL_{min}) + b)dF(w)\right]. \tag{51}$$

Hence λ now reads as

$$\lambda = \frac{\int_{\underline{\beta}}^{\beta_1}\int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)} \Psi'(V(\beta))\, dG(\alpha|\beta)\, dH(\beta) + \int_{w_0}^{w_1}\int_{\alpha^m(w)}^{\alpha^u(w)} \Psi(V(w,\alpha))\, dG(\alpha|w)}{1 - \int_{w_0}^{w_1}\int_{\alpha^u(w)}^{\alpha_1} dGdF - \int_{w_0}^{w_1} g(\alpha^u(w)|w)(T(wL_{min}) + b)dF(w)}. \tag{52}$$

References

BLUNDELL, R., A. BOZIO, AND G. LAROQUE (2011): "Labour Supply and the Extensive Margin," *American Economic Review Papers and Proceedings (forthcoming)*.

BOONE, J., AND L. BOVENBERG (2004): "The optimal taxation of unskilled labor with job search and social assistance," *Journal of Public Economics*, 88(11), 2227–2258.

——— (2006): "Optimal welfare and in-work benefits with search unemployment and observable abilities," *Journal of Economic Theory*, 126(1), 165–193.

BRETT, C., AND J. A. WEYMARK (2003): "Financing education using optimal redistributive taxation," *Journal of Public Economics*, 87(11), 2549–2569.

CHEN, Y. (1991): "Youth labour supply and the minimum hours constraint-the case of single males," *Applied Economics*, 23(1), 229–235.

CHETTY, R. (2009): "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1(1), 451–488.

CHETTY, R., A. GUREN, D. MANOLI, AND A. WEBER (2011): "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins," *American Economic Review Papers and Proceedings*, 101, 471–475.

CHONÉ, P., AND G. LAROQUE (2011a): "Negative marginal tax rates and heterogeneity," *American Economic Review*, 100(5), 2532–47.

CHONÉ, P., AND G. LAROQUE (2011b): "Optimal Taxation in the Extensive Model," *Journal of Economic Theory*, 146(2), 425–453.

COGAN, J. F. (1981): "Fixed Costs and Labor Supply," *Econometrica*, 49(4), 945–963.

COUGHLIN, P. (1986): "Elections and income redistribution," *Public Choice*, 50(1), 27–91.

DIAMOND, P. (1980): "Income Taxation with Fixed Hours of Work," *Journal of Public Economics*, 13(1), 101–110.

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

EISSA, N., AND J. B. LIEBMAN (1996): "Labor Supply Response to the Earned Income Tax Credit," *The Quarterly Journal of Economics*, 111(2), 605–637.

EUWALS, R., AND A. VAN SOEST (1999): "Desired and actual labour supply of unmarried men and women in the Netherlands," *Labour Economics*, 6(1), 95–118.

HAUSMAN, J. (1985): "Taxes and Labor Supply," in *Handbook of Public Economics*, ed. by A. Auerbach, and M. Feldstein. Elsevier, North-Holland.

HECKMAN, J. J. (1993): "What Has Been Learned about Labor Supply in the Past Twenty Years?," *American Economic Review*, 83(2), 116–121.

HELLWIG, M. F. (2007): "A contribution to the theory of optimal utilitarian income taxation," *Journal of Public Economics*, 91(7-8), 1449–1477.

ILMAKUNNAS, S., AND S. PUDNEY (1990): "A model of female labour supply in the presence of hours restrictions," *Journal of Public Economics*, 41(2), 183–210.

IMMERVOLL, H., H. KLEVEN, C. KREINER, AND E. SAEZ (2007): "Welfare Reform in European Countries: A Microsimulation Analysis," *The Economic Journal*, 117(516), 1–44.

JACQUET, L., E. LEHMANN, AND B. VAN DER LINDEN (2010): "Optimal Redistributive Taxation with both Extensive and Intensive Reponses," *IZA Discussion Paper No. 4837*.

LAROQUE, G. (2005): "Income Maintenance and Labor Force Participation," *Econometrica*, 73(2), 341–376.

LINDBECK, A., AND J. WEIBULL (1987): "Balanced-budget redistribution as the outcome of political competition," *Public Choice*, 52(3), 273–297.

MANKIW, N., M. WEINZIERL, AND D. YAGAN (2009): "Optimal Taxation in Theory and Practice," *The Journal of Economic Perspectives*, 23(4), 147–174.

MEGHIR, C., AND D. PHILLIPS (2008): "Labour Supply and Taxes," *IFS Working Paper 08/04*.

MEYER, B. D., AND D. T. ROSENBAUM (2001): "Welfare, The Earned Income Tax Credit, and The Labor Supply of Single Mothers," *Quarterly Journal of Economics*, 116(3), 1063–1114.

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38(2), 175–208.

MOFFITT, R. (1982): "The Tobit Model, Hours of Work and Institutional Constraints," *The Review of Economics and Statistics*, 64(3), 510–515.

SACHIKO, K., AND Y. ISAMU (2011): "Firm's demand for work hours: Evidence from multi-country and matched firm-worker data," *Discussion paper, RIETI*.

SAEZ, E. (2000): "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses," *NBER Working Paper*, 7708.

——— (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *The Review of Economic Studies*, 68(1), 205–229.

——— (2002): "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses," *Quarterly Journal of Economics*, 117(3), 1039–1073.

SCHEUER, F. (2011): "Entrepreneurial Taxation and Occupational Choice," *Working Paper, Stanford University*.

SEADE, J. (1982): "On the Sign of the Optimum Marginal Income Tax," *The Review of Economic Studies*, 49(4), 637–643.

STIGLITZ, J. E. (1982): "Self-Selection and Pareto Efficient Taxation," *Journal of Public Economics*, 17(2), 213–240.

TUMMERS, M., AND I. WOITTIEZ (1991): "A simultaneous wage and labor supply model with hours restrictions," *Journal of Human Resources*, 26(3), 393–423.

TUOMALA, M. (1990): *Optimal Income Tax and Redistribution*. Clarendon Press, Oxford.

VAN SOEST, A., I. WOITTIEZ, AND A. KAPTEYN (1990): "Labor supply, income taxes, and hours restrictions in the Netherlands," *Journal of Human Resources*, 25(3), 517–558.

WERNING, I. (2000): "An Elementary Proof of Positive Optimal Marginal Tax Rates," *Mimeo, MIT*.

——— (2007): "Pareto Efficient Income Taxation," *Mimeo, MIT*.