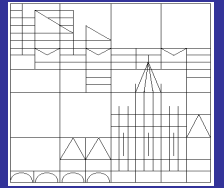




University of Konstanz
Department of Economics



Optimal Participation Taxes and Efficient Transfer Phase-Out

Normann Lorenz and Dominik Sachs

Working Paper Series
2012-37

<http://www.wiwi.uni-konstanz.de/forschung/>

Optimal Participation Taxes and Efficient Transfer Phase-Out

Normann Lorenz* and Dominik Sachs†

December 17, 2012

Abstract

We analyze the optimal design of income transfer programs with a special focus on participation taxes and the marginal tax rates in the phase-out region. The analytical framework incorporates labor supply responses along the intensive and extensive margin, where the latter is due to a minimum hours constraint. All results are expressed in reduced form, i.e. in terms of intensive and extensive labor supply elasticities. We derive a formula for the optimal participation taxes and provide a condition under which negative participation taxes are never part of the optimal tax schedule. Concerning the marginal tax rates in the phase-out region, we develop a test for a tax-transfer system to be beyond the top of the Laffer curve and thus to be (second-best) Pareto inefficient. In such a case there would be room for tax cuts (or increases in transfers) which are self-financing and therefore constitute a Pareto improvement. Applying this test to Germany, our analysis suggests that the structure of marginal tax rates in the transfer phase-out region is (second-best) Pareto inefficient.

JEL-classification: H 21, H 23.

Keywords: Optimal taxation, participation taxes, extensive margin, Laffer curve, multidimensional screening.

*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, Fabian Krüger, Etienne Lehmann, Nick Netzer, Florian Scheuer, Dirk Schindler, Esther Schuch, Jean-François Tremblay and Iván Werning as well as seminar participants in Gerzensee, the 4th RGS Doctoral Conference in Dortmund, the IIPF in Dresden and the ESEM in Málaga for helpful comments and suggestions. Dominik 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. One issue in this debate is whether it is the unemployed or individuals with low income who should receive the largest benefits. Under a Negative Income Tax (NIT), transfers are highest for the unemployed. Individuals with positive income receive lower transfers and thus pay a participation tax when entering the labor market. This contrasts to an Earned Income Tax Credit (EITC): Here individuals with low income receive the highest transfers. Because these transfers exceed those for the unemployed, low income individuals receive a participation subsidy (a negative participation tax) for entering the labor market.

A second issue in this debate concerns the marginal tax rates for those income levels where transfers are phased out. In most real world tax-transfer systems – regardless of whether NIT or EITC – these phase-out rates are very high. On the one hand, one may argue that this is unavoidable if society wants to grant large transfers. On the other hand, high marginal tax rates heavily distort labor supply. As these phase-out rates are close to 100% in many countries, one may suspect that they leave room for Laffer reforms, i.e. tax cuts that are self-financing because of strong labor supply effects.

Using methods of optimal nonlinear taxation, we address these two issues and ask (i) whether a tax-transfer system should levy participation taxes on or provide participation subsidies for individuals with low income and (ii) under which conditions a tax-transfer system is beyond the top of the Laffer curve, so that there is room for tax cuts which increase tax revenue. We derive the following main results: (i) We generalize a well-known theoretical result from the pure extensive margin model going back to Diamond (1980) to a framework with intensive *and* extensive labor supply responses: participation subsidies are never part of the optimal tax schedule if the social marginal utility of the lowest income workers is smaller than the marginal value of public funds.¹ (ii) We develop a test – based only on intensive and extensive labor supply elasticities and the income distribution – that can uncover whether a nonlinear tax schedule is beyond the top of the Laffer curve. Applying this test to Germany, we find that the marginal tax rates in the phase-out region may or may not be inefficiently high (depending on labor supply elasticities), but that they certainly exhibit an inefficient structure so that there is room for Pareto improving reforms.

As a formal starting point, we solve the optimal nonlinear income tax problem in a model with both intensive *and* extensive labor supply responses. As pointed out by numerous empirical studies, extensive responses are large, in particular for individuals with low income.² Addressing them when analyzing the optimal design of income transfer programs is therefore crucial.

One reason for an extensive margin to exist is a minimum hours constraint, as first proposed by Moffitt (1982). Such a constraint can be due to several causes: For example, some tasks require the worker to be present for a certain amount of time, or there may be fixed costs on the side of the firm (e.g., for training or for providing equipment) on which the firm wants to economize.³

¹Throughout the paper we will use the term tax schedule to describe the effective schedule of the entire tax-transfer system incorporating income taxes and all benefit programs.

²See Heckman (1993), Eissa and Liebman (1996), Meyer and Rosenbaum (2001) and Meghir and Phillips (2010).

³Numerous empirical papers provide evidence for a minimum hours constraint. 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 fewer part time jobs than desired by workers in the Netherlands. Ilmakunnas and Pudney (1990) find similar results for Finland. van Soest et al. (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.

We incorporate such a minimum hours constraint in a model without income effects where individuals differ in two dimensions, productivity and preferences for leisure. To keep this two-dimensional screening problem tractable, we focus on a special kind of separable preferences which allows to apply a type-aggregator.⁴

Our first contribution is of methodological nature in that we solve this two-dimensional screening problem and show how to express the optimality conditions in reduced form. The reduced form solution for the marginal tax rates shows a tight connection to the papers of Saez (2002) and Jacquet et al. (2012). Saez (2002) considers a model, where each individual can choose among two different occupations and unemployment. Jacquet et al. (2012) analyze a Mirrlees model, in which the extensive margin arises due to disutility of participation. Concerning optimal marginal tax rates, we show that the findings of these papers carry over to a setting where the extensive margin is due to a minimum hours constraint.

Based on our reduced form solution, we contribute to the EITC versus NIT debate by deriving a formula for the optimal participation taxes in the presence of intensive and extensive labor supply responses. This formula allows to state the condition that participation subsidies are never part of the optimal tax-transfer system if the social marginal utility of the lowest income workers is smaller than the marginal value of public funds. This is a generalization of the result from the pure extensive model, see Diamond (1980), Saez (2002) and Choné and Laroque (2011b). Importantly, this result does not depend on our specific setting with a minimum hours constraint, but holds in general, i.e. also for other frameworks with intensive and extensive labor supply responses.

Concerning the issue of high marginal tax rates in the phase-out region, we ask under what conditions a given nonlinear tax schedule is beyond the top of the Laffer curve. Whereas the concept of the Laffer curve is well understood for a linear tax, no one explicitly derived the conditions for a nonlinear income tax schedule to be efficient prior to Laroque (2005) (extensive margin) and Werning (2007) (intensive margin). For a setting with both intensive and extensive labor supply responses, we first propose a simple test whether the marginal tax rate at a certain income level is above its Laffer value, i.e. whether a decrease of the marginal tax rate at that income level would increase tax revenue. We then show that a tax schedule may be inefficient for more subtle reasons: even if each marginal tax rate itself is below its Laffer value, the structure of marginal tax rates may be inefficient. We therefore develop a stronger version of the test that can also identify such inefficient structures of marginal tax rates. We express both versions in reduced form, so that they only require knowledge of the income distribution and elasticities. Thus, no assumptions concerning the underlying reason for extensive labor supply responses are necessary when applying the test to a certain tax-transfer system.

Finally, we apply this test to the tax-transfer system in Germany (for singles). Whether marginal tax rates in the phase-out region are beyond their Laffer values crucially depends on the values of the intensive elasticities. However, with the stronger version of the test, we identify an inefficiency irrespective of the values of the elasticities. This inefficiency is caused by the structure of the marginal tax rates which heavily decrease at the income level where the transfer is (just) phased out.⁵ Our analysis suggests that a reform that decreases the marginal tax rate below this threshold income level, and increases it above, constitutes a Pareto improvement: As the absolute level of taxes does not increase for any income level, no individual is made worse off, but tax revenue increases due to the induced labor supply responses along the intensive and extensive margin.

⁴The concept of a type-aggregator has also been applied by Brett and Weymark (2003) and Choné and Laroque (2011a).

⁵The inefficiency does not hinge on the discontinuity in the marginal tax rates but exists as well if marginal tax rates decrease smoothly.

Besides the above mentioned papers, this paper is also related to the following studies. Kleven and Kreiner (2006) find that the marginal cost of public funds is higher if in addition to an intensive margin also extensive labor supply responses are taken into account. Boone and Bovenberg (2004) also consider the optimal tax problem in the presence of both margins: individuals have to search for a job and can either be unemployed voluntarily or involuntarily. They elaborate how the government should optimally balance distortions on search incentives with those on work effort incentives.

Kanbur et al. (1994) consider the optimal design of tax-transfer systems for the case that the government's goal is the alleviation of poverty. They show that if it is optimal that everybody works, the marginal tax rate for the lowest income is negative. Pirttilä and Tuomala (2004) extended this result for the case that the government can also levy linear commodity taxes.

Our Pareto efficiency test is related to the test derived by Scheuer (2012), who considers differential taxation of entrepreneurs and workers. In his model the extensive margin captures the decision of being a worker or entrepreneur. da Costa and Pereira (2011) derive the properties of tax schedules that satisfy a minimum equal sacrifice rule and use the Pareto efficiency test of Werning (2007) to test whether these schedules are Pareto efficient.

The application of our Pareto efficiency test to Germany is related to a study by Blundell et al. (2009). For Great Britain and Germany, they calculate the welfare weights that would render the given tax-transfer systems for lone mothers optimal. In a similar vein, Bargain et al. (2011) pursue this approach for 17 EU countries and the US focusing on singles without children.⁶ Both of these studies estimate the relevant elasticities and – for a discretized income distribution with a small number of intervals – then invert the optimal tax formula to calculate the respective welfare weights. We instead apply a more continuous approach and thereby examine the structure of marginal tax rates in greater detail. This makes our approach more powerful in identifying inefficiencies.

The remainder of this paper is organized as follows: In Section 2 we present our model of labor supply (Section 2.1) and the government's problem (Section 2.2). We reformulate the government's problem as a direct mechanism in Section 2.3. We derive the solution and express it in reduced form in Section 2.4. In Section 3 we discuss the properties of the optimal tax schedule with respect to the marginal tax rates (Section 3.1) and the participation taxes (Section 3.2). In Section 4 we derive the Pareto efficiency test (Section 4.1) and apply it to the German tax-transfer system (Section 4.2). Section 5 concludes.

2 The Model

2.1 Individuals' Labor Supply

Individuals' preferences over consumption C and hours of work L are characterized by the quasi-linear utility function

$$U(C, L; \alpha) = C - v(\alpha L),$$

with $v(0) = 0$, $v' \geq 0$, $v'' > 0$. Individuals differ in their productivity w and in α , which measures preferences for leisure. α is assumed to enter the utility function in this way to render

⁶Relatedly, Immervoll et al. (2007) consider two kinds of marginal reforms for several European countries: increasing the welfare benefit and increasing in-work benefits. They find that the latter would increase welfare in many EU countries for a large set of redistributive preferences whereas the former can only be justified by very strong redistributive preferences.

the two-dimensional screening problem tractable. w and α are distributed within $[w_0, w_1]$ and $[\alpha_0, \alpha_1]$ according to a joint density function $d(w, \alpha)$, which we represent by the marginal density $f(w)$ and the conditional density $g(\alpha|w)$: $d(w, \alpha) = f(w)g(\alpha|w)$. The mass of individuals is normalized to one.

Individuals have to pay (possibly negative) taxes T , and because α , w and L are unobservable for the government, these taxes only depend on income $Y = wL$. All individuals with the same income therefore receive the same consumption level $C = Y - T(Y)$.

Without the minimum hours constraint, income and utility depend on w and α only via the one-dimensional aggregate $\beta = \frac{w}{\alpha}$, which can easily be seen by expressing preferences in terms of C and Y :⁷

$$U = C - v(\alpha L) = C - v\left(\frac{\alpha Y}{w}\right) = C - v\left(\frac{Y}{\beta}\right) = U(C, Y; \beta).$$

Note that the smallest and largest value of β are $\beta_0 = w_0/\alpha_1$ and $\beta_1 = w_1/\alpha_0$ respectively (see Figure 1). Let $K(\beta)$ be the distribution function of β with corresponding density $k(\beta)$ that has support $[\beta_0, \beta_1]$. Also, let the conditional density of α in terms of β be $\tilde{g}(\alpha|\beta)$, with corresponding distribution function $\tilde{G}(\alpha|\beta)$.

Along each β -line individuals are identical concerning income and consumption, but – in moving away from the origin – the number of hours an individual works decreases: Because of their higher preference for leisure α , these individuals work less, but because of their higher productivity, they earn the same income and receive the same utility.

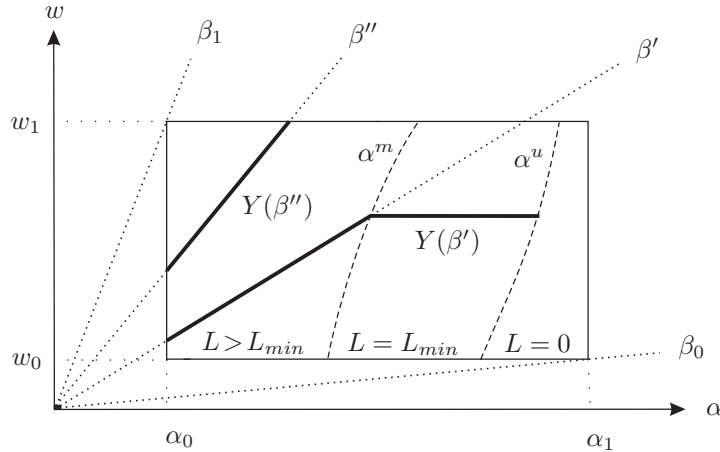


Figure 1: Partition of the type-space by α^m and α^u and iso-income curves

With the minimum hours constraint $L \geq L_{min}$, this no longer holds. As labor supply decreases along each β -line, it will at some point equal L_{min} , and would then fall below it. This, however, is not possible, so these individuals have to work the minimum number of hours. We denote the critical values of α that separate those that are not restricted by the minimum hours constraint ($L > L_{min}$) from those that are ($L = L_{min}$) by $\alpha^m(w)$; we provide a formal definition of this α^m -curve below. Since all individuals to the right of α^m work L_{min} , income in this area is constant along a horizontal line, see Figure 1. Along this horizontal line, α is increasing and

⁷Choné and Laroque (2011a) consider a similar model, but allow for a more general aggregation function than $\beta = w/\alpha$. Also, Brett and Weymark (2003) use such a type-aggregator in a model with endogenous education.

for sufficiently large values of it, individuals prefer not to work at all. We denote this second threshold by $\alpha^u(w)$; again, we provide a formal definition below. In the following we present the model and the derivation of the results for the case depicted in Figure 1, i.e. that both curves are interior and do not cross. This assumption is not necessary to derive the results but greatly simplifies the notation. Also, we could let the minimum hours constraint depend on w , i.e. define a function $L_{min}(w)$. Again, to simplify the notation we refrain from doing so.

For income levels like $Y(\beta')$, the iso-income curve is a kinked line. The group of individuals earning this income consists of two subgroups: Those on the increasing part of the curve can adjust their labor supply freely and those on the horizontal part cannot. This is important for the average elasticity of income with respect to a change in the marginal tax rate: for each income level, this elasticity will depend on the share of these two subgroups.

Finally, let the density of Y be $h(Y)$, with corresponding distribution function $H(Y)$; the formal definition is provided in Appendix A.6. As labor supply decisions depend on the tax-transfer system, this distribution function is endogenous.

2.2 The Government's Problem

The government's aim is to choose the nonlinear tax schedule $T(Y)$ that maximizes social welfare

$$W = \int_{w_0}^{w_1} \int_{\alpha_0}^{\alpha_1} \Psi(V(w, \alpha)) dG(\alpha|w) dF(w), \quad (1)$$

subject to a budget constraint and

$$V(w, \alpha) = \max C - v(\alpha L) \quad \text{s.t.} \quad C \leq wL - T(wL) \quad \text{and} \quad L \geq L_{min} \vee L = 0. \quad (2)$$

We denote the welfare benefit by $b = -T(0)$. Participation taxes then are

$$T_{part}(Y) = T(Y) - T(0) = T(Y) + b.$$

Note that the government may find it optimal to have a discontinuity in the tax schedule at the bottom, i.e. $T(Y_{min}) \neq T(0)$, so that $T_{part}(Y_{min}) \neq 0$. This can well be the case even for $Y_{min} \rightarrow 0$.

$\Psi(\cdot)$ is increasing and concave and may either represent redistributive preferences of the government or a concave transformation of individual utilities that does not change preferences over leisure and consumption. With $\Psi = \Psi(V(w, \alpha))$, all individuals with the same utility have the same impact on welfare. This implies that (e.g. along the increasing part of the iso-income curve) the utility of a 'lazy and able person' is valued the same as that of an 'unable and hardworking' individual. One could, however, easily generalize all our results to the case $\Psi(V(w, \alpha), \alpha)$.

2.3 The Government's Problem as a Direct Mechanism

As is well known, a way to make the problem of optimally choosing a nonlinear tax function tractable is to formulate it as a direct mechanism, where the government chooses the optimal income-consumption bundle $(C(w, \alpha), Y(w, \alpha))$ for each type (w, α) .⁸ The government then

⁸The reader not interested in the detailed derivation of the representation as a mechanism may skip this section and move immediately to Section 2.4, where we present the solution to the government's problem in terms of the tax schedule.

maximizes social welfare (1) subject to the resource constraint (where R denotes exogenous government expenditure)

$$\int_{w_0}^{w_1} \int_{\alpha_0}^{\alpha_1} C(w, \alpha) dG(\alpha|w) dF(w) + R \leq \int_{w_0}^{w_1} \int_{\alpha_0}^{\alpha_1} Y(w, \alpha) dG(\alpha|w) dF(w), \quad (\text{RC})$$

the minimum hours constraint

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

and the incentive compatibility constraints

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

$$\forall w, \alpha, w', \alpha' \text{ with } \frac{Y(w', \alpha')}{w} \geq L_{min} \vee Y(w', \alpha') = 0.$$

Denote (by some abuse of notation) the associated indirect utility function also by $V(w, \alpha)$; it is the ‘direct-mechanism equivalent’ to (2) and defined by:

$$V(w, \alpha) = \max_{w', \alpha'} C(w', \alpha') - v\left(\frac{\alpha' Y(w', \alpha')}{w}\right) \text{ s.t. } \frac{Y(w', \alpha')}{w} \geq L_{min} \vee Y(w', \alpha') = 0.$$

This problem is not straightforward to solve. For example, the set of incentive compatibility constraints is diminished by the fact that for most individuals it is impossible to mimic some of the other types because the income-consumption bundle designated to these other types would require the individual to work less than L_{min} . In the following we show how to rewrite this problem in a tractable manner. For this purpose, we state three lemmas.

Lemma 1. *In any incentive compatible allocation, $Y(w, \alpha)$ must be non-increasing in α .*

Proof. If for $\tilde{\alpha} > \alpha$ we have $Y(w, \tilde{\alpha}) > Y(w, \alpha)$, we must have $C(w, \tilde{\alpha}) > C(w, \alpha)$. But if, of two bundles, $(C(w, \tilde{\alpha}), Y(w, \tilde{\alpha})) \gg (C(w, \alpha), Y(w, \alpha))$, type $(w, \tilde{\alpha})$ prefers the first one, so must type (w, α) because of the lower disutility of labor. \square

Based on this lemma, we can define the threshold functions α^u and α^m ; based on α^m we can then define the value of β that is associated with the lowest income $Y_{min} = w_0 L_{min}$:

Definition 1. *For any incentive compatible allocation, define the threshold function $\alpha^u(w)$ such that $\frac{Y(w, \alpha)}{w} \geq L_{min}$ for $\alpha \leq \alpha^u(w)$ and $\frac{Y(w, \alpha)}{w} = 0$ for $\alpha > \alpha^u(w)$.*

Definition 2. *For any incentive compatible allocation, define the threshold function $\alpha^m(w)$ such that $\frac{Y(w, \alpha)}{w} > L_{min}$ for $\alpha < \alpha^m(w)$ and $\frac{Y(w, \alpha)}{w} \leq L_{min}$ for $\alpha \geq \alpha^m(w)$. This threshold can also be expressed in terms of β , which will sometimes simplify the notation. In this case we denote it by α_β^m ; formally it is given by $\alpha^m(w) = \alpha_\beta^m\left(\frac{w}{\alpha^m(w)}\right)$.*

Definition 3. *For any incentive compatible allocation, define $\underline{\beta} = \frac{w_0}{\alpha^m(w_0)}$, or implicitly by $Y(\underline{\beta}) = w_0 L_{min}$.*

In Figure 1, $\underline{\beta}$ would correspond to the β -line through the intersection of the α^m -curve and the lower border w_0 of the type space. Note that $\underline{\beta}$ constitutes the lower bound for unconstrained workers, i.e. for $\beta > \underline{\beta}$ there is a positive mass of individuals working more than L_{min} .

The next lemma is based on the type aggregator β and allows to define income, consumption and utility for all workers that work more than L_{min} solely as a function of β .

Lemma 2. *If it is optimal for type (w, α) to choose income-consumption bundle (\tilde{C}, \tilde{Y}) with $\frac{\tilde{Y}}{w} > L_{min}$, then it is also optimal for any other type (w', α') with $\frac{w'}{\alpha'} = \frac{w}{\alpha}$ and $\frac{\tilde{Y}}{w'} \geq L_{min}$.*

The next lemma simply follows from the fact that the government can only observe income. It will be used to link workers who work L_{min} with those who work more than L_{min} but earn the same income.

Lemma 3. *In any incentive compatible allocation, whenever $Y(w, \alpha) = Y(w', \alpha')$ for some types (w, α) and (w', α') , then $C(w, \alpha) = C(w', \alpha')$.*

Based on these three lemmas, we show in Appendix A.1 that the government's problem can be rewritten in the following way, which can then be solved using standard Lagrangian techniques:

Proposition 1. *Instead of choosing $\{C(w, \alpha), Y(w, \alpha)\}$ in order to maximize (1) subject to (MHC), (RC) and (IC), the planner can also choose $\{C(\beta), Y(\beta)\}$ for all unconstrained workers, consumption levels for the constrained workers $\{C(w, \alpha)\}$ and consumption levels b for all inactive workers subject to*

(i) a no discrimination constraint

$$C(\beta) = C\left(\frac{Y(\beta)}{L_{min}}, \alpha\right) \quad \forall \beta \text{ and } \alpha \in \left[\alpha^m\left(\frac{Y(\beta)}{L_{min}}\right), \alpha^u\left(\frac{Y(\beta)}{L_{min}}\right)\right], \quad (NDC)$$

(ii) an envelope condition

$$V'(\beta) = v'\left(\frac{Y(\beta)}{\beta}\right) \frac{Y(\beta)}{\beta^2} \quad \forall \beta \geq \underline{\beta}, \quad (EC)$$

(iii) a monotonicity constraint

$$Y'(\beta) \geq 0 \quad \forall \beta \geq \underline{\beta}, \quad (MC)$$

(iv) and the government budget constraint

$$\begin{aligned} \int_{w_0}^{w_1} \int_{\alpha^u(w)}^{\alpha_1} b \, dG(\alpha|w) dF(w) + R &= \int_{\underline{\beta}}^{\beta_1} (Y(\beta) - C(\beta)) \tilde{G}(\alpha_{\beta}^m(\beta)|\beta) dK(\beta) \quad (GBC) \\ &+ \int_{w_0}^{w_1} \int_{\alpha^m(w)}^{\alpha^u(w)} (wL_{min} - C(w, \alpha)) \, dG(\alpha|w) dF(w), \end{aligned}$$

where $\underline{\beta} = \frac{w_0}{\alpha^m(w_0)}$ and the threshold functions $\alpha^u(w)$ and $\alpha_{\beta}^m(\beta)$ satisfy

$$C(w, \alpha) - v(\alpha^u(w), L_{min}) = b \quad \forall w$$

and

$$Y(\beta) = \beta \alpha_{\beta}^m(\beta) L_{min} \quad \forall \beta.$$

Proof. See Appendix A.1 □

The intuition for the restatement of the government's problem is as follows: For all individuals to the left of the α^m -line, incentive constraints can be replaced by an envelope condition and a monotonicity constraint as in a standard Mirrlees problem. Any further incentive compatibility is guaranteed by the definition of the thresholds and the no discrimination constraint, which formalizes the fact that individuals with the same income must be assigned the same consumption as well.

2.4 Solution to the Government's Problem

In Appendix A.2, we state the first order conditions of the government's problem the way it was expressed as a direct mechanism in Proposition 1. Using these first order conditions, in Appendix A.3, we then derive the solution in terms of the optimal tax schedule which can be summarized as follows:

$$\frac{T'(Y(\beta))}{1 - T'(Y(\beta))} \lambda \beta \frac{\varepsilon_{Y,1-T'}}{\varepsilon_{Y,1-T'} + 1} \tilde{G}(\alpha^m(\beta)|\beta) k(\beta) - \mathcal{A}(Y(\beta)) = 0 \quad \forall \beta, \quad (3)$$

where $\varepsilon_{Y,1-T'}$ denotes the elasticity of the unconstrained workers, and with

$$\begin{aligned} \mathcal{A}(Y(\beta)) &= \int_{\beta}^{\beta_1} \int_{\underline{\alpha}(\beta')}^{\alpha^m(\beta')} (\lambda - \Psi'(V(\beta'))) d\tilde{G}(\alpha|\beta') dK(\beta') \\ &+ \int_{\frac{Y(\beta)}{L_{min}}}^{w_1} \left[\int_{\alpha^m(w)}^{\alpha^u(w)} (\lambda - \Psi'(V(w, \alpha))) dG(\alpha|w) \right. \\ &\quad \left. + \lambda g(\alpha^u(w)|w) \frac{\partial \alpha^u(w)}{\partial T(wL_{min})} (T(wL_{min}) + b) \right] dF(w). \end{aligned} \quad (4)$$

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

$$\lambda = \int_{w_0}^{w_1} \int_{\alpha_0}^{\alpha_1} \Psi'(V(w, \alpha)) dG(\alpha|w) dF(w). \quad (5)$$

Further, we have

$$\mathcal{A}(Y(\underline{\beta})) = \mathcal{A}(Y(\beta_1)) = 0. \quad (6)$$

Deriving the conditions for the optimal marginal tax rates and the optimal participation taxes from equations (3) and (4) is rather cumbersome. Also, we want to express the test for Pareto inefficiency in terms of observable labor supply elasticities and the income distribution. We therefore rewrite the above solution in reduced form. To do so, let $\bar{\Psi}'(Y)$ be the average social marginal utility of income for all individuals earning Y . Also, let $\xi(Y)$ be the semi-elasticity of participation, i.e. the increase in the number of unemployed relative to the number of individuals earning income level Y due to an absolute increase in $T(Y)$ (or b). Likewise, let $\bar{\varepsilon}(Y)$ be the average elasticity of income with respect to $1 - T'$ of all individuals earning income Y . Note that we define all (semi-) elasticities in a way that they are positive. Finally, denote the maximum income by $Y_{max} = Y(\beta_1)$.

Proposition 2. *The optimality conditions (3)-(6) can be expressed in reduced form as*

$$\frac{T'(Y)}{1-T'(Y)}\lambda Y\bar{\varepsilon}(Y)h(Y) - \mathcal{A}(Y) = 0 \quad \forall Y \quad (7)$$

with

$$\mathcal{A}(Y) = \int_Y^{Y_{max}} \left[(\lambda - \bar{\Psi}'(\tilde{Y})) - \lambda \xi(\tilde{Y})T_{part}(\tilde{Y}) \right] dH(\tilde{Y}), \quad (8)$$

$$\lambda = \bar{\Psi}'(0)H(0) + \int_{Y_{min}}^{Y_{max}} \bar{\Psi}'(Y)dH(Y), \quad (9)$$

$$\mathcal{A}(Y_{min}) = \mathcal{A}(Y_{max}) = 0, \quad (10)$$

Proof. See Appendix A.4. □

Equation (7) captures the optimality of the tax schedule at each income level Y . It could as well have been derived by the tax perturbation method as in Piketty (1997) and Saez (2001). We briefly state the derivation using this method because it makes the optimality conditions easier to interpret. Consider an infinitesimal increase dT' of the marginal tax rate in an income interval of infinitesimal length dY around income Y . This will have three effects on welfare, a substitution effect, a redistribution effect and a participation effect:

Substitution effect: Individuals within the interval adjust their labor supply along the intensive margin. By the envelope theorem, these labor supply responses only change welfare by their impact on public funds. The mass of individuals affected is $h(Y)dY$. The average increase in income is given by $\frac{\partial Y}{\partial T'}dT' = -\frac{Y\bar{\varepsilon}(Y)}{1-T'(Y)}dT'$, which multiplied by $T'(Y)$ yields the effect on public funds. The substitution effect in terms of welfare then is

$$dW^S = -\lambda \frac{T'(Y)}{1-T'(Y)} Y\bar{\varepsilon}(Y)h(Y)dT'dY. \quad (11)$$

Redistribution effect: The increase of the marginal tax rate would result in a higher overall tax of $dT'dY$ for all individuals earning more than Y and thereby redistribute money from these individuals (valued by $\bar{\Psi}'$) to the government (valued by λ). This redistribution effect on welfare therefore reads as

$$dW^R = dT'dY \int_Y^{Y_{max}} (\lambda - \bar{\Psi}'(\tilde{Y}))dH(\tilde{Y}). \quad (12)$$

Participation effect: Some of the individuals earning more than Y will stop working due to the higher participation tax. For each income level $\tilde{Y} \geq Y$, their mass is captured by $\frac{\partial h(\tilde{Y})}{\partial T_{part}}dT'dY = \xi(\tilde{Y})h(\tilde{Y})dT'dY$. By choosing unemployment over employment, the governments tax revenue is decreased by the participation tax of these individuals $T_{part}(\tilde{Y})$.⁹ This effect on welfare equals

$$dW^P = -dT'dY \int_Y^{Y_{max}} \lambda \xi(\tilde{Y})T_{part}(\tilde{Y})dH(\tilde{Y}). \quad (13)$$

For the tax function to be optimal, we have to have $dW^S + dW^R + dW^P = 0$, yielding (7).

⁹As these individuals are indifferent between working and not working, these labor supply responses only change welfare by their impact on public funds.

Equation (9) states the well known result that the average social marginal utility of income is equal to the marginal value of public funds if there are no income effects. A marginal increase in income for everyone increases welfare by the aggregate social marginal utility of income, and decreases government revenue by one (as the mass of individuals is one), which is valued by λ . For the optimal tax schedule, these effects have to cancel out, yielding (9).

$\mathcal{A}(Y_{min}) = 0$ could also have been derived by a small perturbation of the tax schedule such that all employed pay marginally higher taxes, leaving $T(0) = -b$ constant. This marginal and identical increase in the participation tax for all individuals that are employed would only cause a redistribution and a participation effect as defined in (12) and (13), both integrated over all income levels greater than or equal to Y_{min} . The condition $\mathcal{A}(Y_{min}) = 0$ then follows from $dW^R + dW^P = 0$. Because this reform increases $T_{part}(Y_{min})$ while $T(0)$ stays constant, this condition implicitly determines the optimal ‘size’ of the discontinuity in the tax schedule. As argued by Choné and Laroque (2011b) and Jacquet et al. (2012), it may well be that the government finds it optimal to have negative participation taxes for low income workers induced by such a discontinuity and at the same time have strictly positive marginal tax rates. In section 4.2 we will derive a formula for the optimal participation tax and provide a condition under which participation taxes are strictly positive.

3 Properties of the Optimal Tax Schedule

We first briefly comment on the optimal marginal tax rates in Section 3.1, before we discuss the optimal participation taxes in Section 3.2 in greater detail.

3.1 Marginal Tax Rates

Simply rewriting condition (7) yields the following corollary:¹⁰

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

$$\frac{T'(Y)}{1 - T'(Y)} = \frac{\int_Y^{Y_{max}} [(\lambda - \bar{\Psi}'(\tilde{Y})) - \lambda \xi(\tilde{Y}) T_{part}(\tilde{Y})] dH(\tilde{Y})}{\lambda Y \bar{\varepsilon}(Y) h(Y)}. \quad (14)$$

The denominator captures the substitution effect. The higher the mass of individuals $h(Y)$ whose marginal incentives are distorted and the higher their average elasticity $\bar{\varepsilon}(Y)$ and their productivity, the larger the excess burden and therefore the lower the marginal tax rate.

The first term in the numerator represents the redistribution effect as defined in (12). The greater the aggregated difference between λ and $\bar{\Psi}'$, the higher marginal tax rates should be.

The participation effect as defined in (13) is captured by the second term in the numerator. It is increasing in the mass of individuals responding along the extensive margin $\xi(Y)h(Y)$ and the participation tax. It counteracts the redistribution effect and leads to lower marginal tax rates.

¹⁰As stated in the introduction, this formula resembles the results of Saez (2002) and Jacquet et al. (2012).

We now briefly comment on the sign of the optimal marginal tax rates. First, marginal tax rates need not be zero at the bottom and at the top. According to (10), the numerator of condition (14) equals zero for Y_{min} and Y_{max} . For these two values of Y we then have to have

$$\frac{T'(Y)}{1 - T'(Y)} \lambda Y \bar{\varepsilon}(Y) h(Y) = 0.$$

For $\bar{\varepsilon}(Y)h(Y) > 0$ we get the standard result of no distortion, but if $\bar{\varepsilon}(Y)h(Y) = 0$ we may have $T'(Y) \neq 0$. This case applies in our minimum hours model for both Y_{min} and Y_{max} : Because the increasing part of the iso-income curve is infinitesimally small for Y_{min} , we have $\bar{\varepsilon}(Y_{min}) = 0$. For Y_{max} we have $h(Y_{max}) = 0$ because the length of the respective iso-income curve is infinitesimally small.

Secondly, for interior income levels we prove in Appendix A.5 that marginal tax rates are non-negative if $\bar{\Psi}'(Y)$ is decreasing in income and

$$\frac{\partial}{\partial Y} \left[\frac{\lambda - \bar{\Psi}'(Y)}{\xi(Y)} \right] > 0.$$

Note that this condition resembles the condition derived by Jacquet et al. (2012). This shows that their result carries over to different frameworks with extensive and intensive margin.

3.2 Participation Taxes

We now turn to the question whether the optimal tax-transfer system should levy participation taxes or provide participation subsidies. To do so, we first state the optimality condition for the participation taxes:¹¹

Corollary 2. *For income levels with $\xi(Y) > 0$, optimal participation taxes are given by*

$$T_{part}(Y) = \frac{(\lambda - \bar{\Psi}'(Y)) h(Y) + \frac{\partial}{\partial Y} \left[\frac{T'(Y)}{1 - T'(Y)} h(Y) Y \lambda \bar{\varepsilon}(Y) \right]}{\lambda \xi(Y) h(Y)}. \quad (15)$$

Proof. Because equation (7) holds for all values of Y , we can take its derivative with respect to Y ; rearranging terms then yields (15). \square

To gain an intuitive understanding of this expression, first consider equation (15) without the term $\frac{\partial}{\partial Y} [\cdot]$. We then have the standard interpretation of a model with only extensive labor supply responses: The sign of the optimal participation tax only depends on the social marginal utility of income compared to the marginal value of public funds: For income levels with $\bar{\Psi}' < \lambda$, participation taxes are positive, for those with $\bar{\Psi}' > \lambda$, they are negative (Diamond 1980, Saez

¹¹In our minimum hours model, the extensive margin may be missing for very high income levels, so that $\xi(Y) = 0$. For these income levels, T_{part} cannot be inferred from (15), but is implicitly defined by the set of optimality conditions stated in Proposition 2. However, condition (15), when multiplied by the denominator, holds for all Y .

2002, Choné and Laroque 2011b).¹² This result can most easily be understood by considering an (infinitesimally) small perturbation of the tax schedule as shown in Figure 2, so that the tax at income Y is reduced by dT due to a small decrease of the marginal tax rates in the interval $[Y - dY, Y]$ and a small increase of the marginal tax rates in the interval $[Y, Y + dY]$.¹³ Without intensive labor supply responses, this only has two effects: a redistribution effect of $(\bar{\Psi}'(Y) - \lambda)h(Y)$ as individuals with income Y pay lower taxes, and a participation effect of $T_{part}(Y)\lambda\xi(Y)h(Y)$ as some of the unemployed start working if the participation tax is reduced. For the optimal tax schedule, these two effects on welfare have to add up to zero and therefore the sign of the participation tax is equal to the sign of $\lambda - \bar{\Psi}'$.

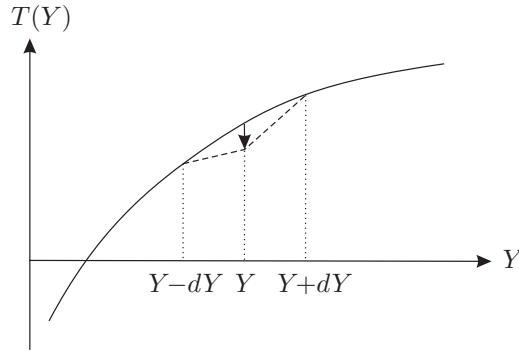


Figure 2: Tax perturbation

With labor supply responses along the intensive margin such a perturbation also has a substitution effect as defined in (11) 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 reduce it. Whether government revenue increases or decreases depends on the difference of these two effects, which in the limit, as $dT \rightarrow 0$, is captured by the derivative of the substitution effect, i.e. $\frac{\partial}{\partial Y}[\cdot]$. This derivative can be smaller than zero if for example the density $h(Y)$ is decreasing quickly. However, for a constant density, a constant elasticity and a constant marginal tax rate, the substitution effect is increasing (so that $\frac{\partial}{\partial Y}[\cdot]$ is positive), which then makes negative participation taxes less likely compared to the pure extensive model. This shows that we can have $\bar{\Psi}' > \lambda$ and still $T_{part} > 0$, so that the result of the pure extensive model does not carry over to a setting with both margins.

¹²Christiansen (2012) also discusses the question of negative participation taxes in an extensive margin model and refers to the important role of labor supply responses of high-skilled for this condition to hold. He also generalizes the result to a general equilibrium framework.

¹³Werning (2007) considers such a reform in a model with only intensive labor supply responses to test whether a given income tax schedule is Pareto inefficient.

As we have shown in Section 3.1, the substitution effect at Y_{min} equals zero since either $\bar{\varepsilon}(Y_{min})h(Y_{min}) = 0$ or $T'(Y_{min}) = 0$. This raises the question if at least at the bottom of the income distribution the result of the pure extensive model holds. Note that the term $\frac{\partial}{\partial Y}[\cdot]$ can be decomposed as

$$\frac{T'(Y)}{1-T'(Y)}\lambda\frac{\partial}{\partial Y}\left[h(Y)Y\bar{\varepsilon}(Y)\right] + h(Y)Y\lambda\bar{\varepsilon}(Y)\frac{\partial}{\partial Y}\left[\frac{T'(Y)}{1-T'(Y)}\right]. \quad (16)$$

For $\bar{\varepsilon}(Y_{min})h(Y_{min}) = 0$, the second term vanishes and the derivative in the first term is unambiguously positive so that the first term as whole is positive (negative) if $\frac{T'(Y_{min})}{1-T'(Y_{min})} > (<) 0$. For $T'(Y_{min}) = 0$, the first term vanishes and the second term is positive (negative) if T' is increasing (decreasing) in Y at Y_{min} . Thus, although the substitution effect equals zero at the bottom of the income distribution, the result from the binary model that the sign of the participation tax depends solely on $\bar{\Psi}'$ relative to λ does not carry over. However, using (16) one can show that the reverse holds:

Proposition 3. *If the social marginal utility at the bottom of the income distribution is smaller than the social marginal value of public funds, i.e. $\bar{\Psi}'(Y_{min}) < \lambda$, and $\bar{\Psi}'(Y)$ decreases in income, then $T_{part}(Y)$ is positive for all $Y \geq Y_{min}$.*

Proof. If $\bar{\Psi}'(Y_{min}) < \lambda$, then $T_{part}(Y_{min})$ can only be negative if (16) is negative. Again we have to distinguish two cases: For $\bar{\varepsilon}(Y_{min})h(Y_{min}) = 0$, the second term of (16) vanishes, so that (16) can only be negative if $T'(Y_{min}) < 0$, because $\frac{\partial}{\partial Y}[h(Y)Y\bar{\varepsilon}(Y)] \geq 0$ for Y_{min} . For $T'(Y_{min}) = 0$, the first term of (16) vanishes, so that (16) can only be negative if $\frac{\partial}{\partial Y}\left[\frac{T'(Y_{min})}{1-T'(Y_{min})}\right] < 0$, which implies $T'(Y_{min} + \epsilon) < 0$ for some small ϵ . In both cases, for $T_{part}(Y_{min})$ to be negative, T' has to be negative for Y equal or close to Y_{min} .

However, because $T(0) \leq 0$ by definition (since individuals without income cannot pay taxes), T_{part} has to be positive for some Y so that the government budget constraint is satisfied. This implies that T' has to turn positive for some value of Y , say \tilde{Y} , where T_{part} is still negative. At \tilde{Y} , $T'(\tilde{Y}) = 0$, $\frac{\partial}{\partial Y}\left[\frac{T'(\tilde{Y})}{1-T'(\tilde{Y})}\right] > 0$, and $\bar{\Psi}'(\tilde{Y}) < \lambda$, so the right hand side of (15) is unambiguously positive, a contradiction to $T_{part}(\tilde{Y})$ still being negative at that point. (For $\xi(\tilde{Y}) = 0$, the numerator of the right hand side of (15) would be positive, while $T_{part}(\tilde{Y})\lambda\xi(\tilde{Y})h(\tilde{Y}) = 0$, again a contradiction.)

If $T_{part}(Y_{min}) \geq 0$ and $T_{part}(\hat{Y}) < 0$ for some $\hat{Y} > Y_{min}$, there must be a $\bar{Y} < \hat{Y}$ such that $T'(\bar{Y}) < 0$ and $T_{part}(\bar{Y}) = 0$. If then T_{part} becomes positive for some $Y > \bar{Y}$, the same reasoning of the previous paragraph applies again. If T_{part} did not become positive, we would have $T_{part}(Y) < 0 \forall Y > \bar{Y}$. But then the right hand side of (14) would be positive for $Y = \bar{Y}$, a contradiction to $T'(\bar{Y}) < 0$. \square

This proposition generalizes a well-known result from the optimal tax model with only participation decisions (going back to Diamond (1980)) to a framework with both intensive and extensive labor supply responses. Whether the condition that the social marginal utility at the bottom of the income distribution is smaller than the social marginal value of public funds is fulfilled depends on the welfare function and, e.g., on the number of inactive workers. The higher this number, the stronger is the impact of their social marginal utility on the marginal value of public funds and the more likely this condition is fulfilled. Also, the more concave the social welfare function, the more likely it is fulfilled. In the extreme case of a Rawlsian welfare function, the condition always holds.

4 A Test for Pareto Inefficiency

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 tax-transfer system is second-best Pareto inefficient.¹⁴ In the following Section 4.1 we derive the test. We then apply this test to Germany in Section 4.2.

4.1 Theoretical Considerations

4.1.1 Inefficiently High Marginal Tax Rates

We first ask whether the marginal tax rate at a certain income level (given the marginal tax rates for the other income levels) is so high that it is beyond its Laffer value. To determine this value, it is helpful to rewrite the optimality condition (7) in the following way:

$$\begin{aligned} \frac{T'(Y)}{1-T'(Y)} \bar{\varepsilon}(Y) h(Y) Y - (1-H(Y)) + \int_Y^{Y_{max}} \xi(\tilde{Y}) T_{part}(\tilde{Y}) dH(\tilde{Y}) \\ = -\frac{1}{\lambda} \int_Y^{Y_{max}} \bar{\Psi}'(\tilde{Y}) dH(\tilde{Y}). \end{aligned} \quad (17)$$

The Laffer value, i.e. the revenue maximizing marginal tax rate is found by ignoring the effect on individual utility, i.e. by setting all welfare weights $\bar{\Psi}'$ to zero. It then immediately follows that $T'(Y)$ is too high if the left hand side of (17) is greater than zero. This yields a first test for inefficiency, which can be applied if the tax schedule, the income distribution and the labor supply elasticities are known:

Proposition 4. *For given intensive elasticities $\bar{\varepsilon}(Y)$, extensive semi-elasticities $\xi(Y)$, an income distribution $H(Y)$ and quasi-linear preferences, whenever a tax schedule satisfies*

$$\frac{T'(Y)}{1-T'(Y)} \bar{\varepsilon}(Y) h(Y) Y - (1-H(Y)) + \int_Y^{Y_{max}} \xi(\tilde{Y}) T_{part}(\tilde{Y}) dH(\tilde{Y}) > 0 \quad (18)$$

for at least some Y , then the tax schedule is second-best Pareto inefficient.

This proposition can be considered as the natural extension of the Laffer argument to nonlinear taxation: With a linear tax schedule, it is the constant marginal tax rate that is too high over the entire schedule; here, it is the marginal tax rate $T'(Y)$ at a specific income level Y . Lowering $T'(Y)$ will increase tax revenue; it will also reduce T for all income levels Y and above, which will make these individuals better off. A small reduction of the marginal tax rate $T'(Y)$ therefore constitutes a Pareto improvement.

This test will identify some of the inefficient tax schedules, but we will now argue that a stronger test exists: Even if each marginal tax rate itself is below the Laffer value, the tax schedule can be inefficient because the structure of marginal tax rates is not efficient. In this case, a different reform will be needed to achieve a Pareto improvement.

¹⁴Saez (2001) first suggested this idea. Werning (2007) elaborates this approach for the Mirrlees model with intensive labor supply responses. We extend this approach to the case of intensive and extensive labor supply responses.

4.1.2 Inefficient Structure of Marginal Tax Rates

To derive the stronger version of the test we use the fact that for each Pareto efficient tax schedule, there exists a set of nonnegative welfare weights so that the tax schedule is the solution to the welfare maximization problem for these weights. If one of these weights has to be negative, the tax schedule cannot be efficient. Taking the derivative of condition (17) yields an expression for these weights:

$$\frac{\partial}{\partial Y} \left[\frac{T'(Y)}{1 - T'(Y)} \bar{\varepsilon}(Y) h(Y) Y \right] + h(Y) - \xi(Y) T_{part}(Y) h(Y) = \frac{\bar{\Psi}'(Y)}{\lambda} h(Y). \quad (19)$$

A negative welfare weight $\bar{\Psi}'(Y)$ and thus a Pareto inefficiency exists if the left hand side of (19) is negative, i.e., if the left hand side of (17) is decreasing in income. This defines the stronger version of the test:

Proposition 5. *Given intensive elasticities $\bar{\varepsilon}(Y)$, extensive semi-elasticities $\xi(Y)$, an income distribution $H(Y)$ and quasi-linear preferences, a tax schedule $T(Y)$ is second-best Pareto inefficient, if*

$$\frac{T'(Y)}{1 - T'(Y)} \bar{\varepsilon}(Y) h(Y) Y - (1 - H(Y)) + \int_Y^{Y_{max}} \xi(\tilde{Y}) T_{part}(\tilde{Y}) dH(\tilde{Y}) \quad (20)$$

is decreasing in Y for at least one Y .

Again, this test can be applied if the tax schedule, the income distribution and the labor supply elasticities are known. Note that it nests the condition for Pareto inefficiency of Proposition 4, i.e., whenever a tax schedule is inefficient according to (18), it is also inefficient according to (20): If the cumulative welfare weights are smaller than zero (so that the right hand side of (17) is positive), then at least one of the welfare weights has to be negative. On the other hand, the weighted sum might still be positive although some of the weights are negative.

If the test indicates that a tax schedule is inefficient, then a reform as depicted in Figure 2 in Section 3.2, conducted at income level Y , will yield a Pareto improvement.¹⁵ Such a reform will be self-financing or even increase tax revenue. Without labor supply responses, this tax cut of course decreases tax revenue, but the labor supply responses will outweigh this loss. Using equation (19) instead of (20) makes it easier to see, when that will be the case.

The mechanical loss in tax revenue is given by $h(Y)$, the mass of individuals affected by the tax cut. The participation effect on public funds induced by the tax reform is captured by the third term on the left hand side of (19): The larger $T_{part}(Y)$ and the larger the participation semi-elasticity $\xi(Y)$, the larger is this participation effect. The argument for the substitution effect is more subtle as the tax reform on the one hand increases marginal tax rates for incomes slightly higher than Y and on the other hand decreases marginal tax rates for incomes slightly lower than Y . In the limit, the overall sign of these intensive labor supply responses is captured by the derivative of the substitution effect captured by the first term on the left hand side of (19). That is, the positive effect on public funds induced by the labor supply increase of those with slightly lower income is more likely to outweigh the other effect if the marginal tax rate, the density $h(Y)$ or the elasticity is decreasing in income.

¹⁵For the case without extensive labor supply responses, such a reform has already been proposed by Werning (2007).

4.1.3 Overcoming Inefficiencies

A favorable property of the proposed tests is that to apply them only the income distribution and elasticities are required. In the terminology of Chetty (2009), the income distribution and the elasticities are sufficient statistics to uncover inefficiencies. However, for overcoming an inefficiency, one has to know how individuals react to large tax reforms and therefore has to make structural assumptions about their labor supply decisions. Nevertheless our analysis provides theory-based guidance for such reforms. Whenever a tax schedule is characterized by inefficiently high marginal tax rates as discussed in Section 4.1.1, we know that a small decrease in these marginal tax rates yields a Pareto improvement. In order to know how strong these decreases have to be to not only yield a Pareto improvement but to completely eliminate the inefficiency, one has to make structural assumptions. Similarly, if the structure of marginal tax rates is inefficient as discussed in Section 4.1.2, we know that a small reform as depicted in Figure 2 yields a Pareto improvement. But again, structural assumptions are required to determine how to eliminate the inefficiency.

Of course there will always exist not only one, but a whole set of Pareto improving reforms. Each of these reforms would yield a different allocation on the Pareto frontier. That is, when deciding how to overcome the inefficiency, one has to abandon the sole ‘efficiency consideration’ and make a choice of how to value the utility of different individuals.

4.2 An Application to Germany

In order to apply the Pareto inefficiency test, the tax-transfer schedule has to be known. For Germany (and likely for other countries as well), it is not immediately apparent what this schedule looks like because it is the result of the interplay of three different systems. We discuss how to construct this schedule and how we estimate the income distribution in the following Section 4.2.1. The results are presented in Section 4.2.2. Policy implications are discussed in Section 4.2.3.

4.2.1 Income Distribution, Tax-Transfer System and Elasticities

As in most countries, the tax-transfer system conditions on marital status as well as on the number of children. As the taxation of families raises a number of additional issues, we focus on singles without children. In addition, eligibility for welfare benefits depends on assets. Therefore, we only consider individuals with sufficiently low assets such that eligibility for welfare benefits is ensured.

The tax-transfer system results from the interplay of three different systems: the income tax schedule, the welfare benefit system including the phase-out region and social insurance contributions. We refrain from presenting the detailed derivation of the schedule and only state the main steps:¹⁶ Gross income determines payments to the social insurance system according to the Social Security Code. Gross income and social insurance contributions then determine the tax liability according to the Personal Income Tax Code. Transfers then depend on gross income, taxes and social insurance contributions. Integrating the three systems, we arrive at the schedule of effective marginal tax rates (for the year 2010) as shown in Figure 3. Marginal tax rates are very high for low incomes. As soon as transfers are phased out, marginal tax rates decrease drastically.¹⁷

¹⁶The detailed derivation is available from the authors upon request.

¹⁷There is a small downward jump in the tax schedule at 400 €, which is why T' tends to $-\infty$ at this income level. As

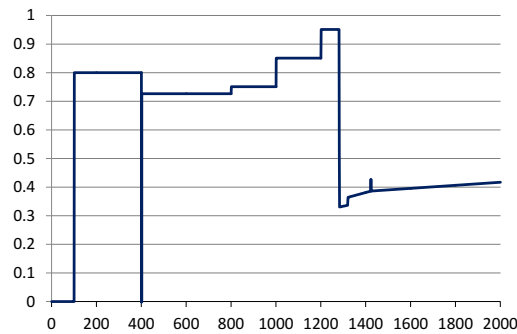


Figure 3: Marginal Tax Rates as Function of Monthly Income for the year 2010

In contrast to other studies (like Sinn et al. 2006), the highest phase-out rate is below 100%. This is because we consider contributions to the pension system not purely as a tax, as there is a Bismarckian pension system in place in Germany, see OECD (2011). Although the rate of return in the pension system is likely to be very low, it seems reasonable to assume that individuals will (on average) receive at least half of their (marginal) contributions as (higher) pensions; this reduces the effective marginal tax rate by about five percentage points.¹⁸

To estimate the income distribution we use data for the year 2010 of the German Socio-Economic Panel (SOEP), which is a representative sample of German households that are interviewed annually, see Wagner et al. (2007). Our sample (of singles, aged 18 to 65, out of education, and with sufficiently low assets) consists of 586 observations. The minimum and maximum value of gross monthly income are 0 and 14.065 Euro. The mean income is 1.844 Euro (2.248 Euro if restricted to positive incomes).

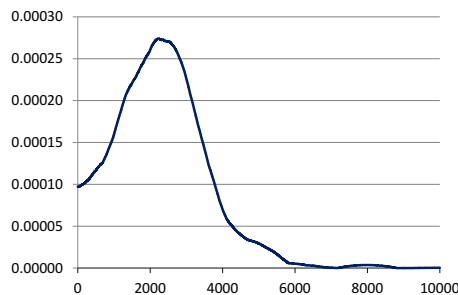


Figure 4: Density of the income distribution for the year 2010

We estimate the density of the income distribution nonparametrically (using the standard SOEP weights), employing an Epanechnikov kernel and Silverman’s rule of thumb to determine the bandwidth, see Fan and Gijbels (2003). Results for the Pareto efficiency test are, however, basically identical for different values of the bandwidth, so we refrain from applying any cross-validation procedure to determine an optimal bandwidth. The distribution of monthly gross incomes is illustrated (up to 10.000 Euro) in Figure 4.

this inefficiency is of second-order importance, we do not further comment on it. Also, there is a small spike at 1.423 €, which is due to the way the tax formula is stated in the tax code. As it arises due to rounding, it can be ignored. Note, that this small spike is also visible in Figure 5.

¹⁸The main result of an inefficient structure of the marginal tax rates is robust with regard to how contributions to the pension system are taken into account.

We do not estimate elasticities ourselves but instead apply a wide range of values of the empirical literature. For the benchmark case we use 0.25 for the extensive elasticities (which we denote by ν), and 0.33 for the intensive elasticities, see Chetty et al. (2011), but our main result holds for a large set of values (see below).

4.2.2 Results

As marginal tax rates are very high in the phase-out region, one might suspect that they are beyond their Laffer value as defined in Section 4.1.1.

Figure 5(a) shows our test function (20) for the benchmark case with intensive elasticities $\bar{\epsilon} = 0.33$ and extensive elasticities $\nu = 0.25$. For the interval where marginal tax rates are about 95%, they are indeed above their Laffer value, since the test function is larger than zero.¹⁹

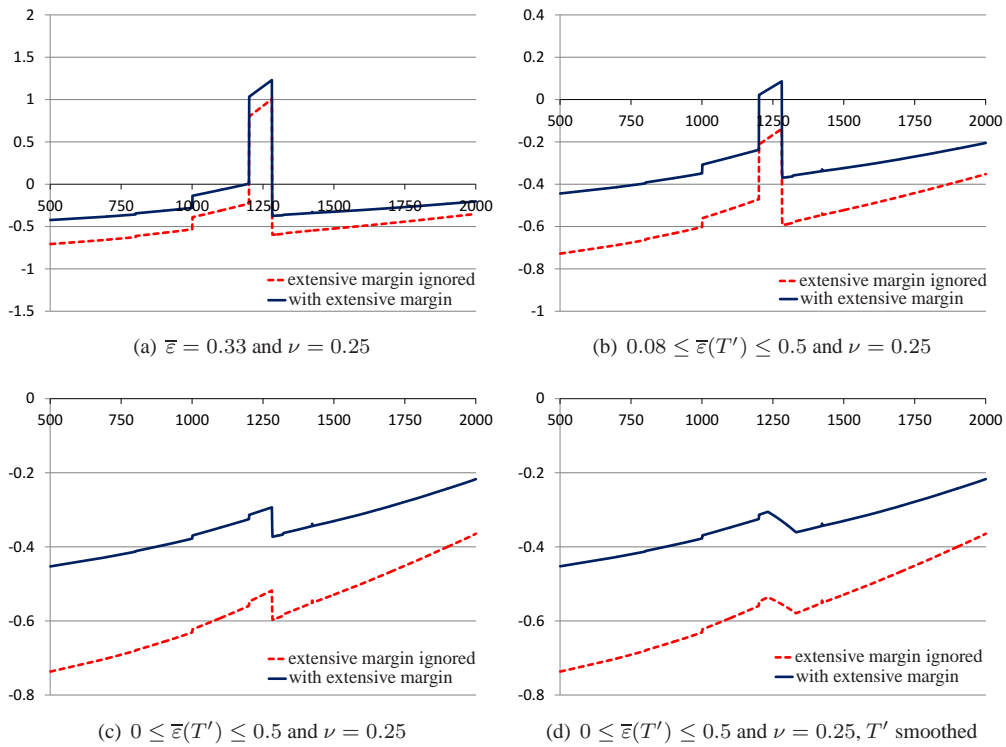


Figure 5: Graph of the test function (left hand side of (17)) for different intensive elasticities $\bar{\epsilon}$ and extensive elasticity $\nu = 0.25$; (a)-(c) original tax schedule, (d) smoothed tax schedule.

This could be considered a strong result, but it may need the following qualification: Assuming an intensive elasticity that does not depend on the value of the marginal tax rate may not be appropriate. With a constant elasticity, the percentage increase in income due to a 1 percentage point increase in T' strongly increases in T' . For example, a decrease in T' from 95% to 94% induces a relative increase in income that is 10 times as high as for a decrease from 50% to

¹⁹As in most data sets, top incomes are underrepresented in the SOEP data we use. Taking this into account would slightly weaken the case for the marginal tax rates being above their Laffer values. However, our main result, that the structure of marginal tax rates is inefficient, is independent of any underrepresentation of high incomes.

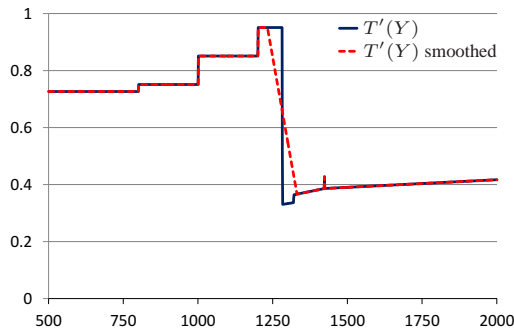


Figure 6: Marginal Tax Rates and smoothed Tax Rates as Function of Monthly Income for the year 2010

49%. This would imply a 10 times as large semi-elasticity. Such huge differences in the semi-elasticities might be considered too large. We therefore also apply our test for the case that the semi-elasticity is constant, and for an intermediate case. To keep the semi-elasticity constant, we let the elasticity decrease linearly in T' from 0.5 to 0, i.e., we assume $\bar{\varepsilon}(T') = 0.5 - 0.5T'$.²⁰ In this case, the inefficiency according to Proposition 4 vanishes, see Figure 5(c), as our test function is now below zero. For the intermediate case, we let the semi-elasticity increase less heavily in T' than is the case with a constant elasticity. When we assume $\bar{\varepsilon}(T') = 0.5 - 0.42T'$ (so that the lower bound for $\bar{\varepsilon}$ is not 0, but 0.08), our test identifies an inefficiency when extensive effects are incorporated, but fails to do so, if they are ignored, see Figure 5(b).

Whether the German tax-transfer system passes the test of Proposition 4 is therefore very sensitive with respect to the elasticities. However, in all three cases (Figure 5(a)-(c)) the test curve is falling, so the test shows an inefficiency according to Proposition 5. Since marginal tax rates drop discontinuously (see Figure 3), one might argue that this is actually trivial. Whenever there is a discontinuity in the marginal tax rates, the test function (20) is going to be characterized by a discontinuous downward jump if elasticities and the income distribution are smooth. We therefore test the following: We smooth the decrease in marginal tax rates (see Figure 6 for a smoothing interval of 100 Euro), and determine how large the smoothing interval has to be (leaving everything else equal), so that the inefficiency disappears. If this interval is small, the inefficiency is of second-order importance. However, this is not the case. Figure 5(d) shows our test function for a smoothing interval of 100 Euro (50 Euro below and above the discontinuity in T') for the elasticity values as in Figure 5(c): The inefficiency clearly stays present. Indeed the inefficiency does not disappear for any smoothing interval smaller than 438 Euro.²¹

The decrease of marginal tax rates after transfers are phased out, as it is observed in many countries, has already been criticized by Kaplow (2007, p. 304). Referring to results from numerical simulations based on utilitarian welfare functions, he argues that marginal tax rates in the phase-out region are too high, and too low afterwards. For Germany we show this to be correct but also make the argument even stronger since the tax-transfer system is (second-best) Pareto inefficient and can therefore not be justified by any welfare function.

²⁰For the intermediate value of $T' = 0.33$ we get the intensive elasticity = 0.33 that we assumed before.

²¹Considering in addition income effects would of course yield different numbers. With income effects, a tax reform as depicted in Figure 2 would slightly reduce labor supply of individuals earning Y , which would decrease tax revenue and therefore make such a reform less likely to be feasible. As the tax schedule remains inefficient even for a large smoothing interval, taking into account income effects would not change our main result but only slightly decrease the extent of the inefficiency, especially because the literature has found income effects to be rather small, see Meghir and Phillips (2010) and the references therein.

4.2.3 Possible Reforms and Policy Implications

The German tax-transfer system has often been criticized for its disincentives to work for individuals with low incomes. One proposal has been to lower marginal tax rates in the phase-out region, financed by a decrease in the welfare benefit (Sinn et al. 2006). For individuals that cannot find a job this proposal also included a guaranteed job offer in the public sector; if accepted, transfers would then be as high as before the reform. Such a reform would increase employment, but its welfare consequences are ambiguous because at least some of the welfare recipients are worse off.

Our analysis clearly suggests that the high marginal tax rates in the phase-out region are indeed hard to justify. If intensive elasticities in this region are above 0.1, they might even be above their Laffer value, which would imply that slightly lowering them would increase tax revenue.

As already mentioned, this result may have to be qualified because high incomes are underrepresented in the SOEP data. In contrast, the inefficiency identified by the second test is independent of how accurately the density of high income is estimated (see equation (19)). Also, the test indicates an inefficiency for a very wide range of elasticities. We therefore conjecture that there is room for a Pareto improving reform, where marginal tax rates are decreased in the phase-out region, and increased for incomes just above this region. Because the absolute level of taxes does not increase for any income level, no individual is made worse off, but tax revenue increases due to the induced labor supply responses along the intensive and extensive margin. To obtain concrete numbers for such a reform so that not only a Pareto improvement, but also a Pareto optimal allocation is achieved one has to make structural assumptions about the elasticities in order to predict labor supply responses to reforms that come with substantial changes in marginal tax rates. One should then also take consumption taxes into account. Whereas this would have no effect on the general pattern of such a Pareto improving reform, it may well influence concrete numbers.

5 Conclusion

We analyzed the optimal design of tax-transfer systems in the presence of intensive and extensive labor supply responses. We derived optimality conditions for the entire tax schedule, but our interest was mainly on that part of the schedule where individuals receive transfers. More specifically, we asked whether participation subsidies and high marginal tax rates in the transfer phase-out region can be grounded in optimal tax theory.

Concerning participation subsidies, we derived a condition for negative participation taxes to be never part of an optimal tax schedule: the social marginal utility of the lowest income worker to be smaller than the marginal value of public funds. We thereby extended the result from Diamond (1980), Saez (2002) and Choné and Laroque (2011b) to the case where in addition to extensive labor supply responses also intensive labor supply responses are considered.

Regarding the issue of high marginal tax rates in the phase-out region, we developed a test for the Pareto inefficiency of a given tax-transfer system. This test is expressed in reduced form and is an extension of Werning (2007) to the case of intensive and extensive labor supply responses. When applied to the German tax-transfer system the results suggest an inefficient structure of marginal tax rates: a decrease of marginal tax rates in the phase-out region combined with an increase of marginal tax rates for slightly higher incomes could yield a Pareto improvement. Using these insights as a starting point for analyzing Pareto improving tax reforms in a structural labor supply

model for Germany in the spirit of Blundell and Shephard (2012) would be an interesting task for future research.

Applying this test to other countries would also be worthwhile. Constructing the schedule of effective marginal tax rates, however, requires detailed knowledge of the interplay of the tax code and all elements of the welfare benefit program (at the federal and the state level in some countries), but once the schedule is known, the test can easily be applied. The extension of such a test to tax-transfer systems for couples and families should also be pursued.²² This would add additional interesting aspects because the marginal tax rate of the primary earner often depends on the earnings of the secondary earner and vice versa.

²²See Cremer et al. (2012), Kleven et al. (2009), Immervoll et al. (2009) and Bach et al. (2012) for recent contributions to the optimal tax treatment of couples.

A Appendix

A.1 Proof of Proposition 1

To prove that (NDC) , (EC) , (MC) and the definition of the thresholds imply incentive compatibility, we use Figure 7, where a representative iso-income curve is illustrated. We first argue that an iso-income curve indeed has a shape as illustrated in Figure 7: Note that by Lemma 2 income and consumption are constant along the increasing part of that line because $\frac{w}{\alpha}$ is constant. By the definition of the $\alpha^m(w)$ -curve, income is the same on the flat part of this kinked line; thus this curve is indeed an iso-income curve. Because of (NDC) , we know that consumption on the flat part must also be equal to consumption on the increasing part, so the iso-income curve is also an iso-consumption-curve. Finally note that by the definition of the $\alpha^u(w)$ -curve, income is zero to the right of it and therefore consumption must be the same for all those types.

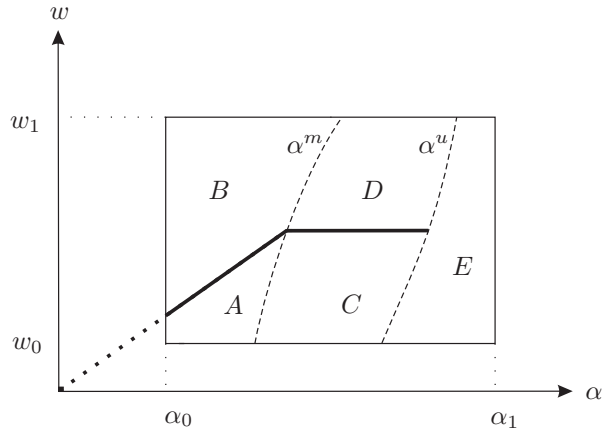


Figure 7: Incentive compatibility constraints

We now show that (NDC) , (EC) , (MC) and the definition of the thresholds imply incentive compatibility for all types on such a kinked line:

Incentive constraints for the increasing part: For the increasing part of the iso-income curve, (EC) and (MC) guarantee that no income-consumption bundle to the left of the $\alpha^m(w)$ -curve is preferred; individuals on the increasing part prefer their income-consumption bundle to any income-consumption bundle in A or B. By the no-discrimination constraint we know that for each income-consumption bundle in D or C there exists an equivalent income-consumption bundle in A or B; individuals on the increasing part thus prefer their income-consumption bundle to any income-consumption bundle in C or D. Since along the flat part of the iso-income curve, utility is decreasing in α , we know that utility at the kink is larger than at the point where the iso-income curve intersects the $\alpha^u(w)$ -curve; thus individuals on the increasing part prefer their income-consumption bundle to the income-consumption bundle in E.

Incentive constraints for the flat part: By the same argument as above, it follows that individuals on the flat part prefer their income-consumption bundle to that in E. By the minimum hours constraint, they cannot choose income-consumption bundles in A and C. Each consumption bundle in B is not preferred by the type on the kink of the curve as argued above. Since income in B is higher and disutility of work is increasing in α along the flat part, no individuals in the

flat part prefer any income-consumption bundle in B (and therefore also none in D) to their own income-consumption bundle.

Finally, by the definition of the $\alpha^u(w)$ -line and the same arguments as above, no individuals in E prefer any income-consumption bundle to the left of the $\alpha^u(w)$ -line.

A.2 The First-Order Conditions of the Government's Problem

As a first step, rewrite the government's objective (1) in terms of $V(\beta)$, $(\alpha^u(w))$, $\alpha^m(w)$, $\alpha_\beta^m(\beta)$, b and $V(w, \alpha)$ as

$$\begin{aligned} W = & \int_{\underline{\beta}}^{\beta_1} \int_{\underline{\alpha}(\beta)}^{\alpha_\beta^m(\beta)} \Psi(V(\beta)) d\tilde{G}(\alpha|\beta) dK(\beta) \\ & + \int_{w_0}^{w_1} \left[\int_{\alpha^m(w)}^{\alpha^u(w)} \Psi(V(w, \alpha)) dG(\alpha|w) + \int_{\alpha^u(w)}^{\alpha_1} \Psi(b) dG(\alpha|w) \right] dF(w). \end{aligned}$$

where $\underline{\alpha}(\beta)$ is the lowest value of α associated with a certain value of beta, i.e. $\underline{\alpha}(\beta) = \frac{w_0}{\beta}$. Using the definition of the indirect utility function, replace $C(\beta)$ by $V(\beta) + v\left(\frac{Y(\beta)}{\beta}\right)$ and $C(w, \alpha)$ by $V(w, \alpha) + v(\alpha L_{min})$. The Lagrangian for the problem then reads as:

$$\begin{aligned} \mathcal{L} = & \int_{\underline{\beta}}^{\beta_1} \int_{\underline{\alpha}(\beta)}^{\alpha_\beta^m(\beta)} \Psi(V(\beta)) d\tilde{G}(\alpha|\beta) dK(\beta) \\ & + \int_{w_0}^{w_1} \int_{\alpha^m(w)}^{\alpha^u(w)} \Psi(V(w, \alpha)) dG(\alpha|w) dF(w) + \int_{w_0}^{w_1} \int_{\alpha^u(w)}^{\alpha_1} \Psi(b) dG(\alpha|w) dF(w) \\ & + \lambda \left\{ \int_{\underline{\beta}}^{\beta_1} \left[Y(\beta) - \left(V(\beta) + v\left(\frac{Y(\beta)}{\beta}\right) \right) \right] \tilde{G}(\alpha_\beta^m(\beta)|\beta) dK(\beta) \right. \\ & \quad \left. + \int_{w_0}^{w_1} \int_{\alpha^m(w)}^{\alpha^u(w)} (wL_{min} - (V(w, \alpha) + v(\alpha L_{min}))) dG(\alpha|w) dF(w) \right. \\ & \quad \left. - \int_{w_0}^{w_1} b(1 - G(\alpha^u(w)|w)) dF(w) \right\} \\ & + \int_{\underline{\beta}}^{\beta_1} \int_{\alpha^m(Y(\beta)/L_{min})}^{\alpha^u(Y(\beta)/L_{min})} \eta(\beta, \alpha) \left[V(\beta) + v\left(\frac{Y(\beta)}{\beta}\right) - V(w, \alpha) - v(\alpha L_{min}) \right] d\alpha d\beta \\ & + \int_{\underline{\beta}}^{\beta_1} \left[\mu(\beta) V'(\beta) - \mu(\beta) v' \left(\frac{Y(\beta)}{\beta} \right) \frac{Y(\beta)}{\beta^2} \right] d\beta, \end{aligned}$$

where λ is the multiplier of the resource constraint, $\eta(\beta, \alpha)$ is the multiplier function of the no discrimination constraint and $\mu(\beta)$ is the multiplier function of the envelope condition. Partially integrating the term $\int_{\underline{\beta}}^{\beta_1} \mu(\beta) V'(\beta) d\beta$ yields $-\int_{\underline{\beta}}^{\beta_1} \mu'(\beta) V(\beta) + \mu(\beta_1) V(\beta_1) - \mu(\underline{\beta}) V(\underline{\beta})$, so that the last line of the Lagrangian can be replaced by

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

The first order conditions are:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial V(\beta)} &= \int_{\underline{\alpha}(\beta)}^{\alpha_{\beta}^m(\beta)} (\Psi'(V(\beta)) - \lambda) d\tilde{G}(\alpha|\beta) k(\beta) - \mu'(\beta) \\ &+ \int_{\alpha^m(Y(\beta)/L_{min})}^{\alpha^u(Y(\beta)/L_{min})} \eta(\beta, \alpha) d\alpha = 0 \end{aligned} \quad (21)$$

$$\left. \frac{\partial \mathcal{L}}{\partial V(w, \alpha)} \right|_{\alpha < \alpha^u} = (\Psi'(V(w, \alpha)) - \lambda) g(\alpha|w) f(w) - \eta(Y^{-1}((wL_{min}), \alpha)) = 0 \quad (22)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial V(w, \alpha^u(w))} &= (\Psi'(V(w, \alpha)) - \lambda) g(\alpha|w) f(w) - \eta(Y^{-1}((wL_{min}), \alpha)) \\ &+ \lambda g(\alpha^u(w)|w) \frac{\partial \alpha^u(w)}{\partial V} (b + wL_{min} - (V(w, \alpha) + v(\alpha^u(w)L_{min}))) = 0 \end{aligned} \quad (23)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial Y(\beta)} &= \lambda \left(1 - v' \left(\frac{Y(\beta)}{\beta} \right) \frac{1}{\beta} \right) \tilde{G}(\alpha_{\beta}^m(\beta)|\beta) k(\beta) \\ &- \mu(\beta) \frac{v' \left(\frac{Y(\beta)}{\beta} \right) + v'' \left(\frac{Y(\beta)}{\beta} \right) \frac{Y(\beta)}{\beta}}{\beta^2} \\ &+ \int_{\alpha^m(Y(\beta)/L_{min})}^{\alpha^u(Y(\beta)/L_{min})} \eta(\beta, \alpha) \left[v' \left(\frac{Y(\beta)}{\beta} \right) \frac{1}{\beta} - \frac{\partial V}{\partial w} \frac{1}{L_{min}} \right] d\alpha = 0 \end{aligned} \quad (24)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial b} &= \int_{w_0}^{w_1} \int_{\alpha^u(w)}^{\alpha_1} \Psi'(b) dG(\alpha|w) dF(w) - \lambda \int_{w_0}^{w_1} \int_{\alpha^u(w)}^{\alpha_1} dG(\alpha|w) dF(w) \\ &- \lambda \int_{w_0}^{w_1} \frac{\partial \alpha^u(w)}{\partial b} g(\alpha^u(w)|w) (b + wL_{min} - V(w, \alpha) - v(\alpha L_{min})) dF(w). \end{aligned} \quad (25)$$

Finally the derivatives with respect to the endpoint conditions are

$$\frac{\partial \mathcal{L}}{\partial V(\beta_1)} = \mu(\beta_1) = 0 \quad (26)$$

and

$$\frac{\partial \mathcal{L}}{\partial V(\underline{\beta})} = \mu(\underline{\beta}) = 0. \quad (27)$$

A.3 Solution to the Government's Problem

First integrating (22) over α^m to α^u and adding (23), then integrating this expression over w_0 to w_1 , finally adding (25) as well as (21) integrated over $\underline{\beta}$ to β_1 yields

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

i.e. equation (5). Integrating (21) yields

$$\begin{aligned} \mu(\beta) &= \int_{\beta}^{\beta_1} \int_{\underline{\alpha}(\beta')}^{\alpha_{\beta'}^m(\beta)} (\lambda - \Psi'(V(\beta'))) d\tilde{G}(\alpha|\beta') dK(\beta') \\ &\quad - \int_{\beta}^{\beta_1} \int_{\alpha^u(Y(\beta')/L_{min})}^{\alpha^m(Y(\beta')/L_{min})} \eta(\beta, \alpha) d\alpha d\beta'. \end{aligned} \quad (28)$$

Inserting (22) and (23) into (28) then results in

$$\begin{aligned} \mu(\beta) &= \int_{\beta}^{\beta_1} \int_{\underline{\alpha}(\beta')}^{\alpha_{\beta'}^m(\beta')} [\lambda - \Psi'(V(\beta'))] d\tilde{G}(\alpha|\beta') dK(\beta') \\ &\quad + \int_{\beta}^{\beta_1} \left[\int_{\alpha^m(Y(\beta')/L_{min})}^{\alpha^u(Y(\beta')/L_{min})} \left[\lambda - \Psi' \left(V \left(\frac{Y(\beta')}{L_{min}}, \alpha \right) \right) \right] d\alpha \right. \\ &\quad \left. + \lambda g(\alpha^u(w)|w) \frac{\partial \alpha^u(w)}{\partial V(w, \alpha^u(w))} (T(wL_{min}) + b) \right] dK(\beta'). \end{aligned}$$

Using $\frac{\partial V(w, \alpha)}{\partial w} = (1 - T'(Y(\beta)))L_{min}$ and $v' \left(\frac{Y(\beta)}{\beta} \right) \frac{1}{\beta} = 1 - T'(Y(\beta))$ to simplify (24) yields:

$$\lambda \left(1 - v' \left(\frac{Y(\beta)}{\beta} \right) \frac{1}{\beta} \right) \tilde{G}(\alpha^m(\beta)|\beta)k(\beta) - \mu(\beta) \frac{v' \left(\frac{Y(\beta)}{\beta} \right) + v'' \left(\frac{Y(\beta)}{\beta} \right) \frac{Y(\beta)}{\beta}}{\beta^2} = 0. \quad (29)$$

Inserting (28) into (29) 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

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

Together with the endpoint conditions (26) and (27) this constitutes the solution.

A.4 Proof of Proposition 2

Let $\bar{\Psi}'(Y)$ be the average marginal utility of income of all individuals earning income Y ; it is given by

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

Using $\bar{\Psi}'(Y)$, the first and second line of (4) can be rewritten as $\int_Y^{Y_{max}} [(\lambda - \bar{\Psi}'(\tilde{Y}))] dH(\tilde{Y})$.

Let $\xi(Y)$ be the semi-elasticity of participation, i.e. the increase in the number of unemployed relative to the number of individuals earning income level Y , $h(Y)$, due to an absolute increase in $T(Y)$ (or b); it is given by

$$\xi(Y(\beta)) = \frac{-\frac{\partial \alpha^u(w_\beta)}{\partial T} g(\alpha^u(w_\beta)|w_\beta) \frac{f(w_\beta)}{L_{min}}}{h(Y(\beta))}.$$

Applying integration by substitution, the term

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

can be rewritten as

$$\int_{Y(\beta)}^{w_1 L_{min}} \underbrace{g\left(\alpha^u\left(\frac{Y}{L_{min}}\right) \middle| \frac{Y}{L_{min}}\right) \frac{\partial \alpha^u\left(\frac{Y}{L_{min}}\right)}{\partial T(Y)} \frac{1}{h(Y)} \frac{f\left(\frac{Y}{L_{min}}\right)}{L_{min}}}_{\xi(Y)} (T(Y) + b) dH(Y).$$

Using this, the third line of (4) can be rewritten as

$$\int_Y^{Y_{max}} -\lambda \xi(\tilde{Y}) T_{part}(\tilde{Y}) dH(\tilde{Y}).$$

Using the definition of $\varepsilon_{Y,\beta}$ and the first order condition $(1 - T')\beta = v'$, we have

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

We can therefore rewrite the first term of (3) as

$$\frac{T'(Y(\beta))}{1 - T'(Y(\beta))} \lambda \beta \frac{\varepsilon_{Y,1-T'}}{\frac{\partial Y}{\partial \beta}} \tilde{G}(\alpha^m(\beta)|\beta) k(\beta). \quad (30)$$

Since the average elasticity $\bar{\varepsilon}(Y)$ for income Y and the elasticity $\varepsilon_{Y,1-T'}$ are linked by

$$\frac{\bar{\varepsilon}(Y)}{\varepsilon_{Y,1-T'}} = \frac{\tilde{G}(\alpha^m(\beta)|\beta) k(\beta)}{h(Y) \frac{\partial Y}{\partial \beta}},$$

we can rewrite (30) as $\frac{T'(Y)}{1-T'(Y)} \lambda Y \bar{\varepsilon}(Y) h(Y)$.

A.5 Proof of Conditions for Nonnegative Marginal Tax Rates

In our minimum hours model we have $\xi(Y) = 0 \forall Y \geq \bar{Y}$ for some $\bar{Y} \in]Y_{min}, Y_{max}[$. In the following we show that – if $\frac{\partial}{\partial Y} \left[\frac{\lambda - \bar{\Psi}'(Y)}{\xi(Y)} \right] > 0$ and $\bar{\Psi}'(Y)$ is decreasing in income – there can be no interval $]Y_1, Y_2[$ with negative marginal tax rates for

- Case 1: $Y_1 < Y_2 < \bar{Y}$,
- Case 2: $Y_1 < \bar{Y} \leq Y_2$,
- Case 3: $\bar{Y} \leq Y_1 < Y_2$.

In a model where the extensive margin is always present, only Case 1 applies.

In all three cases we would have $\mathcal{A}(Y_1) = 0$ and $\mathcal{A}(Y_2) = 0$ because $T'(Y_1) = 0$ and $T'(Y_2) = 0$.²³ We would also have $\mathcal{A}'(Y_1) \leq 0$ and $\mathcal{A}'(Y_2) \geq 0$, because for T' to be negative in $]Y_1, Y_2[$, $\mathcal{A}(Y)$ has to be negative within this interval.

Case 1: In this case we have

$$\mathcal{A}'(Y) = [\bar{\Psi}'(Y) - \lambda + \lambda\xi(Y)T_{part}(Y)]h(Y) \quad (31)$$

for $Y = Y_1$ and $Y = Y_2$. Solving $\mathcal{A}'(Y_1) \leq 0$ for $T_{part}(Y_1)$ and $\mathcal{A}'(Y_2) \geq 0$ for $T_{part}(Y_2)$, and using $T_{part}(Y_1) > T_{part}(Y_2)$ since marginal tax rates are negative, we have

$$\frac{\lambda - \bar{\Psi}'(Y_1)}{\lambda\xi(Y_1)} \geq \frac{\lambda - \bar{\Psi}'(Y_2)}{\lambda\xi(Y_2)},$$

which cannot hold if $\frac{\partial}{\partial Y} \left[\frac{\lambda - \bar{\Psi}'(Y)}{\xi(Y)} \right] > 0$.

Case 2: Using (31), in this case we have

$$\begin{aligned} \bar{\Psi}'(Y_1) - \lambda + \lambda\xi(Y_1)T_{part}(Y_1) &\leq 0 \\ \bar{\Psi}'(Y_2) - \lambda &\geq 0, \end{aligned}$$

which implies $\lambda\xi(Y_1)T_{part}(Y_1) \leq \bar{\Psi}'(Y_2) - \bar{\Psi}'(Y_1)$. Because the right hand side is negative, this requires $T_{part}(Y_1) < 0$.

We now have to distinguish two cases:

If $\bar{\Psi}'(Y_1) < \lambda$, we would then have $T_{part}(Y_1 + \epsilon) < 0$, $\bar{\Psi}'(Y_1 + \epsilon) < \lambda$ and $T'(Y_1 + \epsilon) < 0$, which cannot hold as we show in the proof of Proposition 3.

If $\bar{\Psi}'(Y_1) > \lambda$, we must have $\bar{\Psi}'(\tilde{Y}) = \lambda$ for some $\tilde{Y} \leq \bar{Y}$; if not, $\frac{\partial}{\partial Y} \left[\frac{\lambda - \bar{\Psi}'(Y)}{\xi(Y)} \right] > 0$ would be violated (close to \bar{Y}). Again we have to distinguish two cases:

If $\tilde{Y} < \bar{Y}$ or $\tilde{Y} = \bar{Y} < Y_2$, we would have $T_{part}(\tilde{Y} + \epsilon) < 0$, $\bar{\Psi}'(\tilde{Y} + \epsilon) < \lambda$ and $T'(\tilde{Y} + \epsilon) < 0$, which again cannot hold as we show in the proof of Proposition 3.

If $\tilde{Y} = \bar{Y} = Y_2$, we would have $\xi(\tilde{Y}) = 0$ and $\bar{\Psi}'(\tilde{Y}) = \lambda$. But then, since $\bar{\Psi}'(Y) < \lambda \forall Y > \bar{Y}$, we would have $\mathcal{A}(Y_2) > 0$, see (8).

²³If $Y_1 = Y_{min}$, we may have $T'(Y_{min}) \neq 0$, but nevertheless $\mathcal{A}(Y_{min}) = 0$, see (10). The same applies for Y_{max} , see (10) again.

Case 3: In this case we have

$$\mathcal{A}'(Y) = (\bar{\Psi}'(Y) - \lambda)h(Y)$$

for $Y = Y_1$ and $Y = Y_2$. $\mathcal{A}'(Y_1) \leq 0$ and $\mathcal{A}'(Y_2) \geq 0$ would imply $\bar{\Psi}'(Y_2) \geq \bar{\Psi}'(Y_1)$, a contradiction to decreasing social marginal utility of income.

A.6 Formal Definition of the Income Distribution Function $H(Y)$

The distribution function of Y , $H(Y)$ is given by

$$H(Y(\beta)) = \int_{\underline{\beta}}^{\beta} \tilde{G}(\alpha_{\beta}^m(\beta')|\beta')dK(\beta') + \int_{w_0}^{\frac{Y(\beta)}{L_{min}}} [G(\alpha^u(w)|w) - G(\alpha^m(w)|w)]dF(w).$$

The corresponding density is

$$h(Y(\beta)) = \tilde{G}(\alpha_{\beta}^m(\beta)|\beta)k(\beta)\frac{d\beta}{dY} + (G(\alpha^u(w_{\beta})|w_{\beta}) - G(\alpha^m(w_{\beta})|w_{\beta}))\frac{f(w_{\beta})}{L_{min}},$$

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

References

- BACH, S., G. CORNEO, AND V. STEINER (2012): "Optimal top marginal tax rates under income splitting for couples," *European Economic Review*, 56(6), 1055 – 1069.
- BARGAIN, O., M. DOLLS, D. NEUMANN, A. PEICHL, AND S. SIEGLOCH (2011): "Tax-benefit systems in Europe and the US: between equity and efficiency," *CESifo Working Paper 3534*.
- BLUNDELL, R., M. BREWER, P. HAAN, AND A. SHEPHARD (2009): "Optimal Income Taxation of Lone Mothers: An Empirical Comparison Of The UK and Germany," *The Economic Journal*, 119(535), 101–121.
- BLUNDELL, R., AND A. SHEPHARD (2012): "Employment, Hours of Work and the Optimal Taxation of Low-Income Families," *The Review of Economic Studies*, 79(2), 481–510.
- 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.
- 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.
- CHRISTIANSEN, V. (2012): "Optimal Participation Taxes," *CESifo Working Paper No. 3941*.
- CREMER, H., J.-M. LOZACHMEUR, AND P. PESTIEAU (2012): "Income taxation of couples and the tax unit choice," *Journal of Population Economics*, 25(2), 763–778.
- DA COSTA, C., AND T. PEREIRA (2011): "Sacrifice and Efficiency of the Income Tax Schedule," *Escola de Pós-Graduação em Economia da FGV*.
- DIAMOND, P. (1980): "Income Taxation with Fixed Hours of Work," *Journal of Public Economics*, 13(1), 101–110.
- 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.
- FAN, J., AND I. GIJBELS (2003): *Local Polynomial Modelling and Its Applications*. CRC Press, Boca Raton.

- HECKMAN, J. J. (1993): "What Has Been Learned about Labor Supply in the Past Twenty Years?," *American Economic Review*, 83(2), 116–121.
- 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.
- IMMERVOLL, H., H. KLEVEN, C. KREINER, AND N. VERDELIN (2011): "Optimal Tax and Transfer Programs for Couples with Extensive Labor Supply Responses," *Journal of Public Economics*, 95(11), 1485–1500.
- JACQUET, L., E. LEHMANN, AND B. VAN DER LINDEN (2012): "Signing distortions in optimal tax and other adverse selection problems with random participation," *CESifo Working Paper* 3766.
- KANBUR, R., M. KEEN, AND M. TUOMALA (1994): "Optimal non-linear income taxation for the alleviation of income-poverty," *European Economic Review*, 38(8), 1613 – 1632.
- KAPLOW, L. (2007): "Optimal Income Transfers," *International Tax and Public Finance*, 14(3), 295–325.
- KLEVEN, H., C. KREINER, AND E. SAEZ (2009): "The Optimal Income Taxation of Couples," *Econometrica*, 77(2), 537–560.
- KLEVEN, H. J., AND C. T. KREINER (2006): "The marginal cost of public funds: Hours of work versus labor force participation," *Journal of Public Economics*, 90(10-11), 1955–1973.
- LAROQUE, G. (2005): "Income Maintenance and Labor Force Participation," *Econometrica*, 73(2), 341–376.
- MEGHIR, C., AND D. PHILLIPS (2010): "Labour Supply and Taxes," in *Dimensions of Tax Design: The Mirrlees Review*, ed. by Institute for Fiscal Studies. Oxford University Press.
- 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.
- MOFFITT, R. (1982): "The Tobit Model, Hours of Work and Institutional Constraints," *The Review of Economics and Statistics*, 64(3), 510–515.
- OECD (2011): "Pensions at a Glance 2011: Retirement-Income Systems in OECD and G20 Countries," www.oecd.org/els/social/pensions/PAG.
- PIKETTY, T. (1997): "La redistribution fiscale face au chômage," *Revue Française d'Économie*, 12, 157–201.
- PIRTTILÄ, J., AND M. TUOMALA (2004): "Poverty alleviation and tax policy," *European Economic Review*, 48(5), 1075–1090.
- 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. (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. (2012): “Entrepreneurial Taxation with Endogenous Entry,” *Working Paper, Stanford University*.
- SINN, H.-W., C. HOLZNER, W. MEISTER, W. OCHEL, AND M. WERDING (2006): “Aktivierende Sozialhilfe 2006: Das Kombilohn-Modell des ifo Instituts,” *ifo Schnelldienst*, 59(2), 6–27.
- TUMMERS, M., AND I. WOITTIEZ (1991): “A simultaneous wage and labor supply model with hours restrictions,” *Journal of Human Resources*, 26(3), 393–423.
- 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.
- WAGNER, G., J. FRICK, AND J. SCHUPP (2007): “The German Socio-Economic Panel Study – Scope, Evolution and Enhancements,” *Schmoller’s Jahrbuch*, 127(4), 626–654.
- WERNING, I. (2007): “Pareto Efficient Income Taxation,” *Mimeo, MIT*.