Optimal Redistributive Taxes and Redistributive Preferences in the Netherlands*

Floris T. Zoutman[‡] Bas Jacobs[◊] Egbert L.W. Jongen[§]

October 15, 2013

Abstract

We study the optimal non-linear income and participation tax in the Netherlands using a model with strong empirical foundations, which captures both intensive and extensive tax-base responses. We find that current marginal top rates maximize tax revenue and marginal tax rates at the bottom of the income distribution are too low. Lower participation taxes for low-income earners are generally optimal. An optimal flat tax substantially decreases social welfare as it yields less efficiency, less equity, or both compared to the optimal non-linear tax system.We also calculate social welfare weights under the current tax-benefit system using the inverse optimal-tax method. The Dutch government prefers transferring resources to the non-working poor over the working poor. Social welfare weights increase until modal income implying that middle-income groups lower their tax burden at the expense of low- and high-income groups. Social welfare weights decrease after modal income and become slightly negative for top-income earners. All our findings imply that tax rates in the top bracket should not be increased and net incomes for the working-poor should be substantially raised through EITC-type programs.

Key words: optimal non-linear taxation, intensive and extensive labor supply, Dutch tax and benefit system.
JEL-codes: H21, H23, H24

*We like to thank Miriam Gielen from CPB Netherlands Bureau for Economic Policy Analysis for their assistance with calculating the marginal tax rates used in this paper. We also thank seminar and congress participants at Tinbergen Institute Rotterdam, IIPF Michigan, CESifo Munich, Helsinki Center for Economic Research and UCL Louvain-la-Neuve for useful comments and suggestions. In addition we would like to express our gratitude to Olivier Bargain, Laurence Jacquet, Etienne Lehmann, Bertholt Leeftink, Erzo Luttmer, Luca Micheletto, Andreas Peichl, Pierre Pestieau, Rick van der Ploeg, Paul Tang, Coen Teulings, Emmanuel Saez, Laurent Simula, Trine Vattø and Danny Yagan for their comments, suggestions or assistance with this paper. All remaining errors are our own.

‡ Erasmus University Rotterdam and Tinbergen Institute. E-mail: zoutman@ese.eur.nl.

◊ Erasmus University Rotterdam, Tinbergen Institute, Netspar, and CESifo. Corresponding author: Erasmus School of Economics, Erasmus University Rotterdam, PO box 1738, 3000 DR Rotterdam, The Netherlands. Phone: +31–10–4081452/1441. E-mail: bjacobs@ese.eur.nl. Internet: people.few.eur.nl/bjacobs.

§ CPB Netherlands Bureau for Economic Policy Analysis. E-mail: E.L.W.Jongen@cpb.nl.

1 Introduction

What is the optimal structure of tax rates in a redistributive income tax system? This is a simple and policy-relevant question, but the answer is quite difficult. In his Nobel-prize winning article James Mirrlees (1971) wrote: "One would expect that in any economic system where equality is valued, progressive income taxation would be an important instrument of policy. [...] but there is virtually no relevant economic theory to appeal to, despite the importance of the tax (p.175)". Mirrlees has solved the theoretical problem of how to determine the optimal non-linear income tax and he concluded: "The problem seems to be a rather difficult one, even in the simplest cases "(p.175). Due to its analytical complexity, Mirrlees, and subsequently many authors, resorted to numerical simulations of the model to shed light on the shape of the optimal tax schedule.

This study presents simulations of the optimal non-linear income tax for the Netherlands. To the best of our knowledge it is the first ever to do so.[1] Besides the intensive margin of labor supply, we will also allow for an extensive margin as in Jacquet et al. (2013), which combines Mirrlees (1971) with Diamond (1980). Recent advances in the empirical labor-supply literature point to the importance of the extensive margin for labor-supply decisions, see Blundell et al. (2011). This is also true for the Netherlands as Mastrogiacomo et al. (2013) have shown. Allowing for the participation margin has important implications for the setting of optimal income tax rates and the design of in-work tax credits, as stressed by Saez (2002). Finally, by using the inverse optimal-tax approach developed in Bourguignon and Spadaro (2012), we will derive implicit social welfare weights of the actual tax-benefit system had it been optimized. This allows us to detect inconsistencies in the current tax-benefit system and helps us finding welfare-improving tax reforms.

Our study sheds light on many policy questions that are currently fiercely debated in the Netherlands. For example, Should the Netherlands introduce a flat tax as proposed by Bovenberg and Teulings (2005) and Wetenschappelijk Instituut voor het CDA (2009)? Should the tax rate at the top of the income distribution be raised as suggested by some political parties in the 2012 election platforms, *e.g.* by the Labor Party (PvdA), the Green Left and the Socialist Party?[2] Should social-assistance benefits be reduced, as proposed by *e.g.* the Christian Democratic Party (CDA), and the conservative-liberal party VVD[3], or increased, as proposed by the Socialist Party? Should the Netherlands increase the earned-income tax credit, as proposed by *e.g.* the conservative-liberal party VVD, the Christian Union (CU), the social-liberal party D66, the Labor Party, the Green Left and the Socialist Party? Furthermore, how should rent assistance, health-care subsidies, and subsidies to families with dependent children be phased out with income?[4]

[1] Jacobs (2008) calculates an optimal top marginal tax rate of 50% using empirical estimates by Atkinson and Salverda (2005) for the Pareto parameter for the top tail of the Dutch income distribution. However, he does not analyze the full optimal income tax schedule for the Netherlands.

[2] The Labor Party and Green Left propose to raise the income tax rate from 52 to 60% for taxable income beyond 150 thousand euro. The Socialist Party wants to raise the tax rate to 65%, see CPB and PBL (2012).

[3] In contrast to American-English use of the term 'liberal', this word has no 'left-wing' connotation in the Netherlands (and in many other European countries). To emphasize this distinction we use the adjective 'conservative'. In addition, there is a more left-leaning liberal party in the Netherlands, D66, which we label as 'social-liberal'.

[4] For a complete overview of the proposals for income dependent taxes and subsidies by the political parties in the

The relevance of this paper extends beyond the Dutch case. The Netherlands is a country with a large amount of redistribution via the welfare state, which resembles other European and Scandinavian countries in many respects.[5] Moreover, many policy questions in the Netherlands are discussed elsewhere. Most of the literature, however, has mainly focused on the Anglo-Saxon countries, in particular the US and the UK (Diamond, 1998; Jacquet et al., 2013; Mirrlees, 1971; Saez, 2001; Tuomala, 1984). Our analysis reveals that there are some notable differences in the optimal non-linear tax schedules in comparison to those for the US and the UK.

Our main findings are the following. For the model with only an intensive labor supply margin, our calculations reveal that the current tax system is highly suboptimal. The optimal marginal tax schedule is U-shaped with decreasing marginal tax rates up to median income. However, marginal taxes are roughly increasing over the entire income distribution in the current tax system. The optimal top rate is almost equal to the current top rate for the most redistributive (Rawlsian) social preferences. For any social welfare function attaching a positive welfare weight to the top-income earners, the top rate is set beyond the top of the Laffer curve. Raising the top-rate to 55 or 60% lowers social welfare by both reducing redistribution and economic efficiency. In addition, we find that current marginal tax rates for the low-income groups and the average tax rate for middle-income groups are too low compared to the optimal non-linear tax schedule.

However, when the extensive margin of labor supply is included in the analysis, we find that optimal marginal tax rates are substantially lowered, especially for the low-income earners. Intuitively, by raising participation tax rates, marginal tax rates discourage participation. Because tax revenue declines when participation falls, marginal taxes should optimally be lower. Participation responses are especially important for the lower end of the earnings distribution. For high-income earners the participation responses to income taxation are relatively weak, since not many high-income earners will stop working when the marginal taxes slightly increase. With both intensive and extensive margins, the actual tax schedule is much closer to the optimum than with only an intensive margin. Nevertheless, the optimal tax schedules become more U-shaped when the participation margin is included, since marginal tax rates are lowered especially in the bottom half of the earnings distribution. In addition, compared to the current tax system, optimal marginal tax rates at the bottom are still higher under both Rawlsian and utilitarian social preferences. As regards the top rate, no real changes are found as optimal top rates are not sensitive to the participation margin.

Our findings suggest that the current political system is either not able to redistribute income in the most efficient way or is not maximizing a standard social welfare function exhibiting declining social welfare weights. Policies to lower average taxes for the working poor, for example by raising the earned-income tax credit (EITC), have the potential to raise social welfare. The current government does not redistribute sufficient income to the 'working poor' in comparison with

2012 elections, see CPB and PBL (2012).

[5] A principal-component analysis places the Netherlands in the group with Scandinavian countries, see Dekker and Ederveen (2003). Additionally, Bargain et al. (2013a) demonstrate that inequality aversion in the Netherlands ranks the fourth highest in Europe.

the 'non-working poor'. Whether participation is ultimately taxed or subsidized on a net basis critically depends on the social preferences for income redistribution. Intuitively, in-work subsidies redistribute resources from non-employed to employed workers, and this raises inequality between employed and non-employed workers. Only when the government has relatively weak redistributive social preferences, net participation subsidies are optimal.

The optimal tax system is heavily non-linear, thereby discarding the proposals for a flat tax. Indeed, our calculations suggest that the optimal flat tax always reduces efficiency, equity or both in comparison with the optimal non-linear tax. For the model with only an intensive margin the welfare losses of an optimal flat tax compared to the optimal non-linear tax are 0.4% of GDP for utilitarian social objectives, and increasing until 9% of GDP for Rawlsian social objectives. These findings reveal that the flat tax is a particularly tight strait jacket when social preferences are more redistributive. Intuitively, the flat tax employs no information on individual earnings and income redistribution cannot be effectively targeted to the individuals with the lowest incomes. The flat tax requires much higher marginal tax rates to obtain a given amount of income redistribution. Therefore, the equity-efficiency trade-off worsens and Okun (1975)'s 'leaking bucket' becomes a sieve.

Finally, by computing the social welfare weights implied by the current tax-benefit system, we indeed find that social welfare weights do *not* monotonically decline. Instead, welfare weights first increase and then decrease, become slightly negative for the top-income earners, and are zero in the limit. Social welfare weights discontinuously drop with about one third when individuals earning no income start participating in the labor market. Hence, the current political system does not seem to maximize a standard social welfare function. Instead, the current government redistributes resources away from the working poor towards the middle-income groups. Moreover, it soaks the top-income earners as much as it can, and it even penalizes them by setting too high marginal taxes. Finally, the current political system strongly prefers transferring resources to the non-working rather than to the working poor. This could be taken as evidence that political-economy considerations matter a lot in shaping actual tax schedules. Moreover, the non-working poor are apparently seen as much more deserving of income support than the working poor, for reasons that remain unclear to us.

This paper is structured as follows. Section 2 provides a review of the literature and discuss our paper's contributions. Section 3 derives the model used in our simulations. In Section 4 we introduce the data, compute marginal and participation tax rates, estimate the Pareto parameter for the Dutch income distribution, calibrate the utility function and estimate the distributions of skills and participation costs. In Section 5 we compute the optimal non-linear income tax for the Netherlands for models with only an intensive labor supply margin and with both an intensive and an extensive labor supply margin. In Section 6 we calculate the social welfare weights that are implied by the current Dutch tax and benefit system. In Section 7 we discuss the limitations of our analysis and provides the policy conclusions. Various Appendices contain less essential technical derivations.

2 Earlier literature

Our study aims to contribute to the scientific literature on optimal taxation by presenting advanced simulations of the Mirrlees (1971) model, which is extended with an extensive labor supply margin and income effects along the lines of Jacquet et al. (2013). From his own simulations, Mirrlees concluded: "[P]erhaps the most striking feature of the results is the closeness to linearity of the tax schedules (p.206)". However, subsequent research has shown that this conclusion was premature, the result depended heavily on functional form assumptions for the utility function (Cobb-Douglas) and the income distribution (log-normal). Tuomala (1984) uses a different utility function, which allows for a more realistic elasticity of taxable income and finds declining optimal marginal tax rates with income. Diamond (1998) finds a U-shape for optimal marginal tax rates, using a Pareto distribution for the top incomes. The log-normal distribution implies an upper tail for the income distribution that is too 'thin', resulting in optimal top tax rates that are too low. Saez (2001) also finds a U-shape for optimal marginal tax rates, extending the analysis of Diamond (1998) by *e.g.* allowing for income effects. Our paper contributes in a number of important ways to the existing literature on optimal-tax simulations.

A substantial part of this paper is devoted to estimating the joint distribution of ability and participation costs. In contrast to earlier papers that assumed synthetic skill distributions (Mirrlees, 1971; Tuomala, 1984), we estimate the skill distribution using the structural method pioneered by Saez (2001) and Bourguignon and Spadaro (2000). In particular, we assume that individuals maximize a particular utility function, which is defined over consumption and labor supply, subject to a non-linear budget constraint. We carefully reconstruct the individuals' budget constraints, taking into account income-dependent transfers, numerous tax credits, indirect taxes, and welfare benefits. The first-order conditions and the household budget constraints enable us to retrieve the non-observable skill level for each household. In doing so, we will rely on advanced, recent Dutch estimates for the elasticity of taxable income (Jongen and Stoel, 2013), the intensive labor-supply elasticity and the extensive labor-supply elasticity (Mastrogiacomo et al., 2013). Additionally, this study very precisely estimates a Pareto distribution for the top of the Dutch skill distribution. The Pareto parameter is estimated to be around 3.35, which is among the highest found in the literature, suggesting that it is lonely at the top in the Netherlands, see the overviews in Heady (2010) and Atkinson et al. (2011).

We confirm the findings of Saez (2001) that the optimal tax schedules feature a U-shape. Marginal tax rates at lowest income groups are very high, in the order of 70-80%. Marginal tax rates decline towards the middle-income groups, increase again after middle-income groups, and converge to a constant of about 50% for the top-income earners. The increase in the marginal tax rates after modal income is, however, much more limited than in Saez (2001), which is due to the very thin top tail of the earnings distribution in the Netherlands. Indeed, for the Rawlsian social welfare function, we find only a tiny increase in marginal tax rates, in contrast to the US. For the same reason, optimal top rates in the Netherlands are much lower than those for the US (Saez, 2001) or the UK (Brewer et al., 2010).

We further contribute to the analysis of optimal income tax simulations with both intensive and extensive labor-supply margins. In contrast to Jacquet et al. (2013), we estimate the distribution of participation costs using the first-order conditions for labor-market participation, which are supplemented with data on participation rates and participation taxes under the current tax-benefit system. Like Jacquet et al. (2013), we assume that idiosyncratic participation costs/benefits are separable from leisure and consumption and that the distribution of participation costs is normal. We estimate the parameters of the distribution of participation costs such that participation rates match with skill-specific employment rates. Participation rates in the Netherlands run from 0.51 for low-educated individuals to 0.85 for high-educated individuals. In this way, we are able to fully determine the non-observed joint distribution of ability and participation costs.

Our findings contrast sharply with the baseline simulation of Jacquet et al. (2013). When including the extensive margin, they find that the optimal non-linear tax schedule shifts downwards across all income levels compared to the optimal schedule without the extensive margin. We find that marginal taxes mainly fall in the bottom part of the skill distribution. We believe that this is due to the different specifications of participation costs employed in both studies. Jacquet et al. (2013) do not estimate the distribution of participation costs, but assume that i) participation rates have some non-linear relation with ability, and ii) participation elasticities have a linear relation with ability. These relationships are not empirically estimated. As a result, participation rates are very similar for low- and high-skill types, rising from 0.7 for the lowest to 0.8 for the highest skill types. Consequently, the optimal tax schedule shifts down in Jacquet et al. (2013) for all income levels. Our specification of participation costs implies that net participation costs are much lower for higher-skilled individuals, since they have much higher participation rates. The participation elasticities in our model (about 0.25) are calibrated on empirical values. We believe that this explains why optimal tax schedules mainly shifts down for low-skilled individuals, and not for the high-skilled individuals, when including the participation margin in the Mirrlees model.[6]

The inverse optimal-tax problem is analyzed in Bourguignon and Spadaro (2000), Bourguignon and Spadaro (2012), Blundell et al. (2009), Bargain and Keane (2010) and Bargain et al. (2011). Inconsistencies of actual tax systems compared to optimal tax systems derived from standard social welfare functions are found in most of the literature. Bourguignon and Spadaro (2012) find monotonically declining welfare weights for singles in France when considering only the intensive margin of labor supply. However, welfare weights turn negative for high-income earners. In addition, when they include the extensive margin, welfare weights are not monotonically declining. In particular the working poor receive a negative social welfare weight, while both the unemployed and the middle-income earners obtain a positive weight. They conclude that policy makers underestimate the tax distortions on the extensive margin. Blundell et al. (2009) consider single mothers in Germany and the UK analyzing both intensive and extensive responses. They also find that weights are not monotonically decreasing with a dip in the welfare weight for the working poor.

[6] A sensitivity analysis of Jacquet et al. (2013) indeed confirms that when participation elasticities fall with income, the effect on the optimal tax schedule will be more pronounced at the lower income levels.

Similar results are found in Bargain and Keane (2010) and Bargain et al. (2011), where the authors consider singles in respectively Ireland, and 17 European countries and the US.

We also detect numerous inconsistencies in the current tax-benefit system. In particular, our analysis demonstrates that social welfare weights are not continuously declining, but increasing up to modal income, which could be explained by political-economy considerations. After modal income, they decline as expected, but even turn negative for the very high income earners, indicating that tax rates are set beyond the top of the Laffer curve. Moreover, welfare weights discontinuously drop for workers moving from non-participation to participation, suggesting that income is taxed too highly for the low-income earners.

3 Theory

In this section we first introduce the Mirrlees (1971) model of optimal labor income taxation with intensive labor-supply responses. Then, we analyze the model of Jacquet et al. (2013) with both intensive and extensive labor-supply responses. Finally, this section derives how to determine the social welfare weights implicit in our current tax-benefit system using the inverse optimal-tax method.

3.1 Intensive margin

3.1.1 Individuals

We follow the optimal-tax literature by supposing that heterogeneity in individual types derives from their exogenous ability to earn income (and their participation costs when the participation decision is included, see below). The fundamental insight of Vickrey (1947) and Mirrlees (1971) is that earnings ability is not observable by the government. Due to the non-observability of ability the government needs to resort to distortionary tax instruments, most importantly taxes on labor income, to redistribute income. Taxing labor income is distortionary because it not only taxes the return to ability, but also the fruits of labor effort. Hence, income redistribution leads to the well-known trade-off between equity and efficiency.[7]

Ability is distributed according to probability density function $f(n)$ and corresponding cumulative distribution function $F(n)$, with support $\mathcal{N} \equiv [\underline{n}, \overline{n})$. The upper bound $\overline{n}$ can be infinite. n denotes the number of efficiency units of labor. We follow Mirrlees (1971) by assuming perfect substitution between skill types on the labor market. Hence, by normalizing the wage rate per efficiency unit of labor to unity, we can associate n with the wage rate per hour worked of individual n. Gross labor earnings of an individual with ability n are given by $z_n \equiv nl_n$ where l_n denotes the normalized labor supply of an individual with ability n.

[7] As long as ability differences are the only source of heterogeneity, and preferences of individuals are homogeneous and weakly separable, only the non-linear income tax will be employed for income redistribution Atkinson and Stiglitz (1976).

If we assume that all net income is consumed, the individuals' budget constraint is given by:

$$c_n = z_n - T(z_n), \quad \forall n, \tag{1}$$

where c_n denotes consumption and $T(z_n)$ the tax schedule as a function of gross labor income. $T'(z_n) \equiv \mathrm{d}T(z_n)/\mathrm{d}z_n$ is the marginal tax rate. All individuals have identical preferences over consumption c_n and labor l_n, which are represented by a separable, continuous and twice-continuously differentiable utility function:[8]

$$u_n \equiv v\left(c_n\right) - h\left(l_n\right), v', h' > 0, \quad v'', -h'' \leq 0, \quad \forall n, \tag{2}$$

where $v(\cdot)$ is a concave function representing the utility of consumption, and $h(\cdot)$ is a convex function representing the disutility of labor effort. By substituting the budget constraint (1), into the utility function (2), the maximization problem of the individual can be stated as:

$$\max_{z_n} v\left(z_n - T(z_n)\right) - h\left(\frac{z_n}{n}\right), \quad \forall n, \tag{3}$$

The first-order condition (FOC) of this problem is:

$$\left(1 - T'(z_n)\right) v'(c_n) = \frac{h'(l_n)}{n}, \quad \forall n. \tag{4}$$

The marginal benefits of earning an additional euro on the labor market, as represented by the left-hand side, are equated to the marginal utility cost of labor required to earn the additional euro of income, as represented by the right-hand side. As can be seen, the marginal benefits of work are decreasing in the tax rate.

The allocation is said to be incentive compatible if the following first-order incentive-compatibility constraint holds:

$$\frac{\mathrm{d}u_n}{\mathrm{d}n} = \frac{l_n h'(l_n)}{n}, \quad \forall n. \tag{5}$$

This condition can be derived from totally differentiating utility with respect to ability and using the first-order condition for labor supply.

The incentive-compatibility constraint (5) is a necessary constraint. However, each incentive-compatible allocation must also respect second-order sufficiency conditions for utility maximization (Mirrlees, 1976). This is the case if, in addition, the Spence-Mirrlees and monotonicity constraints are satisfied:

$$\frac{\mathrm{d}\left(\frac{h'(l_n)}{nv'(c_n)}\right)}{\mathrm{d}n} \leq 0, \quad \forall n, \tag{6}$$

$$\frac{\mathrm{d}z_n}{\mathrm{d}n} > 0, \quad \forall n. \tag{7}$$

[8] The assumption of separability in the utility function is made in all simulation studies in the literature. Numerically, it is very difficult, if not impossible, to simulate the optimal tax schedule if the utility function is non-separable. See e.g. Mirrlees (1971) and Saez (2001).

These conditions imply that the utility function features the single-crossing property. Hence, at each bundle of gross and net income, individuals with a higher ability have incentives to self-select into the bundles with higher net and gross income. The Spence-Mirrlees condition is satisfied by most utility functions used in the literature, including the ones that are used in our simulations. The second condition states that income should increase monotonically with ability.[9] Hence, the second condition ensures that self-selection of higher ability types into higher consumption-earnings bundles will also occur. From the monotonicity condition we can derive that it is never optimal to have higher marginal tax rates than 100%, otherwise the monotonicity condition would be violated, since it implies that $\frac{\mathrm{d}c_n}{\mathrm{d}n} > 0$, see Mirrlees (1976).

In our simulations we will use the first-order approach using (5), assuming that the second-order conditions will be satisfied. After having derived the optimal allocation, we will check ex post whether the sufficiency conditions (6) and (7), are indeed met, which is always the case.

3.1.2 Government

The objective of the government is to maximize social welfare. Social welfare is assumed to be described by a Samuelson-Bergson social welfare function, which is a concave sum of individual utilities:

$$\int_{\mathcal{N}} W(u_n)f(n)\mathrm{d}n, \quad W' > 0, \quad W'' \leq 0, \tag{8}$$

Redistribution is socially desirable if either the social marginal value of utility (W') or the private marginal value of income (u_c) are decreasing, i.e., $W'' < 0$ or $u_{cc} < 0$.[10] The government has to respect the economy's resource constraint:

$$\int_{\mathcal{N}} (z_n - c_n)\, f(n) \geq R, \tag{9}$$

where R denotes exogenous government expenditure. As long as the economy's resource constraint and the household budget constraints are met, also the government budget constraint is satisfied by Walras' law.

3.1.3 Optimal income taxation

The optimal allocation is found by maximizing the social welfare function, (8), subject to the resource constraint, (9), and the incentive compatibility constraint, (5). The Appendix derives that the optimal schedule of marginal income taxes then satisfies the following ABC-formula:

$$\frac{T'(z_n)}{1 - T'(z_n)} = A_n B_n C_n, \quad \forall n, \tag{10}$$

[9] See Ebert (1992) for a detailed discussion on this issue.

[10] In the extreme case, where both $W'' = 0$ and $u_{cc} = 0$, the optimal tax problem becomes trivial, as there is no social desire for redistribution and the government will finance all of its spending through non-distortinary lump-sum taxes.

where:

$$A_n \equiv \frac{1}{\varepsilon_n^c}, \quad \varepsilon_n^c \equiv -\frac{\partial l_n}{\partial T'(z_n)} \frac{1-T'(z_n)}{l_n}, \quad \forall n, \tag{11}$$

$$B_n \equiv \frac{v'(c_n) \int_n^{\overline{n}} \frac{1-g_m}{v'(c_m)} f(m)\, \mathrm{d}m}{1-F(n)}, \quad g_n \equiv \frac{W'(u_n) v'(c_n)}{\lambda} \quad \forall n, \tag{12}$$

$$C_n \equiv (1+\varepsilon_n^u) \frac{(1-F(n))}{f(n)n}, \quad \varepsilon_n^u \equiv \frac{\partial l_n}{\partial n} \frac{n}{l_n}, \quad \forall n, \tag{13}$$

where ε_n^c is the compensated *tax* elasticity of labor supply, ε_n^u is the uncompensated wage elasticity of labor supply, and g_n is the social marginal value (in monetary terms) of providing individual n a unit of resources. We shall refer to g_n as the social welfare weight of individual n.

At each point of the income distribution, marginal equity gains and efficiency losses of the marginal tax rate are equalized. Intuitively, the function of the marginal tax rate at an income level z_n is to raise tax revenue from all individuals above z_n. The marginal tax rate at z_n redistributes resources from individuals above z_n to the government. In turn, the government can use this revenue to raise the uniform transfer $-T(0)$ in the tax system. A higher marginal tax rate at z_n thus increases the average tax burden above z_n and lowers the average tax burden on individuals below z_n.

A_n represents the efficiency costs of having a marginal tax at income level z_n. If the marginal tax rate at income level z_n is increased, individuals with income z_n have an incentive to decrease their labor supply. This behavioral response is captured by the compensated elasticity of labor supply with respect to the tax rate ε_n^c.

B_n represents the average redistributional gain of having a marginal tax at income z_n. B_n is equal to the revenue of a euro increase in taxes on individuals above z_n, minus the monetized value of the welfare loss g_n due to extracting an additional euro revenue from these individuals. The difference is represented by the term $1-g_n$. B_n averages this difference over all individuals with an income above z_n.

Term C_n gives weights to terms A_n and B_n via the distribution of earnings ability. The optimal tax rate is determined by the number of individuals paying the marginal tax rate $(1-F(n))$ and the number of individuals whose labor supply choice is distorted $(nf(n))$. The more individuals above income level z_n, the larger the redistributive gains of a higher marginal tax. The more individuals at skill level n, or the larger their wage rates, the larger is the tax base, and, therefore, the larger are the efficiency losses of a higher marginal tax rate.

If we would express the optimal-tax formula in terms of earnings densities, rather than the densities of the ability distribution, the C_n-term would collapse to $C_n = \frac{1-\tilde{F}(z_n)}{\tilde{f}(z_n)z_n}$, where $\tilde{F}(z_n) \equiv F(n)$ is the cumulative earnings distribution, $\tilde{f}(z_n)$ is the earnings density at z_n, and $z_n\tilde{f}(z_n) = (1+\varepsilon_n^u)nf(n)$, see Saez (2001). Hence, the C_n term is entirely determined by the shape of the empirical earnings distribution $\tilde{f}(z_n)$.

There is no closed-form solution for the optimal tax rate. Nevertheless, a few properties of optimal tax schedules can be established analytically. We already derived that the optimal marginal tax rate is never above 100% at any income level. In addition, the marginal tax rate is never below 0%, see the ABC-formula. Indeed, a marginal tax rate below 0% redistributes income in the wrong direction, and thereby lowers social welfare (Seade, 1982).[11] Sadka (1976) and Seade (1977) show that the marginal tax rate at the bottom and the top of the skill distribution should be equal to zero if the skill distribution has a finite top and all individuals provide positive work effort. Intuitively, there are no redistributional gains and only distortions associated with marginal taxes at the endpoints, so that marginal tax rates are zero.[12] As is shown in Diamond (1998), the result of the zero tax rate at the top does not apply if (the top of) the skill distribution is Pareto distributed. Later, we will demonstrate that this is also the case for the Netherlands. Similarly, the zero marginal tax at the bottom is positive, and generally very large, when there is an atom of non-working individuals, which is observed in the real world and assumed in most simulations. No further analytical results can be obtained. Therefore, many authors have resorted to simulations of the optimal non-linear tax schedule. That is what we will do in the remainder of this paper, after we introduced the extensive margin.

3.2 Extensive margin

In the Mirrlees model individuals can only adjust their labor supply on the intensive margin. They can decide to work more or less, but they cannot decide to enter or exit the labor market entirely. In contrast, Diamond (1980) derives the optimal tax schedule where individuals can only adjust their labor supply along the extensive margin, but not on the intensive margin. Saez (2002) and Jacquet et al. (2013) combine the Mirrlees-model with the Diamond-model to analyze the optimal non-linear income tax and the optimal participation tax. In this paper, we will follow the analysis of Jacquet et al. (2013) to find the optimal tax schedule with both intensive and extensive labor supply responses for the Netherlands.

3.2.1 Individuals

The extensive margin is introduced through a random participation model. Each individual has an individual-specific participation utility cost φ of entering the labor market, which reflects the individuals' outside options such as household production or income from the black labor market. We also allow some individuals to have a negative disutility of participation. This could be related to a social stigma of being unemployed. We assume that the disutility of participation is unobservable to the government. φ follows a probability density function conditional on ability n given by $k(\varphi|n)$. The corresponding cumulative distribution function is $K(\varphi|n)$. The support, also potentially conditional on n, is given by $[\underline{\varphi}^n, \overline{\varphi}^n]$.

[11] Note that a zero marginal tax rate at the bottom or the top does not imply that there is no redistribution on average. The amount of redistribution is determined by the average tax individuals pay over all their income, not by the marginal tax they pay over their last euro of income.

[12] Especially, the zero top rate has attracted a lot of attention. However, its practical applicability is limited, because it is a very local result, see Tuomala (1984).

Individuals can decide not to participate and receive unemployment benefits b. We assume that the government can verify the employment status of an individual, and hence, condition non-employment benefits on it. The utility of a non-employed worker is equal to $v(b)$. b can be different from the transfer $-T(0)$ implied by the tax schedule. The utility of an employed individual with ability n and discrete participation cost φ is given by:

$$U_n \equiv v(c_n) - h(l_n) - \varphi, \quad \forall n. \tag{14}$$

An individual decides to participate in the labor market if the maximum utility derived of participation is at least as large as the utility of the unemployment benefits:

$$u_n - \varphi \geq v(b), \quad \forall n. \tag{15}$$

where $u_n \equiv v(c_n) - h(l_n)$. The individual will participate his/her utility cost of working is sufficiently low, or if his/her ability to earn income is sufficiently high.

The participation tax is the net extra amount of tax an individual pays if he/she decides to participate and earns gross income level z_n. The participation tax consists of two components. First, when working the individual is subject to the tax schedule $T(z_n)$, and, second, the individual loses his/her benefits b. The total participation tax is therefore $T(z_n) + b$. A higher participation tax naturally discourages participation. We do not constrain the participation tax to be positive, and, therefore, the government is also allowed to give a participation subsidy when this raises social welfare, i.e., $T(z) + b < 0$ would imply an in-work tax credit.

The incentive-compatibility constraint (5) is unaltered by the introduction of the participation costs. Intuitively, a worker with ability n has to incur participation cost φ irrespective of whether the worker self-selects in the consumption-income bundle for type n or decides to mimic a worker of type m to obtain the consumption-income bundle intended for type m.

3.2.2 Government

The government's objective is a weighted sum of the utility of non-employed and the utility of employed workers:

$$\int_{\mathcal{N}} \left(\int_{\underline{\varphi}^n}^{u_n - v(b)} W(u_n - \varphi) k(\varphi|n) \mathrm{d}\varphi f(n) + W(v(b)) \left(f(n) - \tilde{k}(n)\right) \right) \mathrm{d}n. \tag{16}$$

The bounds of the inner integral are given by equation (15). Therefore, all individuals with φ in $[\underline{\varphi}^n, u_n - v(b)]$, given by $K(u_n - v(b)|n)$, participate, and all individuals with φ in $(u_n - v(b), \overline{\varphi}^n]$ do not participate. There are $f(n)$ individuals with ability n, and, hence, the fraction of individuals in the population that work at skill level n is given by $\tilde{k}(n) \equiv K(u_n - v(b)|n) f(n)$. The fraction of non-employed individuals at skill level n is, therefore, $f(n) - \tilde{k}(n)$, as can be seen in the second term of equation (16).

Correspondingly, the economy's resource constraint is modified to:

$$\int_{\mathcal{N}} \Big((z_n - c_n)\tilde{k}(n) - (f(n) - \tilde{k}(n))b \Big)\, \mathrm{d}n \geq R. \tag{17}$$

First, note that at each skill level n only the working fraction of the population $\tilde{k}(n)$ produces output z_n and consumes c_n. Second, at each skill level n, the non-working population $f(n) - \tilde{k}(n)$ does not produce anything and consumes its unemployment benefits b.

3.2.3 Optimal income taxation

The adjusted ABC-formula for optimal taxation in the presence of intensive and extensive labor-supply responses is given by – see Appendix for the derivation:

$$\frac{T'(z_n)}{1 - T'(z_n)} = A_n B_n C_n, \quad \forall n, \tag{18}$$

where:

$$A_n \equiv \frac{1}{\varepsilon_n^c}, \quad \varepsilon_n^c \equiv -\frac{\partial l_n}{\partial T'(z_n)} \frac{1 - T'(z_n)}{l_n}, \quad \forall n, \tag{19}$$

$$B_n \equiv \frac{v'(c_n) \int_n^{\overline{n}} \left(\frac{1 - g_m^P}{v'(c_m)} - \kappa_m \left(b + T(z_m) \right) \right) \tilde{k}(m) \mathrm{d}m}{\tilde{K}(\overline{n}) - \tilde{K}(n)}, \tag{20}$$

$$\kappa_n \equiv \frac{K'\left(u_n - v(b)|n\right) f(n)}{\tilde{k}(n)}, \quad g_n^P \equiv \frac{\int_{\underline{\varphi}^n}^{u_n - v(b)} \frac{W'(u_n - \varphi) v'(c_n)}{\lambda} k(\varphi|n) \mathrm{d}\varphi}{K\left(u_n - b\right)}, \quad \forall n, \tag{21}$$

$$C_n \equiv (1 + \varepsilon_n^u) \frac{\tilde{K}(\overline{n}) - \tilde{K}(n)}{n\tilde{k}(n)}, \quad \forall n, \tag{22}$$

where g_n^P is the social welfare weight given to *employed* workers, and κ_n is the semi-elasticity of participation with respect to a utility increase for the employed. $\tilde{k}(n)$ is the fraction of employed with ability level n, $\tilde{K}(n)$ is the fraction of employed in the population with ability n or less, and $\tilde{K}(\overline{n})$ is the total fraction of workers participating.

Term A_n and its interpretation is unaltered by the introduction of the extensive margin. In term C_n all occurrences of the distribution of earnings ability ($f(n)$ and $1 - F(n)$) have been replaced by the distribution of employed workers ($\tilde{K}(\overline{n}) - \tilde{K}(n)$ and $\tilde{k}(n)$). Intuitively, weights to term A_n and B_n should be given on the basis of the number of employed workers, because non-employed workers do not pay the marginal tax rate.

The largest difference with the model without an extensive margin is found in term B_n. The extensive margin reduces the average revenue available for redistribution, because a higher marginal tax rate results in revenue losses by discouraging labor-force participation. Suppose the government increases the tax rate at income level z_n such that all individuals above z_n need to pay one euro extra tax. Mechanically this raises the tax revenue for all individuals with income z_n or larger by

1 euro. In addition, the government inflicts a welfare loss on all taxed individuals as represented in normalized welfare weights g_n^P. Finally, a higher marginal tax increases the total tax paid by working individuals, and, thereby, reduces the attractiveness of participation compared to non-participation. The government loses revenue as some individuals decide to exit the labor market, stop paying taxes and start collecting non-employment benefits. The decline in revenue is thus determined by the participation elasticity, κ_n, which governs the reduction in participation, and by the net participation tax $T(z_n)+b$. Term B_n is the average of the difference, $1-g_n^P-\kappa_n(T(z_n)+b)$, over all employed workers with an income level above z_n.

In addition, term B_n may change due to a second-order effect. The optimal marginal tax rates decrease due to the participation margin, which is captured by $\kappa_n(T(z_n)+b)$. Therefore, this reduces the amount of income redistribution. In turn, the decrease in redistribution might raise the value of redistribution B_n at some income levels as g_n^P decreases when individuals are taxed less. This second-order effect might actually lead to an increase in the marginal tax rate at some income levels. B_n decreases at low- and medium-income levels, because the participation elasticity κ_n is large among these income groups. However, B_n increases at high-income levels because the participation elasticity at these income levels is typically small. If g_n^P falls enough to offset the effect of $\kappa_n(T(z_n)+b)$, then B_n rises and the tax rate at high-income earners could increase. This will never occur under Rawlsian preferences, since social welfare weights for employed workers g^P are constant and equal to zero.

3.2.4 Optimal participation taxation and transfers

Above, we derive the optimal marginal tax rates. However, the government also optimizes the optimal non-employment benefits b. The latter determines the optimal participation taxes at each point in the income distribution. The optimal non-employment benefit b is set such that the following equation is satisfied – see Appendix for the derivation:

$$\int_{\mathcal{N}} \kappa_m(T(z_m)+b)\tilde{k}(m)\mathrm{d}m = \int_{\mathcal{N}} \frac{(1-g_m^P)}{v'(c_m)}\tilde{k}(m)\mathrm{d}m, \quad \forall n. \tag{23}$$

This equation implicitly defines the total, aggregate participation distortion over the entire working population. The left-hand side gives the distortion in participation of providing a higher non-employment benefit b, which is captured by the participation elasticity κ_n, times the participation tax $T(z_n)+b$, aggregated over all households. The right-hand side gives the total distributional benefits of providing higher non-employment benefits. Distributional benefits occur if $g_n^P<1$ at skill level n, while they yield distributional losses if $g_n^P>1$ at skill level n. Hence, the larger are the distributional benefits of transferring resources to the non-working part of the population, the larger will be the participation distortions.

The optimal intercept of the tax function $T(0)$ is determined implicitly by ensuring that the weighted average of the marginal social welfare weights sums to one (as we have derived before), see the Appendix:

$$\frac{(g_0-1)}{v'(b)}(1-\tilde{K}(\overline{n}))=\int_{\mathcal{N}}\frac{(1-g_m^P)}{v'(c_n)}\tilde{k}(m)\mathrm{d}m, \tag{24}$$

where $g_0 \equiv W'(v(b))v'(b)/\lambda$ denotes the marginal social welfare weight of non-employed individuals. This equation ensures that the marginal euro is valued equally by the public and private sector as all the social welfare weights g_n sum to one. Equivalently, this equation states that the marginal cost of public funds equals one at the optimal tax system. Distributional benefits of redistribution cancel against deadweight losses at the optimal tax system, see Jacobs (2013).

3.3 Inverse optimal-tax problem

In the last part of this paper, we will calculate the social welfare weights g_n^P at each skill level n using the inverse optimal-tax problem, see also Bourguignon and Spadaro (2000), Bourguignon and Spadaro (2012), Blundell et al. (2009), Bargain and Keane (2010) and Bargain et al. (2011). That is, by supposing the government has maximized social welfare by optimally designing its tax-benefit system, we can use the current tax-transfer system to back out the social welfare weights corresponding to the social welfare function.

The implied social welfare weights for employed workers can be found by solving the optimal tax formula, (18), for normalized welfare weights g_n^P. First, we rewrite equation (18) for B_n:

$$B_n=\frac{T'(z_n)}{1-T'(z_n)}\frac{1}{A_nC_n}. \tag{25}$$

Second, insert the definition for A_n, B_n, and C_n from equations (19), (20), and (22), and simplify:

$$\int_n^{\overline{n}}\left(\frac{(1-g_m^P)}{v'(c_m)}-\kappa_m(b+T(z_m))\right)\tilde{k}\,(m)\,\mathrm{d}m=\frac{T'(z_n)}{1-T'(z_n)}\frac{\varepsilon_n^c n\tilde{k}(n)}{(1+\varepsilon_n^u)\,v'(c_n)}. \tag{26}$$

Next, differentiate both sides of the equation with respect to n, and apply Leibniz' rule to the left-hand side:

$$g_n^P=1-\kappa_n(b+T(z_n))v'(c_n)+\frac{\mathrm{d}}{\mathrm{d}n}\left[\frac{T'(z_n)}{1-T'(z_n)}\frac{\varepsilon_n^c n\tilde{k}(n)}{(1+\varepsilon_n^u)\,v'(c_n)}\right]\frac{v'(c_n)}{\tilde{k}(n)}. \tag{27}$$

We cannot obtain an analytical solution for the differential on the right-hand side. Therefore, we numerically approximate the expression in our calculations.

In addition, we can derive the weight of the unemployed by solving equation (24) for g_0:

$$g_0=1+\frac{v'(b)}{(1-\tilde{K}(\overline{n}))}\int_{\mathcal{N}}\frac{(1-g_m^P)\tilde{k}(m)\mathrm{d}m}{v'(c_m)}. \tag{28}$$

Applying equations (27) and (28) to our data yield the welfare weights that are implied by the current Dutch tax-benefit system. We calculate welfare weights on the intensive margin by setting κ_n equal to zero and replacing all instances of $\tilde{k}(n)$ with $f(n)$ in equation (28).

4 Data and calibration

To compute the optimal non-linear tax-benefit system we need the following ingredients: i) the distribution of skills and participation costs/benefits, which determine the amount of income inequality and the number of non-participating individuals; ii) the utility function, which governs the behavioral impact of taxes and transfers; and iii) the social welfare function, which gives the social preferences to redistribute income. In this section, we define labor income, and determine the corresponding earnings distribution. Since there are only few observations on earnings for the top tail of the earnings distribution, we estimate the top of the income distribution with a Pareto distribution, which gives an excellent fit for top incomes. We then define marginal tax rates, and determine the distribution of marginal tax rates. We use the data on labor income and marginal tax rates together with a utility function consistent with empirical studies to retrieve the ability distribution. Together with data on participation by level of education (as a proxy for skill) we use these to calibrate the distribution of idiosyncratic participation costs/benefits. We, finally, determine the revenue requirement of the government and the transfer to non-employed.

4.1 Income distribution

Following Brewer et al. (2010), we use labor costs rather than gross wages, because the former includes all premiums paid by employers and employees.[13] Most of these premiums eventually flow back to workers in the form of deferred payments in the states of unemployment, disability or retirement. As long as employers' premiums are a constant fraction of gross wages, using either gross wages or labor costs to calculate the ability distribution only affects the mean of the skill distribution, but not its basic shape. However, in the Netherlands labor costs are not proportional to gross wages, since premiums are collected only over earnings between certain thresholds, where the thresholds differ for the different premiums.

We use the data from the Dutch Income Panel Investigation (IPO in Dutch) from 2002 to determine the earnings distribution in the Netherlands.[14,15] The data are gathered by Statistics Netherlands from individual tax returns. The sample consists of 175,876 individuals in 2002. We consider individuals aged 23 until 65. We ignore individuals that are in school or studying, because their earnings are not a good indicator of their earning ability. We also exclude all individuals with a non-positive gross labor income, because we cannot determine their earnings capacity. Our final data set consists of a sample of 94,859 individuals.

Figure 1 gives a Gaussian kernel density estimate of the income distribution up to 200,000 euro (99% of the sample in 2002). The solid line gives labor costs and the dashed line gross wages.

[13] Although labor costs are already a broad definition of individual compensation, there are still some types of compensation missing like the use of a lease car, favorable mortgage loan rates, and so on. We do not have data on these types of fringe benefits.

[14] IPO data have previously been used by Atkinson and Salverda (2005) to determine the top income share (up to 1999).

[15] We use IPO 2002, because it has been checked by several researchers, and has been cleaned of various mistakes, such as forgotten commas.

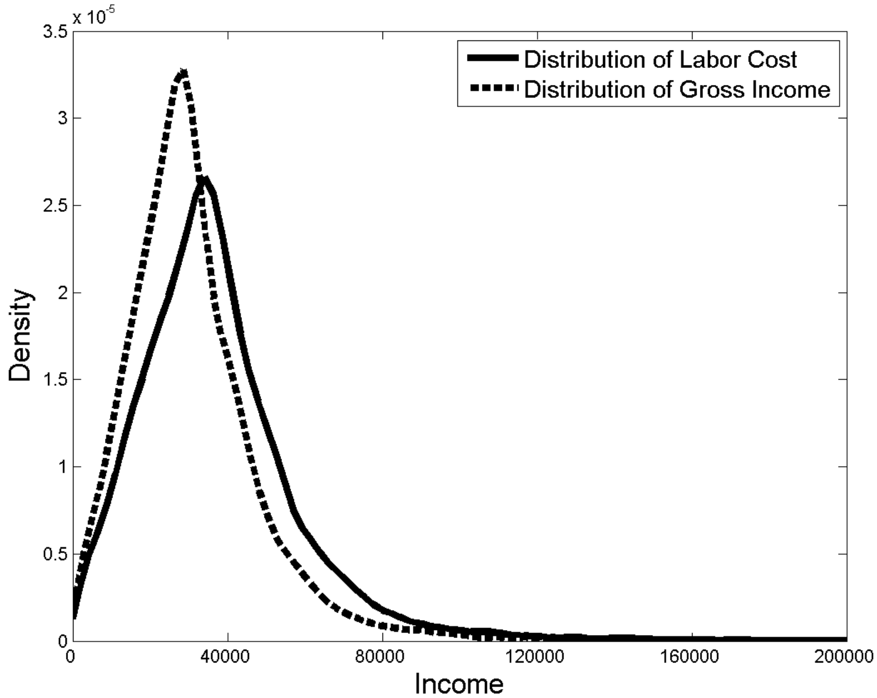

Figure 1: A kernel density estimate of income in the Netherlands in 2002

Mean labor costs are (approximately) 35,000 euro, and the median is 31,000 euro, so the earnings distribution is skewed to the right. The mode is 33,000 euro. Table 1 gives the descriptive statistics for the distribution of labor costs and gross wages. Gross wages are lower than labor costs because the latter includes employers' premiums.

4.2 Estimation of the Pareto tail

The kernel estimate is an accurate estimate of the true density of income for most income levels. However, because the sample does not include many observations in the right tail, we make a distributional assumption for this part of the distribution. Like many papers in the optimal-tax literature we assume the right tail to be Pareto distributed, see for example Saez (2001) and Jacquet et al. (2013). Also, Clementi and Gallegati (2005a) and Clementi and Gallegati (2005b) find evidence of a Pareto distributed right tail in Germany, Italy, the US and the UK. Below we will demonstrate that the Pareto distribution fits the top-income data extremely well.

Different sources of income are taxed under separate regimes in the Netherlands. This separate tax treatment could bias our estimates for the Pareto tail. Therefore, we explore various income definitions in our estimations. Labor income of workers, fictitiuous labor income of self-employed and fictitiuous labor income of director-shareholders of closely-held firms are all taxed under the progressive labor income tax ('Box-1'). Capital income of director-shareholders of closely-held companies in excess of fictitiuous labor income, retained profits, dividends and capital gains on

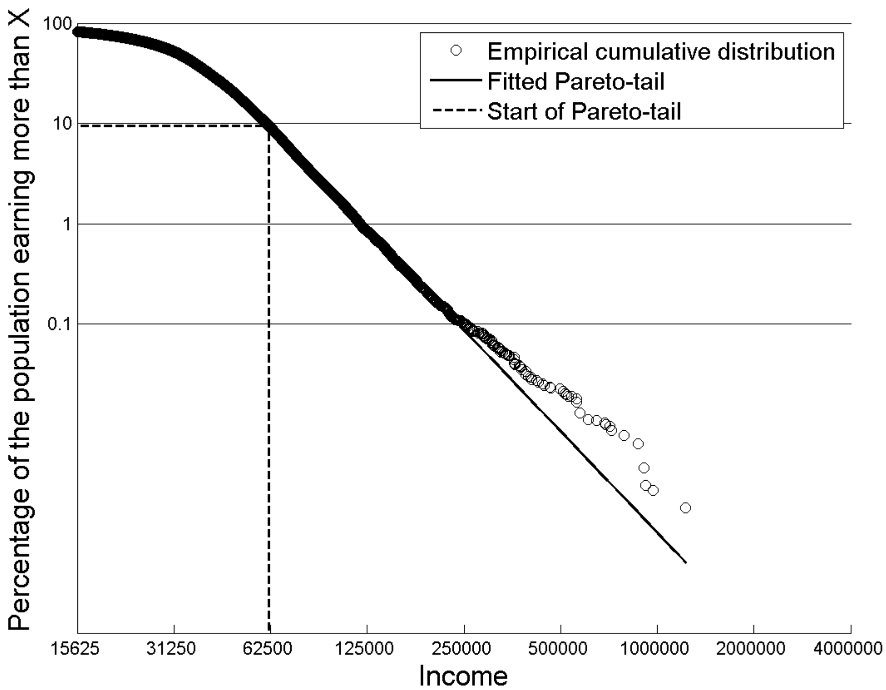

Figure 2: The fit of the Pareto tail

shares which form a dominant holding are taxed at a 25% rate ('Box-2').[16] Director-shareholders of closely-held companies will therefore allocate income over Box-1 and Box-2 to minimize their tax burden. As a result, part of Box-2 income might be considered as income from labor effort. Table 1 also gives descriptive statistics for the sum of Box-1 and Box-2 income. In our main analysis we will focus on labor income taxation (in Box-1).

The cumulative distribution function of the Pareto distribution is given by:

$$F(z) = 1 - \left(\frac{\hat{z}}{z}\right)^{\alpha}, \tag{29}$$

where α is the Pareto parameter, z is gross income, and $\hat{z}$ denotes the cut-off level after which the Pareto distribution applies. We estimate the parameters of the Pareto tail using the method developed by Clauset et al. (2009). In particular, for a given $\hat{z}$, we choose α such that it maximizes the likelihood function. Subsequently, we choose $\hat{z}$ by minimizing the Kolmogorov-Smirnov (KS) statistic. The KS-statistic measures the maximum distance between the estimated and the empirical cumulative distribution function.[17]

Table 2 provides the estimation results for the Netherlands using IPO 2002. We report the estimates of the Pareto parameter with the corresponding asymptotic standard errors and boot-

[16]See Bovenberg and Cnossen (2001) for an overview of the system of tax boxes in the Netherlands.

[17]Using simulated data, Clauset et al. (2009) show that this method outperforms other methods proposed in the literature.

Table 1: Summary statistics for the income distribution in 2002

Income definition	Mean	Median	Standard deviation
Labor costs	34,487	31,253	26,350
Gross wages	28,691	25,892	22,202
Box-1 and Box-2 income[a]	34,774	31,309	27,286

[a]See the main text for the definition of Box-1 and Box-2 income.

Table 2: Estimates of the Pareto parameter α and starting point $\hat{z}$ in 2002

Income definition	α	SE[a]	CI[b]	$\hat{z}$	CI[b]	R^2
Labor cost	3.35	0.037	[3.24,3.44]	61,793	[56,624,77,182]	.995
Gross wages	3.22	0.029	[3.15,3.32]	45,040	[39,565,61,388]	.997
Box-1 and Box-2 income[c]	3.18	0.029	[3.08,3.26]	55,448	[48,115,77,238]	.997
Labor cost 2006	3.30	0.061	[3.17,3.48]	61,943	[56,926,73814]	.991

[a]Asymptotic standard errors.
[b]Bootstrapped 95% confidence interval.
[c]See the main text for the definition of Box-1 and Box-2 income.

strapped 95% confidence intervals, the estimates for the starting point of the Pareto distribution with the corresponding bootstrapped 95% confidence intervals, and the R^2-measures of fit. For our preferred definition of income, labor costs, we find a Pareto parameter of 3.35 with a 95% confidence interval of $[3.24, 3.44]$. The starting point is estimated at approximately 62,000 euro with a 95% confidence interval of approximately $[57; 77]$. The fit of the Pareto tail is extremely good, with an R^2 of .995. Since $1 - F(z) = (\hat{z}/z)^{\alpha}$, we can write for the earnings distribution $\ln(1 - F(z)) = \alpha \ln \hat{z} - \alpha \ln z$. Therefore, a plot on a log-log scale with $1 - F(z)$ on the vertical axis and labor earnings on the horizontal axis should be a straight line with a slope of $-\alpha$ if the tail is Pareto distributed.

Figure 2 shows this plot for our sample. The dots are one minus the empirical cumulative distribution function for each earnings level. The dashed line is the estimated Pareto tail. As can be seen, the relationship between the logarithm of z and the logarithm of one minus the empirical cumulative distribution function is indeed extremely close to being linear.

We investigate whether ignoring Box-2 income leads us to overestimate the Pareto parameter. According to IPO data, only .39% of individuals has income in Box-2.[18] Table 1 shows that the mean of the sum of Box-1 and Box-2 income is still higher than mean labor costs, even though it ignores employers' premiums, and the same is true for the median. Nevertheless, the point estimates of the Pareto parameter using gross wages and the sum of Box-1 and Box-2 income are very close, although somewhat lower, at 3.22 and 3.18, respectively. The starting points are estimated to be lower, since mean labor costs are 20% higher than mean gross wages. For gross wages the point estimate of the starting point of the Pareto tail (approximately 45,000 euro) is very close to the starting point of the top income tax bracket (approximately 48,000 euro) in 2002 (more on the

[18]The data on income in Box-2 are right censored at 250,000 euro. 62 of the 489 individuals (13%) with Box-2 income in IPO are right censored. This may further lead us to somewhat underestimate the thickness of the right tail of the income distribution.

parameters of the Dutch tax system are provided later).

Our empirical findings are very close to the estimates from Atkinson and Salverda (2005), although the latter are based on aggregate household income, including capital income, whereas we use individual, labor income. Atkinson et al. (2011) report Pareto-parameter estimates for 20 countries. Like Atkinson and Salverda (2005), these estimates are all based on aggregate tax statistics.[19] Clearly, Pareto parameters vary much across countries. Notably, the Netherlands features the highest Pareto parameter of all studies covered in this study. In a separate study for Denmark, Kleven and Kreiner (2006) report the Danish Pareto parameter to be 3.5, which is the highest estimate in the world that we are aware of. Therefore, it is lonely at the top in the Netherlands.

4.3 Marginal tax rates

Marginal taxes measure the difference between the increase in gross and net income when an individual's gross income increases by a small, marginal amount. However, determining the additional amount of net income after an increase in earnings is a complex task for various reasons. First, tax systems feature all kinds of non-linearities induced by income-dependent tax credits and income support. Next to income tax rates, we take into account income-dependent transfers, all important (income-dependent) tax credits, indirect taxes, child-care support, income dependent public health-care insurance, rent support and welfare benefits. To that end, we employ the MIMOSI model, a sophisticated tax-benefit calculator, which contains all relevant details of the tax and benefit system in the Netherlands.[20]

Second, one needs to include indirect taxes and subsidies, such as the value added tax, but also the tax treatment of owner-occupied housing. Indirect taxes distort the choice over income and leisure just like direct taxes do. According to Statistics Netherlands (2013) (net) indirect taxes were 11.7% of total consumption in 2002. Bettendorf et al. (2012) demonstrate that indirect taxes are very close to proportional in total consumption in the Netherlands. Hence, we assume indirect taxes are flat.

Third, we need to determine the net income component of premiums. Most studies treat all premiums as taxes, *e.g.* Saez (2001), Gruber and Saez (2002), Brewer et al. (2010), and in their study of marginal tax rates in the Netherlands Gielen et al. (2009). However, this is not correct for the Netherlands. Individual benefits (unemployment, disability, pension) are linked to individual contributions made by either employees or their employers. Hence, not all premiums are taxes, and one needs to treat premiums for unemployment, disability and pensions as deferred wage income. However, determining the marginal tax rate on deferred wage incomes is complicated for a number of reasons.

[19] The estimates by Atkinson et al. (2011) are based on total income, including not only labor income, but also capital income. Capital gains are excluded from their (and our) income definition. Naturally, the estimates gathered in Atkinson et al. (2011) are only as good as the aggregate income tax statistics from which they are computed. These authors provide an extensive discussion of the potential caveats.

[20] See Gielen et al. (2009) for a recent analysis of changes in marginal tax rates over the past decade using MIMOSI.

First, replacement rates during unemployment, disability and retirement (state and occupational pensions combined) are around 70 percent of earned income. Hence, individuals typically experience a substantial drop in income, and hence the marginal tax rate, when they receive the deferred wage income. Therefore, we assume that deferred wage incomes are taxed in one tax bracket below the current bracket (except for individuals in the first tax bracket). Second, it is hard to determine whether premiums are used to redistribute from high-income to low-income earners, and thereby contribute to the marginal wedge. The unemployment and disability schemes redistribute income from high income to low income workers.[21] However, the Dutch pension scheme redistributes income from low-income to high-income individuals.[22] Due to these complexities we decided ignore redistribution in premiums. Third, assets accumulated for (and blocked until) retirement are not subject to wealth taxes. Given sufficient separability in preferences between consumption and labor (as we assume), the exemption for the wealth tax does not directly affect labor supply incentives. Therefore, we also ignore intertemporal considerations in premiums.

A potentially important element missing from our calculations is the tax deductibility of interest on mortgage loans. Individuals that earn more income are more likely to own a (more expensive) house. Hence, one could argue that the mortgage-rent deduction acts as an indirect subsidy on labor. However, this reasoning assumes a perfectly elastic supply of housing. When housing supply is not perfectly elastic, as in the Netherlands, larger demand for housing translates into higher housing prices, which reduce the incentive to supply labor.[23] Based on the Dutch situation, we roughly calculate that the effect of housing subsidies on the total tax wedge is relatively minor (roughly 2.5%, see previous footnote). Therefore, we decided to ignore housing subsidies in the calculation of marginal taxes.

Despite the latter limitation, our calculations of marginal tax rates are much more advanced than in any other study we are aware of. For example, Saez (2001) and Jacquet et al. (2013) assume a linear tax system to recover the ability distribution. However, marginal tax rates are quite nonlinear, as we will demonstrate below, and this could bias the estimation of the ability distribution.

[21] de Koning et al. (2006) calculate the implicit redistribution in unemployment insurance from low income to high income individuals, using panel data for 12 years. They divide the population in three skill groups, all of equal size. Over a period of 12 years, the lowest 33% of the population uses 46% of all benefit days used, whereas the highest 33% uses only 20% of all benefit days used. de Koning et al. (2006) also calculate the implicit redistribution in disability insurance from low-income to high-income individuals. Over a period of 12 years, the lowest 33% of the population uses 59% of all benefit days used, whereas the highest 33% uses 17% of all benefit days used.

[22] Bonenkamp (2009) calculates redistribution in occupational pensions in the Netherlands on a lifetime basis. He calculates the present discounted value of pension contributions and pension benefits for four skill groups by gender (and cohort). The inter-educational redistribution for the lowest skill groups is − 17% of premiums for low skilled males and − 13% of premiums for low skilled females. High-skilled males receive a subsidy of 6% on a lifetime basis, and high-skilled females 1% (controlling for 'cross-gender' redistribution).

[23] Denote the elasticity of housing demand by $\varepsilon^d \equiv \frac{\mathrm{d}h^d}{\mathrm{d}p^*}\frac{p^*}{h^d}$, and the elasticity of housing supply by $\varepsilon^s \equiv \frac{\mathrm{d}h^s}{\mathrm{d}p}\frac{p}{h^s}$, where $p^* \equiv (1-s)\,p$ denotes the net housing price, p the gross housing price and s the housing subsidy. Then, standard tax-incidence analysis shows that $\frac{\mathrm{d}p^*}{p^*} = -\frac{\varepsilon_s}{(\varepsilon_s+\varepsilon_d)}\frac{\mathrm{d}s}{s}$. van Ewijk et al. (2006) estimate that $\varepsilon_s = \varepsilon_d = 0.7$, hence the net housing price falls 0.5% when the subsidy increases with 1%. On average, net housing expenditures are about 25% of total income (CPB, 2010b), and this pattern is rather flat over all income groups. The tax advantage in housing is about 20% of total housing costs, so this reduces the tax wedge on labor by at most 2.5% ($1/2 \times 20\% \times 25\%$).

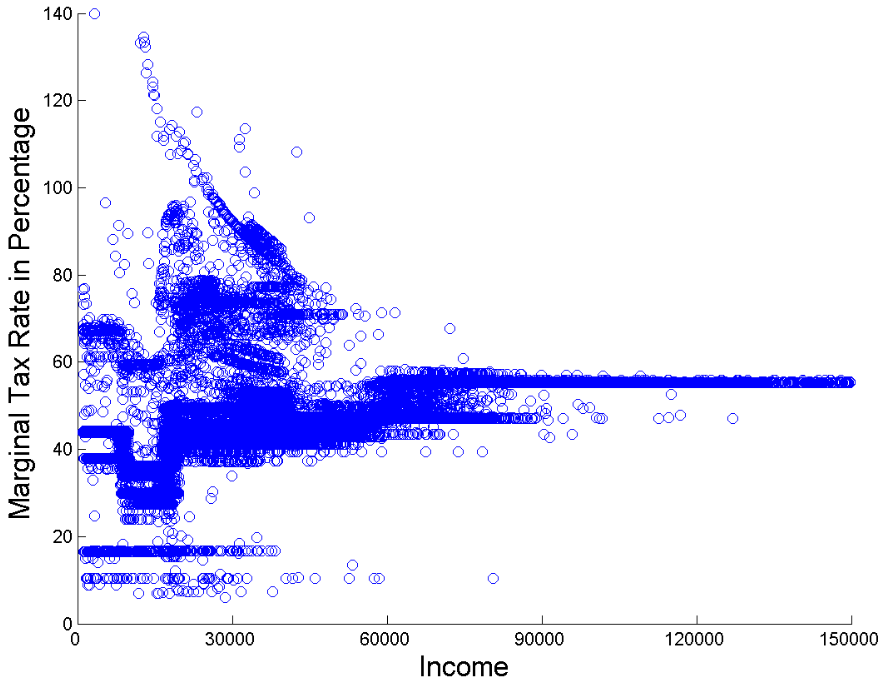

Figure 3: Scatterplot of marginal tax rates by income

There is huge variation in marginal tax rates at each income level, see Figure 3 where we made a scatter plot of marginal taxes against income. However, the model only works with a single marginal tax rate at each income level, hence we use a kernel estimate to smooth out the differences. Figure 4 gives the kernel estimate for effective marginal tax rates in the Dutch income distribution for all workers participating in the labor market. To understand the patterns in Figure 4, Table 3 provides some parameters of the Dutch tax system in 2002.

In 2002, the Dutch tax system had four tax brackets for labor income, based on individual (not household) income, with rates rising from somewhat below 33% at the bottom to 52% at the top. This explains why marginal tax rates are typically lower for individuals with low income than for individuals with high income.

But there are also a number of noticeable deviations from these rates, which result mostly from targeted subsidies and tax credits. For the lowest incomes, marginal tax rates are initially significantly higher than the first tax bracket because a number of income-support schemes are phased out with income, in particular rent subsidies and a general child tax credit.[24] Marginal tax rates are much lower in the segment where the earned-income tax credits (EITCs) are phased in (see Table 3). The end of the phase-in range for the EITCs (almost) coincides with the start of the second tax bracket, and marginal tax rates jump up by some 15%-points between 15,000 and 20,000 euro.

[24]The exact subsidy levels and taper rates vary with household characteristics other than income, and are therefore not reported in Table 3.

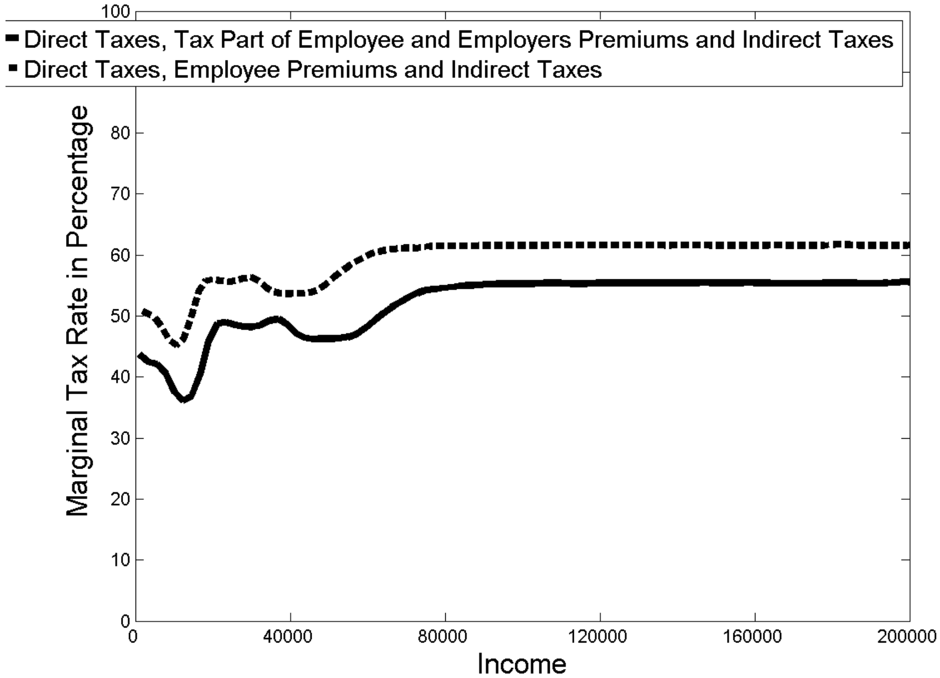

Figure 4: Kernel of marginal tax rates by income

Table 3: Tax brackets and tax credits in 2002

	Start	End	Percentage	Maximum amount
Tax brackets				
First tax bracket	0	15,331	32.35	4,960
Second tax bracket	15,331	27,847	37.85	4,737
Third tax bracket	27,847	47,745	42.00	8,357
Fourth tax bracket	47,745	∞	52.00	∞
Tax credits				
General tax credit	0	∞	0	1,647
Earned-income tax credit				
- First part	0	7,692	1.73	133
- Second part	7,692	15,375	10.62	949
Single parent tax credit	0	∞	0	1,301
Earned-income single-parent tax credit	0	30,256	4.30	1,301

Another noticeable jump can be observed for individuals with a gross income close to 40,000 euro. Individuals below this income threshold are eligible for the public health-insurance scheme with relatively low contribution rates, whereas individuals above this threshold are required to have private health-care insurance with relatively high premiums. For some households close to the threshold this results in very high marginal tax rates.[25]

4.4 Government revenue requirement and benefit level

We assume that the government has to collect 9.5% of total output to finance government consumption net of income redistribution. This is the sum of expenditures on public administration, the police, the justice system, defense and infrastructure minus non-tax revenues (for example from the sales of natural gas) as a percentage of GDP in 2002 (CPB, 2010a, Annex 9).[26] This is in the same order of magnitude as Tuomala (2010), who assumes (maximum) government consumption of 10% of GDP. With the revenue requirement set at 10% of total labor income, the tax system is budgetary neutral with a benefit level of approximately 12,000 euro. This is somewhat higher than the current level of net welfare benefits in 2002 amounting to 9,014 euro for a single-person household. However, we ignored some other forms of social assistance at the local level ('Bijzondere Bijstand'), exemptions from local taxes, and transfers in kind (discounts for arts, public transport, etc.), training, public employment, and labor-market programs, which also act as support schemes for the non-employed.

4.5 Elasticity of income with respect to marginal taxes

An important determinant of optimal income tax rates is the elasticity of the tax base. We use recent Dutch estimates of the participation elasticity and the elasticity of taxable labor income to calibrate the extensive and intensive margin responses of the tax base in the model. We discuss these estimates and the calibration method below.

4.5.1 Elasticity of labor supply

Traditionally, economists have analyzed at the impact of taxes on labor supply to measure the distortions from income taxation. Table 4 gives an overview of recent estimates of labor-supply elasticities in the Netherlands. Mastrogiacomo et al. (2013) estimate a structural discrete-choice model for a number of subgroups using data for the period 1999-2005. Our calibration year (2002) is in the middle of this sample. These authors present estimates for the uncompensated wage elasticity of total hours worked, the participation rate and hours per worker.

[25] In 2003 this health-care system has been replaced by a uniform, obligatory basic health-insurance scheme, which is financed by a payroll tax and 'lump-sum' premiums paid by individuals. Individuals can voluntarily top up the basic health-insurance scheme with supplementary insurance packages.

[26] We have experimented with different levels of the government's revenue requirement, but we do not find that this induces significant changes in the optimal marginal tax schedules.

Table 4 reveals that the total uncompensated labor supply elasticity of men in couples is rather small. Elasticities are larger for women in couples, in particular when small children are present. Single parents have the highest labor-supply elasticities, and elasticities of singles are in between single parents and individuals in couples. Looking at the decomposition of these elasticities into participation (extensive) and hours per worker (intensive) responses, we find that most of the response is on the participation margin.[27] These findings are in line with the findings of related empirical studies for the Netherlands, see Mastrogiacomo et al. (2013, Table 15).

The ranking of the elasticities by household types and the extensive versus the intensive margin are in line with the findings of empirical studies abroad. Once more, see Mastrogiacomo et al. (2013, Section 5) and the excellent overview in Bargain et al. (2013b). Weighting the elasticities of Mastrogiacomo et al. (2013) for the different groups by their respective sizes on the Dutch labor market, we obtain an average total-hours elasticity of 0.30, an average participation elasticity of 0.25, and an hours-per-worker elasticity of 0.06.

Table 4: Estimates of the uncompensated labor-supply elasticity

Group	Obs.	Total hours		Participation		Hours per worker	
		Men	Women	Men	Women	Men	Women
Couples with children	72,000	0.14***	0.50***	0.14***	0.38***	0.01	0.12***
		(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.02)
Couples w/o children	72,000	0.07***	0.27***	0.07***	0.22***	0.00	0.05***
		(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
Singles	24,000	0.39***	0.47***	0.33***	0.39***	0.06***	0.08***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Single parents	24,000	0.45***	0.62***	0.32***	0.43***	0.12***	0.18***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Source: Mastrogiacomo et al. (2013, Table 15, Table A.12). Bootstrapped standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Unfortunately, Mastrogiacomo et al. (2013) do not consider income elasticities, as data on unearned income are lacking. However, some other recent studies use data on unearned income to estimate the unearned-income elasticity for the Netherlands. We convert these unearned-income elasticities into income elasticities using the average share of unearned income in total income in the descriptive statistics. For Vermeulen (2005) we then obtain an income elasticity ranging from -0.01 and -0.02 for single men and women to -0.10 and -0.14 for men and women in couples, resepctively. For Bloemen (2009) we find an income elasticity of -0.10 to -0.18 for men in couples and -0.12 to +0.10 for women in couples. Finally, for Bloemen (2010) we find an income elasticity of -0.22 and

[27] Chetty (2012) has recently argued that optimization frictions may mask part of responses on the intensive margin. Furthermore, the elasticities reported in Mastrogiacomo et al. (2013) are so-called unconditional intensive-margin elasticities. These are simulated by increasing gross wages by 10%, which is common in the structural discrete-choice literature on labor supply. Simulated in this way, the elasticities capture both the response in hours by those already working and a composition effect because new entrants may work different hours than those already working. The conditional intensive-margin elasticity of those already working is actually higher than the unconditional intensive-margin elasticity, though still smaller than the extensive-margin elasticity.

Table 5: Estimates of the uncompensated elasticity of taxable labor income

Group	>10,000	50,000–100,000	>50,000
All workers	0.24***	0.26***	0.46***
	(0.01)	(0.04)	(0.07)
Observations	157,510	11,346	12,196
Singles	0.36***	0.36*	0.58*
	(0.04)	(0.21)	(0.30)
Observations	18,061	507	530

Source: Jongen and Stoel (2013, Table 4, Table 5) and additional estimates for singles using the same data set (details available on request).
*** p<0.01, ** p<0.05, * p<0.1.

-0.16 for men and women in couples, respectively. The average earned income elasticity for men is –0.12, and for women, ignoring the positive value for married women in Bloemen (2009), is –0.11.

4.5.2 Elasticity of taxable income

Following the seminal work by Feldstein (1995) recent empirical studies have looked at the elasticity of taxable income (ETI). The ETI may capture a wider range of behavioral responses to income taxes, such as changes in work effort, occupational choice, human capital investment, tax avoidance, tax evasion, and migration. ETI studies typically focus on the employed, hence ETI-estimates are conceptually close to the intensive-margin elasticities in the labor-supply literature.

Table 5 presents recent estimates of the elasticity of taxable labor income in the Netherlands from Jongen and Stoel (2013). They use data for the period 1999-2005 and exploit the 2001 tax reform to estimate the elasticity of taxable labor income. In their base specification, the estimated ETI is 0.24 for all workers. This baseline only employs workers with taxable labor income above 10,000 euro to circumvent problems with strong mean reversion in incomes at the bottom of the income distribution (Gruber and Saez, 2002). The estimated ETI for higher incomes (ranging from 50,000–100,000 euro) is slightly larger than for all workers (>10,000 euro), though the difference between the estimates is not statistically significant. Including also the highest income earners (>50,000 euro) leads to a much larger ETI-estimate, but this estimate is rather imprecise due to a relatively small number of observations.[28] These ETI-estimates are in line findings abroad. In their overview paper Saez et al. (2012) suggest a range of 0.12 to 0.40.

For singles, Jongen and Stoel (2013) find a somewhat larger ETI. Again, the ETI for incomes between 50,000-100,000 euro is is very close to the ETI for all workers, and the ETI for the highest incomes (>50,000 euro) is much larger, though imprecisely estimated.

Jongen and Stoel (2013) also lack information on unearned income and therefore do not estimate income elasticities ('income effects') for the elasticity of taxable labor income. In their overview

[28]Furthermore, the estimated ETI for incomes above 50,000 euro is sensititve to the controls for exogenous income growth, see Jongen and Stoel (2013), whereas the estimated ETI for all workers and for 50,000-100,000 euro is rather stable for different controls for exogenous income growth.

Saez et al. (2012) suggest that income elasticities are rather small. The income elasticity for taxable income ranges from essentially zero (Kleven and Schultz, 2012) to -.14 in Gruber and Saez (2002).

4.5.3 Elasticities in simulations

In our simulations we consider a baseline scenario, and two alternative scenarios, based on different estimates for the elasticities. The key assumptions in these scenario's are summarized in Table 6. In the baseline case we assume a compensated wage elasticity of earnings supply equal to .35, an income elasticity of .10, and hence an uncompensated wage elasticity of earnings supply equal to .25 based on the findings in Table 5 on recent ETI-studies for the Netherlands. Furthermore, we assume a participation elasticity of .25, also based on recent evidence for the Netherlands, see Table 4. We also consider a robustness scenario with 50% higher elasticities, i.e. a compensated wage elasticity of .53, an income elasticity of .15, an uncompensated wage elasticity of .38, and a participation elasticity of .38. And, for completeness, we also consider the opposite case with 50% lower elasticities: .18 for the compensated wage elasticity, .05 for the income elasticity, .13 for the uncompensated wage elasticity and .13 for the participation elasticity. These robustness checks facilitate our comparisons with Saez (2001), Brewer et al. (2010) and Jacquet et al. (2013).

Table 6: Elasticities used in the simulation

	Compensated wage elasticity	Income elasticity	Uncompensated wage elasticity	Participation elasticity
Baseline scenario	0.35	0.10	0.25	0.25
Low-elasticity scenario	0.18	0.05	0.13	0.13
High-elasticity scenario	0.53	0.15	0.38	0.38

Table 7: Employment Rates for Different Education Levels

Level of Education	Net Employment Rate	Share in Population
Only elementary school	36.90	11.99
Some high school	53.50	25.79
High school	56.80	10.26
Low-level college	71.20	15.84
Mid-level college	79.10	14.95
Bachelor degree	80.40	13.88
Master degree or higher	84.40	7.28

4.6 Utility and welfare functions

We consider optimal tax rates for the two polar, Benthamite (utilitarian) and Rawlsian (maxi-min) cases of the social welfare function:

$$\begin{aligned} \text{Bentham} &: \int_{\mathcal{N}} W(U_n)\mathrm{d}F(n) = \int_{\mathcal{N}} U_n \mathrm{d}F(n), \\ \text{Rawls} &: \int_{\mathcal{N}} W(U_n)\mathrm{d}F(n) = U_{\underline{n}}. \end{aligned} \tag{30}$$

Recall that $U_n \equiv u_n - \varphi$, which equals u_n when only the intensive labor-supply margin is included.

We assume a functional form for the utility function, which encompasses most of the utility functions encountered in the literature:

$$u_n = \frac{c_n^{1-\alpha}}{1-\alpha} - \gamma \frac{l_n^{1+1/\varepsilon}}{1+1/\varepsilon}, \qquad \alpha, \gamma, \varepsilon > 0. \tag{31}$$

Our utility function allows for income effects and is also used by Mankiw et al. (2009). When $\alpha = \frac{1}{\varepsilon}$ this specification is in line with the CES functions used by Mirrlees (1971) and Tuomala (1984). α and ε are calibrated so as to match the compensated and uncompensated elasticity of the scenarios described in Table 6, which is in the spirit of Chetty (2006).[29] Parameter γ is an innocuous scaling parameter, which hardly affects the resulting optimal tax rates. We adjust it to keep the mean of the ability distribution fixed in the different scenarios.

Table 8: Calibrated parameters for the utility function

Parameter values	Base	Low Elasticity	High Elasticity
α	0.46	0.48	0.45
ε	0.38	0.18	0.60
γ	1981.67	13503.12	1082.11
μ_k	55.95	0.00	82.42
σ_k	271.27	511.00	189.98

Table 8 displays the values of the parameters for the utility function. As can be seen, parameter α is almost constant in all scenarios. Hence, the elasticity of the marginal utility of income is the same across the simulations. Therefore, the difference in optimal tax rates in the scenarios should be attributed to the differences in the elasticities, and not a changed preference for redistribution via the curvature of the private utility function.

4.7 Determination of the ability distribution

The determination of the distributions of ability and participation costs/benefits is not straightforward, since they are not directly observable. We assume that the data on earnings and partici-

[29] As long as the ratio $\varepsilon^c/\varepsilon^u$ is fixed, the calibrated α is almost the same for different elasticities. This is a useful property, since then we can isolate the effect of a change in the elasticities without changing the redistributional concerns. All our scenarios therefore have the same ratio $\varepsilon^c/\varepsilon^u$.

pation are a choice process that follows from our assumed utility function and the distribution of participation costs/benefits. Given observed choices for earnings and labor force participation, and assuming separability between the leisure and consumption component of utility and participation costs/benefits, we are able to identify the distributions of skills.

In particular, conditional upon participation in the labor market, and using the definition of gross labor earnings $z_n \equiv nl_n$, we can invert the first order condition for optimal labor supply (4) to express ability n as a function of marginal tax rates and income. The solution for ability is:

$$n = \left(\frac{\gamma z_n^{1/\varepsilon}}{(1 - T'(z_n))c_n^{-\alpha}} \right)^{\frac{\varepsilon}{\varepsilon+1}}. \tag{32}$$

Using information on gross earnings z_n, consumption c_n, and marginal tax rates $T'(z_n)$ we are able to compute ability of each working individual in the data-set. Note that the consumption level follows from the difference between gross earnings and total taxes paid: $c_n = z_n - T(z_n)$.

4.8 Calibration of the distribution of participation costs and benefits

We estimate the distribution of participation costs using information on the employment rate and the participation elasticity. Ideally, we would like to have data on the employment rate for each level of ability n, but no such data are available. However, we do have data on employment rates by 7 levels of education. These data are given in Table 7. By assuming that the cumulative distribution of education corresponds to the cumulative distribution of ability, the education-specific employment rates allow us to estimate the distribution of participation costs by skill type. Since we assume that everyone has the same utility function, we also assume that the distribution of the disutility of participation is independent of ability, i.e. $k(\varphi|n) = k(\varphi)$. Under this assumption, the theoretical model predicts an ability-specific participation rate $\hat{E}(n_1, n_2)$ for all individuals between skill levels n_1 and n_2 for any pair $\{n_1, n_2\}$, with $n_2 > n_1$ equal to:

$$\hat{E}(n_1, n_2) = \frac{\tilde{K}(n_2) - \tilde{K}(n_1)}{F(n_2) - F(n_1)} = \frac{\int_{n_1}^{n_2} \tilde{k}\,(m)\,\mathrm{d}m}{F(n_2) - F(n_1)} = \frac{\int_{n_1}^{n_2} K\,(u_m - v(b))\,f\,(m)\,\mathrm{d}m}{F(n_2) - F(n_1)}, \tag{33}$$

In addition, the elasticity of participation with respect to the gross wage rate has recently been estimated, as discussed in section 4.5.1. Our model can also be used to predict the value of this participation elasticity.

Note that labor-force participation at ability level n is given by $K(u_n - v(b))$. In addition, the gross wage rate is equal to n. Therefore, the participation elasticity ε_n^P at skill level n with respect to the gross wage rate is given by:

$$\varepsilon_n^P = \frac{\partial K(u_n - v(b))}{\partial n} \frac{n}{K(u_n - v(b))} = \frac{\partial u_n}{\partial n} \frac{nk(u_n - v(b))}{K(u_n - v(b))}, \tag{34}$$

where the final step follows from the envelope theorem. Hence, the predicted average participation elasticity in the economy is given by:

$$\hat{\varepsilon}^P = \int_{\underline{n}}^{\overline{n}} \varepsilon_m^P f(m)\mathrm{d}m = \int_{\underline{n}}^{\overline{n}} \frac{\partial u_m}{\partial m} \frac{mk(u_m - v(b))}{K(u_m - v(b))} f(m)\mathrm{d}m. \tag{35}$$

In equations (33) and (35) u_n, $\frac{\partial u_n}{\partial n}$, $v(b)$, $F(n)$ and $f(n)$ can be inferred from the data. Furthermore, we assume that $k(\varphi)$, is normally distributed with mean μ_k and standard deviation σ_k: $\varphi \sim N(\mu_k, \sigma_k^2)$. Finally, assume that data exist on both $E(n_1, n_2)$ and $\hat{\varepsilon}^P$. In that case, we can write the error term between the model-predicted employment rate and the true employment rate, and the model-predicted participation elasticity and the true elasticty by:

$$\begin{aligned} \epsilon_{n_2} &= \hat{E}(n_1, n_2) - E(n_1, n_2), & (36) \\ \epsilon_p &= \hat{\varepsilon}^P - \varepsilon^P. & (37) \end{aligned}$$

We choose parameters μ_k and σ_k such that the absolute value of the weighted error terms is minimized using non-linear least squares. The weighting procedure in estimating the distribution of participation costs is simple. Table 7 provides observations of skill-specific employment rates for seven different education levels. We only have one estimate of the participation elasticity. Hence, we give each error-term for the skill-specific employment rate a weight equal to 1, and we give the error term for the participation elasticity a weight equal to 7.

We do need to take into account a selection bias in our estimation of the participation-cost distribution. The reason is that we measure the skill distribution by observing the labor-market choices of employed individuals only. In the data we only observe the density of ability *conditional on employment* $f(n|I)$, where I is an indicator variable equal to 1 when an individual is employed, and zero otherwise. However, our model predicts a specific relationship between skill and employment; better skilled individuals are more likely to participate, because they receive a higher wage. Since we are interested in the unconditional skill density $f(n)$, we need to reweigh the skill distribution taking this selection bias into account. This correction is very similar to the procedure described in Heckman (1979).[30]

Bayes' Law provides the relationship between the two densities:

$$f(n) = \frac{f(n|I)\, p(I)}{p(I|n)}. \tag{38}$$

The unconditional probability of employment $p(I)$ equals the total employment rate in the population $E(\underline{n}, \overline{n})$, which can be derived from Table 7. The conditional probability of employment conditional on ability then equals the cut-off level of the disutility of participation below which all individuals with ability n work: $p(I|n) = K(u_n - v(b)|n)$. Using these results, we can recalibrate

[30] If we would not correct for this selection bias, our estimates would underestimate the number of individuals at the low end of the earnings distribution with 60% and overestimate the number of individuals at the top end with 30%.

the skill distribution to get a correct measure of the skill distribution for the entire population:

$$f(n) = \frac{f(n|I)\,E(\underline{n},\overline{n})}{K(u_n - v(b)|n)}. \tag{39}$$

Note that we can only adjust for the estimation bias if we have the unconditional distribution $k(\varphi)$. However, we can only find the distribution $k(\varphi)$ through non-linear least squares if we have derived $f(n)$, which needed to be determined in the first place. We resolve this indeterminacy as follows. First, we guess that the unconditional distribution of ability $f(n)$ is equal to the conditional distribution of ability $f(n|I)$. Based on this initial guess, we estimate the parameters of the distribution $k(\varphi)$. After retrieving the distribution $k(\varphi)$ we can update our initial estimate for $f(n)$. With the updated estimate for $f(n)$ we can again re-estimate the parameters of $k(\varphi)$, etc. We exit the updating procedure when the parameters of $k(\varphi)$ converge. Table 8 provides the estimated values of γ, μ_k and σ_k, which are hard to interpret, as they depend on the cardinal properties of the utility function.

5 Optimal tax schedule

Having discussed the calibration of the model, we now turn to the optimal tax profiles for different elasticities and different social welfare functions. We first consider the optimal tax schedules when individuals can only respond on the intensive margin, and subsequently consider the optimal tax schedules when individuals can respond both on the intensive and the extensive margin. However, we start with a discussion of the optimal top rate, which is virtually identical in both cases, as non-participation is basically not relevant for individuals with a top income.

5.1 Top rate

For both social welfare functions, the marginal tax rate converges to a constant at the top. Table 9 reports the resulting optimal top rates for the different assumptions about the elasticities and the social welfare function.

Table 9: Optimal effective marginal top rates

Uncompensated/compensated elasticity	.13/.18	.25/.35	.38/.53
Rawlsian	65%	56%	49%
utilitarian	60%	48%	40%

Source: Figures taken from simulations, see later in paper.

The Rawlsian (maxi-min) government aims to maximize tax revenue from the top income earners, it wants to 'soak the rich' by setting their tax rate at the top of the Laffer curve. Table 9 shows that the current effective top tax rate of 55.4% (which includes indirect taxes) is virtually identical to the baseline value (55.6%). Increasing the current top rate from 52% to 60% (excluding indirect taxes) – as some political parties have suggested – thus results in revenue losses. Higher top rates

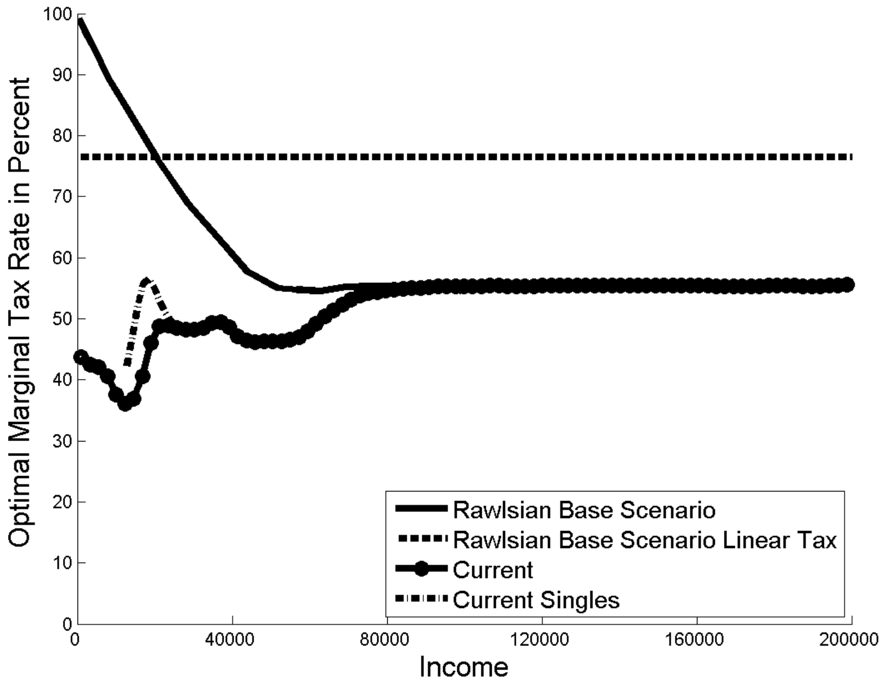

Figure 5: The optimal tax schedule with Rawlsian (maxi-min) social preferences

then result in both less income redistribution and larger deadweight losses. Consequently, both equity and efficiency are reduced. Only at a low elasticity of the tax base, a higher top rate could be optimal.

A Benthamite (utilitarian) social welfare function attaches a positive welfare weight to high-income earners; the euro extracted from the top-income earners results in a utility loss, which is valued by the government. The monetized valued of this utility loss needs to deducted from tax revenues to determine the optimal top rate, see the theory section. In this case, the optimal effective marginal top rate is 48% in the baseline. Hence, the current effective top rate of 55.4% would be set too high. Again, only with very low elasticities a higher top rate would be optimal under utilitarian social preferences.

5.2 Intensive margin

Next, we consider the entire profile of optimal marginal tax rates. Figures 5 and 6 show the optimal linear and non-linear tax schedule under Rawlsian and utilitarian preferences, respectively. For comparison they also show the actual tax schedule.

For the Rawlsian social welfare function, we find that optimal marginal tax rates are generally decreasing. After modal income there is a very tiny increase in marginal tax rates. This contrasts with Saez (2001) and Brewer et al. (2010) who find an inverse U-shape for a Rawlsian social welfare function. The intuition is that the skill distribution behaves differently in the Netherlands in

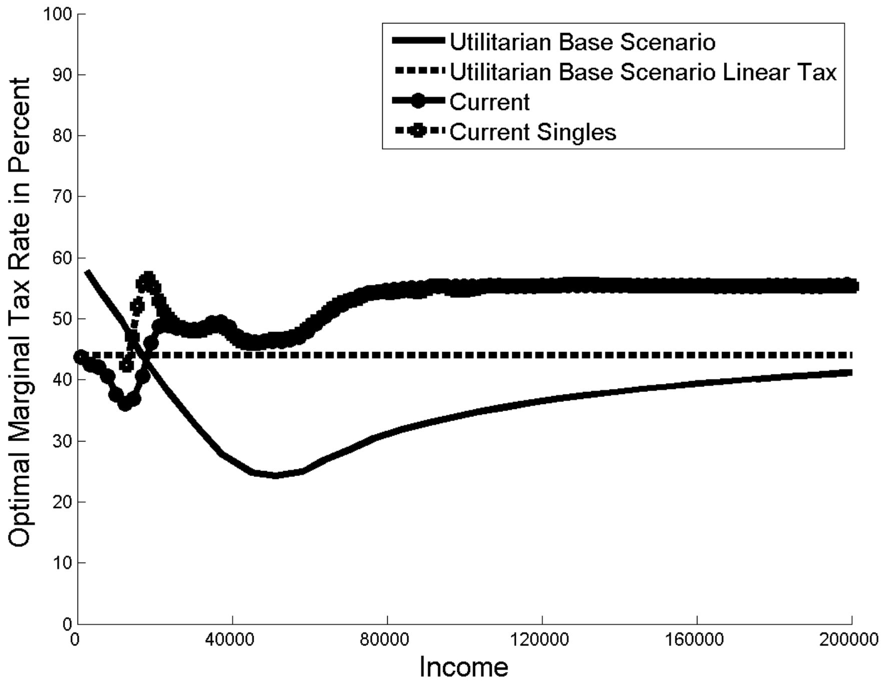

Figure 6: The optimal tax schedule for Bethamite (utilitarian) social preferences

comparison with the US or the UK.

Note that with the Rawlsian social welfare function, the B-term of the optimal tax formula is unity, see Section 3. The A-term does not play a role, since the elasticity is constant across the entire earnings distribution. Hence, all the changes in optimal taxes after modal income are generated by the C-term, which is determined by the earnings distribution. The top tail of the earnings distribution in the Netherlands is much thinner than in both the US and the UK. Hence, setting higher marginal tax rates produces relatively small distributional benefits compared to deadweight losses. Indeed, the plot for the Rawlsian optimal tax schedule implies that the C-term becomes roughly constant after modal income. The U-shape in the optimal marginal tax rates under the utilitarian government is almost entirely driven by the B-term, as the C-term remains relatively constant and the A-term does not play a role. Since average distributional benefits are always rising with income, B rises with income, hence marginal tax rates increase after modal income (Diamond, 1998). These findings are in line with Saez (2001) and Brewer et al. (2010).

Optimal utilitarian average tax rates are set below the current tax rate everywhere, except at the bottom of the income distribution. Even with utilitarian social preferences marginal tax rates at the bottom are too low. Higher marginal tax rates at the bottom allow the government to redistribute more income from lower-middle incomes to the lowest incomes. The lower marginal tax rates everywhere else indicate that the utilitarian government is less interested in income redistribution amoung all other groups in comparison to the current system. Hence, apparently social preferences implicit in the actual system are not very utilitarian (we consider the social welfare weights implicit in the actual system more closely below).

A comparison with Saez (2001) and Brewer et al. (2010) further shows that optimal marginal tax rates are generally lower in the Netherlands than in the UK and the US. Abilities are more equally distributed in the Netherlands than in the US and the UK so that the gains of redistribution are typically lower. Distortions of income redistribution are similar, since elasticities of labor supply are comparable. Hence, and optimal taxes are lower in the Netherlands.

Comparing the actual Dutch tax schedule with the optimal tax schedule, we see that the optimal marginal tax rates under Rawlsian social preferences are typically higher, except at the top. We see that the difference between actual and optimal tax rates is largest at the bottom of the income distribution, where the optimal tax rate is close to 100%. The efficiency loss of the high marginal tax rate at the bottom is small, because it only affects a small group of individuals. Hence, the optimal tax rate at the bottom is an efficient way to redistribute income from the rich and middle income groups to the poor. Below we will see if this conclusion remains valid once we allow for an extensive margin decision. A Rawlsian government sets marginal tax rates that are generally declining for the earnings distribution. In addition, it basically 'soaks' all the middle and high incomes to maximize government revenue, so as to give the highest feasible transfer to the poorest people in society. Indeed, such a transfer can only be financed if it is phased out through very high marginal tax rates.

We are not only interested in marginal tax rates but also in total taxes. Who gains and who loses under each tax-benefit system? Figure 5.2 illustrates the allocations that follow from the optimal non-linear tax schedules and the current tax schedule. In addition it shows the laissez-faire allocation in which gross income equals net income. The intercept is the transfer the government provides to individuals without gross income. The slope equals 1 minus the marginal tax rate. Individuals left of the laissez-faire allocation receive net income support from the government, whereas individuals right of the laissez-faire allocation are net tax payers.

With Rawlsian social objectives we find a much higher optimal transfer to non-employed individuals than in the actual system. In addition, individuals who earn up to about 15,000 euros are better off under the optimal Rawlsian tax schedule than they currently are. And, up to about 22,000 euros individuals receive net-income support from the government. The utilitarian government provides about the same transfer to non-employed individuals as in the actual tax-benefit system. Individuals below the average income level are worse of under the utilitarian allocation than the actual system, and individuals above the average income are better off.

Figure 8 plots the average tax rates. Both the current tax schedule and the two optimal tax schedules are strictly progressive as can be witnessed from the fact that the average tax is strictly increasing. The optimal utilitarian social planner would increase average taxes at the bottom and decrease taxes at the top. The difference between the current average tax rate and the optimal utilitarian average tax rate is largest at the top. On the other hand, the Rawlsian social planner would decrease the average tax rate for the lowest income earners and increase the average tax rate

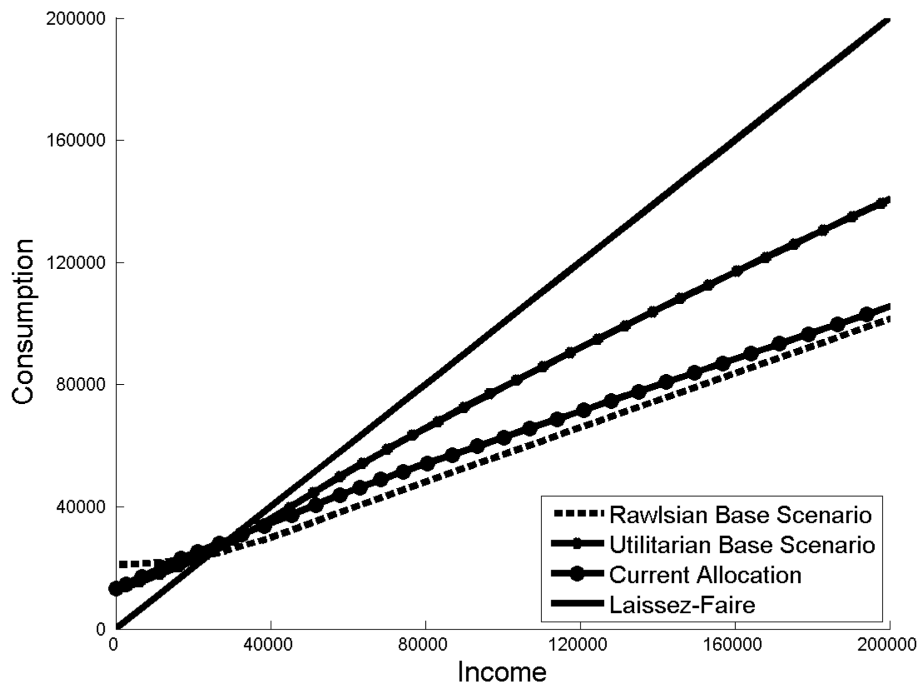

Figure 7: The allocation for the different social preferences

for the middle- and top-income earners. The difference is largest for middle-income earners. In addition, the average tax burden for high-income earners up to 200,000 euro is still higher under the optimal Rawlsian tax schedule even though the marginal tax rate at the top of the current tax schedule is set at the top of the Laffer-curve. Only the individuals earning more than 250,000 euro will face a lower average tax rate, but less than 0.1% of the population has an income above that level in the Netherlands.

5.2.1 Sensitivity analysis

Figures 9 and 10 display the optimal tax rates under a scenario with low elasticities, $\varepsilon^c = 0.18$ and $\varepsilon^u = 0.13$, and a scenario with high elasticities, $\varepsilon^c = 0.53$ and $\varepsilon^u = 0.38$. As noted before, when the elasticity is very low, and we employ a Rawlsian social welfare function, the top rate is too low. But for the high-elasticity case it is too high. These figures also demonstrate that the actual marginal tax rate at the bottom is always too low, even if the elasticity of taxable income is large and we assume a utilitarian social welfare function. Below we consider whether this conclusion still holds when we introduce an extensive margin.

5.3 Intensive and extensive margin

Figures 11 and 12 give the optimal marginal tax rates under both intensive and extensive labor-supply responses. From these figures we see that the introduction of an extensive labor-supply

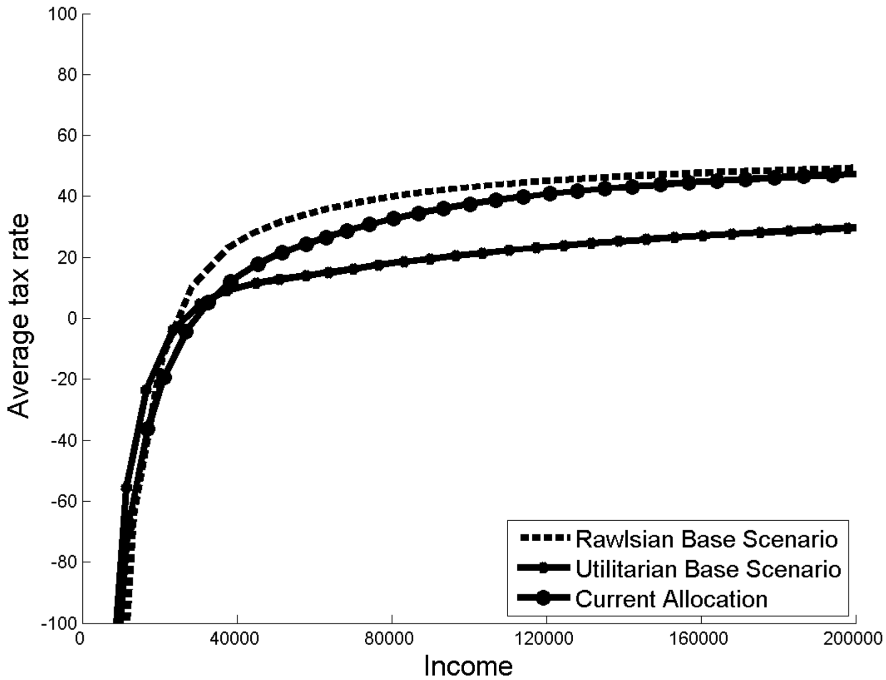

Figure 8: The average tax rates for different social preferences

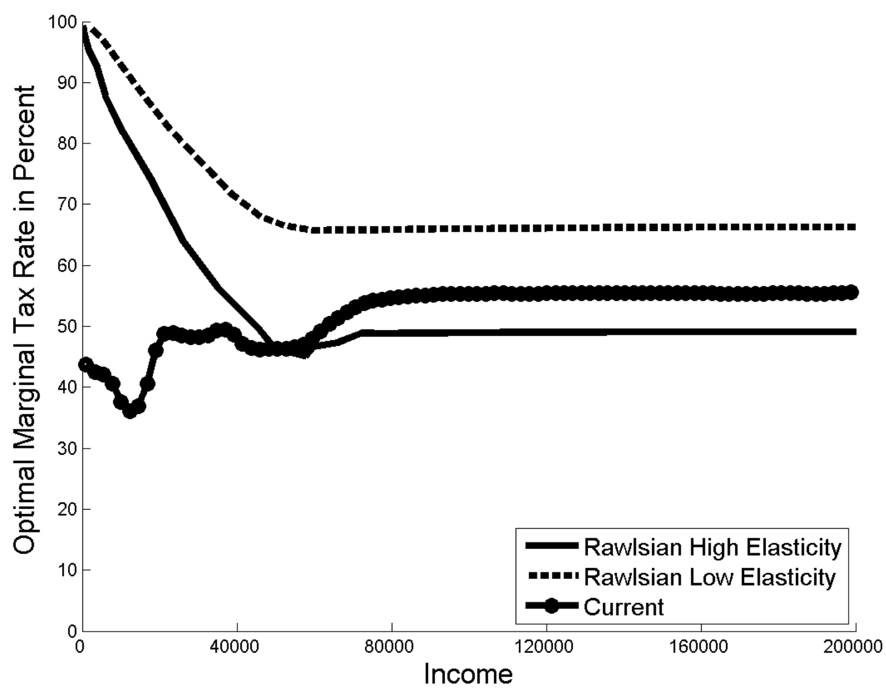

Figure 9: The optimal tax schedule with Rawlsian social preferences. Low: $\varepsilon^c = 0.18$ and $\varepsilon^u = 0.13$. High: $\varepsilon^c = 0.53$ and $\varepsilon^u = 0.38$.

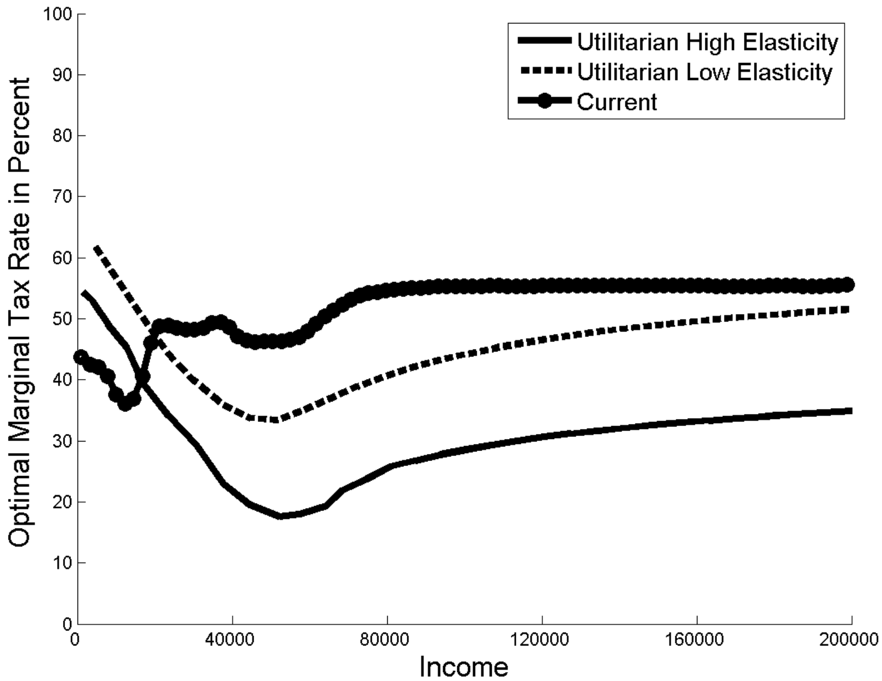

Figure 10: The optimal tax schedule with utilitarian social preferences. Low: $\varepsilon^c = 0.18$ and $\varepsilon^u = 0.13$. High: $\varepsilon^c = 0.53$ and $\varepsilon^u = 0.38$.

response hardly affects the optimal top rate. High-income earners do not really respond on the extensive margin. The extensive labor-supply response reduces the marginal tax rates especially for low- and middle-income earners.

For low-income earners the optimal marginal tax rate drops significantly due to the introduction of the extensive margin, especially under Rawlsian social preferences. A high marginal tax rate at the bottom increases the average tax rate for middle-income earners. This induces middle-income earners to leave the labor market, which reduces government revenues. Hence, marginal tax rates are optimally set lower. Nevertheless, when compared to the actual system, marginal tax rates at the bottom should still be higher, even under utilitarian social preferences. For middle-income earners we see that the marginal tax rates in the current tax system are lower than actual rates, even under Rawlsian preferences. From this we could conclude that in the current tax system the marginal tax rates for middle-income groups are too high.

Our findings contrast sharply with those of Jacquet et al. (2013). The main difference between their simulations and ours is that the participation elasticities in our model are highly non-linear and hump-shaped with income, see Figure 13. Although the average participation elasticity is calibrated at .25 in the baseline, the participation elasticity is endogenously determined by the distribution of participation costs. We can fit observed education-specific employment rates only when participation elasticities are low for the lowest-skilled workers (due to high non-employment benefits) and for the highest-skilled workers (due to very high earnings compared to non-employment

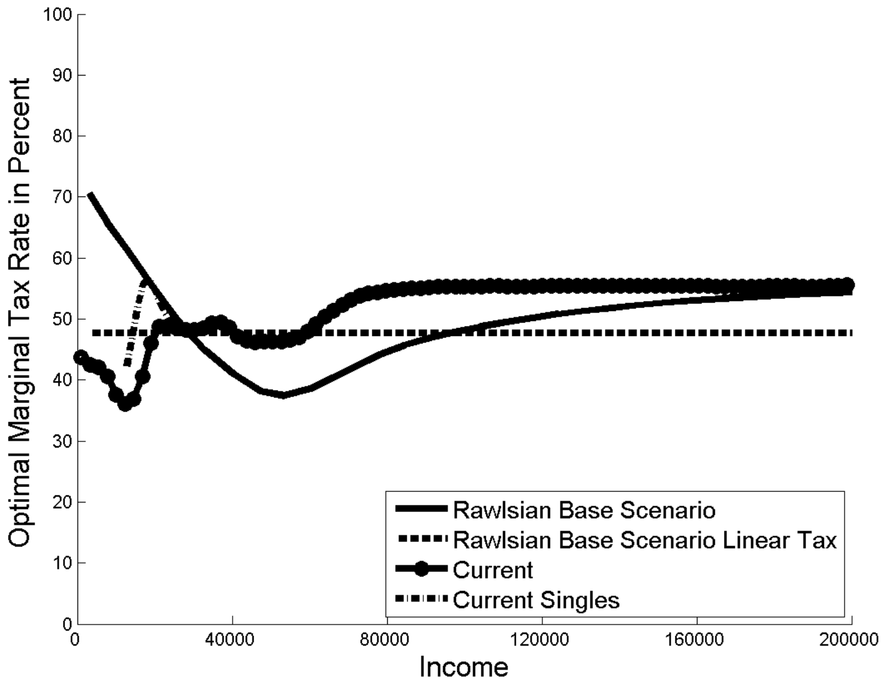

Figure 11: The optimal tax schedule with Rawlsian social preferences, with intensive and extensive labor-supply responses

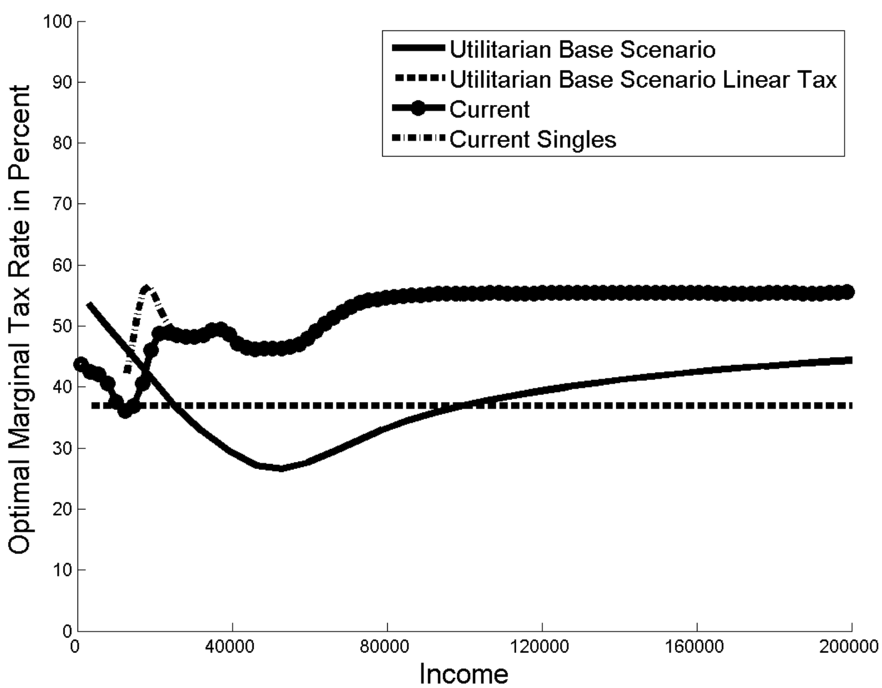

Figure 12: The optimal tax schedule with utilitarian social preferences, with intensive and extensive labor-supply responses

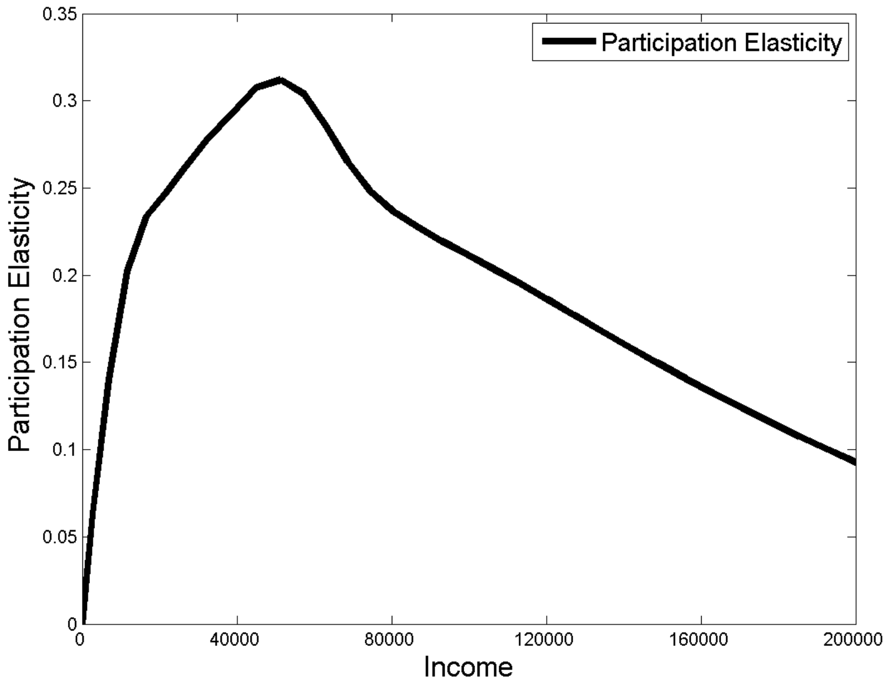

Figure 13: Participation elasticity by income

benefits).

In the baseline simulations of Jacquet et al. (2013), the participation elasticity is roughly flat over the entire earnings distribution. These authors do not estimate the distribution of participation costs and calibrate the model to real-world data, but plainly assume that participation elasticity declines from 0.5 to 0.4 from the lowest to the highest income level. As a result, introducing an extensive margin results in much lower optimal tax rates over the entire earnings distribution, including the top. We think that our estimation of the distribution of participation costs is better founded, and produces empirically more plausible participation elasticities, see also the literature review.

Figure 14 shows the participation tax corresponding to the optimal tax schedules in 11 and 12, and the actual schedule. Recall from Section 2 that the participation tax $(T(z_n) + b)$ measures the transfer to the government if an individual decides to enter the labor market, pay taxes and forgo non-employment benefits.

For a utilitarian social welfare function a positive participation subsidy (about 2,250 euro) is optimal for the workers earning a very low income. Such a subsidy redistributes resources to the working poor, which still have a large social welfare weight. For a Rawlsian social welfare function, it is always optimal to tax participation on a net basis, even for those workers with low earnings. The participation tax is then about 9,600 euro. The Rawlsian government only cares about the worst-off in society, which are the non-employed. As a result, redistribution to the working poor

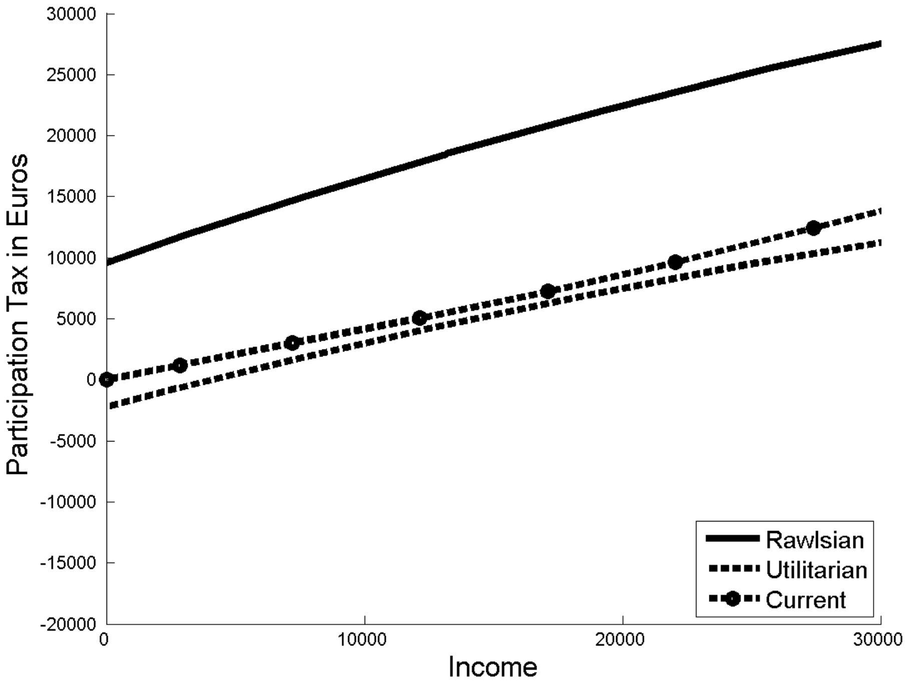

Figure 14: The participation tax $T(z) + b$ with utilitarian and Rawlsian preferences

does not raise social welfare, as it implies less redistribution towards the non-working poor. The participation tax with the Rawlsian government is thus quite high even for very low earnings.

Figure 15 gives the optimal average tax rates. Both the optimal utilitarian tax schedule and the optimal Rawlsian tax schedule are no longer strictly progressive. The average tax rate slightly decreases a little above median income levels. Participation of these groups is very important for government revenue. In particular, the participation elasticity is highest among the middle-income groups, see Figure 13. By slightly decreasing the average tax rate, the government boosts labor-force participation, which yields larger government revenue. As expected, the utilitarian social planner increases the tax burden for poor individuals and decreases the tax burden for the rich. The Rawlsian social planner increases the tax burden for all participating individuals, in order to reach maximum support levels for the unemployed.

Figure 16 depicts the cumulative employment rate up to each ability level for the current tax schedule and the two optimal tax schedules. As can be seen, total employment decreases by about 20% when the Rawlsian tax schedule would be implemented. On the other hand, the optimal utilitarian tax schedule increases total employment by around 10%. The slope of both lines is about equal, which indicates that the large differences in employment are caused by differences in benefit levels. The difference in marginal tax rates is less important.

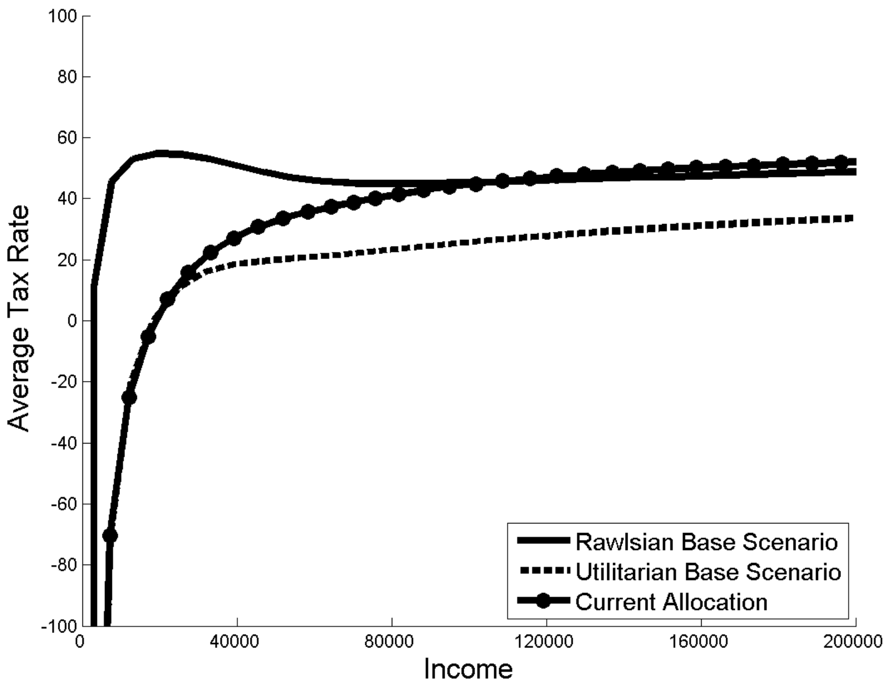

Figure 15: The average tax rates for different social preferences

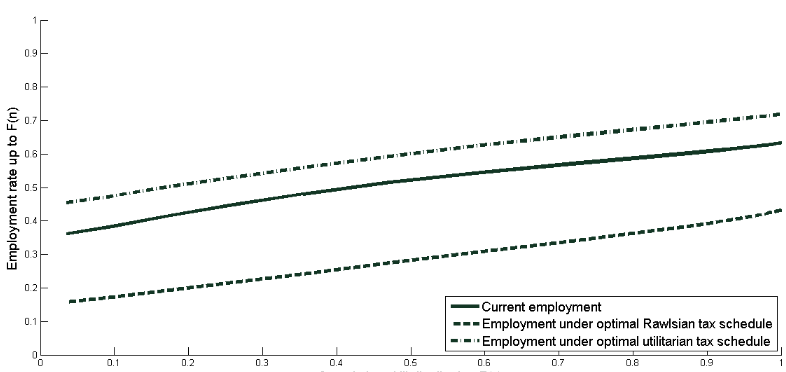

Figure 16: Cumulative employment under the current and optimal tax schedules

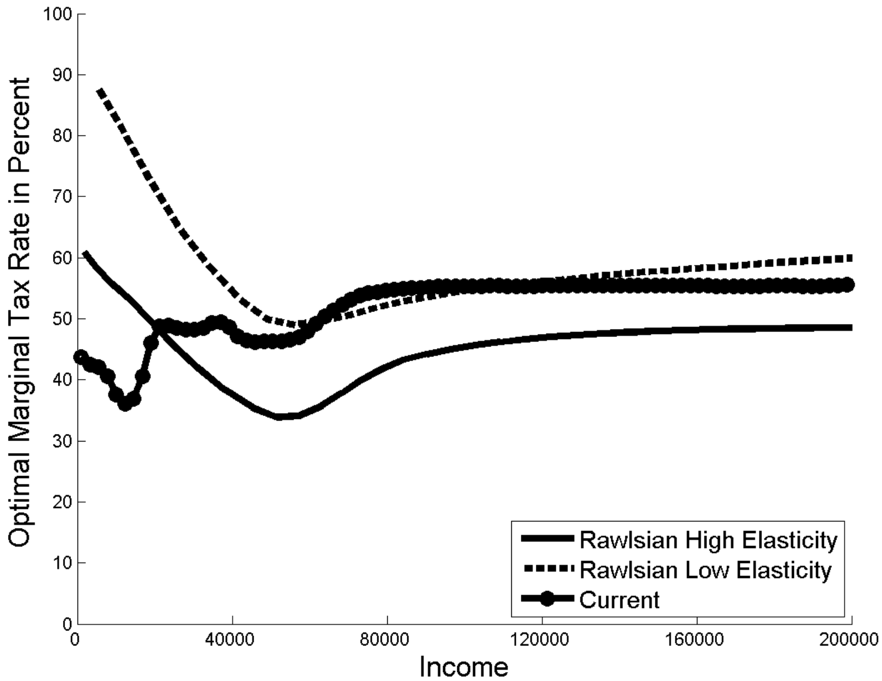

Figure 17: The optimal tax schedule under Rawlsian social preferences, with intensive and extensive labor-supply responses. Low: $\varepsilon^c = 0.18$, $\varepsilon^u = 0.13$ and $\varepsilon^P = 0.13$. High: $\varepsilon^c = 0.53$, $\varepsilon^u = 0.38$ and $\varepsilon^P = 0.38$.

5.3.1 Sensitivity analysis

We should note that the results with an extensive margin in general should be interpreted with the appropriate care. We have only limited knowledge on participation rates by skill, and on participation elasticities by skill. Consequently, there is significant uncertainty surrounding our estimates for the distribution of participation costs. Therefore, we also conducted alternative simulations with different elasticities. Figures 17, 18 give the optimal tax rates under a scenario with low intensive elasticities($\varepsilon^c = 0.18$ and $\varepsilon^u = 0.13$) and a low extensive elasticity ($\varepsilon^P = 0.13$), and a scenario with high intensive elasticities ($\varepsilon^c = 0.53$ and $\varepsilon^u = 0.38$) and a high extensive elasticity ($\varepsilon^P = 0.38$).

From Figures 17 and 18 we conclude once more that the current tax rate at the top is above (below) the optimal tax rate if the elasticity of taxable income is high (low) and the government has a utilitarian (Rawlsian) preferences. The result that the tax rate at the bottom is too low remains valid even for a utilitarian social welfare function and a high elasticity.

5.4 Single earners vs. all earners

We explore the robustness of one of our main conclusions – that current marginal tax rates are too low for the low-income earners – by focusing at single-earning individuals. Although our analysis

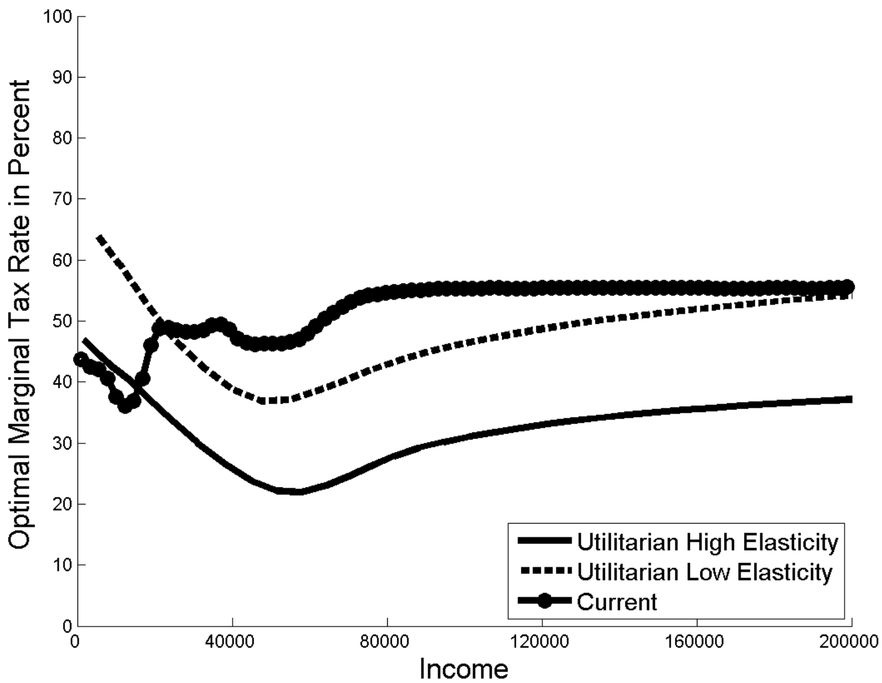

Figure 18: The optimal tax schedule with utilitarian social preferences, with intensive and extensive labor-supply responses. Low: $\varepsilon^c = 0.18$, $\varepsilon^u = 0.13$ and $\varepsilon^P = 0.13$. High: $\varepsilon^c = 0.53$, $\varepsilon^u = 0.38$ and $\varepsilon^P = 0.38$.

applies to the average of all tax payers, it might not apply to all types of tax payers. Since single-earner households are generally more reliant on income-dependent support than other tax payers are, their current marginal tax rates at low incomes are larger than for the average earner.

The kernel estimate of the non-linear tax schedule indeed masks a lot of heterogeneity, as the scatter plot in Figure 3 reveals. Single earners are located at the upper-left corner of Figure 3, but their marginal tax rates are completely smoothed out in the kernel estimate of the tax schedule. This is due to a large group of secondary earners, in the bottom-left corner, having low earnings, and facing low marginal tax rates, since they are not receiving income-dependent support.

Therefore, we not only plotted the current effective marginal tax schedule for all tax payers in the graphs with optimal tax schedules, but also the current effective marginal tax schedule for single earners, see Figures 5, 6, 11 and 12. We thus smoothed the marginal tax rates for single-earning individuals only.[31] The tax schedules of the average income earner and the single-income earner are very close, but the tax schedule for the single-income earner at very low earnings indeed features higher marginal tax rates.

Consequently, when designing policy reforms, one should keep in mind that the single-earning individuals already face marginal tax rates that are closer to the optimal non-linear schedule than most other tax payers. Moreover, the marginal tax rates of the current tax schedule would be somewhat too high for single earners under utilitarian social preferences, with a very weak social preference for redistribution. Still, marginal taxes would still be much too low for single earners for a very redistributive Rawlsian social welfare function. Hence, even for moderately redistributional concerns our conclusion would survive.

A final caveat is that primary earners have much lower labor-supply elasticities than secondary earners, see also the review of our labor-supply stimates. This may also undermine our conclusion that low-income earners face too low marginal tax rates, since we assumed that all individuals have the same labor-supply elasticity. Quite some elastic secondary earners in small part-time jobs could be located at the lower end of the earnings distribution. Hence, optimal taxes could be lower than in our simulations.[32] Nevertheless, the ETI-estimates discussed earlier also revealed that the elasticities of taxable income are roughly flat over the entire earnings distribution. Hence, it remains unclear whether our simulations are indeed biased.

[31] We did not re-estimate the skill distribution and re-compute optimal tax schedules for single earners. The reason is that the actual Dutch tax schedule is individualized and not dependent on whether individuals are single earners or not. Hence, the optimal tax schedule would remain the same. Moreover, if we re-estimate the skill distribution based on the marginal tax schedule for single-earners and then re-compute optimal tax schedules, the resulting tax schedules are indistinguishable from the reported ones, since the marginal tax schedules for single-earners are very close to the marginal tax schedules for all income earners.

[32] Our simulation model cannot cope with preference heterogeneity resulting in different labor-supply elasticities for different groups of income earners. Consequently, future research should explore the sensitivity of our conclusions with respect to more elastic secondary earners.

6 Flat tax

The flat income tax consists of a flat tax rate, which finances a non-individualized lump-sum transfer ($-T(0)$) in the model with an intensive margin only. In the model with both labor-supply margins the flat tax finances both the transfer for the working population ($-T(0)$) and the non-employment benefit (b).

We derive that a flat tax is clearly not desirable. The optimal linear tax rates are always higher than the income-weighted marginal tax rates under the optimal non-linear schedule with an intensive margin only, as the dashed lines in Figures 5 and 6 demonstrate. The optimal flat tax is still higher when both intensive and extensive margins are included for the Rawlsian government, but almost correspond to the weighted average of non-linear tax rates with a utilitarian government, see Figures 11 and 12. Moreover, the amount of income transferred to the working and non-working poor is always lower under the optimal flat tax compared to the non-linear tax in all simulations, but one, see Table 10. Only in the utilitarian case with both intensive and extensive labor-supply margins the optimal transfer provided to the non-working poor is slightly higher under the flat tax. Hence, the flat tax entails either less efficiency or less equity or both.

Intuitively, in order to organize a given amount of redistribution, the linear tax always requires higher marginal tax rates, since the lump-sum transfers are provided to everyone, irrespective of income. The flat tax cannot precisely target transfers to different income groups so that the leaking bucket of Okun is leaking more when a flat income tax is employed rather than the non-linear income tax. The flat income tax is therefore an inferior instrument for income redistribution.

Table 10: Comparison optimal non-linear tax with optimal flat tax and current tax system

	Intensive margin only			Intensive + extensive margin		
	$-T(0)$	DWL	Welfare loss	$-T(0)$	b	DWL
Rawlsian						
Optimal non-linear	21,105	0.82	0.000	1,891	11,490	0.27
Optimal linear	18,587	1.00	0.090	795	10,801	0.49
Current	13,268	0.37	0.278	8,086	8,086	0.41
utilitarian						
Optimal non-linear	12,689	0.15	0.000	9,055	6,800	0.17
Optimal linear	9,220	0.16	0.004	4,045	6,879	0.18
Current	13,268	0.37	0.075	8,086	8,086	0.41

We calculate the marginal deadweight losses of the optimal non-linear tax, the optimal linear tax and the current tax-benefit system in Table 10.[33] Clearly, the marginal deadweight losses are always lower under an optimal non-linear tax system in comparison to the optimal flat tax. Moreover, our simulations with both extensive and intensive labor-supply margins demonstrate that moving from

[33]The general formula for the marginal deadweight loss under a non-linear income tax is derived in the appendix and equals: $\int_{\mathcal{N}} \varepsilon_n^c \frac{T'(z_n)}{1-T'(z_n)} n l_n \tilde{k}(n) \, \mathrm{d}n \left(\int_{\mathcal{N}} n l_n \tilde{k}(n) \, \mathrm{d}n \right)^{-1}$.

the current tax-benefit system to the optimal non-linear tax lowers the marginal deadweight loss from about 41 cent per euro, to 27 cent per euro in the Rawlsian case to 17 cent per euro in the utilitarian case. Hence, reforming the current tax-benefit system towards the optimal non-linear system almost halves the distortions of the tax system for any social preference for redistribution.

We also compute the welfare loss of the flat tax system in comparison with optimal non-linear tax system, see Table 10. Similarly, we also calculate the welfare loss of the current tax system in comparison with the optimal non-linear tax system. Due to computational complexities, we were only able to do so with model with an intensive margin.[34] Our welfare measure is the compensating variation: how much resources can be taken out of the economy with optimal non-linear taxes to achieve the same level of social welfare as in the economy with an optimal flat tax or the current tax-benefit system? The welfare cost of moving from the optimal non-linear taxes to the optimal flat tax with utilitarian social preferences is 0.4% of GDP, which is relatively modest. But, one needs to recall that the utilitarian government is only weakly redistributive. The welfare loss for the Rawlsian government is a very large 9% of GDP. Intuitively, when social preferences are more redistributive, the flat tax is more of a strait jacket to the government to achieve its redistributional objectives. The welfare difference between the optimal non-linear tax system and the current tax-benefit system is 7.5% of GDP under utilitarian social preferences and an astonishing 28% of GPD under Rawlsian social preferences. This, again, demonstrates the sub-optimality of the current tax-benefit system in achieving social objectives. An important caveat is in order here. These welfare analyses are conducted for the model with an intensive margin only. Hence, they should be interpreted as an upper bound of the potential welfare losses of not implementing the optimal non-linear tax schedule.

7 Social welfare weights current tax-benefit system

In the previous sections, we have derived the optimal tax schedule for given social preferences. In this section, we invert the question: under what social preferences is the current tax schedule optimal? Under any Bergson-Samuelson social welfare function, welfare weights are monotonically declining in income. In addition, all welfare weights are non-negative. However, political-economy considerations might induce politicians to set a tax schedule which attaches more weight to the middle-income groups. Policy makers might also have under- or overestimated distortions associated with the current tax schedule. We investigate whether inconsistencies in social welfare weights are present in the Netherlands.

7.1 Results

In figure 7.1, we plot the social welfare weights implied by the current tax-benefit system had it been optimized. We do so for the model with an intensive margin only and the model with both intensive and extensive margins. The social welfare weights under the extensive margin reveal the

[34]In a future version of this paper we hope to report these results.

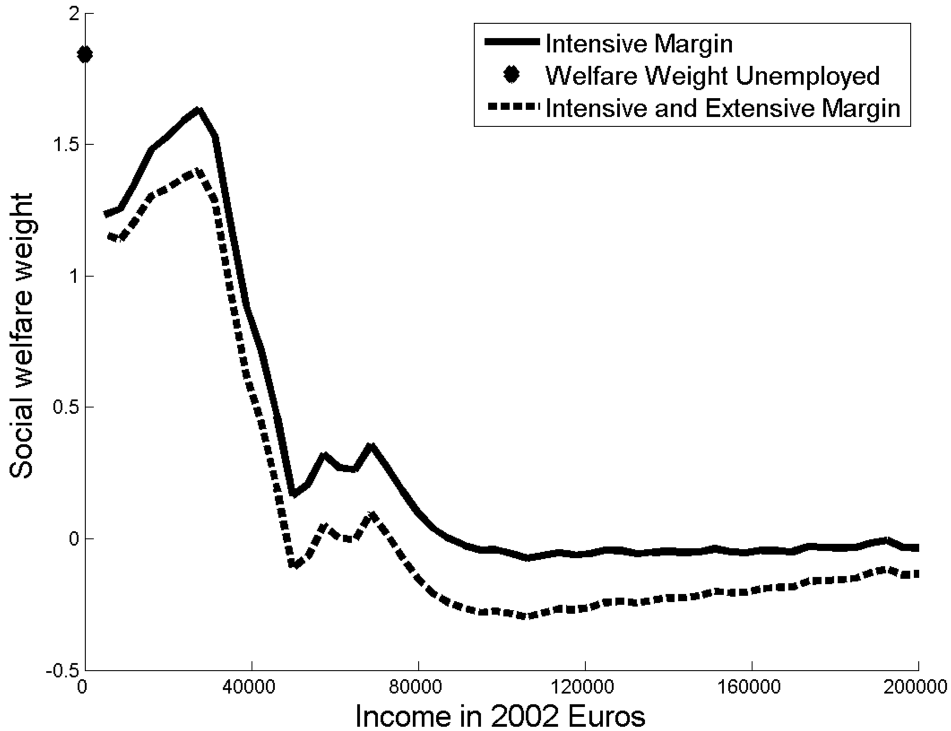

Figure 19: Social welfare weight intensive labor supply responses and under both responses

same patterns as the social welfare weights with an intensive margin only, although the increase in the welfare weights between low-income workers and middle-income workers is less prominent. As can be seen, social welfare weights under both labor-supply responses are generally below the social welfare weights under the intensive margin. For a given tax rate, distortions are larger if individuals can also respond on the extensive margin, and social welfare weights are lower. There are three clear inconsistencies in these patterns of social welfare weights.

First, in both graphs the social welfare weights are increasing until modal income. Indeed, the political system attaches the largest social welfare weights to the middle-income groups. This implies that the current government positively values taxing the working poor to redistribute more resources towards the middle-income groups. This is inconsistent with any standard social welfare function, which attaches a lower welfare weight to middle-income earners than to the working poor. These results suggest that political-economy considerations can be important in explaining current tax schedules. Indeed, the densely populated middle-income groups constitute the largest fraction of the Dutch electorate.

Second, for the high income levels, the welfare weights are slightly negative, and return slightly above zero in the very limit. Apparently, the current government values penalizing the high-income earners. The reason is that the current tax rate in the top bracket is set beyond the top of the

Laffer curve for most top-income tax payers.[35] Such a policy produces no redistributional benefits and only distortions. Negative welfare weights at the top of the income distribution are in line with findings in Bourguignon and Spadaro (2012) for France.

Third, there is a large discontinuous drop in the welfare weight for the poor as they start working and earning income. In particular, the current government values a euro transferred to the non-working poor 1.5 times as high as transferring the same euro to the working poor. This is an anomaly as it suggests that the government views the non-working poor as much more deserving of income support than the working poor even if they have the same income. Low welfare weights for the working poor have been consistently found in other studies, see e.g. Bourguignon and Spadaro (2012) and Bargain et al. (2011).

Figures 20 and 21 show the welfare weights in the case of a high and low labor-supply elasticity. As can be seen from the figures, welfare weights become close to monotonically decreasing if the elasticity of taxable income is very low. In addition, all welfare weights will then be positive. A possible explanation for the anomalies in our baseline simulation is that policy-makers underestimate the efficiency costs of taxes and do not optimize tax-benefit systems accordingly. On the other hand, if the earnings-supply elasticity is high, the non-monotonicity in the welfare weights is even more striking. Also, the welfare weights at the bottom are much lower and welfare weights for top-income earners are even more negative.

8 Directions for future research

We assumed that all worker types are perfect substitutes and that there are, therefore, no general-equilibrium effects on the wage structure as a result of redistribution policy, or otherwise. Rothschild and Scheuer (2013) extend the Stiglitz (1982) model of optimal income taxation with endogenous wages to an infinite number of skill types. These authors demonstrate that redistributive governments should exploit general-equilibrium effects on the wage structure by setting less progressive marginal tax schedules. Optimal tax rates would increase at the bottom and decrease high at the top. When applied to the Netherlands, this would presumably render the current tax-benefit system even more sub-optimal than our analysis has demonstrated.

Recent developments in behavioral economics point to a number of potential weaknesses of our analysis. It might be that individuals are engaged in 'rat races' (Akerlof, 1976) and 'keeping up with the Joneses' (Layard, 1980). Distortionary income taxes then not only entail deadweight losses, but also yield benefits by taming the rat race or correcting status-seeking behavior. Total distortions of income taxation are then smaller, and optimal taxes increase. See also Kanbur et al. (2006). By the same token, Alesina et al. (2005) argue that there could be rivalry in leisure as well. This raises distortions of income taxation, since not only labor-supply choices are distorted, but also a 'leisure multiplier' is put in motion. Hence, taxes should optimally be set lower. Gerritsen (2013) shows that utility-maximizing individuals might not maximize well-being, and, hence, suffer

[35] Only for tax-payers with incomes above 200,000 euro the social welfare weights turn marginally positive. This explains why the revenue-maximizing top-rate is still marginally above the current top rate.

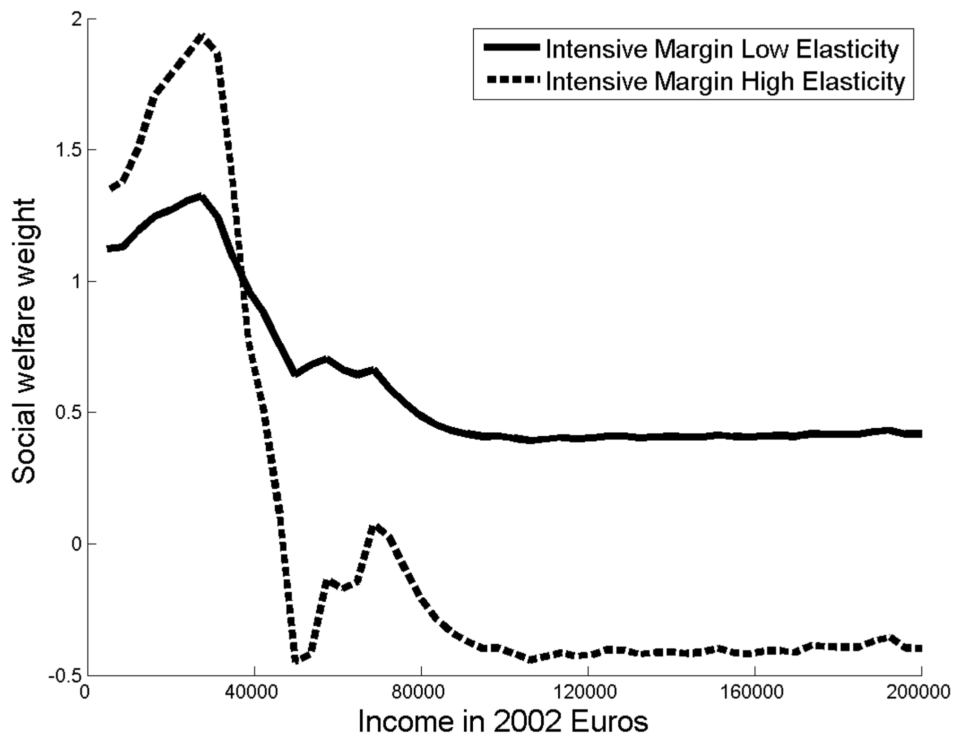

Figure 20: Social welfare weight under intensive labor supply responses

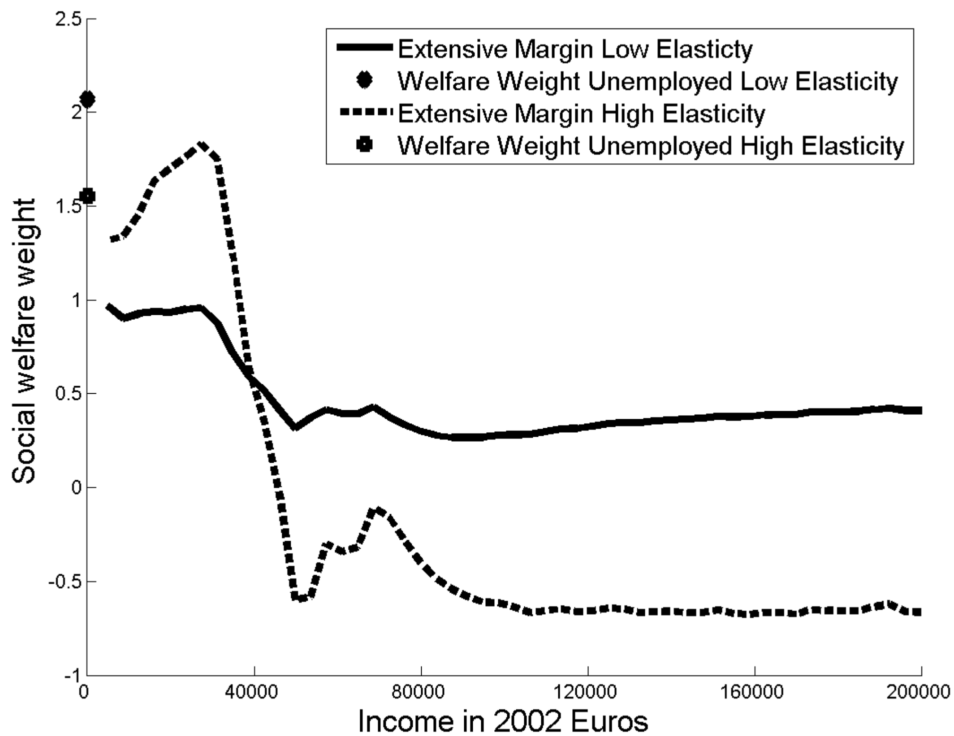

Figure 21: Social welfare weight under both intensive and extensive labor supply responses

from internalities. He finds that marginal tax rates should be lowered for the poor to give them stronger incentives to work more, whereas they should be increased to the rich, to give them stronger incentives to enjoy more leisure.

In our analysis we ignore that many individuals live in multi-person households. We thereby ignore, for example, intrahousehold redistribution and economies of scale. An analysis of optimal family taxation, following the lead by *e.g.* Boskin and Sheshinski (1983), Apps and Rees (1998), Schroyen (2003), Alesina et al. (2011) and Kleven et al. (2009), is beyond the scope of this paper. Indeed, it would be useful to explore conditioning tax schedules on the income of primary and secondary earners, since the latter are typically more elastic. This is left for future research.

9 Conclusions

This study analyzed the optimal redistributive tax and transfer system in the Netherlands using realistically calibrated models with both intensive and extensive margins of labor supply. We found that the optimal non-linear tax schedule features a U-shape. This contrasts sharply with the current schedule of effective marginal tax rates in the Netherlands; tax rates are gradually increasing with income. Although the optimal marginal tax rates at the bottom fall significantly when an extensive margin is introduced, we find that marginal tax rates are too low at the bottom of the earnings distribution compared to the current tax schedule for all the social welfare functions we analyzed. Higher marginal tax rates until modal income thus help to redistribute more income towards the working and non-working poor. Marginal tax rates for the middle-income earners are too high. Also, the top tax rate appears to be set too high, and even on the wrong side of the Laffer-curve. The observed patterns in marginal taxes suggest that the middle-income earners are undertaxed, at the expense of the top-income earners and the working poor.

A central finding in all our simulations is that the working poor should pay much lower average taxes. However, this does not imply that they receive a net subsidy to work. A large participation subsidy is found only under weakly redistributive social objectives. Already for moderately redistributive preferences we find that there should always be a net participation tax for the working poor.

A flat income tax schedule is never found to be optimal. Indeed, all simulations demonstrate the inferiority of the flat tax to redistribute income. Under an optimal flat tax, marginal tax rates are higher, transfers/benefits are lower or both. Hence, the equity-efficiency trade-off worsens substantially. Simulations of the model with an intensive margin only demonstrate that an optimal flat tax gives substantial welfare losses compared to the optimal non-linear tax, running from 0.4% of GDP for utilitarian to 9% of GDP for Rawlsian social preferences. The flat tax is a particularly costly strait jacket for strongly redistributive governments. The marginal deadweight loss of the current tax system (41 cents per additional euro revenue) is roughly cut in half when the optimal non-linear tax schedule would be implemented for any social desire to redistribute income. A flat tax renders the leaking bucket of Okun (1975) a sieve. Hence, political discussions about a flat tax are therefore an economic non-starter.

The social welfare weights underlying the current tax-benefit system give rise to similar conclusions. Dutch social welfare weights are increasing with income until median income. The government thus prefers transferring resources to middle-income earners rather than the working poor. Moreover, social welfare weights for top-income earners are slightly negative. This implies that the current government likes to penalize top-income earners by setting too high marginal tax rates. Finally, the government attaches a much larger welfare weight to the non-working poor than the working poor. Why the working poor are apparently less deserving of income support than the non-working poor – even if they have the same income – remains unclear to us.

The policy implications of our research are clear. The government should lower the tax burden on the working poor, by raising the tax burden on the middle- and higher-income groups. This can raise social welfare under all standard social welfare criteria we analyzed. This is typically not a Pareto improvement, since middle- and higher-income earners need to pay higher taxes. However, tax reforms are feasible where the welfare gains for the low-income groups outweigh the welfare losses for the middle- and higher-income groups. Substantial increases of the EITC therefore appear to be socially desirable. By exactly how much is a political judgement. The top rate should not be increased further, as it would only increase deadweight losses while reducing tax revenue available for income redistribution.

Appendix

Optimal income taxation with intensive margin only

We will solve the optimal income tax using Lagrangian methods. Multiply the incentive constraint with θ_n and apply integration by parts to $\theta_n \frac{\mathrm{d}u_n}{\mathrm{d}n}$ so as to find:

$$\int_{\mathcal{N}} \left(-\theta_n \frac{z_n h'(z_n/n)}{n^2} - u_n \frac{\mathrm{d}\theta_n}{\mathrm{d}n} \right) \mathrm{d}n + \theta_{\overline{n}} u_{\overline{n}} - \theta_{\underline{n}} u_{\underline{n}} = 0. \tag{40}$$

Now, set up the optimal-tax problem as a Lagrangian with c_n, z_n, and u_n as control variables. We furthermore introduce λ as the Lagrange multiplier of the economy's resource constraint. $\eta_n f(n)$ denotes the composite Lagrange multiplier of the utility constraint at n (we have harmlessly pre-multiplied each multiplier η_n with $f(n)$ to avoid some additional notation). θ_n is the Lagrange multiplier of the incentive-compatibility constraint at n:[36]

$$\begin{aligned}\mathcal{L} \equiv & \int_{\mathcal{N}} (W(u_n) + \lambda (z_n - c_n - R)) f(n) \mathrm{d}n + \int_{\mathcal{N}} \eta_n (v(c_n) - h(z_n/n) - u_n) f(n) \mathrm{d}n \\ & - \int_{\mathcal{N}} \left(\theta_n \frac{z_n h'(z_n/n)}{n^2} + u_n \frac{\mathrm{d}\theta_n}{\mathrm{d}n} \right) \mathrm{d}n + \theta_{\overline{n}} u_{\overline{n}} - \theta_{\underline{n}} u_{\underline{n}}. \end{aligned} \tag{41}$$

[36] We need the latter constraint because all variables in the utility function c and z as well as utility itself u are considered choice variables for the government in this optimization procedure. Alternatively, one may invert the utility function and write consumption as a function of the allocation: $c(z, u)$, which is usually done in the literature.

The first-order and transversality conditions for this control problem are given by:

$$\frac{\partial \mathcal{L}}{\partial c_n} = 0 \quad : \quad -\lambda f(n) + \eta_n f(n) v'(c_n) = 0, \quad \forall n, \tag{42}$$

$$\frac{\partial \mathcal{L}}{\partial z_n} = 0 \quad : \quad \lambda f(n) - \eta_n \frac{h'(l_n)}{n} - \theta_n \frac{h'(l_n) + l_n h''(l_n)}{n^2} = 0, \quad \forall n, \tag{43}$$

$$\frac{\partial \mathcal{L}}{\partial u_n} = 0 \quad : \quad W'(u_n) f(n) - \eta_n - \frac{\mathrm{d}\theta_n}{\mathrm{d}n} = 0, \quad \forall n \neq \underline{n}, \overline{n}, \tag{44}$$

$$\lim_{n \to \underline{n}} \theta_n = 0, \quad \lim_{n \to \overline{n}} \theta_n = 0. \tag{45}$$

We omitted restating the incentive-compatibility and resource constraints. We now derive the optimal tax formula as reported in Saez (2001).

First, solve (42) for η_n to find $\eta_n = \frac{\lambda}{v'(c_n)}$, and substitute this into (43) and simplify:

$$1 - \frac{h'(l_n)}{n v'(c_n)} = \frac{\theta_n \left(h'(l_n) + l_n h''(l_n) \right)}{\lambda f(n) n^2}. \tag{46}$$

Substitute the individuals' FOC (4) into (46) and simplify the resulting equation:

$$\frac{T'(z_n)}{1 - T'(z_n)} = \left(1 + \frac{l_n h''(l_n)}{h'(l_n)} \right) \frac{\theta_n v'(c_n)/\lambda}{(1 - F(n))} \frac{1 - F(n)}{f(n) n}. \tag{47}$$

For the utility function we used, the compensated and uncompensated labor supply elasticities are given by (see last Appendix):

$$\varepsilon_n^c \equiv -\frac{\partial l_n}{\partial \tau} \frac{1 - T'}{l_n} = \frac{v'}{\frac{l h'' v'}{h'} - \frac{l h' v''}{v'} + v' n l \frac{T''}{1 - T'}}, \tag{48}$$

$$\varepsilon_n^u \equiv \frac{\partial l_n}{\partial n} \frac{n}{l_n} = \frac{v' + \frac{l h' v''}{v'} - v' n l \frac{T''}{1 - T'}}{\frac{l h'' v'}{h'} - \frac{l h' v''}{v'} + v' n l \frac{T''}{1 - T'}}. \tag{49}$$

Therefore, we find

$$\frac{1 + \varepsilon_n^u}{\varepsilon_n^c} = 1 + \frac{l_n h''(l_n)}{h'(l_n)}. \tag{50}$$

In addition, integrating equation (44), and using a transversality condition, a solution for θ_n can be derived:

$$\theta_n = \int_n^{\overline{n}} \left(\frac{\lambda}{v'(c_m)} - W'(u_m) \right) f(m) \, \mathrm{d}m. \tag{51}$$

By introducing normalized social welfare weight $g_n \equiv \frac{W'(u_n) v'(c_n)}{\lambda}$, which denotes the monetized welfare gain of providing one euro to individual n, this expression can be simplified:

$$\theta_n = \lambda \int_n^{\overline{n}} \frac{(1 - g_m)}{v'(c_m)} f(m) \, \mathrm{d}m. \tag{52}$$

Substituting results (50) and (52) into (47). The final constraints on θ are the transversality conditions. Note that the transversality conditions imply that the distortion on labor supply at the top and the bottom should equal zero.

Optimal income taxation with both intensive and extensive margins

The incentive-compatibility constraint is unaffected by introducing the extensive margin. We will solve the optimal income tax again using a Lagrangian, which uses c_n, z_n, and u_n as control variables. λ is the Lagrange multiplier of the economy's resource constraint, $\eta_n f(n)$ denotes the composite Lagrange multipliers on the utility constraint for each n, and θ_n is the Lagrange multiplier of the incentive-compatibility constraint at n. The Lagrangian can be written as:

$$\begin{aligned}
\mathcal{L} \;=\; & \int_{\mathcal{N}} \left(\int_{\underline{\varphi}^n}^{u_n - v(b)} W(u_n - \varphi) k(\varphi|n) \mathrm{d}\varphi f(n) + W\left(v(b)\right)\left(f(n) - K(u_n - v(b)|n) f(n)\right) \right) \mathrm{d}n \\
& + \int_{\mathcal{N}} \lambda \left((z_n - c_n) K(u_n - v(b)|n) f(n) - (f(n) - K(u_n - v(b)|n) f(n)) b) - R f(n) \right) \mathrm{d}n \\
& + \int_{\mathcal{N}} \eta_n \left(v(c_n) - h(z_n/n) - u_n \right) f(n) \mathrm{d}n \\
& - \int_{\mathcal{N}} \left(\theta_n \frac{z_n h'\left(z_n/n\right)}{n^2} + u_n \frac{\mathrm{d}\theta_n}{\mathrm{d}n} \right) \mathrm{d}n + \theta_{\overline{n}} u_{\overline{n}} - \theta_{\underline{n}} u_{\underline{n}}.
\end{aligned} \tag{53}$$

where we substituted the definition for $\tilde{k}(n) \equiv K(u_n - v(b)|n) f(n)$. The first-order and transversality conditions for this control problem are given by:

$$\frac{\partial \mathcal{L}}{\partial c_n} = 0 \;\;:\;\; -\lambda \tilde{k}(n) + \eta_n v'(c_n) f(n) = 0, \quad \forall n, \tag{54}$$

$$\frac{\partial \mathcal{L}}{\partial z_n} = 0 \;\;:\;\; \lambda \tilde{k}(n) - \eta_n \frac{h'(l_n)}{n} - \theta_n \frac{h'(l_n) + l_n h''(l_n)}{n^2} = 0, \quad \forall n, \tag{55}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial u_n} = 0 \;\;:\;\; & \int_{\underline{\varphi}^n}^{u_n - v(b)} W'(u_n - \varphi) k(\varphi|n) \mathrm{d}\varphi f(n) + \\
& \lambda \kappa_n \left(T(z_n) + b\right) \tilde{k}(n) - \eta_n - \frac{\mathrm{d}\theta_n}{\mathrm{d}n} = 0, \quad \forall n \neq \underline{n}, \overline{n},
\end{aligned} \tag{56}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial b} = 0 \;\;:\;\; & \int_{\mathcal{N}} \left(W'(v(b)) v'(b) (f(n) - \tilde{k}(n)) - \lambda (f(n) - \tilde{k}(n)) \right) \mathrm{d}n, \\
& -\lambda \int_{\mathcal{N}} \kappa_n v'(b) \left(T(z_n) + b\right) \tilde{k}(n) \mathrm{d}n = 0,
\end{aligned} \tag{57}$$

$$\lim_{n \to \underline{n}} \theta_n = 0, \quad \lim_{n \to \overline{n}} \theta_n = 0, \tag{58}$$

where we used $\kappa_n \equiv \frac{K'(u_n - v(b)|n) f(n)}{\tilde{k}(n)}$, which denotes the semi-elasticity of participation with respect to a utility increase for employed. We employed Leibniz' rule in the first-order conditions for u_n and b to find the derivatives of the Lagrangian with respect to the the bound $(u_n - v(b))$ of the integrals. We also used $T(z_n) = z_n - c_n$ to simplify the first-order conditions for u_n and b. We omitted restating the incentive-compatibility and resource constraints.

We can solve for the modified ABC-formula using the same procedure as for the intensive margin. First, solve for η_n using equation (54): $\eta_n = \frac{\lambda \tilde{k}(n)}{v'(c_n) f(n)}$ and substitute this into (55) and simplify:

$$1 - \frac{h'(l_n)}{nv'(c_n)} = \frac{\theta_n \left(h'(l_n) + l_n h''(l_n)\right)}{\lambda \tilde{k}(n) n^2}. \tag{59}$$

Substitute (59) in first-order condition (4) and rewrite:

$$\frac{T'(z_n)}{1 - T'(z_n)} = \left(1 + \frac{l_n h''(l_n)}{h'(l_n)}\right) \frac{\theta_n v'(c_n)/\lambda}{\tilde{K}(\overline{n}) - \tilde{K}(n)} \frac{\tilde{K}(\overline{n}) - \tilde{K}(n)}{n \tilde{k}(n)}, \tag{60}$$

where $\tilde{K}(n) = \int_{\underline{n}}^{n} \tilde{k}(m)\, \mathrm{d}m$ is the fraction of employed workers in the population with skill level n or less.

Integrate equation (56), and use the transversality condition, to find the solution for θ_n:

$$\theta_n = \int_n^{\overline{n}} \lambda \left(\frac{1}{v'(c_m)} - \kappa_m (T(z_m) + b)\right) \tilde{k}(m)\, \mathrm{d}m - \int_n^{\overline{n}} \int_{\underline{\varphi}^n}^{u_m - v(b)} W'(u_m - \varphi)\, k(\varphi|n) \mathrm{d}\varphi f(m)\, \mathrm{d}m. \tag{61}$$

Use the expected, conditional welfare weight of an individual with ability n
$g_n^P \equiv \int_{\underline{\varphi}^n}^{u_n - v(b)} \frac{W'(u_n - \varphi) v'(c_n)}{\lambda} k(\varphi|n) \mathrm{d}\varphi / K(u_n - b)$ and simplify (61):

$$\theta_n = \int_n^{\overline{n}} \lambda \left(\frac{1 - g_m^P}{v'(c_m)} - \kappa_m (T(z_m) + b)\right) \tilde{k}(m)\, \mathrm{d}m, \tag{62}$$

Finally, combine expressions (62), (60), and (50) to obtain the adjusted ABC-formula (18).

In addition, equation (57) describes an optimality condition for unemployment benefits b. It can be simplified by solving the integrals and introducing the marginal social welfare weight g_0 of unemployed individuals: $g_0 \equiv W'(v(b)) v'(b)/\lambda$. Use g_0 to simplify (57):

$$(g_0 - 1)(1 - \tilde{K}(\overline{n})) = v'(b) \int_{\mathcal{N}} \kappa_m (T(z_m) + b) \tilde{k}(m) \mathrm{d}m. \tag{63}$$

Simplify the right-hand side by imposing the transversality condition at the top:

$$\theta_{\overline{n}} = \int_{\mathcal{N}} \lambda \left(\frac{g_m^P - 1}{v'(c_m)} + \kappa_m (T(z_m) + b)\right) \tilde{k}(m) \mathrm{d}m = 0. \tag{64}$$

From (64) then follows the expression for the optimal participation tax:

$$v'(b) \int_{\mathcal{N}} \kappa_n (T(z_n) + b) \tilde{k}(m) \mathrm{d}m = v'(b) \int_{\mathcal{N}} \frac{(1 - g_m^P) \tilde{k}(m) \mathrm{d}m}{v'(c_m)}. \tag{65}$$

Use (65) to simplify (63):

$$\frac{(g_0 - 1)(1 - \tilde{K}(\overline{n}))}{v'(b)} = \int_{\mathcal{N}} \frac{(1 - g_m^P) \tilde{k}(m) \mathrm{d}m}{v'(c_n)}. \tag{66}$$

Deriving behavioral elasticities

This appendix derives the exact elasticities – taking into account the non-linearity of the tax system, as in Jacquet et al. (2013). Individuals maximize utility $u(c,l)$ subject to their budget constraint $c = nl - T(nl)$. The first-order condition (FOC) is given by $n(1-T')u_c(.) + u_l(.) = 0$. Define the following function

$$\begin{aligned} Y(l,n,\tau,\rho) \quad &\equiv \quad n\left(1 - T'(nl) + \tau\right) u_c\left(nl - T(nl) + \tau(nl - nl_n) + \rho, l\right) \\ &\quad + u_l\left(nl - T(nl) + \tau(nl - n\hat{l}) + \rho, l\right). \end{aligned} \tag{67}$$

$Y(z,n,\tau,\rho)$ measures the shift of the first-order condition of the household when the marginal tax rate exogenously increases with τ (i.e., for any level of earnings) or when the household receives an exogenous amount of income ρ, irrespective of the amount of work effort. The first-order condition of the household is equivalent to $Y(z_n,n,0,0) = 0$. Introducing the second term, $\tau(nl - n\hat{l})$, has the following intuition. Suppose we raise the marginal tax rate – irrespective of income level nl – and we evaluate the impact at $n\hat{l}$ (the optimum choice for $\hat{l}$ of household n), then this marginal tax increase does not change income, only the marginal incentives to supply labor. ρ represents the income effect: suppose that we give the household a marginal increase in income of ρ, starting from $\rho = 0$, what will happen to labor supply?

We find the following derivatives, using the first-order condition $-u_l = n(1-T')u_c$:

$$Y_l(l_n,n,0,0) \quad = \quad u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\frac{u_l}{u_c}u_{cl} + nu_l\frac{T''}{1-T'}, \tag{68}$$

$$Y_n(l_n,n,0,0) \quad = \quad \left(-u_l/l + nu_l\frac{T''}{1-T'} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - \left(\frac{u_l}{u_c}\right)u_{lc}\right)\frac{l}{n}, \tag{69}$$

$$Y_\tau(l_n,n,0,0) \quad = \quad nu_c, \tag{70}$$

$$Y_\rho(l_n,n,0,0) \quad = \quad n\left(1-T'\right)u_{cc} + u_{lc} = \frac{u_{lc}u_c - u_l u_{cc}}{u_c}. \tag{71}$$

Now, by applying the envelope theorem we find

$$\frac{\partial l}{\partial x} = -\frac{Y_x}{Y_l}, \quad x = n,\tau,\rho \tag{72}$$

Hence, the uncompensated *wage* elasticity of labor supply ε^u is equal to:

$$\varepsilon^u \equiv \frac{\partial l}{\partial n}\frac{n}{l} = \frac{u_l/l + \left(\frac{u_l}{u_c}\right)u_{lc} - \left(\frac{u_l}{u_c}\right)^2 u_{cc} - nu_l\frac{T''}{1-T'}}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\frac{u_l}{u_c}u_{cl} + nu_l\frac{T''}{1-T'}}. \tag{73}$$

The compensated *wage* elasticity of labor supply ζ^c is given by:

$$\zeta^c \equiv \frac{\partial l}{\partial n}\frac{n}{l} = \frac{u_l/l - nu_l\frac{T''}{1-T'}}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\frac{u_l}{u_c}u_{cl} - nu_l\frac{T''}{1-T'}}. \tag{74}$$

And, the compensated *tax* elasticity ε^c is:

$$\varepsilon^c \equiv -\frac{\partial l}{\partial \tau}\frac{1-T'}{l} = \frac{u_l/l}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\frac{u_l}{u_c}u_{cl} + nu_l\frac{T''}{1-T'}}. \tag{75}$$

Note that the compensated *wage* elasticity of labor supply and the compensated *tax* elasticity of labor supply are not identical due to the non-linearity in the tax system. Increasing the marginal tax rate τ amounts to increasing the marginal tax, irrespective of the income level, whereas increasing the wage rate also changes the marginal tax rates as a result of the non-linearities in the tax system.

The income elasticity of labor supply is defined by the Slutsky equation ($\eta \equiv \varepsilon^u - \zeta^c$):

$$\eta = \left(1-T'\right) n\frac{\partial l}{\partial \rho} = \frac{\frac{-u_l}{u_c}\left(\frac{u_l}{u_c}u_{cc} - u_{lc}\right)}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\frac{u_l}{u_c}u_{cl} + nu_l\frac{T''}{1-T'}}. \tag{76}$$

All the elasticities depend on the second derivatives of the tax function. Hence, in contrast to Saez (2001), the second-derivatives cannot be ignored in the expressions of the elasticities if tax systems are non-linear. We thus confirm Blomquist and Simula (2010). If $T'' > 0$ distortions of taxes are lower – ceteris paribus. However, if $T'' < 0$ the reverse is true. The reason is that if marginal tax rates are increasing ($T'' > 0$) the labor supply response dampens out, but if the marginal tax rates are decreasing ($T'' < 0$) the labor supply response is magnified by the non-linearity in the tax schedule.

Note that we can derive that

$$1 + \frac{lu_{ll}}{u_l} - \frac{lu_{lc}}{u_c} = \frac{1+\varepsilon^u}{\varepsilon^c} \tag{77}$$

Thus, the term $1 + \frac{lu_{ll}}{u_l} - \frac{lu_{lc}}{u_c}$ equals one plus the uncompensated *wage* elasticity of labor supply, divided by the compensated *tax* elasticity of labor supply. The former does include the impact of the non-linear tax schedule, whereas the latter does not. Only when the tax system is linear, this expression reduces to $\frac{1+\varepsilon^u}{\zeta^c}$ as in Saez (2001).

For the specific utility function $u(c,l) \equiv v(c) - h(l)$ we obtain the following elasticities:

$$\varepsilon^u = \frac{v' + \frac{lh'v''}{v'} - v'nl\frac{T''}{1-T'}}{\frac{lh''v'}{h'} - \frac{lh'v''}{v'} + v'nl\frac{T''}{1-T'}}, \tag{78}$$

$$\varepsilon^c = \frac{v'}{\frac{lh''v'}{h'} - \frac{lh'v''}{v'} + v'nl\frac{T''}{1-T'}}. \tag{79}$$

Marginal dead weight loss non-linear income tax

To determine the deadweight loss of a non-linear tax schedule $T(z_n)$ with $T' \equiv \mathrm{d}T(\cdot)/\mathrm{d}z_n$, suppose that we increase the marginal tax rates at each and every point of the tax schedule with $\mathrm{d}T'$, how large is the marginal deadweight loss of that tax increase? To answer this question, we conduct the following hypothetical thought experiment. Each household n gets perfectly compensated via a household-specific lump-sum transfer T_n so that its utility remains unaffected.[37] This implies that our deadweight loss measure is based on the compensating variation. The marginal deadweight loss then equals the net loss in public revenue so as to keep everyone's utility constant. Note that this hypothetical tax reform does not affect the participation margin, since the benefit given to non-working individuals b remains constant and the utility of all working individuals u_n does not change.

Indirect utility of all working individuals can be written as a function $v((1 - T'(nl_n))n, \tilde{T}_n)$ of the net marginal wage rate $(1 - T')n$, and so-called virtual income $\tilde{T}_n$. Virtual income is defined as

$$\tilde{T}_n \equiv nl_n - T(nl_n) - (1 - T'(nl_n))nl_n. \tag{80}$$

So that the household budget constraint can be written as:

$$c_n = nl_n - T(nl_n) = (1 - T'(nl_n))nl_n + \tilde{T}_n. \tag{81}$$

Note that virtual income works like the intercept of the tax function if the marginal tax rate T' had been constant. From applying Roy's identity we find that

$$\frac{\partial v_n}{\partial \tilde{T}_n} = \lambda_n, \quad \frac{\partial v_n}{\partial T'} = -\lambda_n nl_n, \tag{82}$$

The change in taxes $\mathrm{d}T'$ and lump-sum income $\mathrm{d}T_n$ for each household n, which leaves private utility unaffected satisfies:

$$\mathrm{d}v_n = \lambda_n \mathrm{d}T_n - \lambda_n nl_n \mathrm{d}T' = 0. \tag{83}$$

where we used the derivatives of indirect utility here. The transfers T_n play the same role as the virtual income $\tilde{T}_n$. Hence, the derivatives of indirect utility with respect to $\tilde{T}_n$ or T_n are identical. Consequently, when each individual gets a perfect compensation for the tax change, we have

$$\mathrm{d}T_n = nl_n \mathrm{d}T'. \tag{84}$$

What is the effect of this tax policy on the public budget? There are three effects. i) For each working individual n, the government loses revenue $\mathrm{d}T_n$. ii) When the tax rate increases, the government also gains revenue $nl_n\mathrm{d}T'$. iii) The individual will change its (compensated) labor supply in response to higher taxation. This results in a decline of total tax revenue for the government

[37] Of course this instrument does not exist, since it boils down to an individualized lump-sum tax. However, this thought-exercise allows us to calculate the excess burden of the tax.

with $T'n\frac{\partial l_n^c}{\partial t}\mathrm{d}T'$.

The change in total public revenue $\mathrm{d}R_n$ per individual is the sum of these three effects:

$$\mathrm{d}R_n = -\mathrm{d}T_n + nl_n\mathrm{d}T' + T'n\frac{\partial l_n^c}{\partial T'}\mathrm{d}T' = T'n\frac{\partial l_n^c}{\partial T'}\mathrm{d}T'. \tag{85}$$

Note that the first two terms sum to zero, since each household gets perfectly compensated: $\mathrm{d}T_n = nl_n\mathrm{d}T'$, see above. Therefore, the total revenue loss for the government on individual n is

$$\frac{\mathrm{d}R_n}{\mathrm{d}T'} = nT'\frac{\partial l_n^c}{\partial T'} = \frac{T'}{1-T'}nl_n\frac{\partial l_n^c}{\partial T'}\frac{1-T'}{l_n} = -\frac{T'}{1-T'}nl_n\varepsilon_n^c. \tag{86}$$

Finally, summing the revenue losses $\frac{\mathrm{d}R_n}{\mathrm{d}t}$ over all working households and dividing this sum by total taxable income $\int_{\mathcal{N}} nl_n\tilde{k}\,(n)\,\mathrm{d}n$ yields the total marginal excess burden as a fraction of taxed income:

$$MEB \equiv \frac{\int_{\mathcal{N}} -\frac{\mathrm{d}G_n}{\mathrm{d}t}\tilde{k}\,(n)\,\mathrm{d}n}{\int_{\mathcal{N}} nl_n\tilde{k}\,(n)\,\mathrm{d}n} = \frac{\int_{\mathcal{N}} \frac{T'(z_n)}{1-T'(z_n)}nl_n\varepsilon^c\tilde{k}\,(n)\,\mathrm{d}n}{\int_{\mathcal{N}} nl_n\tilde{k}\,(n)\,\mathrm{d}n}, \tag{87}$$

Note that this deadweight loss formula is applicable to any tax schedule, including the optimal one. If the compensated tax elasticity of labor supply (ε_n^c) is constant across skills, as we assume, then we find that the marginal deadweight loss is a function of the income-weighted marginal tax rates $\frac{T'(z_n)}{1-T'(z_n)}$:

$$MEB \equiv \varepsilon_n^c\frac{\int_{\mathcal{N}} \frac{T'(z_n)}{1-T'(z_n)}nl_n\tilde{k}\,(n)\,\mathrm{d}n}{\int_{\mathcal{N}} nl_n\tilde{k}\,(n)\,\mathrm{d}n}. \tag{88}$$

References

George A. Akerlof. The economics of the caste and of the rat race and other woeful tales. *Quarterly Journal of Economics*, 90(4):599–617, 1976.

Alberto Alesina, Edward Glaeser, and Bruce Sacerdote. Work and leisure in the United States and Europe: Why so different? *NBER Macroeconomics Annual*, 20:1–64, 2005.

Alberto Alesina, Andrea Ichino, and Loukas Karabarbounis. Gender-based taxation and the division of family chores. *American Economic Journal: Economic Policy*, 3(2):1–40, 2011.

Patricia Apps and Ray Rees. Taxation and the household. *Journal of Public Economics*, 35(3): 355–369, 1998.

Anthony B. Atkinson and Wiemer Salverda. Top incomes in the Netherlands and the United Kingdom over the twentieth century. *Journal of the European Economic Association*, 3(4):883–913, 2005.

Anthony B. Atkinson and Joseph E. Stiglitz. The design of tax structure: direct versus indirect taxation. *Journal of Public Economics*, 6(1-2):55–75, 1976.

Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez. Top incomes in the long run of history. *Journal of Economic Literature*, 49(1):3–71, 2011.

Olivier Bargain and Claire Keane. Tax-benefit-revealed redistributive preferences over time: Ireland 1987-2005. *Labour*, 24:141–167, 2010.

Olivier Bargain, Mathias Dolls, Dirk Nuemann, Andreas Peichl, and Sebastian Siegloch. Tax-benefit systems in Europe and the US: between equity and efficiency. IZA Discussion Paper no. 5440, Bonn, January 2011.

Olivier Bargain, Mathias Dolls, Dirk Neumann, Andreas Peichl, and Sebastian Siegloch. Comparing inequality aversion across countries when labor supply responses differ. *International Tax and Public Finance*, 2013a. forthcoming.

Olivier Bargain, Kristian Orsini, and Andreas Peichl. Labor supply elasticities in Europe and the US. *Journal of Human Resources*, forthcoming, 2013b.

Leon Bettendorf, Sijbren Cnossen, and Casper van Ewijk. Btw-verhoging treft hoge en lage inkomens even sterk. *Me Judice*, 25 april, 2012.

Hans G. Bloemen. An empirical model of collective household labour supply with non-participation. *Economic Journal*, 120(543):183–214, 2009.

Hans G. Bloemen. Income taxation in an empirical collective household labour supply model with discrete hours. Tinbergen Institute Discussion Paper 2010-010/3, Amsterdam, 2010.

Søren Blomquist and Laurent Simula. Marginal deadweight loss when the income tax is nonlinear. mimeo, Uppsala University, 2010.

Richard Blundell, Mike Brewer, Peter Haan, and Andrew Shephard. Optimal income taxation of lone mothers: An empirical comparison of the UK and Germany. *Economic Journal*, 119: 101–121, 2009.

Richard Blundell, Antoine Bozi, and Guy Laroque. Labour supply responses and the extensive margin: The US, UK and France. mimeo, 2011.

Jan Bonenkamp. Measuring lifetime redistribution in Dutch occupational pensions. *De Economist*, 157(1):49–77, 2009.

Michael J. Boskin and Eytan Sheshinski. Optimal tax treatment of the family: Married couples. *Journal of Public Economics*, 20(3):281–297, 1983.

Francois Bourguignon and Amedeo Spadaro. Social preferences revealed through effective marginal tax rates. DELTA Working Paper 29, Paris, 2000.

François Bourguignon and Amedeo Spadaro. Tax–benefit revealed social preferences. *Journal of Economic Inequality*, 10(1):75–108, 2012.

Lans Bovenberg and Sybrand Cnossen. Fundamental tax reform in the Netherlands. *International Tax and Public Finance*, 8(4):471–484, 2001.

Lans Bovenberg and Coen Teulings. De vlaktaks in concreto. *Economisch Statistische Berichten*, 90(4455):112–114, 2005.

Mike Brewer, Emmanuel Saez, and Andrew Shephard. Means-testing and tax rates on earnings. In James A. Mirrlees, Stuart Adam, Timothy J. Besley, Richard Blundell, Steven Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth D. Myles, and James M. Poterba, editors, *The Mirrlees Review – Dimensions of Tax Design*, chapter 3, pages 202–274. Oxford University Press, Oxford, 2010.

Raj Chetty. A new method of estimating risk aversion. *American Economic Review*, 96(5):1821–1834, 2006.

Raj Chetty. Bounds on elasticities with optimization frictions: a synthesis of micro and macro evidence on labor supply. *Econometrica*, 80(3):969–1018, 2012.

Aaron Clauset, Cosma R. Shalizi, and Mark E.J. Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

Fabio Clementi and Mauro Gallegati. Pareto's law of income distribution: Evidence for Germany, the United Kingdom, and the United States. In Arnab Chatterjee, Sudhakar Yarlagadda, and Bikas K. Chakrabarti, editors, *Econophysics of Wealth Distributions*, chapter 1, pages 3–14. Springer, Milan, 2005a.

Fabio Clementi and Mauro Gallegati. Power law tails in the Italian personal income distribution. *Physica A: Statistical Mechanics and its Applications*, 350(2-4):427–438, 2005b.

CPB. *Centraal Economisch Plan 2010*. CPB Netherlands Bureau for Economic Policy Analysis, The Hague, 2010a.

CPB. Hervorming van het Nederlandse woonbeleid. CPB Netherlands Bureau for Economic Policy Analysis, The Hague, 2010b.

CPB and PBL. *Charted Choices 2013-2017: Effects of Nine Election Platforms on the Economy and the Environment*. Number 5 in CPB Book. CPB Netherlands Bureau for Economic Policy Analysis, The Hague, 2012.

Jaap de Koning, Hassel Kroes, and Alex van der Steen. Patronen van werk en gebruik van sociale regelingen. SEOR working paper, Rotterdam, 2006.

Paul Dekker and Sjef Ederveen. *Europese Verkenning 1: Sociaal Europa 1*. SCP Netherlands Institute for Social Research, The Hague, 2003.

Peter A. Diamond. Income taxation with fixed hours of work. *Journal of Public Economics*, 13(1): 101–110, 1980.

Peter A. Diamond. Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates. *American Economic Review*, 88(1):83–95, 1998.

Udo Ebert. A reexamination of the optimal nonlinear income tax. *Journal of Public Economics*, 49(1):47–73, 1992.

Martin Feldstein. The effect of marginal tax rates on taxable income: A panel study of the 1986 Tax Reform Act. *Journal of Political Economy*, 103(3):551–572, 1995.

Aart Gerritsen. Optimal taxation when people do not maximize well-being. mimeo, Erasmus University Rotterdam, 2013.

Miriam Gielen, Joke Goes, Marcel Lever, and Rocus Van Opstal. Ontwikkeling en verdeling van de marginale druk 2001-2011 (development and distribution of marginal tax burdens 2001-2011 in Dutch). CPB Document 195, The Hague, 2009.

Jonathan Gruber and Emmanuel Saez. The elasticity of taxable income: evidence and implications. *Journal of Public Economics*, 84(1):1–32, 2002.

Christopher Heady. Trends in top incomes and income inequality and their implications for tax policy. OECD Working Paper (concept) WP2(2010)6, Paris, 2010.

James J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1):153–161, 1979.

Bas Jacobs. *De Prijs van Gelijkheid (The Price of Equality in Dutch)*. Bert Bakker, Amsterdam, 2008.

Bas Jacobs. The marginal cosf of public funds is one at the optimal tax system. Erasmus University Rotterdam, 2013.

Laurence Jacquet, Etienne Lehmann, and Bruno Van Der Linden. Optimal redistributive taxation with both extensive and intensive responses. *Journal of Economic Theory*, 148(5):1770–1805, 2013.

Egbert L.W. Jongen and Maaike Stoel. Estimating the elasticity of taxable labour income in the Netherlands. CPB Background Document, CPB Netherlands Bureau for Economic Research, 2013. May 29.

Ravi Kanbur, Matti Tuomala, and Jukka Pirttilae. Non-welfarist optimal taxation and behavioral economics. *Journal of Economic Surveys*, 20(5):1467–6419, 2006.

Henrik Kleven and Esben Anton Schultz. Estimating taxable income responses using Danish tax reforms. EPRU Working Paper Series, Economic Policy Research Unit (EPRU), University of Copenhagen, 2012.

Henrik J. Kleven, Claus T. Kreiner, and Emmanuel Saez. The optimal income taxation of couples. *Econometrica*, 77(2):537–560, 2009.

Henrik Jacobsen Kleven and Claus Thustrup Kreiner. *Skat, Arbejde og Lighed*, chapter Beskatning af Arbejdsindkomst i Danmark (Taxation of Labor Income in Denmark). Gyldendal, Copenhagen, 2006. Ch. 7.

Richard Layard. Human satisfactions and public policy. *Economic Journal*, 90(360):737–750, 1980.

N. Gregory Mankiw, Matthew Weinzierl, and Danny Yagan. Optimal taxation in theory and practice. *Journal of Economic Perspectives*, 23(4):147–174, 2009.

Mauro Mastrogiacomo, Nicole M. Bosch, Miriam D.A.C. Gielen, and Egbert L.W. Jongen. A structural analysis of labour supply elasticities in the Netherlands. CPB Discussion Paper 235, The Hague, March 2013.

James A. Mirrlees. An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, 38(2):175–208, 1971.

James A. Mirrlees. Optimal tax theory: A synthesis. *Journal of Public Economics*, 6(4):327–358, 1976.

Arthur M. Okun. *Equality and Efficiency: The Big Tradeoff*. Brookings Institution Press, 1975.

Casey Rothschild and Florian Scheuer. Redistributive taxation in the Roy model. *Quarterly Journal of Economics*, 128(2):623–668, 2013.

Efraim Sadka. On income distribution, incentive effects and optimal income taxation. *Review of Economic Studies*, 43(2):261–267, 1976.

Emmanuel Saez. Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68(1):205–229, 2001.

Emmanuel Saez. Optimal income transfer programs: intensive versus extensive labor supply responses. *Quarterly Journal of Economics*, 117(3):1039–1073, 2002.

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

Fred Schroyen. Redistributive taxation and the household: The case of individual filings. *Journal of Public Economics*, 87(11):2527–2547, 2003.

Jesus K. Seade. On the shape of optimal tax schedules. *Journal of Public Economics*, 7(2):203–235, 1977.

Jesus K. Seade. On the sign of the optimum marginal income tax. *Review of Economic Studies*, 49 (4):637–643, 1982.

Statistics Netherlands. Input-output tables the Netherlands. www.cbs.nl, 2013.

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

Matti Tuomala. On the optimal income taxation: Some further numerical results. *Journal of Public Economics*, 23:351–366, 1984.

Matti Tuomala. On optimal non-linear income taxation: Numerical results revisited. *International Tax and Public Finance*, 17(3):259–270, 2010.

Casper van Ewijk, Martin Koning, Marcel Lever, and Ruud A. de Mooij. Economische effecten van aanpassing fiscale behandeling eigen woning. CPB bijzondere publicaties 62, Den Haag, 2006.

Frederic Vermeulen. And the winner is ... An empirical evaluation of unitary and collective labour supply models. *Empirical Economics*, 30(3):711–734, 2005.

William S. Vickrey. *Agenda for Progressive Taxation*. Ronald Press, New York, 1947.

Wetenschappelijk Instituut voor het CDA. Een sociale vlaktaks, naar werkbare en begrijpelijke inkomstenbelastingen. Wetenschappelijk Instituut voor het CDA, The Hague, 2009.