Using Elasticities to Derive Optimal Income Tax Rates

Emmanuel Saez*

April 1998. This Version: May 2, 1999

Abstract

This paper derives optimal income tax formulas using the concepts of compensated and uncompensated elasticities of earnings with respect to tax rates. This method of derivation casts new light on the original Mirrlees formulas of optimal taxation and can be easily extended to a heterogeneous population of taxpayers. A simple formula for optimal marginal rates for high income earners is derived as a function of the two elasticities of earnings and the thickness of the income distribution. The relative share of income effects and uncompensated elasticity for a given compensated elasticity, which is not taken into account in deadweight burden computations, is shown to be an important element for optimal taxation. The link between the distribution of skills and the income distribution in the Mirrlees model is also investigated. Empirical earnings distributions are examined using tax returns data. Optimal income tax simulations are presented using empirical wage income distributions and a range of realistic elasticity parameters. (*JEL* H21)

1 Introduction

There is a controversial debate about the degree of progressivity that the income tax should have. This debate is not limited to the economic research area but attracts much attention in the political sphere and among the public in general. At the center of the debate lies the equity-efficiency trade-off. Progressivity allows the government to redistribute from rich to poor because high incomes end up paying for a disproportionate share of public spending. But

*Graduate student at MIT, Department of Economics, Building E52, 77 Massachusetts Avenue, Cambridge, MA 02139. email:saez@mit.edu. I thank Peter Diamond, Esther Duflo, Roger Guesnerie, Michael Kremer, Thomas Piketty and James Poterba and numerous seminar participants for helpful comments. Financial support from the Alfred P. Sloan Foundation is thankfully acknowledged.

progressive taxation and high marginal tax rates have efficiency costs. High rates may affect the incentives to work and may therefore reduce the tax base (or even total tax receipts in the most extreme case), producing very large deadweight losses. The modern setup for analyzing the equity-efficiency tradeoff using a general nonlinear income tax was built by Mirrlees (1971). Since then, the theory of optimal income taxation based on the original Mirrlees's framework has been considerably developed. The implications for policy, however, are limited for two main reasons. First, optimal income tax schedules have few general properties: we know that optimal rates must lie between 0 and 1 and that they equal zero at the top and the bottom. These properties are of little practical relevance for tax policy. In particular the zero marginal rate at the top is a very local result which applies only at the very top and is not robust when uncertainty is introduced in the model; it is therefore of no practical interest. Moreover, numerical simulations tend to show that tax schedules are very sensitive to utility functions chosen (see for example Tuomala (1990), Chapter 6).

Second, optimal income taxation has interested mostly theorists and has not changed the way applied public finance economists think about the equity-efficiency tradeoff. Theorists are mostly interested in general qualitative properties of utility functions and tax schedules whereas elasticities are the key concept in applied studies. There has been no systematic attempt to derive results in optimal taxation which could be easily used in applied studies. Most of the empirical literature on the behavioral effects of income taxation tries to estimate elasticities of income (such as wage income, capital gains or overall taxable income) with respect to marginal rates. Once elasticities are computed, optimal taxation theory is often ignored and tax reform discussions are centered on the concept of deadweight burden.[1] Therefore, most discussions of tax reforms focus only on the costs of taxation but are unable to weight both costs and benefits to decide whether taxes are too high or too low.

This paper argues that there is a simple link between optimal tax formulas and elasticities of income familiar to empirical studies. The aim of an optimal income tax (in addition to meeting government's revenue needs) is to redistribute income to the poor. The income tax, however, produces distortions and may have negative effects on labor supply and thus can reduce income and even total taxes collected. Therefore, what is important to know is whether the wealthy

[1]The deadweight burden is a measure of the inefficiency of taxation. The approximation commonly used, known as Harbarger's triangle formula, is proportional to the compensated elasticity of income with respect to marginal tax rates.

continue to work when tax rates increase (without utility compensation); the uncompensated elasticity[2] is thus likely to play a bigger role than the compensated elasticity in optimal tax formulas. In other words, this paper shows that the precise division of compensated effects into uncompensated effects and income effects plays a major role in optimal taxation. However, the empirical literature has rarely paid much attention to this division because it focused almost exclusively on deadweight burden approximations.

Recently, Diamond (1998) has taken an important step toward the narrowing of the gap between optimal taxation theory and practical policy recommendations by considering quasi-linear utility functions and analyzing precisely the influence of elasticities of labor supply and the shape of the wage rate distribution on the optimal tax schedule. Using quasi-linear utility functions is equivalent to assuming no income effects and thus Diamond (1998) could not examine the role of income effects. It turns out that his results can be considerably generalized and that a very simple formula for high income tax rates can be derived in terms of both the compensated and uncompensated tax rate elasticities of incomes and the thickness of the top tail of the income distribution. Expressing the first order condition for optimal rates in terms of elasticities simplifies considerably the general Mirrlees formula and gives a much better understanding of the key economic effects that underlie it. Moreover, the optimal tax formulas derived using elasticities can be easily extended to a heterogeneous population.

Empirical studies provide a wide range of elasticity estimates but the thickness of the tail of the income distribution has not been studied extensively for practical taxation purposes because it does not enter the deadweight burden approximation formula and thus has not been considered as a crucial element when discussing tax policy. This paper also examines the empirical distributions of earned income using tax returns data and displays simulations of optimal income tax schedules using empirical distributions of income and making realistic assumptions about elasticity parameters.

The paper is organized as follows. Section 2 reviews the main results of the optimal income tax literature. Section 3 first recalls the usual results about elasticities of earnings. It then derives a simple formula for optimal high income tax rates. The optimal linear income tax is also examined. Section 4 presents the theoretical results of this paper in the framework of the Mirrlees model. The general Mirrlees first order condition for optimal rates is reexamined

[2]The uncompensated elasticity is equal to the compensated elasticity minus revenue effects by the Slutsky equation. See Section 3.1.

in terms of elasticities. The relation between the distribution of skills and the distribution of incomes is examined and optimal asymptotic tax rates are derived. Section 5 discusses the elasticity results of the empirical taxation literature and presents empirical results about wage income distributions along with numerical simulations of optimal tax rates. Section 6 concludes and discusses policy implications. The main results of this paper can be understood without relying explicitly on the Mirrlees framework of optimal income taxation. Section 4 is more technical but can be skipped without affecting the understanding of the subsequent sections.

2 Literature Review

The Mirrlees framework captures the most important features of the tax design problem. The economy is competitive and households differ only in the levels of skills in employment. Households supply labor elastically and thus taxation has efficiency costs. The government wants to maximize a social welfare function but cannot observe skills; it must therefore rely on a distortionary nonlinear income tax to meet both its revenue requirements and redistribute income.

General results about optimal tax schedules are fairly limited. Tuomala (1990) (Chapter 6) and Myles (1995) (Chapter 5) present most of the formal results. Mirrlees (1971) showed that there is no gain from having marginal tax rates above 100 percent because nobody will choose to have such a rate at the margin. Under reasonable assumptions for the utility function, optimal marginal rates cannot be negative either. Mirrlees (1971) presented these properties and Seade (1982) clarified the conditions under which they hold.

The most striking and well known result is that the marginal tax rate should be zero at the income level of the top skill if the distribution of skills is bounded (Sadka (1976) and Seade (1977)). The argument for this result is intuitive: if the rate faced by the top earner is larger than zero, then, extending the tax schedule to higher incomes with a zero tax rate would lead the top earner to work more and would not reduce tax revenue and thus would lead to a Pareto improvement. Numerical simulations (see for example Tuomala (1990)) have shown, however, that this result is very local. Optimal rates do not approach zero until very close to the top and thus this result is of little practical interest. Mirrlees (1971) did not derive this simple result because he considered unbounded distributions of skills. He nonetheless presented precise conjectures about asymptotic optimal rates in the case of utility functions separable in consumption and labor (Mirrlees (1971), p.189). The optimal asymptotic formulas he derived were simple; they showed clearly that optimal asymptotic rates depend positively on the thickness of the tail of the skill distribution. Nonetheless, these conjectures have remained practically unnoticed in the subsequent optimal income tax literature. This can be explained by two reasons. First, Mirrlees conjectures depend not only on the distribution of skills (which is already unobservable empirically) but also on abstract properties of the utility function with no obvious intuitive meaning. Second the zero top rate result was probably considered for a long time as the definitive result because commonsense would suggest that a finite distribution of skills is closer to the reality than an unbounded one. This paper generalizes and gives a simple interpretation of the early Mirrlees conjectures. Moreover, the empirical results will show that in fact unbounded distributions are of much more interest than bounded distributions to approximate optimal tax rates for high income earners.

In addition to the zero top result, a few more results have been derived for the bottom of the skill distribution. If everybody works (and supplies labor bounded away from zero) then Seade (1977) showed that the bottom rate is also zero. However, if there is an atom of non workers then the bottom tax rate is positive (Ebert (1992)). This later case is probably the most relevant empirically.

Recently, Atkinson (1990) using quasi-linear utility functions with constant labor supply elasticity noticed that the top rate converges to a simple limit when the skill distribution is Pareto distributed. Diamond (1998) extended this particular case and began to examine empirical distributions. Moreover, he obtained simple results about the pattern of the marginal rates as a function of simple properties of the distribution of skills.

Piketty (1997) considered the same quasi-linear utility case and derived Diamond's optimal tax formulas for the Rawlsian criterion without setting a formal program of maximization. He considered instead small local changes in marginal rates and used directly the elasticity of labor supply to derive the behavioral effects of this small reform. The optimal rate can be derived using the fact that at the optimum, the small tax reform should lead to zero first order effect on tax receipts. My paper clarifies and generalizes this alternative method of derivation of optimal taxes.[3]

Another strand of the public economics literature has developed similar elasticity methods to calculate the marginal costs of public funds. The main purpose of this literature was to

[3] I am indebted to Thomas Piketty for his suggestions and help in deriving my results using this alternative method.

develop tools more sophisticated than simple deadweight burden computations to evaluate the efficiency costs of different kinds of tax reforms and the optimal provision of public goods (see for example Mayshar (1991), Ballard and Fullerton (1992) and Dahlby (1998)). Because this literature was mainly interested in assessing the efficiency of existing tax schedules and not in computing optimal tax schedules, the links between this literature and the optimal income tax literature have been very limited. I will show that the methods of this literature can be useful to derive results in optimal taxation and that, in particular, Dahlby (1998) has come close to my results for high income rates.

Starting with Mirrlees (1971), considerable effort has gone into simulations of optimal tax schedules. Following Stern (1976), attention has been paid on a careful calibration of the elasticity of labor supply. Most simulation results are surveyed in Tuomala (1990). It has been noticed that the level of inequality of the distribution of skills and the elasticities of labor supply[4] significantly affect optimal schedules. Nevertheless, simulations did not lead researchers to conjecture or prove a general result for top rates because most simulations use a log-normal distribution of skills which matches roughly the single moded empirical distribution but has also an unrealistically thin top tail and leads to marginal rates converging to zero (Mirrlees (1971)).

Nobody has tried to use empirical distributions of income to perform simulations because the link between skills and realized incomes was never investigated in depth. This study shows that for high income earners, a simple relation can be derived between the distribution of skills and the distribution of incomes. As a result, it is possible to use empirical distributions of income to perform simulations of optimal tax rates which may provide useful practical policy recommendations.

3 Optimal Tax Rates: a Simple Approach

The aim of this Section is to show that the familiar concepts of compensated and uncompensated elasticities of earnings with respect to marginal tax rates can be useful to derive in a simple way interesting results about optimal tax rates. I first consider the problem of the optimal tax rate for high income earners and then the problem of the optimal linear tax.

To deal with the first problem, I consider that the government sets a flat marginal rate τ

[4] The numerical simulations focus on the elasticity of substitution between labor and consumption instead of uncompensated and compensated elasticities of labor supply.

above a given (high) income level $\bar{z}$ and then I consider the effects of a small increase in τ on tax receipts for the government and on social welfare. The behavioral responses can be easily derived using the elasticities. The government sets the optimal tax rate τ such that a small increase in tax rates has no first order effects on total social welfare.[5]

The problem of the optimal linear tax can be solved in a similar way by considering small increases in the optimal flat rate and in the lump sum amount redistributed to every taxpayer. Before presenting the results, I recall the definitions of the elasticities which are used throughout the paper.

3.1 Elasticity concepts

I consider a standard two good model. A taxpayer maximizes an individual utility function $u = u(c, z)$ which depends positively on consumption c and negatively on earnings z. The utility function represents strictly convex preferences. This framework is a simple extension of the standard labor supply model where utility depends on consumption and labor supply and where earnings is equal to labor supply times an exogenous pre-tax wage rate.[6] Assuming that the individual is on a linear portion of the tax schedule, the budget constraint can be written as $c = z(1-\tau) + R$, where τ is the marginal tax rate and R is defined as virtual income. Virtual income is the post-tax income that the individual would get if his earnings were equal to zero was allowed to stay on the "virtual" linear schedule. The first order condition of the individual maximization program, $(1-\tau)u_c + u_z = 0$, defines implicitly a Marshallian (uncompensated) earnings supply function $z = z(1-\tau, R)$ which depends on (one minus) the marginal tax rate τ and the virtual income R. From this earnings supply function, the usual concepts of elasticity of earnings and marginal propensity to earn out of non wage income[7] can be defined. The uncompensated elasticity (denoted by ζ^u) is defined such that:

[5] Dahlby (1998) considered piecewise linear tax schedules and used the same kind of methodology to compute the effects of a general tax rate reform on taxes paid by a "representative" individual in each tax bracket. By specializing his results to a reform affecting only the tax rate of the top bracket, he derived a formula for the tax rate maximizing taxes paid by the "representative" individual of the top bracket. In this Section, I study carefully the issue of aggregation across individuals and show how this method can lead to interesting optimal tax rate results.

[6] My formulation is more general because it allows for potential endogeneity between the wage rate and labor supply.

[7] See Pencavel (1986) for a more detailed presentation.

$$\zeta^u = \frac{1-\tau}{z}\frac{\partial z}{\partial(1-\tau)} \tag{1}$$

The marginal propensity to earn out of non wage income (denoted by mpe) is such that:

$$mpe = (1-\tau)\frac{\partial z}{\partial R} \tag{2}$$

The Hicksian (compensated) earnings function can be defined as the earnings level which minimizes cost $c - z$ needed to reach a given utility level u for a given tax rate τ. I denote it by $z^c = z^c(1-\tau, u)$. The compensated elasticity of earnings ζ^c is defined by:

$$\zeta^c = \frac{1-\tau}{z}\frac{\partial z}{\partial(1-\tau)}\ |_u \tag{3}$$

The two elasticity concepts and the revenue effects are related by the Slutsky equation:

$$\zeta^c = \zeta^u - mpe \tag{4}$$

The compensated elasticity is always non-negative and mpe is non positive if leisure is not an inferior good, an assumption I make from now on. The sign of the uncompensated elasticity is ambiguous but the uncompensated elasticity is always smaller (or equal) than the compensated elasticity. Note that these definitions are identical to usual definitions of elasticities of labor supply if one assumes that the wage rate w is exogenous and that earnings z is equal to wl where l represent hours of work.

3.2 High income optimal tax rates

I assume in this subsection that the government wants to set a constant linear rate τ of taxation above a given (high) level of income $\bar{z}$. I normalize without loss of generality the population with income above $\bar{z}$ to one and I denote by $h(z)$ the density of the income distribution. The goal of this subsection is to find out the optimal τ for the government.

I consider a small increase $d\tau$ in the top tax rate τ for incomes above $\bar{z}$. Clearly, this tax change does not affect taxpayers with income below $\bar{z}$. The tax change can be decomposed into two parts (see Figure 1); first, an overall uncompensated increase $d\tau$ in marginal rates (starting from 0 and not just from $\bar{z}$), second, an overall increase in virtual income $dR = \bar{z}d\tau$. For a given individual earning income z (above $\bar{z}$), total taxes paid are equal to $T(z) = \tau[z(1-\tau, R) - \bar{z}] + T(\bar{z})$. The small tax reform produces the following effect on his tax liability:

$$\frac{\partial T(z)}{\partial \tau} = (z - \bar{z}) + \tau \left[-\frac{\partial z}{\partial (1-\tau)} + \frac{\partial z}{\partial R} \bar{z} \right] \tag{5}$$

Therefore, this tax change has two effects on tax liability. First, there is a mechanical effect (first term in parentheses in equation (5)) and second, there is a behavioral effect (second term in square brackets in equation (5)). Let us examine these two effects successively.

- **Mechanical effect**

The mechanical effect (denoted by M) represents the increase in tax receipts if there were no behavioral responses. A taxpayer with income z (above $\bar{z}$) would pay $(z - \bar{z})d\tau$ additional taxes. This is the first term in equation (5). Therefore, summing over the population above $\bar{z}$ and denoting the mean of incomes above $\bar{z}$ by z_m, the total mechanical effect M is equal to,

$$M = [z_m - \bar{z}]d\tau \tag{6}$$

- **Behavioral Response**

The behavioral response effect (denoted by B) can in turn be decomposed into the two effects displayed in Figure 1: first, an uncompensated elastic effect (first term in the square bracket expression in equation (5)) and second, an income effect (second term in the square bracket expression in equation (5)). The uncompensated effect is the behavioral response of taxpayers to the increase in tax rate $d\tau$. By definition of the uncompensated elasticity, the response of a taxpayer earning z is equal to $-\zeta^u z d\tau/(1-\tau)$. The income effect is due to the increase in virtual income $dR = \bar{z}d\tau$. By definition of mpe, the response of an individual earning z is equal to $mpe\,\bar{z}d\tau/(1-\tau)$. The total behavioral response dz of an individual is the sum of these two effects:

$$dz = -(\zeta^u z - mpe\,\bar{z})\frac{d\tau}{1-\tau} \tag{7}$$

If $z \gg \bar{z}$ the income effect component is negligible and the response is fully uncompensated. If $z \simeq \bar{z}$, then, using the Slutsky equation (4), the response is approximately equal to $-\zeta^c z d\tau/(1-\tau)$; the response is therefore fully compensated. Equation (7) is important to bear in mind when tax reforms are used to estimate elasticities.[8]

[8] In particular, if a tax reform adds a bracket at income level $\bar{z}$, comparing the responses of taxpayers just below $\bar{z}$ and just above $\bar{z}$ allows a simple estimation of ζ^c.

The reduction in income dz displayed in equation (7) implies a reduction in tax receipts equal to τdz. The total reduction in tax receipts due to the behavioral responses is simply the sum of the terms τdz over all individuals earning more than z,

$$B = -\bar{\zeta}^u z_m \frac{\tau d\tau}{1-\tau} + \overline{mpe} \frac{\bar{z}\,\tau d\tau}{1-\tau} \tag{8}$$

where $\bar{\zeta}^u = \int_{\bar{z}}^{\infty} \zeta_{(z)}^u z h(z) dz / z_m$ is a weighted average of the uncompensated elasticity. The elasticity term $\zeta_{(z)}^u$ inside the integral represents the average elasticity over individuals earning income z. $\overline{mpe} = \int_{\bar{z}}^{\infty} mpe_{(z)} h(z) dz$ is the average of $mpe_{(z)}$.[9] Note that $\overline{mpe}$ and $\bar{\zeta}^u$ are not averaged with the same weights. It is not necessary to assume that people earning the same income have the same elasticity; the relevant parameters are simply the average elasticities at given income levels.

Adding equations (6) and (8), the overall effect of the tax reform on government's revenue is obtained,

$$M + B = \left[\frac{z_m}{\bar{z}} - 1 - \frac{\tau}{1-\tau} (\bar{\zeta}^u \frac{z_m}{\bar{z}} - \overline{mpe}) \right] \bar{z} d\tau \tag{9}$$

The tax reform raises revenue if and only if the expression in square brackets is positive. If the government values much more an additional dollar given to the poorest people than the same additional dollar given to the top bracket taxpayers,[10] then it will raise the maximum amount of taxes from the top bracket taxpayers. In that case, it will set the top rate τ such that the expression in square brackets in equation (9) is equal to zero.

In the general case, let us consider $\bar{g}$ which is the ratio of social marginal utility for the top bracket taxpayers to the marginal value of public funds for the government. In other words, $\bar{g}$ is defined such that the government is indifferent between $\bar{g}$ more dollars of public funds and one more dollar consumed by the taxpayers with income above $\bar{z}$. $\bar{g}$ can be considered as a parameter reflecting the redistributive goals of the government.

Each additional dollar raised by the government because of the tax reform reduces social welfare of people in the top bracket by $\bar{g}$ and thus is valued only $1 - \bar{g}$ by the government. First order behavioral changes in earnings lead only to second order effects on welfare (this is the usual consequence of the envelope theorem). As a result, the loss of one dollar in taxes due to

[9] $mpe_{(z)}$ is the average income effect for individuals earning z.

[10] This is of course the case with the Rawlsian criterion. This is also the case with a utilitarian criterion if one considers utility functions with marginal utility of consumption declining to zero as consumption tends to infinity.

behavioral effects is valued one dollar (and not $1-\bar{g}$ dollars) by the government. Consequently, the government wants to set the rate τ such that, $(1-\bar{g})M + B = 0$. Thus the optimal rate is such that,

$$\frac{\tau}{1-\tau} = \frac{(1-\bar{g})(z_m/\bar{z}-1)}{\bar{\zeta}^u z_m/\bar{z} - \overline{mpe}} \tag{10}$$

which leads to,

$$\tau = \frac{1-\bar{g}}{1-\bar{g}+\bar{\zeta}^u(\frac{z_m}{\bar{z}})/(\frac{z_m}{\bar{z}}-1) - \overline{mpe}/(\frac{z_m}{\bar{z}}-1)} \tag{11}$$

This equation gives a strikingly simple answer to the problem of the optimal marginal rate for high income earners. This formula applies to heterogeneous populations. The relevant parameters are the weighted average elasticities $\bar{\zeta}^u$ and average income effects $\overline{mpe}$ which can be estimated empirically. The optimal rate τ is a decreasing function of $\bar{g}$, $\bar{\zeta}^u$ and $-\overline{mpe}$ (absolute size of income effects) and an increasing function of $z_m/\bar{z}$.

The ratio $z_m/\bar{z}$ of the mean of incomes above $\bar{z}$ to the income level $\bar{z}$, is larger than one. From now on, I call this ratio the conditional mean income ratio. If the tail of the income distribution follows a Pareto distribution with parameter $a > 1$ ($Prob(Income > z) = C/z^a$) then the density of incomes $h(z)$ is then to aC/z^{1+a}. In that case, it is easy to show that $z_m/\bar{z}$ is constant and equal to $a/(a-1)$.[11]

If the tail of the distribution is thinner than any Pareto distribution (e.g., such as a log-normal or an exponential distribution) then $z_m/\bar{z}$ tends to 1 and therefore we can consider that in this case $a = \infty$. Section 5 will show that, empirically, $z_m/\bar{z}$ is strikingly stable over a very large range of incomes. Therefore, the tails of empirical earnings distributions can be remarkably well approximated by Pareto distributions.[12] The parameter a is approximately equal to 2.

Let me now consider the asymptotics of equation (11). Assuming that $z_m/\bar{z}$ converges to a value (say m_∞), I can define a (between one and infinity) such that $a/(a-1) = m_\infty$. a can be considered as the limiting "Pareto" parameter of the income distribution. If $\zeta^u_{(z)}$ and $mpe_{(z)}$ converge to limiting values (denoted also by $\bar{\zeta}^u$ and $\overline{mpe}$) when z tends to infinity, the Slutsky

[11] a must be larger than one to rule out infinite aggregate income $\int zh(z)dz$.

[12] This is of course not a new finding. Pareto discovered this empirical regularity more than a century ago (see Pareto (1965)). That is why these power law densities are called Pareto distributions.

equation[13] implies that $\zeta^c_{(z)}$ converges to $\bar{\zeta}^c$ such that $\overline{mpe} = \bar{\zeta}^u - \bar{\zeta}^c$. In this case, (11) can be rewritten as a function of a and the limiting values of the elasticities $\bar{\zeta}^c$ and $\bar{\zeta}^u$:

$$\bar{\tau} = \frac{1 - \bar{g}}{1 - \bar{g} + \bar{\zeta}^u + \bar{\zeta}^c(a - 1)} \tag{12}$$

When these parameters do converge, the government wants to set roughly the same linear rate $\bar{\tau}$ above any large income level and thus $\bar{\tau}$ is indeed the optimal non-linear asymptotic rate of the Mirrlees problem.[14] I show in Section 4 that the parameter a is independent of $\bar{\tau}$ as long as $\bar{\tau} < 1$. The intuition is the following: when elasticities are constant, changing the tax rate has the same multiplicative effect on the incomes of each high income taxpayer and therefore the ratio $z_m/\bar{z}$ is unchanged. Empirically, a does not seem to vary with level of the top rate. I come back to this point in Section 5 but a thorough empirical investigation of this issue is left for future research.

$\bar{\tau}$ is decreasing in the four parameters $\bar{\zeta}^c$, $\bar{\zeta}^u$, a and $\bar{g}$. This is hardly surprising. Interestingly, for a given compensated elasticity $\bar{\zeta}^c$, the precise division into income effects and uncompensated rate effects matters. The higher are absolute income effects ($-\overline{mpe}$) relative to uncompensated effects ($\bar{\zeta}^u$), the higher is the asymptotic tax rate $\bar{\tau}$. This result confirms the intuition developed in the introduction: what matters most for optimal taxation is whether taxpayers continue to work when tax rates increase (without utility compensation).

The top rate $\bar{\tau}$ also depends negatively of the thickness of the top tail distribution measured by the Pareto parameter a or the limiting value of $z_m/\bar{z}$. This is also an intuitive result: if the distribution is thin then raising the top rate for high income earners will raise little additional revenue because the mechanical effect M depends on the difference between z_m and $\bar{z}$ while the distortions are proportional to z_m (for the uncompensated effect) and $\bar{z}$ (for the income effect) and thus are high at high income levels. If the distribution of income is bounded, then close to the top, $\bar{z}$ is close to z_m and so the conditional mean income ratio tends to one and thus the top rate must be equal to zero (see equation (10)). This is the classical zero top rate result derived by Sadka (1976) and Seade (1977). If the tail is infinite but thinner than any Pareto distribution (i.e., $a = \infty$) then the asymptotic rate must also be zero.

[13] It is not possible to use directly the Slutsky equation in (11) because ζ^u and mpe are not averaged with the same weights.

[14] This point is proved rigorously in the Mirrlees model in Section 4.

3.3 Optimal Linear Rate

The analysis above can be easily applied to the case of the optimal linear tax rate. Many papers (beginning with Sheshinski (1972)) have studied this case but no paper has derived the optimal rate using directly the concepts of elasticities of earnings and marginal propensity to earn out of non wage income.[15]

In the case of an optimal linear tax, the government imposes a budget constraint of the form: $c = (1-\tau)z + R$ by choosing the tax rate τ and a lump-sum level R. I note $H(z)$ the distribution of income and $h(z)$ its density function. I note $MS(z)$ the social marginal utility of consumption for individuals with income z and p the social value of public funds.

Consider first an increase of the tax rate from τ to $\tau + d\tau$, then, an individual with income z will pay $zd\tau$ additional taxes (mechanical effect), valued only $z(1 - MS(z)/p)d\tau$ by the government. Moreover, the individual will change its earnings by $dz = -\zeta^u z d\tau/(1-\tau)$ which changes the amount of taxes it pays by τdz. The effect aggregated over the population must be null at the optimum and therefore:

$$1 - \int \frac{MS(z)}{p} \frac{z}{z_M} h(z)dz = \frac{\tau}{1-\tau} \bar{\zeta}^u \tag{13}$$

where $z_M = \int z h(z)dz$ denotes average income and $\bar{\zeta}^u = \int \zeta^u z h(z)dz/z_M$ is a weighted average of the uncompensated elasticity.

Next, suppose that the government increases the lump sum R by dR, then the tax collected on a given individual earning z decreases by dR but the social loss is only $(1 - MS(z)/p)dR$. The individual changes its earnings by $dz = mpe\, dR/(1-\tau)$ which changes the amount of taxes it pays by τdz. The overall aggregated effect must be null at the optimum and thus,

$$1 - \frac{\overline{MS}}{p} = \frac{\tau}{1-\tau} \overline{mpe} \tag{14}$$

where $\overline{MS} = \int MS(z)h(z)dz$ denotes average social marginal utility and $\overline{mpe} = \int mpe\, h(z)dz$ is the average of mpe.

Equations (13) and (14) can be combined to eliminate p, and to obtain the following formula for the optimal tax rate τ^*:

[15] The only exception is Piketty (1997) who derived the optimal linear rate for the simple case of the Rawlsian criterion with the method used here.

$$\tau^* = \frac{1-\bar{G}}{1-\bar{G}+\bar{\zeta}^u - \bar{G}\,\overline{mpe}} \tag{15}$$

where $\bar{G}$ is defined such that,

$$\bar{G} = \int \frac{MS(z)}{\overline{MS}} \frac{z}{z_M} h(z)dz$$

If $MS(z)$ is decreasing (which is a reasonable assumption if the government has redistributive goals), then $\bar{G} < 1$.[16] $\bar{G}$ is the smaller, the greater is inequality and the greater are the redistributive goals of the government. In the Rawlsian case, $MS(z) = 0$ for every z positive which implies $\bar{G} = 0$. Using equation (13), we obtain $\tau^* = 1/(1+\bar{\zeta}^u)$. In any case, $\bar{G}$ can be considered as a parameter chosen by the government according to its preferences. Once a distribution of incomes $H(z)$ is given, the government chooses the function $MS(z)$ and thus can compute $\bar{G}$.[17] Unsurprisingly, the optimal linear rate is decreasing in $\bar{G}$, in the size of the uncompensated elasticity and in the absolute size of income effects.

These results can be derived in the classical model of optimal linear taxation. The interpretations of the optimal rate formula are often close to the one presented here (see for example Atkinson-Stiglitz (1980), pp. 407-408). However, I presented the results without referring to a distribution of skills to show that formula (15) can be applied in a much more general framework with heterogeneous agents. Similarly to the previous subsection, the only thing that matters is average elasticities; these average elasticities can be measured empirically.

3.4 Conclusion

This Section has shown that considering small reforms around the optimum and deriving the behavioral responses using elasticity concepts is a natural way to derive optimal tax rate results. Formulas for optimal rates (12) and (15) show that the pattern of elasticities as well as the shape of the income distribution and the redistributive goals of the government are the relevant parameters. In particular, though ζ^c is a sufficient statistic to approximate the deadweight loss of taxation, same values of ζ^c can lead to very different optimal tax rates. The bigger the income effects relative to uncompensated elasticity, the higher is the optimal tax rate.

[16]This is because $\bar{G}\overline{MS}$ is an average of $MS(z)$ with weights $zh(z)/z_M$ which overweighs high z and thus $\bar{G}\overline{MS}$ is smaller than $\overline{MS}$.

[17]Note however that in the optimal linear tax case, $\bar{G}$ is positive except in the extreme Rawlsian case whereas in the asymptotic non-linear tax case of Section 3.2, $\bar{g}$ could be zero even with a utilitarian criterion.

In the next Section, I show that the systematic use of elasticity concepts in the general Mirrlees model is fruitful. First, by considering as in this Section a small tax reform (a small local increase in marginal rates), it is possible to derive the general Mirrlees formula for optimal tax rates without referring to adverse selection theory. This derivation allows a better grasp on the different effects at play than blind mathematical optimization and can be easily extended to heterogeneous populations. Second, it will be shown that the income distribution and the skill distribution are closely related through the uncompensated elasticity. This result is of crucial importance to perform numerical simulations (presented in Section 5) using empirical earnings distributions.

4 Optimal Tax Rates: General Results

4.1 The Mirrlees model

In the model, all individuals have identical preferences. The utility function depends on composite consumption c and labor l and is noted $u(c, l)$. I assume that preferences are well behaved and that u is regular (at least of class C^4). The individuals differ only in their skill level (denoted by n) which measures their marginal productivity. If an individual with skill n supplies labor or "effort" l, he earns nl. The distribution of skills is written $F(n)$, with density $f(n)$ and support in $[0, \infty)$. f is also assumed to be regular (at least of class C^2). The consumption choice of an individual with skill n is denoted by (c_n, l_n) and I write $z_n = nl_n$ for its earnings and u_n for its utility level $u(c_n, l_n)$. The government does not observe n or l_n but only earnings z_n. Thus it is restricted to setting taxes as a function only of earnings: $c = z - T(z)$. The government maximizes the following social welfare function:

$$W = \int_0^\infty G(u_n) f(n) dn \tag{16}$$

where G is an increasing and concave function of utility. The government maximizes W subject to a resource constraint and an incentive compatibility constraint. The resource constraint states that aggregate consumption must be less than aggregate production minus government expenditures, E:

$$\int_0^\infty c_n f(n) dn \leq \int_0^\infty z_n f(n) dn - E \tag{17}$$

The incentive compatibility constraint is that the selected labor supply l_n maximizes utility, given the tax function, $u(nl - T(nl), l)$. Assuming that the tax schedule T is regular, the optimal choice of l implies the following first order condition:

$$n(1 - T'(z_n))u_c + u_l = 0 \tag{18}$$

This equation holds true as long as the individual chooses to supply a positive amount of labor. This first order condition leads to:

$$\dot{u}_n = -\frac{lu_l}{n} \tag{19}$$

where a dot means (total) differentiation with respect to the skill level n.

Following Mirrlees (1971), in the maximization program of the government, u_n is regarded as the state variable, l_n as the control variable while c_n is determined implicitly as a function of u_n and l_n from the equation $u_n = u(c_n, l_n)$. Therefore, the program of the government is simply to maximize equation (16) by choosing l_n and u_n subject to equations (17) and (19). Forming a Hamiltonian for this expression, we have:

$$H = [G(u_n) - p(c_n - nl_n)]f(n) - \phi(n)\frac{l_n u_l(c_n, l_n)}{n} \tag{20}$$

where p and $\phi(n)$ are multipliers. p is the Lagrange multiplier of the government's budget constraint and thus can be interpreted as the marginal value of public funds. From the first order conditions of maximization, we obtain the classical first order condition for optimal rates (see Mirrlees (1971), equation (33)):

$$(n + \frac{u_l^{(n)}}{u_c^{(n)}})f(n) = \frac{\psi_l^{(n)}}{n}\int_n^{\infty}[\frac{1}{u_c^{(m)}} - \frac{G'(u_m)}{p}]T_{nm}f(m)dm \tag{21}$$

where $T_{nm} = \exp[-\int_n^m \frac{l_s u_{cl}(c_s, l_s)}{s u_c(c_s, l_s)} ds]$. ψ is defined such that $\psi(u, l) = -lu_l(c, l)$ where c is a function of (u, l) such that $u = u(c, l)$. An superscript (n) means that the corresponding function is estimated at (c_n, l_n, u_n). The derivation of (21) is recalled in appendix.

Theorem 2 in Mirrlees (1971) (pp.183-4) states under what conditions formula (21) is satisfied at the optimum. The most important assumption is the single-crossing condition (condition (B) in Mirrlees (1971)) which is equivalent to the uncompensated elasticity being greater than minus one. This condition is very likely to hold empirically and I will also assume from now on that $\zeta^u > -1$.

Even when the single crossing property is satisfied, the first order condition (21) may not characterize the optimum. The complication comes from the need to check that individual labor supply choices satisfying the first order condition (18) are globally optimal choices. Mirrlees showed that the first order condition for individual maximization implies global maximization if and only if the earnings function z_n is non-decreasing in the skill level n. If equation (21) leads to earnings z_n decreasing over some skill ranges then this cannot be the optimum solution and therefore there must be bunching at some income level (a range of workers with skills n lying in $[n_1, n_2]$ choosing the same income level $\bar{z}$). When bunching happens, (21) no longer holds but $\dot{u}_n = -lu_l/n$ remains true. Theorem 2 in Mirrlees (1971) states that (21) holds at every point n where z_n is increasing.[18] Seade (1982) showed that if leisure is not an inferior good (i.e., $mpe \leq 0$) then T' cannot be negative at the optimum. I assumed in Section 3 that $mpe \leq 0$ and continue to do so in this Section.

In this model, redistribution takes place through a guaranteed income level that is taxed away as earnings increase (negative income tax). Optimal marginal tax rates are defined by equation (21). Therefore, the welfare program is fully integrated to the tax program.

4.2 Optimal Marginal Rates

The general Mirrlees first order condition (21) depends in a complicated way on the derivatives of the utility function $u(c, l)$ which are almost impossible to measure empirically. Therefore, it has been impossible to infer directly from the general equation (21) practical quantitative results about marginal rate patterns. Moreover, equation (21) has always been derived using powerful but blind Hamiltonian optimization. Thus, the optimal taxation literature has never been able to elucidate the key economic effects which lead to the general formula (21). In this subsection, I rewrite equation (21) as a function of elasticities of earnings and show precisely the key behavioral effects which lead to this rewritten equation. I first present a simple preliminary result that is a useful step to understand the relation between the income distribution and the distribution of skills in the Mirrlees economy.

Lemma 1 *For any regular tax schedule T (such that T'' exists) not necessarily optimal, the earnings function z_n is non-decreasing and satisfies the following equation,*

[18] Gaps in the distribution of incomes can also happen in case of multiple maxima in the maximization of the Hamiltonian with respect to l. Gaps do not arise generically and can be ruled out under weak assumptions (see Mirrlees (1971)). I will therefore assume from now on, that the equilibrium distribution of incomes has no gaps.

$$\frac{\dot{z}_n}{z_n} = \frac{n\dot{l}_n + l_n}{nl_n} = \frac{1+\zeta^u_{(n)}}{n} - \dot{z}_n \frac{T''_{(n)}}{1-T'_{(n)}} \zeta^c_{(n)} \tag{22}$$

If equation (22) leads to $\dot{z}_n < 0$ then z_n is discontinuous and (22) does not hold.

The proof, which is routine algebra, is presented in appendix. In the case of a linear tax ($T'' = 0$) the earnings equation (22) becomes the familiar equation $dz/z = (1 + \zeta^u)dn/n$. In the general case, a correction term in T'' which represents the effect of the change in marginal rates is present.

The first order condition (21) can be reorganized in order to express optimal tax rates in terms of the elasticities of earnings. This rearrangement of terms is a generalization of the one introduced in Diamond (1998) in the case of quasi-linear utility functions.

Proposition 1 *The first order condition (21) can be rewritten as follows:*

$$\frac{T'(z_n)}{1 - T'(z_n)} = A(n)B(n) \tag{23}$$

where

$$A(n) = \left(\frac{\zeta^u_{(n)} + 1}{\zeta^c_{(n)}}\right) \left(\frac{1 - F(n)}{nf(n)}\right) \tag{24}$$

$$B(n) = \int_n^\infty [1 - \frac{G'(u_m)u_c^{(m)}}{p}]\, S_{nm} \frac{f(m)}{1 - F(n)} dm \tag{25}$$

where

$$S_{nm} = \exp\left[\int_n^m (1 - \frac{\zeta^u_{(s)}}{\zeta^c_{(s)}}) \frac{dz_s}{z_s}\right] \tag{26}$$

The formal proof of this proposition, which starts with equation (21) and is routine algebra, is presented in appendix. This proof, however, does not show the economic effects which lead to formula (23). It is possible, though, to derive this equation by considering small variations in marginal rates around the optimum as in Section 3. This derivation, though complicated, shows precisely how the key effects come into play to lead to formula (23) and therefore is presented in detail. Formula (23) is commented in the light of this direct derivation just after the proof.

Direct Proof of Proposition 1

I note $H(z)$ the distribution of incomes at the optimum and $h(z)$ the corresponding density function. I note again $MS(z)$ the marginal social value of consumption for a taxpayer with income z (i.e., this is exactly $G'(u)u_c$ is Mirrlees notation). p is the marginal value of public funds. I consider the effect of the following small tax reform: marginal rates are increased by an amount $d\tau$ for incomes between $\bar{z}$ and $\bar{z} + d\bar{z}$.[19]

This tax reform has three effects on tax receipts: a mechanical effect, an elasticity effect for taxpayers with income between $\bar{z}$ and $\bar{z} + d\bar{z}$, and an income effect for taxpayers with income above $\bar{z}$.

- **Mechanical Effect**

This effect represents the increase in tax receipts if there were no behavioral responses. Every taxpayer with income z above $\bar{z}$ pays $d\tau d\bar{z}$ additional taxes which are valued $(1 - MS(z)/p)d\tau d\bar{z}$ by the government therefore the overall net effect M is equal to:[20]

$$M = d\tau d\bar{z} \int_{\bar{z}}^{\infty} [1 - \frac{MS(z)}{p}] h(z) dz$$

- **Elastic Effect**

The increase $d\tau$ for a taxpayer with income z between $\bar{z}$ and $\bar{z} + d\bar{z}$ has an elastic effect which produces a small change in income (denoted by dz). This change is the consequence of two effects. First, there is a direct compensated effect due to the exogenous increase $d\tau$. The compensated elasticity is the relevant one here because the change $d\tau$ takes place at level $\bar{z}$ just below z (see the discussion following equation (7) in the previous Section). Second, there is an indirect effect due to the shift of the taxpayer on the tax schedule by dz which induces an endogenous additional change in marginal rates equal to $dT' = T'' dz$. Therefore, the behavioral equation can be written as follows,

$$dz = -\zeta^c \bar{z} \frac{d\tau + dT'}{1 - T'}$$

which implies,

$$dz = -\zeta^c \bar{z} \frac{d\tau}{1 - T' + \zeta^c \bar{z} T''}$$

[19] I also assume that $d\tau$ is second order compared to $d\bar{z}$ so that bunching (and inversely gaps in the income distribution) around $\bar{z}$ or $\bar{z} + d\bar{z}$ induced by the discontinuous change in marginal rates are negligible.

[20] The tax reform has also an effect on $h(z)$ but this is a second order effect in the computation of M.

By Lemma 1, $1 - T' + \zeta^c \bar{z} T'' = (1 - T')(1 + \zeta^u) z / (n \dot{z}) > 0$. When the Single Crossing condition $1 + \zeta^u > 0$ holds, $1 - T' + \zeta^c \bar{z} T'' > 0$ if and only if $\dot{z} > 0$. As reviewed above, $\dot{z} \geq 0$ is a necessary and sufficient condition for the individual choice given by the individual first order condition to be a global maximum. Thus $1 - T' + \zeta^c \bar{z} T'' \geq 0$ is also necessary and sufficient to insure global optimization of the individual choice. I assume in this heuristic proof that $1 - T' + \zeta^c \bar{z} T'' > 0$ for any $\bar{z}$ in order to avoid dealing with bunching issues.[21]

In order to simplify notations, I introduce $h^*(\bar{z})$ which is the density of incomes that would take place at $\bar{z}$ if the tax schedule $T(.)$ were replaced by the linear tax schedule tangent to $T(.)$ at level $\bar{z}$.[22] I call the density $h^*(\bar{z})$ the virtual density. Densities h and h^* are related through the skill density $f(n)$ such that $h^*(\bar{z}) \dot{z}^* = h(\bar{z}) \dot{z} = f(n)$ where $\dot{z}^*$ is the derivative of earnings with respect to n at point $\bar{z}$ if the tax schedule T is replaced by the tangent linear tax schedule. Using Lemma 1, I have, $\dot{z}^* / \bar{z} = (1 + \zeta^u)/n$ and $\dot{z}/\bar{z} = (1 + \zeta^u)/n - \dot{z} \zeta^c T''/(1 - T')$ which implies:

$$\frac{h^*(\bar{z})}{1 - T'(\bar{z})} = \frac{h(\bar{z})}{1 - T'(\bar{z}) + \zeta^c_{(\bar{z})} \bar{z} T''(\bar{z})} \tag{27}$$

where $\zeta^c_{(\bar{z})}$ is the compensated elasticity at income level $\bar{z}$. Using $h^*(\bar{z})$, the overall effect on tax receipts (denoted by E) can be simply written as:

$$E = -\zeta^c_{(\bar{z})} \bar{z} \frac{T'}{1 - T'} h^*(\bar{z}) d\tau d\bar{z}$$

- **Income Effect**

A taxpayer with income z above $\bar{z}$ pays $-dR = d\tau d\bar{z}$ additional taxes. This produces an income response dz which is again due to two effects. First, there is the direct income effect (equal to $mpe\, dR/(1 - T')$). Second, there is an indirect elastic effect due to the change in marginal rates $dT' = T'' dz$ induced by the shift dz along the tax schedule. Therefore,

$$dz = -\zeta^c z \frac{T'' dz}{1 - T'} - mpe \frac{d\tau d\bar{z}}{1 - T'}$$

which implies,

$$dz = -mpe \frac{d\tau d\bar{z}}{1 - T' + z \zeta^c T''} \tag{28}$$

[21] This condition is always satisfied at points where $T''(\bar{z}) \geq 0$.

[22] This linear tax schedule is characterized by the tax rate $\tau = T'(\bar{z})$ and the virtual income $R = \bar{z} - T(\bar{z}) - \bar{z}(1 - \tau)$.

Introducing again the density $h^*(z)$ and summing (28) over all taxpayers with income larger than $\bar{z}$, I obtain the total behavioral effect I due to income effects:

$$I = d\tau d\bar{z} \int_{\bar{z}}^{\infty} -mpe_{(z)} \frac{T'}{1-T'} h^*(z) dz$$

At the optimum, the sum of the three effects M, E and I must be zero which implies,

$$\frac{T'}{1-T'} = \frac{1}{\zeta^c} \left(\frac{1-H(\bar{z})}{\bar{z}h^*(\bar{z})} \right) \left[\int_{\bar{z}}^{\infty} \left(1 - \frac{MS(z)}{p}\right) \frac{h(z)}{1-H(\bar{z})} dz + \int_{\bar{z}}^{\infty} -mpe \frac{T'}{1-T'} \frac{h^*(z)}{1-H(\bar{z})} dz \right] \tag{29}$$

Equation (29) can be considered as a first order linear differential equation and can be integrated (see appendix) using the standard method to obtain:

$$\frac{T'(\bar{z})}{1-T'(\bar{z})} = \frac{1}{\zeta^c_{(\bar{z})}} \left(\frac{1-H(\bar{z})}{\bar{z}h^*(\bar{z})} \right) \int_{\bar{z}}^{\infty} \left(1 - \frac{MS(z)}{p}\right) \exp\left[\int_{\bar{z}}^{z} (1 - \frac{\zeta^u_{(z')}}{\zeta^c_{(z')}}) \frac{dz'}{z'} \right] \frac{h(z)}{1-H(\bar{z})} dz \tag{30}$$

Changing variables from $\bar{z}$ to n, and using the fact, proved above, that $\bar{z}h^*(\bar{z})(1+\zeta^u) = nf(n)$, it is straightforward to obtain the equation of Proposition 1. Therefore, when changing variables from $\bar{z}$ to n, an additionnal term $1+\zeta^u$ appears on the righthand side to form the term $A(n)$ of Proposition 1. This counterintuitive term (higher uncompensated elasticity should not lead to higher marginal rates) should in fact be incorporated into the skill ratio $(1-F)/(nf)$ to lead to the income ratio $(1-H)/(\bar{z}h^*)$ which is easier to relate to the empirical income distribution. Of course, the virtual density h^* is not identical to the actual density h. However, because the density h at the optimum is endogenous (in the sense that changes in the tax schedule affect the income distribution), there is very little inconvenience in using h^* rather than h. Using Lemma 1, one can observe that nonlinear tax schedules produce a deformation of the earnings distribution h. Using h^* is a way to get rid of this deformation component. In that sense, h^* is more closely related than h to the skill distribution which represents intrinsic inequalities.

Last, let me mention that the multiplier p is such that the integral term in (30) must be zero when $\bar{z} = 0$. This can be proved by considering that a small change in the lump sum given to everybody $(-T(0))$ has no first order effect on total welfare. $\mathcal{QED}$

Interpretation of Proposition 1

In the light of this heuristic proof, let us analyze the decomposition of optimal tax rates presented in Proposition 1 or equivalently equation (30). Analyzing equation (30) (or (23)), it appears that three elements determine optimal income tax rates: elasticity (and income) effects, the shape of the income (or skill) distribution and social marginal weights. These elements enter the optimal tax formula in relatively independent ways and thus can be examined independently.

- **Shape of Income Distribution**

The shape of the income distribution affects the optimal rate at level $\bar{z}$ mainly through the term $(1 - H(\bar{z}))/(\bar{z}h(\bar{z}))$. This is intuitive: the elastic distortion at $\bar{z}$ induced by a marginal rate increase at that level is proportional to income at that level times number of people at that income level ($\bar{z}h(\bar{z})$) while the gain in tax receipts is proportional to the number of people above $\bar{z}$ (i.e., $1 - H(\bar{z})$). In other words, a high marginal rate at a given income level $\bar{z}$ induces a negative behavioral response at that level but allows the government to raise more taxes from all taxpayers above $\bar{z}$. Therefore, the government should apply high marginal rates at levels where the density of taxpayers is low compared to the number of taxpayers with higher income. Unsurprisingly, the ratio $(1-H)/(zh)$ is constant and equal to $1/a$ when $H(z)$ is Pareto distributed with parameter a. This ratio tends to zero when the top tail is thinner than any Pareto distribution. Next subsection examines the asymptotics of formula (30). The empirical shape of the ratio $(1 - H)/(zh)$ is studied in Section 5.

- **Elastic and Income effects**

Behavioral effects enter the formula for optimal rates in two ways. First, increasing marginal rates at level $\bar{z}$ induces a compensated response from taxpayers earning $\bar{z}$. Therefore, $\zeta^c_{(\bar{z})}$ enters negatively the optimal tax rate at income level $\bar{z}$. Second, this marginal rate change increases the tax burden of all taxpayers with income above $\bar{z}$. This effect induces these taxpayers to work more through income effects which is good for tax receipts. Therefore, this income effect leads to higher marginal rates (everything else being equal) through the term S_{nm} (or equivalently the exponential term in (30)) which is bigger than one.[23] Note that this term is identically equal to one when there are no income effects (this case was studied by Diamond (1998)). The heuristic proof shows clearly why negative tax rates are never optimal. If the tax rate were negative in some range then increasing it a little bit in that range would decrease earnings in that range (because of the substitution effect) but this behavioral response would increase tax receipts because the tax rate is negative in that range. Therefore, this small tax reform would

[23]The term $1 - G'(u)u_c/p = 1 - MS(z)/p$ is in general increasing in income and is thus always positive above some income level.

unambiguously increase welfare.

- **Social Marginal Welfare Weights**

The social marginal weights (denoted by $MS(z)/p$ in terms of the marginal value of public funds) enter the optimal tax formula through the term $(1 - MS(z)/p)$ inside the integral. The intuition is the following: increasing marginal rates locally at level $\bar{z}$ increases the tax burden of all taxpayers with income above $\bar{z}$. Each additional dollar raised by the government over taxpayers with income z is valued $(1 - MS(z)/p)$. This expression is decreasing with z (as long as the government has redistributive goals). Therefore, redistributive goals is unsurprisingly an element tending to make the tax schedule progressive. If the government had no redistributive goals, then it would choose the same marginal welfare weights for everybody. The formula for the optimal income tax would clearly be qualitatively very close to the general case with redistributive concerns. In particular, the shape of the income distribution and the size of both substitution and income effects would matter for the optimal income tax with no redistributive goals.[24]

The original Mirrlees's derivation relies heavily on the fact that there exists a unidimensional skill parameter which characterizes each taxpayer. As a result, that derivation gives no clue about how to extend the non-linear tax formula to a heterogeneous population in a simple way. The direct proof using elasticities shows that there is no need to introduce an exogenous skill distribution. Formula (30) is valid for any heterogeneous population as long as $\zeta^u_{(z)}$ and $\zeta^c_{(z)}$ are considered as average elasticities at income level z.[25] Therefore, the skill distribution in the Mirrlees model should not be considered as a real economic element (which one should try to measure empirically) but rather as a useful simplification device to perform computations and numerical simulations. The skill distribution should simply be chosen so that the resulting income distribution be close to the empirical income distribution.[26] Mirrlees (1976) and (1986) tried to extend his 1971 formula to heterogeneous populations where individuals are characterized by a multidimensional parameter instead of a single dimensional skill parameter. He

[24] Saez (1999) investigates this point more deeply. I show that the income tax which minimizes deadweight burden is in fact an optimal income tax with particular welfare weights. In the absence of income effects, these weights are the same for everybody.

[25] Equation (27) linking the virtual density h^* to the actual density h can be generalized to the case of heterogeneous populations.

[26] This route is followed in Section 5.

adopted the same approach as he used in his 1971 study and derived first order conditions for the optimal tax schedule. However, these conditions were even more complicated than in the unidimensional case and thus it proved impossible to obtain results or interpret the first order conditions in that general case. It is nonetheless possible to manipulate the first order conditions of the general case considered in Mirrlees (1976) and (1986) in order to recover formula (30). Therefore, the elasticity method of the heuristic proof is a powerful tool to understand the economics of optimal income taxation and is certainly a necessary step to take to extend in a fruitful way the model to heterogeneous populations. This general derivation is out of the scope of the present paper and will be presented in future work.

Though this is not attempted in this paper, let me sketch how formula (30) could be used to perform numerical simulations without need to rely on an exogenous skill distribution. Making assumptions about the pattern of elasticities,[27] selecting a function $MS(.)$ reflecting the redistributive tastes of the government, and using the empirical income distribution to obtain $H(.)$, equation (30) could be used to compute a tax schedule $T'(.)$. Of course, this tax schedule would not be optimal because $H(.)$ is an endogenous function (a tax reform affects income distribution through behavioral responses). Nevertheless, this computed $T'(.)$ could yield interesting information for tax reform. Using this estimated T', a new income distribution $H(.)$ could then be derived leading to a new estimate for T'. This algorithm may converge to the optimal tax schedule. This avenue of research is out of the scope of the present paper but may deserve further investigation.

Formula (30) could also be used to pursue a *positive* analysis of actual tax schedules. Considering the actual tax schedule $T(.)$ and the actual income distribution $H(.)$, and making assumptions about the patterns of elasticities $\zeta^u_{(z)}$ and $\zeta^c_{(z)}$, it is also possible to use equation (30) to infer the marginal social weights $MS(z)/p$. Even if the government does not really maximize welfare, it may be interesting to know what are the implicit weights that the government is using. For example, if some of the weights appear to be negative then the tax schedule is not second-best Pareto efficient. Alternatively, a government maximizing median voter utility would choose for the weights $MS(.)$ a Dirac distribution centered at the median income level, this would produce a jump in marginal tax rates at the median income level. This type of analysis could also be used to assess how different tax reform proposals would map into a change in the weights $MS(z)/p$. This line of research is left for future work.

[27] I review in Section 5 the empirical results about these elasticities.

The remaining part of this Section examines the asymptotics of optimal marginal rates in the framework of the Mirrlees model. I first examine the link between the skill distribution and the income distribution (understanding this link is crucial to perform the numerical simulations of Section 5 using the empirical earnings distribution). I then rederive the formula for high income optimal rates of Section 3 by examining the asymptotics of the general formula for optimal rates discussed above. Readers less interested in technicalities can skip Section 4.3 and go directly to Section 5.

4.3 Optimal Asymptotic Rates

4.3.1 From the skill distribution to the income distribution

Section 3 has shown that Pareto distributions provide a benchmark of central importance to understand optimal asymptotic rates. The optimal rate depends on the limiting behavior of the tail of the income distribution. This limiting behavior can be characterized by a limiting "Pareto" parameter. This subsection first defines this limiting "Pareto" behavior in a rigorous way. I then show that the limiting "Pareto" parameters of the skill distribution and of the earnings distribution are linked through the asymptotic uncompensated elasticity.

$F(n)$ is called a Pareto distribution with parameter γ if and only if $F(n) = 1 - C/n^\gamma$ for some constant C. Its density function is equal to $f(n) = \gamma C/n^{1+\gamma}$. A Pareto density is always decreasing while empirical distributions are in general unimodal (first increasing and then decreasing). Therefore, Pareto distributions are useful to approximate empirical distributions only above the mode. If $F(n)$ is a Pareto distribution with parameter γ, then:

$$nf'(n)/f(n) = -(1+\gamma) \tag{31}$$

For any $\alpha < \gamma$,

$$\int_n^\infty m^\alpha f(m)dm/[n^\alpha(1-F(n))] = \gamma/(\gamma-\alpha) \tag{32}$$

In particular, the mean above any level $\bar{n}$ divided by $\bar{n}$ (i.e., the conditional mean ratio $E(n|n > \bar{n})/\bar{n}$) is constant and equal to $\gamma/(\gamma-1)$. From these two properties characterizing Pareto distributions, I consider two corresponding definitions of the asymptotic relationship between any given distribution and a Pareto distribution.

If $F(n)$ is a regular (at least C^2) distribution function with support in $[0, +\infty)$ and density function $F'(n) = f(n)$ such that (for some $\gamma > 0$ possibly infinite):

$$\lim_{n\to\infty} nf'(n)/f(n) = -(1+\gamma) \tag{33}$$

then I say that F behaves *strongly* like a Pareto distribution with parameter γ.

If $F(n)$ is a distribution that satisfies:

$$\lim_{n\to\infty} \int_n^\infty m^\alpha f(m)dm/[n^\alpha(1-F(n))] = \gamma/(\gamma-\alpha) \tag{34}$$

for any $\alpha < \gamma$ then I say that F behaves *weakly* like a Pareto distribution with parameter γ.

These definitions are constructed in such a way that property (33) implies (34). The reverse implication is not necessarily true if f is not regular enough. The proof is easy and presented in appendix. Now, the following proposition linking the skill distribution and the income distribution can be proved.

Proposition 2 *Suppose that the distribution of skills $f(n)$ behaves weakly like a Pareto distribution (property (34)) with parameter γ (possibly infinite). Suppose that the tax rate schedule (not necessarily optimal) T' tends to $\bar{\tau} < 1$ as n tends to infinity. Suppose also that T' is such that there is no bunching nor gaps above some income level.*

Suppose that the compensated and uncompensated elasticities converge to $\bar{\zeta}^c \geq 0$ and $\bar{\zeta}^u > -1$ as n tends to infinity. In the case $\bar{\zeta}^c = 0$ assume in addition that $\zeta^c_{(n)} \downarrow 0$ for n large.

Then the distribution of earnings behaves weakly like a Pareto distribution (property (34)) with parameter $a = \gamma/(1+\bar{\zeta}^u)$.

The formal proof is presented in appendix. The idea of the proof is easy to understand. From the definition of the uncompensated elasticity, we have $dl/l = \zeta^u dw/w$. Assuming that taxpayers face a linear tax schedule (with constant virtual income R and net of tax wage rates $w = n(1-\tau)$) and that ζ^u is constant and equal to $\bar{\zeta}^u$, it is possible to integrate the above equation over wage rates to obtain $l(n) \simeq Cn^{\bar{\zeta}^u}$ which implies $z_n = nl(n) \simeq Cn^{1+\bar{\zeta}^u}$. If the wage rates are roughly Pareto distributed above a given wage rate level (i.e., $Prob(n > \bar{n}) \simeq C/\bar{n}^\gamma$), then:

$$Prob(income > z) \simeq Prob(Cn^{1+\bar{\zeta}^u} > z) = Prob(n > (z/C)^{\frac{1}{1+\bar{\zeta}^u}}) \simeq C'/z^a$$

where $a = \gamma/(1+\bar{\zeta}^u)$. Therefore the distribution of incomes is also roughly Pareto distributed with parameter $a = \gamma/(1+\bar{\zeta}^u)$ instead of γ.

This result is important because while it is very difficult to observe distributions of skills, observing empirical distributions of wages is much easier. It must be noted that the "Pareto" parameter of the income distribution does not depend on the limiting tax rate $\bar{\tau}$. Therefore, a can be inferred directly from the observation of empirical earnings distributions. Surprisingly, the optimal taxation literature has not noticed this simple result. This may explain why researchers did not try to calibrate numerical simulations to empirical income distributions. They almost always used log-normal skill distributions which match roughly unimodal empirical distributions but approximate very poorly empirical distributions at the tails (both top and bottom tails). Moreover, changing the elasticity parameter without changing the skill distribution, as usually done in numerical simulations, might be misleading. As evidenced in Proposition 2, changing the elasticities modifies the resulting income distribution and thus might affect optimal rates also through this indirect effect. I come back to this point in Section 5.

4.3.2 Asymptotic Rates

Proposition 3 *Assume that along the optimal tax schedule, the elasticities $\zeta^c_{(n)}$ and $\zeta^u_{(n)}$ and income effects $mpe_{(n)}$ converge to values denoted by $\bar{\zeta}^c$, $\bar{\zeta}^u$ and $\overline{mpe}$ when n tends to infinity. Assume that $\overline{mpe} > -1$ (and therefore $\bar{\zeta}^u > -1$). Assume that the ratio of the social marginal utility to marginal value of public funds $G'(u_n)u_c^{(n)}/p$ converges to $\bar{g}$ as n tends to infinity. Assume that the distribution of skills $f(n)$ behaves strongly like a Pareto distribution (property (33)) with parameter γ (possibly infinite). Assume also that there is no bunching above some income level.*

Then, the optimal tax rate T' tends to a limit $\bar{\tau}$ such that:

$$\bar{\tau} = \frac{1-\bar{g}}{1-\bar{g}+\bar{\zeta}^u+\bar{\zeta}^c(a-1)} \tag{35}$$

where, $a = \gamma/(1+\bar{\zeta}^u)$ is the "Pareto" parameter of the tail of the income distribution (as in Proposition 2). If formula (35) leads a value bigger than one for $\bar{\tau}$ then it must be understood that T' tends to one.

The full proof is in appendix. However, using Proposition 1 and 2 and Lemma 1, I can give an idea of the proof. If we admit that T' converges then the term involving T'' in (22) becomes negligible and therefore dz_s/z_s can be replaced by $(1+\zeta^u)ds/s$ in (26). Now assuming that $G'(u)u_c/p$ is constant and equal to $\bar{g}$, that the elasticities are constant, and that $f(n)$ is exactly Pareto distributed, straightforward calculations using (31) and (32) show that T' is exactly equal to $\bar{\tau}$ of Proposition 3.

Therefore, this sketch shows that if the skills are exactly Pareto distributed, the elasticities exactly constant, and the social marginal value constant above a given level of skills then the government would apply a constant marginal rate above this level of skills. Thus formula (35) is likely to be relevant over a broad range of incomes. I come back to this issue in more detail in Section 5.

5 Empirical Results and Simulations

This Section is divided in two parts. First, I examine empirical distributions of wages and discuss elasticity estimates found in the applied literature in order to present asymptotic optimal rates for a range of realistic parameters. Second, I perform numerical simulations to compute optimal tax schedules in the Mirrlees model using empirical earnings distributions.

5.1 Optimal high income tax rates

5.1.1 Empirical elasticities

Labor supply studies have consistently found small or negative uncompensated elasticities of male hours of work (see Pencavel (1986), p.69 and p.73). These studies find in general uncompensated elasticities slightly below 0 (around -0.1) and compensated elasticities slightly higher than zero (around 0.1). Non-linear budget set studies which tend to find larger compensated elasticities have also found small uncompensated elasticities (see Hausman (1985), p.241). The estimates for uncompensated elasticities are also around 0 but the compensated elasticities are usually between 0.2 and 0.5. The labor supply elasticity of women has been found in general higher than the one for men (e.g. Eissa (1995)). Elasticity estimates range in general from 0.5 to 1. However, it should be noticed that the relevant elasticity for a couple is the average elasticity with weights equal to the income share of each member. Even if the elasticity of the second earner is high, the total elasticity of the couple is likely to remain small because the share of the second earner's income is usually small.

Nevertheless, we have seen that for the optimal tax problem, what matters is the total elasticity of earnings and not only the elasticity of hours of work. The former should be higher than the later because hours of work are not the only dimension of "effort". Individuals can vary their labor supply not only by changing hours but also the intensity of work or the types of job they enter in[28]. Several recent empirical studies have found large elasticities of taxable income with respect to net of tax rates (Lindsey (1987), Feldstein (1995), Navratil (1995) and Auten and Carroll (1997)). The elasticities estimated by these authors are around (or even above) one.

These high elasticity results have been criticized on several grounds. First, these studies compare the increase in incomes of high income earners (who experienced large marginal rate cuts) to the increase in incomes of middle or low income earners (who experienced much smaller marginal tax cuts). This methodology amounts therefore to attributing the widening in inequalities to the tax reform. Second, the tax cuts of the 1980s introduced many changes in tax rules which affected the incentives for reporting taxable income. In particular, the incentives for shifting labor income to capital income or for shifting personal income to corporate income may have been substantially reduced by the tax reforms. This issue is investigated in Auerbach and Slemrod (1997) and Gordon and Slemrod (1999). Saez (1997) estimates compensated elasticities of reported income with respect to tax rates using the bracket creep in the US from 1979 to 1981. Although this tax change induced smaller tax rate changes than the tax reforms of the 1980s, it does not suffer from the two problems mentioned. This study finds much lower income elasticity estimates between 0 and 0.5. Last, the tax cuts studies are unable to distinguish between permanent shifts to the form of compensation and temporary shifts to the timing of compensation. This issue was pointed out in Slemrod (1995). Goolsbee (1997) investigates this point using the tax rate increases for high income earners enacted in 1993 and compensation data on corporate executives from 1991 to 1995. He shows convincingly that the tax reform led to a large income shifting from 1993 to 1992 to escape higher tax rates, implying a very large short term elasticity (above one); however, the elasticity after one year is small (at most 0.4 and probably close to zero).

Contrary to most labor supply studies, tax reform studies are in general unable to estimate both substitution and income effects. The elasticities estimated are therefore a mix of compensated and uncompensated effects. In summary, the elasticities of total earnings for high income earners are still poorly known. They are likely to be smaller than those found in the studies of Lindsey (1987) and Feldstein (1995) and may not be significantly larger than those of middle

[28]Feldstein (1995) explains this point in more detail.

income earners.

5.1.2 Empirical wage income distributions

Section 4.1 showed that the conditional mean income ratio (i.e. $E(z|z > \bar{z})/\bar{z}$) is an important element for optimal tax rates. I have computed this function using data on wage earnings from individual tax returns. The Internal Revenue Service (IRS) constructs each year a large cross-section of tax returns (about 100,000 observations per year). These datasets overweight wealthy taxpayers and therefore are one of the most valuable source of information about high income earners. As almost all wealthy taxpayers are married filing jointly, I focus only on this class of taxpayers. As I consider taxation of labor income, I focus mostly on wage income.[29] I define narrowly wage income as income reported on the line "wages, salaries and tips" of the US income tax form.

Figures 2 and 3 plot the values of the conditional mean income ratios as a function of $\bar{z}$ for two different ranges of income. Figure 2 is for incomes between 0 to 500,000 dollars (all Figures are expressed in 1992 dollars and represent yearly income) and Figure 3 for incomes between \$10,000 to \$30 million using a semi-log scale. The Figures show that the conditional mean income ratio is strikingly stable over the tail of the income distribution. The value is around 2.3 for 1992 and 2.1 for 1993. If anything, the curve seems to be slightly increasing from \$100,000 to \$5 million. The plots on Figure 3 become noisy above \$10 million because the number of taxpayers above that level is very small and crossing only one taxpayer has a non trivial discrete effect on the curves. As discussed in Section 3, the ratio must be equal to one at the level of the highest income. However, Figure 3 shows that even at income level \$30 million, the ratio is still around 2. For example, if the second top income taxpayer earns half as much as the top taxpayer then the ratio is equal to 2 at the level of the second top earner. Consequently, the zero top result only applies to the very highest taxpayer and is therefore of no practical interest. Empirical distributions give much support to the assumption that the conditional mean income ratio converges as income increases. In fact, above 150,000 dollars, this ratio can be considered as roughly constant and thus the theory developed in Section 3 is relevant over

[29] It is well known that wealthy taxpayers tend to shift labor income to capital income in order to pay less taxes (see Slemrod (1996)). Note however that after the Tax Reform Act of 1986 and until the tax increases of 1993, tax rates on labor and capital were very similar and therefore the incentives for income shifting were probably much lower than they had been before.

a broad range of incomes. As seen in Section 3, nearly constant conditional mean income ratio means that the income distribution can be well approximated by a Pareto distribution with parameter $a = (z_m/\bar{z})/(z_m/\bar{z} - 1)$ and therefore formula (12) can be applied. Pareto parameters for the wage income distribution are estimated between 1.8 and 2.2 (depending on years[30]).

The mean ratio declines quickly until \$60,000 and then increases from 1.7 to 2.2 until \$130,000. Therefore, if elasticities were roughly constant above \$60,000, the results of Section 3 show that the optimal linear tax rate $\bar{\tau}$ that the government would like to set above the income level $\bar{z}$ is increasing over the range 60,000 to 130,000 dollars. This suggests that the optimal non-linear tax rate is likely to be increasing over that range. I examine this point in detail later on.

The IRS has constructed tax returns files since year 1960. Therefore, it is possible to plot the conditional mean income ratios for many different years and various types of incomes. Because of limited space, I present only two additional Figures. On Figure 4, I plot conditional mean income ratios for years 1987 to 1993 and wage income between 0 and 1,000,000 dollars (incomes are expressed in 1992 dollars). The vertical scale has been expanded so as to stress the differences between the different years. The conditional mean income ratios vary from year to year from a low 1.85 (in 1987) to a high 2.25 (in 1992). In year 1987, the TRA of 1986 was not yet fully phased in and the top tax rate was 38.5% (instead of 28% in 1988). From 1988 to 1992, the top rate was relatively stable (28% in 1988 and 1989 and 31% in 1990, 1991 and 1992). In 1993, the top rate was increased to 39.6% (the top rate for capital gains remained at 28%). The ratio is the lowest for 1987 and one of the highest for 1988, suggesting income shifting from 1987 to 1988 to avoid the high 1987 top rate. The ratio for 1992 (which was the last year before OBRA 1993 significantly increased the top rate) is the highest one, suggesting again a shift from 1993 to 1992 to avoid high rates.

Figure 5 presents the same plots for Adjusted Gross Income (AGI is a measure of total income including both capital and labor income). The ratios are higher than for wages (from a low 2.4 to a high 2.7). 1987 is one of the lowest years.[31] Year 1988 is by far the highest year, supporting the shifting interpretation. The difference between 1992 and 1993 is much smaller for AGI than for wages. The 1993 tax increase did not affect capital gain taxes and thus shifting labor income toward capital income may have decreased the conditional mean income ratio for

[30]Feenberg and Poterba (1993) have estimated Pareto parameters between 1.5 and 2.5 for the top distribution of Adjusted Gross Income over the period 1951-1990.

[31]Only 1991, which was a sharp recession year, is lower.

wages without much affecting the AGI ratio. Looking at the conditional mean income ratios provides interesting information about high income taxpayers' responses to marginal rates and suggests that most of the response is due to short run intertemporal shifts of income around tax reforms years. Extending this study to other years and other types of incomes is left for future research.

It is also interesting to plot the empirical ratio $[1 - H(z)]/zh(z)$, which I call from now on the hazard ratio. This ratio has been shown to be highly relevant for computing optimal tax rates in the general non-linear case. This ratio is exactly equal to $1/a$ if $H(z)$ is Pareto distributed with parameter a. Figure 6 presents the graphs of the ratio $(1 - H)/(zh)$ and of $1 - 1/E(\tilde{z}|\tilde{z} > z)$ (the later one is plotted in dashed line and is given for reference because it also tends to $1/a$). The hazard ratio $(1 - H)/(zh)$ is noisier than the conditional mean income ratio which is not surprising. Asymptotic values are roughly the same for incomes above \$200,000. Both curves are U-shaped but the pattern of the two curves below the \$200,000 income level are different: the hazard ratio is much higher for low incomes, it decreases faster until income level \$80,000; the hazard ratio then increases faster until \$200,000. From \$80,000 to \$200,000, the hazard ratio increases from 0.32 to 0.55. This pattern suggests that optimal rates should be also U-shaped: high marginal rates for low incomes, decreasing marginal rates until \$80,000 and then increase in marginal rates until level \$200,000. This particular pattern of the hazard ratio confirms the previous intuition that increasing marginal rates at high income levels are justified from an optimal taxation point of view if elasticities are constant.

5.1.3 Estimates of high income optimal tax rates

Table 1 presents optimal asymptotic rates using formula (12) for a range of realistic values for the Pareto parameter of the income distribution, $\bar{\zeta}^u$ and $\bar{\zeta}^c$, (the asymptotic elasticities) and $\bar{g}$ (ratio of social marginal utility of income for infinite income to the marginal value of public funds[32]). The Pareto parameter takes 3 values: 1.5, 2 and 2.5. Empirical wage distributions have a Pareto parameter close to 2 and AGI distributions have a parameter closer to 1.5. In the 1960s and 1970s the Pareto parameter of wages and AGI distributions were slightly higher (around

[32]Diamond (1998) presented a table of asymptotic rates in function of the Pareto parameter a of the *skill* distribution, the elasticity of earnings (in the case he considers, compensated and uncompensated elasticities are identical) and the ratio $\bar{g}$. He looked at a wider range of Pareto parameters but confused a and $1 + a$ in selecting examples.

2.5). Uncompensated elasticity takes three values: 0, 0.2 and 0.5. Compensated elasticity takes 3 values: 0.2, 0.5 and 0.8. Two values are chosen for $\bar{g}$: 0 and 0.25.

Except in the cases of high elasticities, the optimal rates are fairly high. Comparing the rows in Table 1, it appears that the Pareto parameter has a big impact on the optimal rate. Comparing columns (2), (5) and (7) (or columns (3), (6), (8)), we see that at fixed compensated elasticity, the optimal rate is very sensitive to the uncompensated elasticity. This confirms the intuition that deadweight burden computations, which depend only on compensated elasticities, may be misleading when discussing tax reforms.

The most convincing elasticity estimates from the empirical literature suggest that the long-term compensated elasticity should not be bigger than 0.5 and that the uncompensated elasticity is probably even smaller. Table 1 suggests that in this case, the optimal top rate on labor income should not be lower that 50% and maybe as high as 80%.

5.2 Numerical simulations of tax schedules

I now present simulations using the distribution of wages of 1992. I use utility functions with constant compensated elasticity ζ^c. Fixing the compensated elasticity has several advantages. First, the compensated elasticity is the key parameter of most empirical studies and therefore, having this parameter fixed over the whole population provides a good benchmark for simulations. Second, deadweight burdens are very easy to compute for utility functions with constant compensated elasticity functions. I derive in appendix the general form of utility functions with constant compensated elasticity. In the simulations, I use two types of utility functions with constant elasticities.

With utility functions of Type I, there are no income effects and therefore compensated and uncompensated elasticities are the same. The utility function takes to following form:

$$u = \log(c - \frac{l^{1+k}}{1+k}) \tag{36}$$

The elasticity is equal to $1/k$. This case was examined by Atkinson (1990) and Diamond (1998). Maximization of this utility function with a linear budget constraint $c = n(1-\tau) + R$ leads to the following first order condition: $l = (n(1-\tau))^\zeta$. Therefore, labor supply l tends to infinity at rate n^ζ. Moreover, positive tax rates reduce labor supply by a factor $(1-\tau)^\zeta$ and therefore have a large negative impact on output.

Type II utility functions are such that,

$$u = \log(c) - \log(1 + \frac{l^{1+k}}{1+k}) \tag{37}$$

The compensated elasticity is equal to $1/k$ but there are income effects. The uncompensated elasticity ζ^u can be shown to tend to zero when n tends to infinity. Realistically, when n increases to infinity, l can be shown to tend to a finite limit equal to $\bar{l} = [(1+k)/k]^{1/(k+1)}$ whatever the linear tax rate τ is. Therefore, taxes have not such a negative impact on output compared to the previous utility function.

I use the wage income distribution of year 1992 to perform numerical simulations. The skill distribution is calibrated such that given the utility function and the *actual* tax schedule, the resulting income distribution replicates the *empirical* wage income distribution. The original Mirrlees (1971) method of computation will be used. The main difficulty here comes from the fact that the empirical distribution is used. The details of the numerical computations are presented in appendix.

Optimal rates are computed such that the ratio of government spending E to aggregate production is equal to 0.25. Optimal rates simulations are performed for the two types of utility functions, two different social welfare criteria (Utilitarian and Rawlsian) and two compensated elasticity parameters ($\zeta^c = 0.25$ and $\zeta^c = 0.5$). Because for both types of utility functions, $u_c \to 0$ as $n \to \infty$, $\bar{g}$ is always equal to zero and thus the asymptotic rates are the same with both welfare criteria. The social marginal weights $MS(z)$ are roughly decreasing at the rate $1/z$.

Results are reported on Figures 7 to 10. Optimal marginal rates are plotted for yearly wage incomes between 0 and 300,000 dollars. The curves represent the optimal non-linear marginal rates and the dotted horizontal lines represent the optimal linear rates (see below). As expected, the precise level of the rates depends on the elasticities and on the type of the utility function. In all cases, however, the optimal rates are clearly U-shaped.[33] Optimal rates are decreasing from \$10,000 to \$75,000 and then increase until income level \$200,000. Above \$200,000 the optimal rates are close to their asymptotic level.

As expected, the Rawlsian criterion leads to higher marginal rates (note that Rawlsian marginal rates at the bottom are equal to one). The difference in rates between the two welfare criteria is larger at low incomes and decreases smoothly toward 0 (the asymptotic rates are

[33] The rate at the bottom is not zero because labor supply tends to zero as the skill n tends to zero, violating one of the assumptions of Seade (1977).

the same). As a consequence, the U-shape is less pronounced for the Rawlsian criterion than for the Utilitarian criterion (compare Figures 7 and 9 and Figures 8 and 10). Unsurprisingly, higher elasticities lead to lower marginal rates. Note also that higher elasticities imply a more pronounced U-shape and therefore a more non-linear tax schedule.

I have also reported on the Figures the optimal *linear* rates computed for the same utility functions, welfare criteria and skill distribution. The optimal linear rates are also computed so that government spending over production be equal to 0.25. The optimal rates are represented by the horizontal dotted lines (the upper one corresponding to $\zeta^c = 0.25$ and the lower one to $\zeta^c = 0.5$). Table 2 reports the optimal average rates[34] in the non-linear case along with the optimal linear rate.[35] The guaranteed consumption levels of people with skill zero (who supply zero labor and thus earn zero income) in terms of average income are also reported. As average incomes differ in the linear and non-linear cases, I report (in parenthesis), below the guaranteed income level for the linear case, the ratio of the guaranteed income for the linear case to the guaranteed income for the non-linear case: this ratio allows a simple comparison between the absolute levels of consumption of the least skilled individuals in the linear and non-linear case.

The average marginal rates are lower in the non-linear cases than in the linear cases. The guaranteed levels of consumption are slightly higher in relative terms in the linear cases (than in the non-linear cases) but as production is lower in the linear cases, the absolute levels are similar. Therefore, non-linear taxation is significantly more efficient than linear taxation to redistribute income. In particular, it is better from an efficiency point of view to have high marginal rates at the bottom (which corresponds to the phasing out of the guaranteed income level). It should be noted also that the linear rate is higher than the non-linear asymptotic rate in the Rawlsian case but the reverse is true in the utilitarian case. With a utilitarian criterion, high income earners face higher marginal tax rates (and therefore end up paying more taxes) in the non-linear case than in the linear case.

Mirrlees (1971) found much smaller optimal marginal rates in the simulations he presented. Rates were slightly decreasing along the income distribution and around 20% to 30%. The smaller rates he found were the consequence of two effects. First, the utility function he chose ($u = \log(c) + \log(1-l)$) implies high elasticities. Income effects are constant with $mpe = -0.5$ and compensated elasticities are large with ζ^c decreasing from around 1 (at the bottom decile)

[34]The average is weighted by incomes (i.e. $\int zT'(z)h(z)dz / \int zh(z)dz$).

[35]The asymptotic rate in the non-linear case is reported in parenthesis.

to 0.5 (at the top decile). These high elasticities lead to low optimal tax rates. Second, the log-normal distribution for skills implies that the hazard ration $(1-H(z))/(zh(z))$ is decreasing over the income distribution and tends to zero as income tends to infinity. This implied a decreasing pattern of optimal rates.

Subsequently, Tuomala (1990) presented simulations of optimal rates using utility functions with smaller elasticities.[36] Unsurprisingly, he found higher tax rates but because he still used a log-normal distribution of skills. The pattern of optimal rates was still regressive, from around 60% at the bottom to around 25% at 99th percentile. Calibrating carefully the skill distribution on the empirical income distribution is thus of much importance to obtain reliable results with numerical simulations. In particular, using log-normal skill distribution always leads to regressive tax schedules-especially at the high end of income distribution.

6 Conclusion

This paper has made an attempt to understand optimal taxation of income using the concepts of compensated and uncompensated elasticities of labor income with respect to marginal tax rates. This approach has proved fruitful on various grounds.

First, a simple formula for optimal asymptotic rates has been derived depending on four key parameters: the compensated and uncompensated average elasticities of high income earners, the conditional mean income ratio (which is the ratio of the mean of incomes above a given level to this level of income), and the redistributive tastes of the government.

The empirical literature on the behavioral effects of taxation has failed to generate a consensus on the size of the elasticities of labor supply. The conditional mean income ratio is much easier to estimate; because of its importance for optimal taxation, this ratio deserves further and more extensive investigation. Empirical distributions of income show that this parameter is roughly constant over a very broad range of high incomes. Therefore, the asymptotic formula for marginal rates is much more relevant empirically than the well known zero marginal top rate result holding for bounded distributions of income. Using elasticity estimates from the empirical

[36]As in Stern (1976) for the linear tax case, Tuomala (1990) used the concept of elasticity of substitution between consumption and leisure to calibrate utility functions. This concept does not map in any simple way into the concepts of income effects and elasticities used in the present paper. Tuomala's utility function implies that compensated elasticity are around 0.5 but income effects are large ($mpe \simeq -1$) implying negative uncompensated elasticities.

literature, the formula for asymptotic top rates suggests that marginal rates for labor income should not be lower than 50% and may be as high as 80%.

Second, it has been shown that optimal tax formulas (both linear and non-linear) can be derived without referring to adverse selection theory by just examining the effects of small tax reforms on reported income and welfare. This method has the advantage of showing precisely how the different economic effects (welfare effects, elasticity effects and income effects) come into play and which are the relevant parameters for optimal taxation. Deriving optimal rates using the original Mirrlees approach gives no hint about the different effects at play and therefore makes the interpretation of the formulas of optimal taxes much more difficult.[37] Moreover, the original Mirrlees approach relies heavily on the fact that all individuals differ only through their skills and thus cannot be generalized to a heterogeneous population. The elasticity method used throughout this paper can be extended much more easily to deal with a heterogeneous population of taxpayers: the same formulas apply once elasticities are considered as the average elasticities over the population at given income levels.

Third, the use of elasticity concepts clarifies the relationship between the distribution of skills and the distribution of incomes. In particular, the Pareto parameters of the income distribution and of the skill distribution are linked through the asymptotic uncompensated elasticity. Numerical simulations could therefore be performed using empirical distribution of wages. The simulations showed that a U-shaped pattern for marginal rates may well be optimal. Marginal rates should be high at low income levels, decrease until the middle class is reached and then increase until it converges to the asymptotic level (which is roughly attained at a level of \$250,000 per year for a household).

My analysis can be extended in a number of ways. First, empirical income distributions deserve further examination. The hazard ratio $(1 - H(z))/(zh(z))$ and the conditional mean income ratio $E(z|z > \bar{z})/\bar{z}$ are particularly interesting because optimal rates are closely related to these ratios. They could be compared across countries and over years. Second, the general framework under which the approach used here to derive optimal tax rates is valid, needs still to be worked out precisely. In particular, knowing whether formula (30) could be implemented using a convenient algorithm would be interesting. This would allow for the estimation of optimal non-linear rates without relying on the specific framework of the Mirrlees model. Last,

[37] This may explain why the theory of optimal income taxation has remained almost ignored by the applied literature in public economics.

it might be fruitful to apply the same methodology to other tax and redistribution problems. In particular, the issue of optimal tax rates at the bottom of income distribution deserves more attention in order to cast light on the important problem of designing income maintenance programs.

Appendix A: Proofs of the Results of Section 4

Derivation of the Mirrlees's FOC for optimal rates (21)

Recall that c is defined implicitly as a function of u and l by $u = u(c, l)$. Therefore, $\partial c/\partial u = 1/u_c$ and $\partial c/\partial l = -u_l/u_c$ The first order conditions for the maximization of the Hamiltonian are given by:

$$\dot{\phi}(n) = -\frac{\partial H}{\partial u} = -[G'(u_n) - \frac{p}{u_c^{(n)}}]f(n) + \phi(n)\frac{l_n u_{cl}^{(n)}}{n u_c^{(n)}} \tag{38}$$

l_n is chosen so as to maximize H:

$$0 = \frac{\partial H}{\partial l} = p[n + \frac{u_l^{(n)}}{u_c^{(n)}}]f(n) + \phi(n)\frac{\psi_l^{(n)}}{n} \tag{39}$$

Equation (38) is a linear differential equation in $\phi(n)$ which can be integrated using the standard method and the transversality conditions $\phi(0) = \phi(\infty) = 0$:

$$\phi(n) = -\int_n^\infty [\frac{p}{u_c^{(m)}} - G'(u_m)] \exp[-\int_n^m \frac{l_s u_{cl}^{(s)}}{s u_c^{(s)}} ds] f(m) dm$$

Replacing the integrated expression of $\phi(n)$ into (39) gives immediately (21). $\mathcal{QED}$

Proof of Lemma 1

$\dot{z}_n/z_n = (l_n + n\dot{l}_n)/(nl_n)$ and $l_n = l(w_n, R_n)$ where $w_n = n(1 - T')$ is the net-of-tax wage rate and $R_n = nl_n - T(nl_n) - nl_n(1 - T')$ is the virtual income of an individual with skill n. $l(w, R)$ is the uncompensated labor supply function introduced in Section 3. Therefore,

$$\dot{l}_n = \frac{\partial l}{\partial w}[1 - T' - n(n\dot{l}_n + l_n)T''] + \frac{\partial l}{\partial R}(n\dot{l}_n + l_n)(nl_n T'')$$

and rearranging,

$$\dot{l}_n = \frac{w_n}{l}\frac{\partial l}{\partial w}\frac{l}{n} + [w_n\frac{\partial l}{\partial R} - \frac{w_n}{l}\frac{\partial l}{\partial w}]\frac{nl_n T''}{n(1 - T')}[l_n + n\dot{l}_n]$$

Using the definitions (1) and (2) along with the Slutsky equation (4), I obtain:

$$\dot{l}_n = \zeta^u \frac{l_n}{n} - \dot{z}_n \frac{l_n T''}{1 - T'}\zeta^c$$

and therefore,

$$\frac{\dot{z}_n}{z_n} = \frac{n\dot{l}_n + l_n}{nl_n} = \frac{1+\zeta^u}{n} - \dot{z}_n \frac{T''}{1-T'}\zeta^c$$

which is exactly (22). The second order condition for individual maximization is $\dot{z}_n \geq 0$. Therefore, if (22) leads to $\dot{z}_n < 0$, this means that T' decreases too fast producing a discontinuity in the income distribution. $\mathcal{QED}$

Proof of Proposition 1

In order to express optimal marginal rates in function of elasticities, I first derive formulas for ζ^u, ζ^c and mpe as a function of the utility function u and its derivatives. The uncompensated labor supply $l(w, R)$ is derived implicitly from the first order condition of the individual maximization program: $wu_c + u_l$. Differentiating this equation with respect to l, w and R leads to:

$$[u_{cc}w^2 + 2u_{cl}w + u_{ll}]dl + [u_c + u_{cc}wl + u_{cl}l]dw + [u_{cc}w + u_{lc}]dR = 0$$

Replacing w by $-u_l/u_c$, the following formulas for ζ^u and mpe are obtained:

$$\zeta^u = \frac{u_l/l - (u_l/u_c)^2 u_{cc} + (u_l/u_c)u_{cl}}{u_{ll} + (u_l/u_c)^2 u_{cc} - 2(u_l/u_c)u_{cl}} \tag{40}$$

$$mpe = \frac{-(u_l/u_c)^2 u_{cc} + (u_l/u_c)u_{cl}}{u_{ll} + (u_l/u_c)^2 u_{cc} - 2(u_l/u_c)u_{cl}}$$

and using the Slutsky equation (6),

$$\zeta^c = \frac{u_l/l}{u_{ll} + (u_l/u_c)^2 u_{cc} - 2(u_l/u_c)u_{cl}} \tag{41}$$

The first order condition of the individual (18) leads to $n + u_l/u_c = nT' = -(u_l/u_c)T'/(1-T')$. Therefore (21) can first be rewritten as follows:

$$\frac{T'}{1-T'} = -\frac{\psi_l}{u_l}\left(\frac{1-F(n)}{nf(n)}\right)\int_n^\infty [1 - \frac{G'(u_m)u_c^{(m)}}{p}]\frac{u_c^{(n)}}{u_c^{(m)}} T_{nm} \left(\frac{f(m)}{1-F(n)}\right) dm \tag{42}$$

The first part of (42) is equal to $A(n)$ iff $-\psi_l/u_l = (1+\zeta^u)/\zeta^c$. ψ is defined such that $\psi(u,l) = -lu_l(c,l)$ where c is a function of (c,l) such that $u = u(c,l)$. Therefore:

$$\psi_l = -u_l - lu_{ll} - lu_{cl}\frac{\partial c}{\partial l} = -u_l - lu_{ll} + lu_{cl}\frac{u_l}{u_c}$$

Now using (40) and (41), it is easy to see that: $(1+\zeta^u)/\zeta^c = 1 + lu_{ll}/u_l - lu_{cl}/u_c$ and therefore indeed $-\psi_l/u_l = (1+\zeta^u)/\zeta^c$.

The second part of (42) is equal to $B(n)$ if it is shown that:

$$T_{nm}\frac{u_c^{(n)}}{u_c^{(m)}} = \exp[\int_n^m (1 - \frac{\zeta^u_{(s)}}{\zeta^c_{(s)}})\frac{\dot{z}_s}{z_s}ds]$$

By definition of T_{nm} and expressing $u_c^{(n)}/u_c^{(m)}$ as an integral:

$$T_{nm}u_c^{(n)}/u_c^{(m)} = \exp[\int_n^m (-\frac{d\log(u_c^{(s)})}{ds} - \frac{l_s u_{cl}^{(s)}}{su_c^{(s)}})ds] \tag{43}$$

I note $H(s) = -(du_c^{(s)}/ds + l_s u_{cl}^{(s)}/s)/u_c^{(s)}$ the expression in (43) inside the integral.

Now, $u_c^{(s)} = u_c(c_s, l_s)$, therefore

$$du_c^{(s)}/ds = u_{cc}^{(s)}\dot{c}_s + u_{cl}^{(s)}\dot{l}_s \tag{44}$$

From (19),

$$u_c^{(s)}\dot{c}_s + u_l^{(s)}\dot{l}_s = \dot{u}_s = -l_s u_l^{(s)}/s \tag{45}$$

Substituting $\dot{c}_s$ from (45) into (44), I obtain:

$$du_c^{(s)}/ds = -[s\dot{l}_s + l_s]u_l u_{cc}/(su_c) + u_{cl}\dot{l}_s$$

Substituting this expression for $du_c^{(s)}/ds$ in $H(s)$ and using again the expressions (40), (41), we have finally:

$$H(s) = [lu_l u_{cc}/u_c^2 - lu_{cl}/u_c]\left(\frac{l_s + s\dot{l}_s}{sl_s}\right) = \left(\frac{\zeta^c - \zeta^u}{\zeta^c}\right)\frac{\dot{z}_s}{z_s}$$

which finishes the proof. Note that on bunching intervals included in (n, m), $\dot{z}_s = \dot{c}_s = 0$, $H(s) = 0$, and all the preceding equations remain true, and thus the proof goes through. $\mathcal{QED}$

Derivation of the formula for optimal rates (30) from formula (29)

I note,

$$K(z) = \int_z^\infty -mpe\frac{T'}{1-T'}h^*(z')dz'$$

Equation (29) can be considered as a first order differential equation in $K(z)$:

$$K'(\bar{z}) = D(\bar{z})[C(\bar{z}) + K(\bar{z})]$$

where $C(\bar{z}) = \int_{\bar{z}}^{\infty}[1 - \frac{MS(z)}{p}]h(z)dz$ and $D(\bar{z}) = mpe/(\bar{z}\zeta^c)$. Routine integration using the method of the variation of the constant and taking into account that $K(\infty) = 0$, leads to:

$$K(\bar{z}) = -\int_{\bar{z}}^{\infty} D(z)C(z)\exp[-\int_{\bar{z}}^{z} D(z')dz']dz$$

Integration by parts leads to:

$$K(\bar{z}) = -\int_{\bar{z}}^{\infty} C'(z)\exp[-\int_{\bar{z}}^{z} D(z')dz']dz - C(\bar{z}) \tag{46}$$

Differentiation of (46) leads directly to (30). $\mathcal{QED}$

Proof of the implications between the limiting Pareto definitions: (33) and (34)

Suppose the density $f(n)$ satisfies (33). Consider a small $\epsilon > 0$. Then for n large enough,

$$\frac{-(1+\gamma)-\epsilon}{n} \leq f'(n)/f(n) \leq \frac{-(1+\gamma)+\epsilon}{n} \tag{47}$$

Integrating (47) from n to m leads to

$$(m/n)^{-(1+\gamma)-\epsilon} \leq f(m)/f(n) \leq (m/n)^{-(1+\gamma)+\epsilon} \tag{48}$$

Integration of (48) times $m^\alpha/n^{\alpha+1}$ over m from n to ∞ leads to:

$$1/(\gamma - \alpha + \epsilon) \leq \int_n^{\infty} m^\alpha f(m)dm/[n^{\alpha+1}f(n)] \leq 1/(\gamma - \alpha - \epsilon)$$

which implies the following property for ϵ tending to zero,

$$\int_n^{\infty} m^\alpha f(m)dm/[n^{\alpha+1}f(n)] = 1/(\gamma - \alpha) \tag{49}$$

Formula (49) is also true for $\gamma = \infty$. This can be proved in a similar way by considering that $f'(n)/f(n) \leq -A/n$ for arbitrary large values A. Assume that property (49) holds for $f(n)$ and $\gamma < \infty$. Then taking ratios for any α and $\alpha = 1$ gives immediately (34). If $\gamma = \infty$ then property (49) shows that for every α, $\int_n^{\infty} m^\alpha f(m)dm$ converges and therefore tends to zero as n

tends to infinity. Moreover, $\int_n^\infty m^\alpha f(m)dm/[n^\alpha(1-F(n))] \geq 1$. Therefore $n^\alpha(1-F(n))$ tends also to zero as n tends to infinity. Integration by parts gives,

$$\int_n^\infty m^\alpha f(m)dm = n^\alpha(1-F(n)) + \alpha \int_n^\infty m^{\alpha-1}[1-F(m)]dm \tag{50}$$

and therefore, using (49) with $\alpha = 0$, I have $1-F(m) \ll mf(m)$ for m large. Thus, (50) implies that $\int_n^\infty m^\alpha f(m)dm/[n^\alpha(1-F(n))]$ tends to one as n tends to infinity. $\mathcal{QED}$

Proof of Proposition 2

I note $H(z)$ the distribution function of earnings and $h(z)$ its density function. I want to show that $I = \int_{z_n}^\infty z_m^\alpha h(z_m)dz_m/[z_n^\alpha(1-H(z_n))]$ converges to $a/(a-\alpha)$ as z_n (or equivalently[38] n) tends to infinity. Changing variables and expressing $\log(z_m/z_n)$ as an integral, I can rewrite I as follows:

$$I = \int_n^\infty \exp[\alpha \int_n^m \frac{\dot{z}_s}{z_s} ds] f(m)dm/[1-F(n)] \tag{51}$$

I assumed in Proposition 2 that there is no bunching nor gaps, thus I can replace $\dot{z}_s/z_s$ in (51) using Lemma 1,

$$\int_n^m \frac{\dot{z}_s}{z_s} ds = \int_n^m \left[\frac{1+\zeta^u}{s} - \dot{z}_s \frac{T''}{1-T'}\zeta^c\right] ds \tag{52}$$

If I assume that the elasticities ζ^c and ζ^u are constant and equal to $\bar{\zeta}^c$ and $\bar{\zeta}^u$ above a given value n, I can compute exactly the integral on the right hand side of (52):

$$\int_n^m \frac{\dot{z}_s}{z_s} ds = (1+\bar{\zeta}^u)\log(\frac{m}{n}) + \bar{\zeta}^c \log\left(\frac{1-T'_{(m)}}{1-T'_{(n)}}\right) \tag{53}$$

Therefore,

$$I = \int_n^\infty \left(\frac{m}{n}\right)^{\alpha(1+\bar{\zeta}^u)} \left(\frac{1-T'_{(m)}}{1-T'_{(n)}}\right)^{\alpha\bar{\zeta}^c} \frac{f(m)}{1-F(n)} dm \tag{54}$$

Now using the assumption that T' converges to $\bar{\tau}$ (remember $\bar{\tau} < 1$), and using property (34) with $\alpha(1+\bar{\zeta}^u)$, I have:

$$\lim_{n\to\infty} I = \frac{\gamma}{\gamma - \alpha(1+\bar{\zeta}^u)} \tag{55}$$

which shows that $H(z)$ satisfies the weak Pareto property (34) with parameter $a = \gamma/(1+\bar{\zeta}^u)$.

[38] This point is clear by considering equation (53) below.

The proof assuming only that elasticities converge is similar but more technical. I can compute the two terms of the integral in (52) only approximately. Let ϵ be any (small) positive number. For n large enough and any $s \geq n$ (remember $\bar{\zeta}^u > -1$),

$$0 < (1 + \bar{\zeta}^u - \epsilon)/s \leq (1 + \zeta^u)/s \leq (1 + \bar{\zeta}^u + \epsilon)/s \tag{56}$$

Integrating (56) from n to m, I obtain,

$$\left(\frac{m}{n}\right)^{\alpha(1+\bar{\zeta}^u-\epsilon)} \leq \exp[\int_n^m \frac{\alpha(1+\zeta^u)}{s} ds] \leq \left(\frac{m}{n}\right)^{\alpha(1+\bar{\zeta}^u+\epsilon)} \tag{57}$$

The second term in (52) is harder to control. I note $\delta(s) = -\dot{z}_s T''/(1 - T')$. I have used above that $\int_n^m \delta(s)ds = \log[(1-T'_{(m)})/(1-T'_{(n)})]$ tends to zero as n tends to infinity. The proof is already done in the case ζ^c constant, therefore my goal is to bound $D = \int_n^m \zeta^c \delta(s)ds - \bar{\zeta}^c \int_n^m \delta(s)ds$.

- First case: $\bar{\zeta}^c > 0$

Because z_n is increasing, by Lemma 1's result, for n large enough, there is a constant C such that: $\delta(s) \geq -C/s$. Now, I write $\delta(s) = \delta(s)^+ - \delta(s)^-$ where δ^+ and δ^- are the positive and negative[39] parts of δ. $0 \leq \delta(s)^- \leq C/s$ and thus $0 \leq \int_n^m \delta(s)^- ds \leq C \log(m/n)$.

As $\int_n^m \delta(s)ds = \log[(1 - T'_{(m)})/(1 - T'_{(n)})]$, I have also, for n large enough,

$$0 \leq \int_n^m \delta(s)^+ ds \leq \log[(1 - T'_{(m)})/(1 - T'_{(n)})] + C \log(m/n) \tag{58}$$

Now, for n large enough, because $\zeta^c \to \bar{\zeta}^c$,

$$|D| \leq \int_n^m |\zeta^c - \bar{\zeta}^c|[\delta(s)^- + \delta(s)^+]ds \leq \epsilon \log(m/n) + \epsilon \log[(1 - T'_{(m)})/(1 - T'_{(n)})] \tag{59}$$

which implies for n large enough (remember T' converges to $\bar{\tau} < 1$),

$$-\epsilon \log(\frac{m}{n}) - \epsilon \leq \int_n^m \zeta^c \delta(s)ds \leq \epsilon \log(\frac{m}{n}) + \epsilon \tag{60}$$

and therefore,

$$(1 - 2\alpha\epsilon) \left(\frac{m}{n}\right)^{-\epsilon\alpha} \leq \exp[\int_n^m \alpha\zeta^c \delta(s)ds] \leq (1 + 2\alpha\epsilon) \left(\frac{m}{n}\right)^{\epsilon\alpha} \tag{61}$$

Multiplying equations (57) and (61) and integrating over m from n to ∞ leads to result (55).

- Second case: $\bar{\zeta}^c = 0$. I assumed in this case that $\zeta^c \downarrow 0$. Using integration by parts,

[39] That is, $\delta^+ = max(0, \delta)$ and $\delta^- = max(0, -\delta)$ and therefore $|\delta| = \delta^+ + \delta^-$.

$$\int_n^m \zeta^c_{(s)}\delta(s)ds = \zeta^c_{(m)}\int_n^m \delta(s)ds - \int_n^m \dot{\zeta}^c_{(s)}[\int_n^s \delta(u)du]ds \tag{62}$$

The first term in (62) is clearly converging to 0. Because $\dot{\zeta}^c_{(s)} \leq 0$, the second term can be bounded as follows,

$$|\int_n^m \dot{\zeta}^c_{(s)}[\int_n^s \delta(u)du]ds| \leq \int_n^m -\dot{\zeta}^c_{(s)}|\int_n^s \delta(u)du|ds \leq C\int_n^m -\dot{\zeta}^c_{(s)}ds = C[\zeta^c_{(n)} - \zeta^c_{(m)}]$$

which tends to zero as n tends to infinity. Therefore, in this second case, we have, $\int_n^m \zeta^c_{(s)}\delta(s)ds \to 0$ and therefore an inequality of the kind of (61) can be obtained and the same proof can go through. $\mathcal{QED}$

Proof of Proposition 3

- Case: $\bar{\zeta}^c = 0$

Because the exponential term inside $B(n)$ is bigger (or equal) to one (see (25)) and $G'(u)u_c/p$ tends to $\bar{g} < 1$, for n large enough, $B(n) \geq (1-\bar{g})/2$. Now because ζ^c tends to zero, $A(n)$ tends to infinity;[40] therefore T' tends to one.

- Case: $\bar{\zeta}^c > 0$

I assume first that $\gamma < \infty$ and that the formula for $\bar{\tau}$ in Proposition 3 is such that $\bar{\tau} < 1$. Using Lemma 1's result, and noting again $\delta(n) = -\dot{z}_n T''/(1-T')$,

$$\log(S_{nm}) = \int_n^m (1 - \frac{\zeta^u}{\zeta^c})\frac{\dot{z}_s}{z_s}ds = \int_n^m (1 - \frac{\zeta^u}{\zeta^c})\left[\frac{1+\zeta^u}{s} + \zeta^c\delta(s)\right]ds$$

The same computations as in Proposition 2 (case $\bar{\zeta}^c > 0$) lead to:

$$\int_n^m (1 - \frac{\zeta^u}{\zeta^c})\zeta^c\delta(s)ds = o(1)\log(\frac{m}{n}) + [-\overline{mpe} + o(1)]\log\left(\frac{1-T'_{(m)}}{1-T'_{(n)}}\right)$$

where $o(1)$ are real functions (of m and n) tending to zero as n tends to infinity. Now,

$$S_{nm} = \left(\frac{m}{n}\right)^{(1-\bar{\zeta}^u/\bar{\zeta}^c)(1+\bar{\zeta}^u)+o(1)}\left(\frac{1-T'_{(m)}}{1-T'_{(n)}}\right)^{-\overline{mpe}+o(1)} \tag{63}$$

For ease of notation, let $\bar{H} = (1-\bar{\zeta}^u/\bar{\zeta}^c)(1+\bar{\zeta}^u)$. Using property (33) for $F(n)$, I have (see equations (47) and (48) and remember $\gamma < \infty$),

[40] Assuming $\gamma < \infty$.

$$\frac{f(m)}{f(n)} = \left(\frac{m}{n}\right)^{-(1+\gamma)+o(1)} \tag{64}$$

Now, using (63), (64) and the expression for T' in Proposition 1,

$$\frac{T'_{(n)}}{1-T'_{(n)}} = (1+o(1))\left(\frac{1+\bar{\zeta}^u}{\bar{\zeta}^c}\right)(1-\bar{g})\int_n^\infty \left(\frac{m}{n}\right)^{\bar{H}-(1+\gamma)+o(1)} \left(\frac{1-T'_{(m)}}{1-T'_{(n)}}\right)^{-\overline{mpe}+o(1)} \frac{dm}{n} \tag{65}$$

Routine algebra shows that,

$$\frac{\bar{\tau}}{1-\bar{\tau}} = \left(\frac{1+\bar{\zeta}^u}{\bar{\zeta}^c}\right)\left(\frac{1-\bar{g}}{\gamma-\bar{H}}\right)$$

where $\bar{\tau}$ is the expression for the asymptotic optimal rate stated in Proposition 3.[41] Therefore, equation (65) can be rewritten as follows:

$$\frac{T'_{(n)}}{(1-T'_{(n)})^{1-\alpha+o(1)}} = (1+o(1))\left(\frac{\bar{\tau}}{1-\bar{\tau}}\right)\int_n^\infty \beta\left(\frac{m}{n}\right)^{-1-\beta+o(1)} (1-T'_{(m)})^{\alpha+o(1)}\frac{dm}{n} \tag{66}$$

where $\beta = \gamma - \bar{H} > 0$ and $0 \leq \alpha = -\overline{mpe} < 1$. Equation (66) implies that (use $1 - T'_{(m)} \leq 1$),

$$\frac{T'_{(n)}}{(1-T'_{(n)})^{1-\alpha}} \leq \left(\frac{\bar{\tau}}{1-\bar{\tau}}\right)(1+o(1))$$

which implies that there exists some (small) $\delta > 0$ such that $T' < 1-\delta$ for n large enough. Now, using once again equation (66) with the inequality $1 - T' > \delta$, we can see that T' is bounded away from 0. Therefore, for some small $\delta > 0$, I have $\delta < T' < 1-\delta$ for n large enough.

Equation (66) is thus bounded (away from 0 and infinity); therefore, I can rewrite (66) as follows:

$$\frac{T'_{(n)}}{(1-T'_{(n)})^{1-\alpha}} = (1+o(1))\left(\frac{\bar{\tau}}{1-\bar{\tau}}\right)\int_n^\infty \beta\left(\frac{m}{n}\right)^{-1-\beta} (1-T'_{(m)})^{\alpha}\frac{dm}{n} \tag{67}$$

Note that the $o(1)$ terms have been pulled outside the integral.[42] I introduce now the following function:

[41] Note that the assumption $\bar{\tau} < 1$ is equivalent to $\gamma > \bar{H}$.

[42] This can be proved by showing that the difference between the left hand sides of equations (66) and (67) tends to zero as n tends to infinity.

$$q(n) = \int_n^\infty \beta \left(\frac{m}{n}\right)^{-1-\beta} \left(\frac{1-T'_{(m)}}{1-\bar{\tau}}\right)^\alpha \frac{dm}{n} \tag{68}$$

The derivative of $q(n)$ is equal to:

$$\dot{q}(n) = \frac{\beta}{n}\left[q(n) - \left(\frac{1-T'_{(m)}}{1-\bar{\tau}}\right)^\alpha\right]$$

Now, introducing the increasing function φ which is the inverse of the increasing function $x \to x/(1-x)^{1-\alpha}$, equation (67) can be rewritten as follows:

$$T'_{(n)} = \varphi[(1+o(1))\varphi^{-1}(\bar{\tau})q(n)] \tag{69}$$

Therefore, $q(n)$ satisfies the following differential equation:

$$\dot{q}(n) = \frac{\beta}{n}\left[q(n) - \frac{1}{(1-\bar{\tau})^\alpha}\left(1 - \varphi[(1+o(1))\varphi^{-1}(\bar{\tau})q(n)]\right)^\alpha\right] \tag{70}$$

First, note that the function $V(q) = q - (1-\varphi[\varphi^{-1}(\bar{\tau})q])^\alpha/(1-\bar{\tau})^\alpha$ is increasing in q (because φ is increasing) and takes value zero at $q=1$. Therefore if $o(1) \equiv 0$ then $q(n) \equiv 1$ is an unstable equilibrium point of the differential equation (70).[43] This property is used to show that even with the $o(1)$, term $q(n)$ tends to one as n tends to infinity.

Suppose that $q(n)$ does not converge to one. Then, there is some $\epsilon > 0$ such that for any $N > 0$, there is some $\bar{n} \geq N$ such that either $q(\bar{n}) < 1 - 2\epsilon$ or $q(\bar{n}) > 1 + 2\epsilon$. Consider first N large such that, $|o(1)| < \epsilon$ for all $n \geq N$. Suppose first, that $q(\bar{n}) < 1 - 2\epsilon$ for some $\bar{n} \geq N$ and thus $(1+o(1))q(\bar{n}) < 1 - \epsilon$.

But now, equation (70) and the fact that function $V(q)$ is increasing, implies that there is a small $\mu = -\beta V(1-\epsilon) > 0$ such that $\dot{q}(\bar{n}) < -\mu/\bar{n}$. Therefore q is decreasing at $\bar{n}$ and in fact $q(n)$ will never get larger than $1 - 2\epsilon$ for any $n \geq \bar{n}$. If it did, q should have increased at some point $n^* \geq \bar{n}$ such that $q(n^*) \leq 1 - 2\epsilon$ (this is clear by considering a graph) which is impossible. Therefore, $\dot{q}(n) < -\mu/n$ for all $n \geq \bar{n}$. This implies that $q(n)$ tends to $-\infty$ which cannot be because $q(n) \geq 0$ by definition (68).

In the same way, if $q(\bar{n}) > 1 + 2\epsilon$ for some $\bar{n} \geq N$, then for a small $\mu > 0$ and all $n \geq \bar{n}$, $\dot{q}(n) > \mu/n$, which implies that $q(n)$ tends to ∞. But because $1 - T' \leq 1$, definition (68) of $q(n)$

[43] The equilibrium is unstable because if $q(n) > 1$ then $\dot{q}(n) > 0$ and thus $q(n)$ gets further away from the equilibrium. The similar converse property holds if $q(n) < 1$.

shows that this cannot be true either. Therefore, $q(n)$ does converge to 1 as n tends to infinity. Using (69), it is clear then that $T' \to \bar{\tau}$.

The case $\gamma = \infty$ is easier and can be proved by considering that for any large A, $f(m)/f(n) \leq (m/n)^{-1-A}$ (for n large enough). Equation (65) then holds as an inequality (with A in place of γ). First, this inequality shows that $T'_{(n)}$ is bounded away from 1; second, the right-hand side is shown to converge to 0 as A increases implying that T' converges to zero.

Let me last show that, in the case $\gamma < \bar{H}$, $T' \to 1$. Suppose not, then there is some $\epsilon > 0$ and a sequence (n_k) increasing to infinity such that $1 - T'_{(n_k)} > 2\epsilon$ for all k. As $\bar{\zeta}^c > 0$, I can use that, as in the proof of Proposition 2, there is some $C > 0$ such that $\delta(s) \geq -C/s$ (for s large enough); thus (for $n \leq m$ large enough),

$$\frac{1 - T'_{(m)}}{1 - T'_{(n)}} \geq \left(\frac{m}{n}\right)^{-C}$$

Therefore, for all k large enough and any $0 \leq s \leq 1$,

$$1 - T'_{(n_k+s)} \geq (1 - T'_{(n_k)}) \left(\frac{n_k + s}{n_k}\right)^{-C} \geq \epsilon \tag{71}$$

Then, inequality (71) can be used to get a positive lower bound of $(1 - T'_{(m)})^{-\overline{mpe}+o(1)}$ over an infinity of intervals $(n_k, n_k + 1)$ in the integral appearing in (65). This integral is therefore larger than an infinite diverging sum and thus diverges. This implies that $T'_{(n)} = 1$ for n large enough which is a contradiction. $\mathcal{QED}$

Appendix B: Technical Results of Section 5

Utility Functions with Constant Compensated Elasticity

Consider a given indifference curve giving utility u. Along this indifference curve, consumption can be considered as a function of labor: $c = c(l, u)$. The budget set is $c = wl + R$. The bundle (c, l) maximizing the agent's utility is such that:

$$\frac{\partial c(l, u)}{\partial l} = w \tag{72}$$

The compensated elasticity is constant (denoted again by ζ^c) if, when the wage rate w increases by 1 percent and the consumer stays on the same indifference curve, the labor supply increases by ζ^c percent. Therefore:

$$\zeta^c \frac{\Delta w}{w} = \frac{\Delta l}{l} \tag{73}$$

Now,

$$\Delta w = \frac{\partial c(l+\Delta l, u)}{\partial l} - \frac{\partial c(l,u)}{\partial l} = \frac{\partial^2 c}{\partial l^2}\Delta l \tag{74}$$

Therefore, plugging (72) and (74) into (73):

$$\frac{\partial^2 c/\partial l^2}{\partial c/\partial l} = 1/(\zeta^c l) \tag{75}$$

For ease of notation, let $k = 1/\zeta^c$. Now equation (75) can be easily twice integrated along indifference curves to get finally:

$$c = A(u)\frac{l^{1+k}}{1+k} + B(u) \tag{76}$$

where $A(u)$ and $B(u)$ are the integration constants. Well behaved indifference curves cannot overlap and therefore by considering the case $l = 0$ in the above equation, $B(u)$ must be strictly increasing and non-negative (to rule out negative consumption). By a recardinalization of u, I assume without loss of generality that $B(u) = e^u$ (I impose $u = -\infty$ when $c = 0$). Now, by considering large values of l, to rule out once again overlapping of indifference curves, it must be the case that $A(u)$ is non-decreasing. Therefore equation (76) defines an implicit utility function $u = u(c, l)$ because for each non-negative values of (c, l) there is a unique u solving (77). Thus the general form of utility functions with constant compensated elasticity is the following:

$$c = A(u)\frac{l^{1+k}}{1+k} + e^u \tag{77}$$

where $k = 1/\zeta^c$ and $A(u)$ is non-decreasing. These utility functions are separable in consumption and labor in two cases: either if $A(u)$ is constant or proportional to e^u. The case $A(u)$ constant leads to a utility function with no income effects. These two types of separable utility functions are used in the simulations. $\mathcal{QED}$

Numerical Simulations

To simplify computations, I consider the separable form of the utility functions Type I and II. For Type I, $u = c - l^{k+1}/(k+1)$, and $G(u) = \log(u)$ (in the utilitarian case). For Type II, $u = \log(c) - \log[1 + l^{k+1}/(k+1)]$ and $G(u) = u$ (in the utilitarian case).

For both types of utility functions, optimal rates are computed by solving a system of two differential equations in $u(n)$ and $vr(n)$ where u is the utility level and vr is defined such that:

$$vr(n) = \frac{1}{\psi_l^{(n)}}(n + \frac{u_l^{(n)}}{u_c^{(n)}}) \tag{78}$$

Because of separability, $\psi_l = -u_l - lu_{ll}$ and equation (21) can be written as:

$$vr(n) = \frac{1}{nf(n)} \int_n^\infty [\frac{1}{u_c^{(m)}} - \frac{G'(u_m)}{p}] f(m) dm$$

Therefore, the system of differential equations can be written as follows:

$$\dot{vr} = -\frac{vr}{n}(1 + \frac{nf'}{f}) - \frac{1}{nu_c} + \frac{G'(u)}{pn}$$

$$\dot{u} = -\frac{lu_l}{n}$$

l and c are implicit functions of u and vr (defined by equations (78) and $u = u(c, l)$). The system of differential equations used to solve optimal rates depends on $f(n)$ through the expression[44] $nf'(n)/f(n)$ which is noisy when taken from empirical data. nf'/f is smoothed using Kernel density methods with large bandwidth. $f(n)$ is derived from the empirical distribution of wage income in such a way that the distribution of income $z(n) = nl(n)$ inferred from $f(n)$ with flat taxes (reproducing roughly the real tax schedule) matches the empirical distribution. nf'/f is taken constant above a large income level (above \$1.5 milion) and such that the Pareto parameter of the income distribution be equal to 1.9.[45]

The differential system is solved using numerical integration methods. In the utilitarian case, a value is assumed for p, then $u(0)$ ($vr(0)$ can be computed as a function of p and $u(0)$) is chosen such that the system converges to the theoretical asymptotic values. p is adapted through trial and error until government surplus over aggregate production is equal to 0.25. In the Rawlsian case, $G' \equiv 0$ and p is not defined, thus a value is assumed for $u(0)$, then $vr(0)$ is chosen such that the solution converges.[46] $u(0)$ is adapted until government surplus over aggregate production is equal to 0.25. I check that the optimal solutions lead to increasing earnings z_n.

[44] This expression should be constant in the case of a perfectly Pareto distributed skill density (see (31)).

[45] This matches the empirical wage income distribution of year 1992. Moreover, knowing the asymptotic values simplifies considerably the numerical computations.

[46] In theory, $vr(0) = \infty$; therefore in the numerical simulation, the lowest skill is taken small but positive so that the initial value of vr be well defined.

References

[1] Atkinson, Anthony B. “Public Economics and the Economic Public.” *European Economic Review*, 1990, 34(2-3), pp. 225-248.

[2] Atkinson, Anthony and Joseph Stiglitz. *Lectures on Public Economics*, London: McGraw-Hill, 1980.

[3] Auerbach, Alan and Joel Slemrod. “The Economic Effects of the Tax Reform Act of 1986.” *Journal of Economic Literature*, 1997, 35(2), pp. 589-632.

[4] Auten, Gerald and Robert Carroll. “The Effect of Income Taxes on Household Behavior.” Office of Tax Analysis, U.S. Department of the Treasury, mimeo 1997, forthcoming *Review of Economics and Statistics.*

[5] Ballard, Charles L. and Don Fullerton. “Distortionary Taxes and the Provision of Public Goods.” *Journal of Economic Perspectives*, Summer 1992, 6(3), pp. 117-31.

[6] Dahlby, Bev. “Progressive Taxation and the Social Marginal Cost of Public Funds.” *Journal of Public Economics*, 1998, 67(1), pp. 105-122.

[7] Diamond, Peter. “Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates.” *American Economic Review*, 1998, 88(1), pp. 83-95.

[8] Ebert, Udo. “A Reexamination of the Optimal Nonlinear Income Tax.” *Journal of Public Economics*, 1992, 49(1), pp. 47-73.

[9] Eissa, Nada. “Taxation and Labor Supply of Married Women: the Tax Reform Act of 1986 as a Natural Experiment.” NBER Working Paper No. 5023, 1995.

[10] Feenberg, Daniel and James Poterba. “Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns.” in *Tax Policy and the Economy*, vol. 7, ed. James Poterba. Cambridge: MIT Press, 1993.

[11] Feldstein, Martin. “The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act.” *Journal of Political Economy*, 1995, 103(3), pp. 551-572.

[12] Goolsbee, Austan. “What Happens When You Tax the Rich? Evidence from Executive Compensation.”, NBER Working Paper, No. 6333, 1997.

[13] Gordon, Roger and Joel Slemrod. "Are 'Real' Responses to Taxes Simply Income Shifting Between Corporate and Personal Tax Bases?" in *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, ed. J. Slemrod, Cambridge University Press, forthcoming 1999.

[14] Hausman, Jerry, "Taxes and Labor Supply." in A.J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, Amsterdam: North-Holland, 1985.

[15] Lindsey, Lawrence. "Individual Taxpayer Response to Tax Cuts: 1982-1984, with Implications for the Revenue Maximizing Tax Rate." *Journal of Public Economics*, 1987, (33), pp. 173-206.

[16] Mayshar, Joram. "On Measuring the Marginal Cost of Funds Analytically." *American Economic Review*, 1991, 81(5), pp. 1329-35.

[17] Mirrlees, James A. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic studies*, 1971, 38, pp. 175-208.

[18] Mirrlees, James A. "Optimal Tax Theory: a Synthesis." *Journal of Public Economics*, 1976, 6, pp. 327-58.

[19] Mirrlees, James A. "The Theory of Optimal Taxation." in K.J. Arrow and M.D. Intrilligator (eds.), *Handbook of Mathematical Economics*, Amsterdam: North-Holland, 1986.

[20] Myles, Gareth D. *Public Economics.* Cambridge: Cambridge University Press, 1995.

[21] Navratil, John. "The Economic Recovery Tax Act of 1981: Evidence on Individual Taxpayer Behavior from Panel Tax Return Data." Unpublished Harvard Thesis 1995.

[22] Pareto, Vilfredo. *Ecrits sur la Courbe de la Répartition de la Richesse*, Genève, Librairie Droz, 1965.

[23] Pencavel, John. "Labor Supply of Men." in O. Ashenfelter and R. Layard (eds.), *Handbook of Labor Economics*, 1986, Amsterdam: North-Holland, pp. 3-102.

[24] Piketty, Thomas. "La Redistribution Fiscale face au Chômage." *Revue Française d'Economie*, 1997, 12(1), pp. 157-201.

[25] Sadka, Efraim. "On Income Distribution, Incentive Effects and Optimal Income Taxation." *Review of Economic Studies*, 1976, 43(2), pp. 261-8.

[26] Saez, Emmanuel. "The Effect of Marginal Tax Rates on Income: A Panel Study of 'Bracket Creep'.", unpublished MIT mimeograph, 1997.

[27] Saez, Emmanuel. "A Characterization of the Income Tax Schedule Minimizing Deadweight Burden.", unpublished MIT mimeograph, 1999.

[28] Seade, Jesus K. "On the Shape of Optimal Tax Schedules." *Journal of Public Economics*, 1977, 7(2), pp. 203-236.

[29] Seade, Jesus K. "On the Sign of the Optimum Marginal Income Tax." *Review of Economic Studies*, 1982, 49(4), pp. 637-643.

[30] Sheshinski, Eytan. "The Optimal Linear Income Tax." *Review of Economic Studies*, 1972, 39, pp. 297-302.

[31] Slemrod, Joel. "Income Creation or Income Shifting? Behavioral Responses to the Tax Reform Act of 1986." *American Economic Review*, 1995, 85(2), pp. 175-80.

[32] Slemrod, Joel. "High Income Families and the Tax Changes of the 1980s", in *Empirical Foundations of Household Taxation*, eds. M. Feldstein and J. Poterba, University of Chicago, 1996.

[33] Stern, Nicolas H. "On the Specification of Models of Optimal Taxation." *Journal of Public Economics*, 1976, 6, pp. 123-62.

[34] Tuomala, Matti. *Optimal Income Tax and Redistribution.* Oxford: Clarendon Press, 1990.

Figure 1: Tax Reform Decomposition

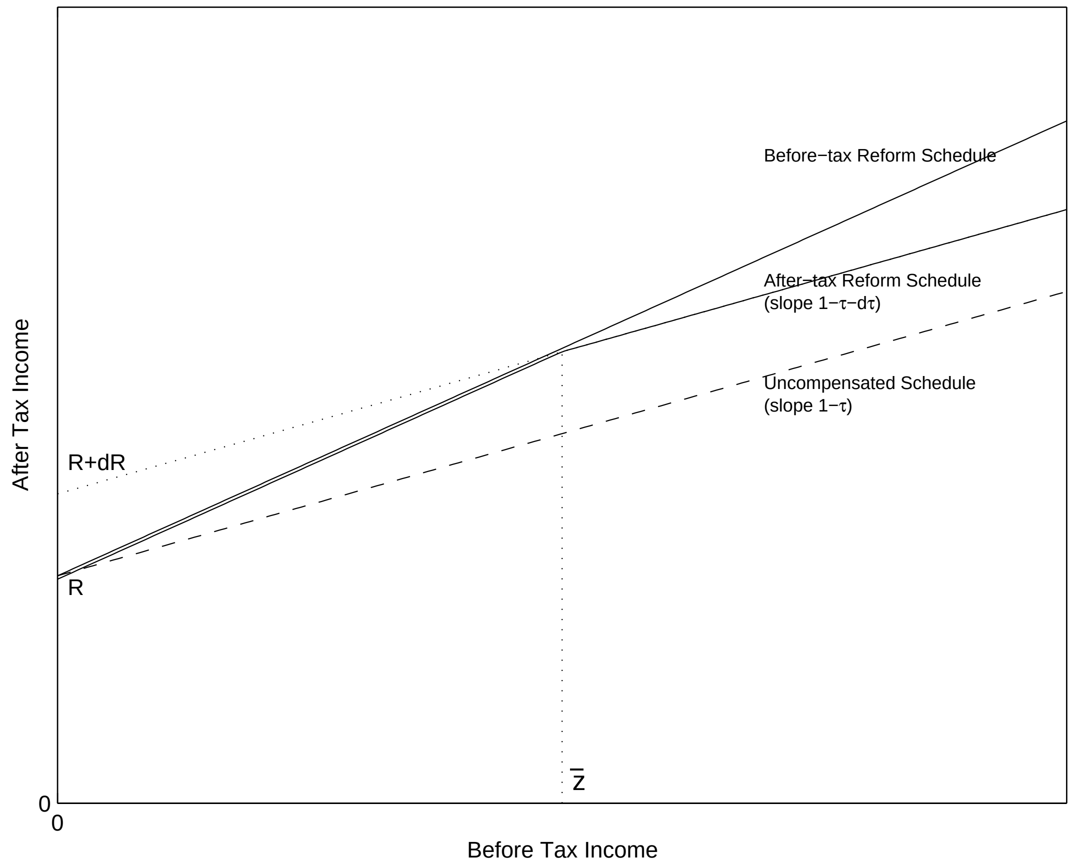

Figure 2: Conditional Mean Income ratios for wages, Years 1992, 1993

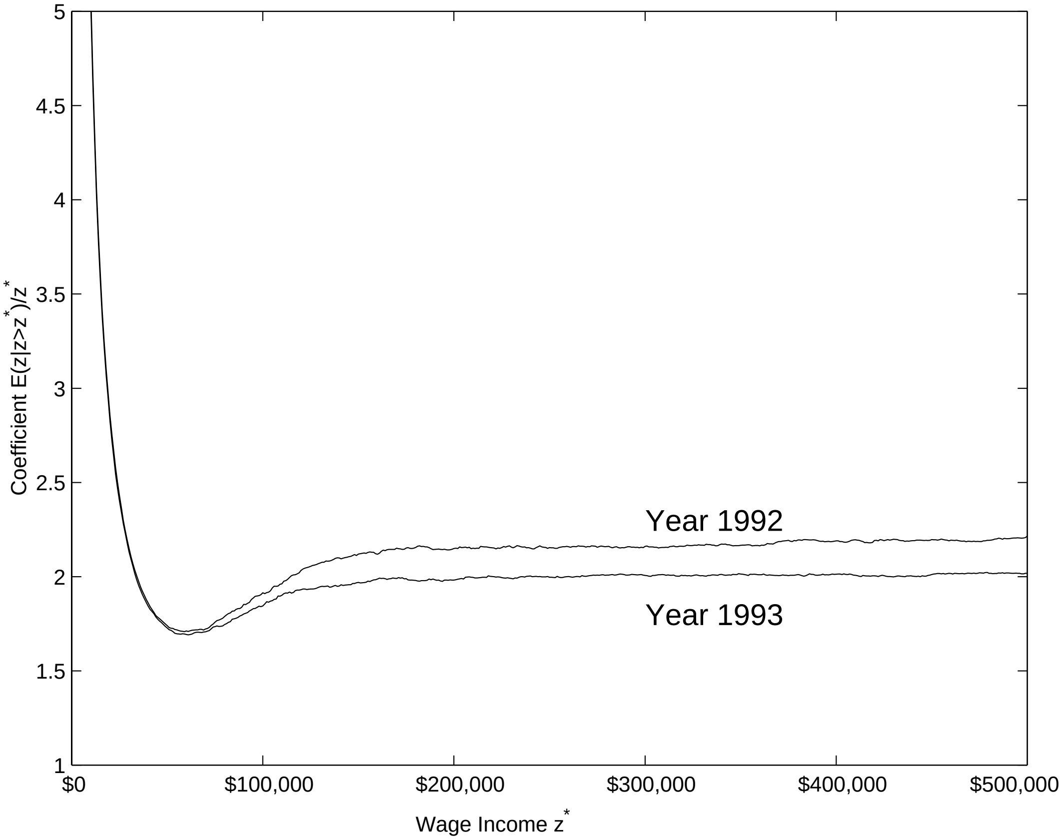

Figure 3: Conditional Mean Income ratios for wages, 1992, 1993 (semilog scale)

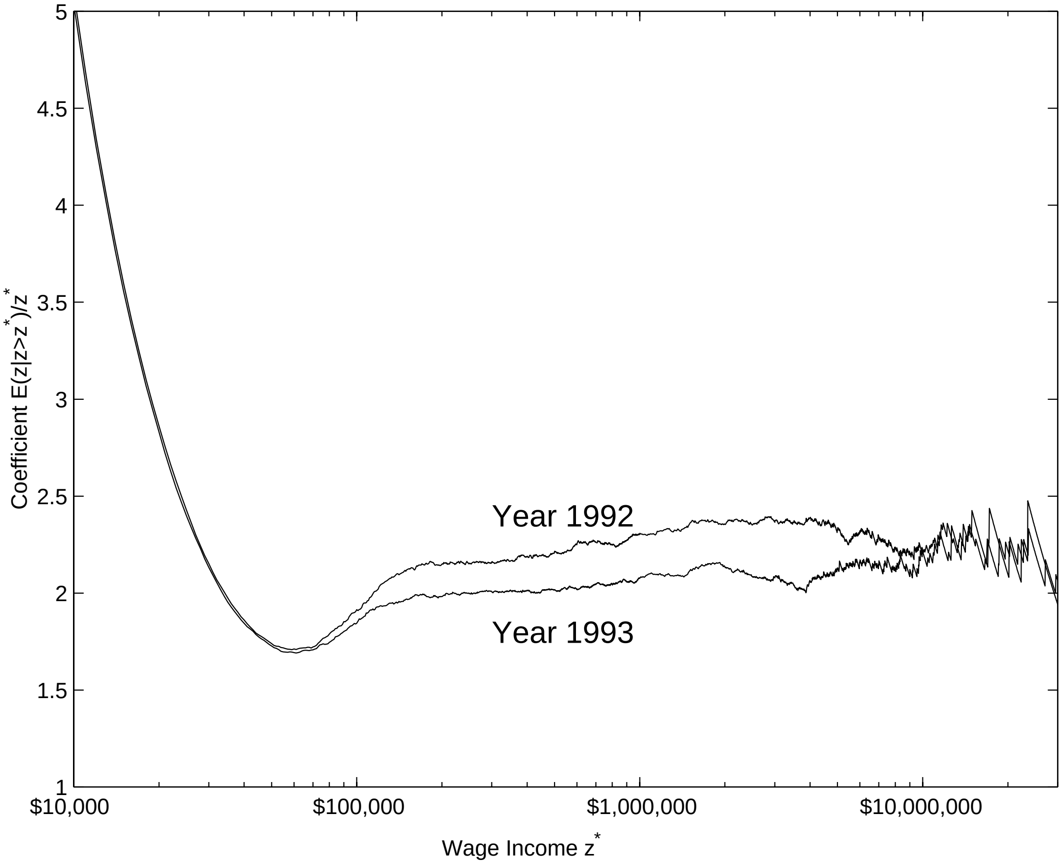

Figure 4: Conditional Mean Income Ratios, Years 1987 to 1993

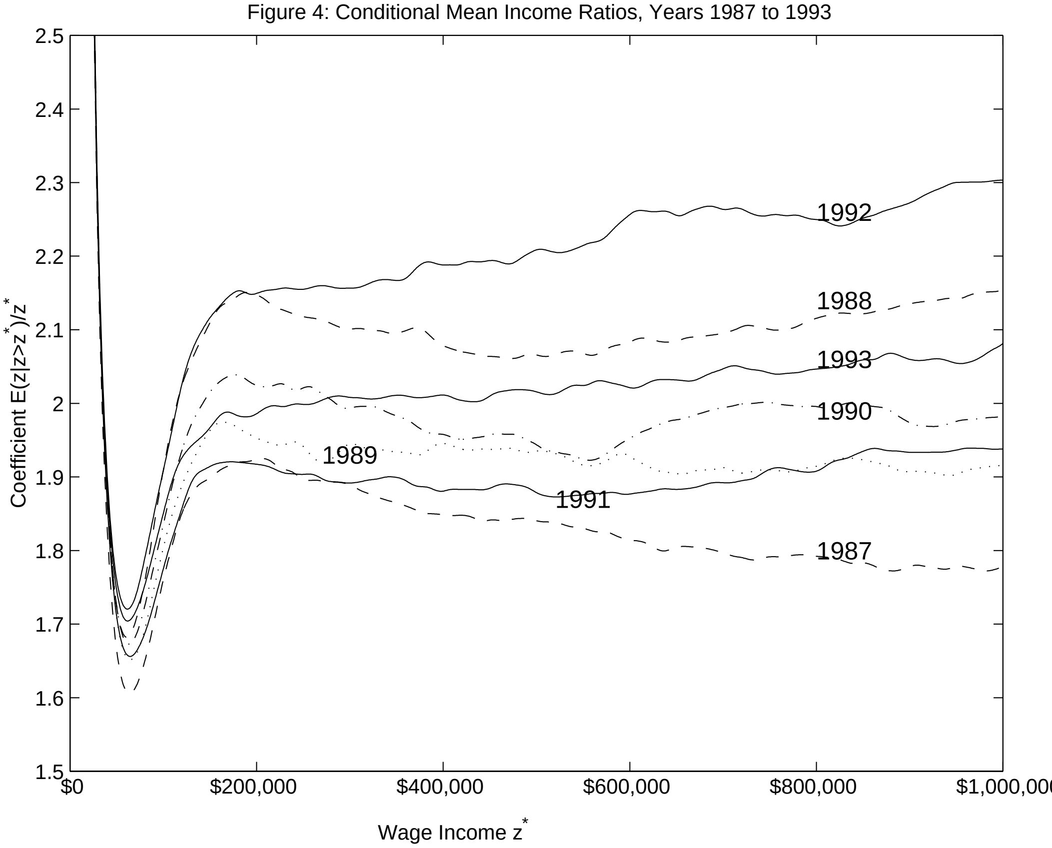

Figure 5: Conditional Mean Income Ratios, Years 1987 to 1993

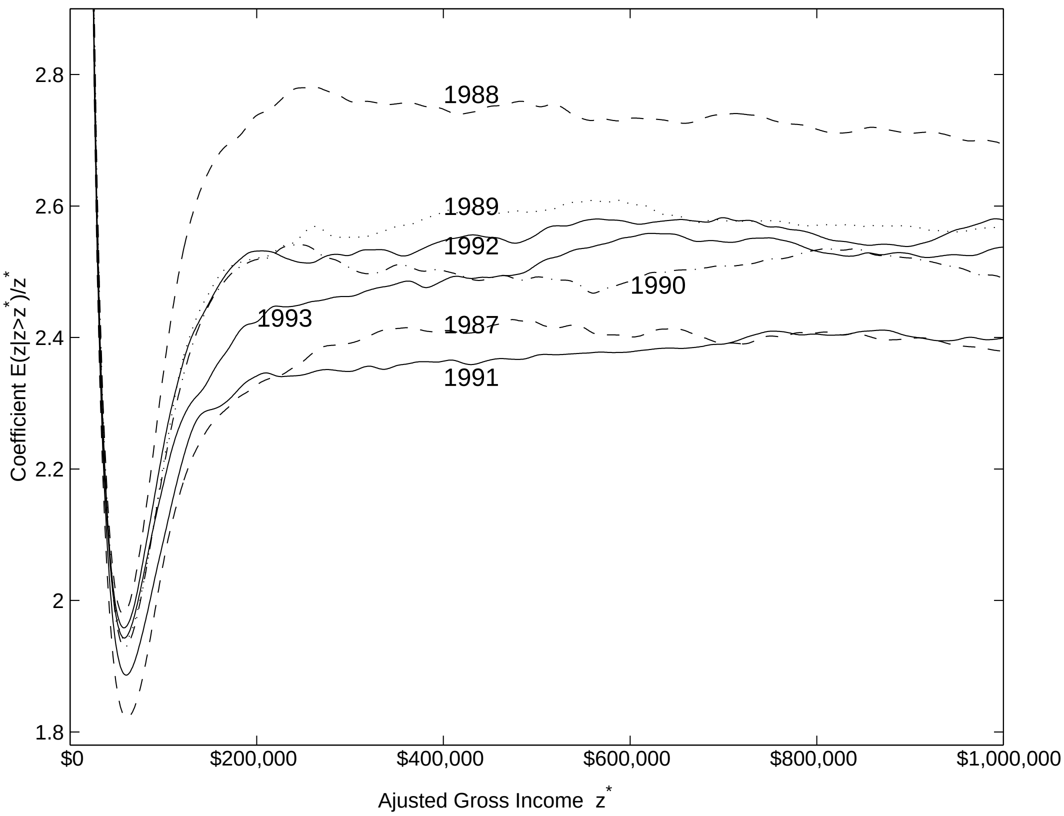

Figure 6: Hazard Ratio [1−H(z)]/[zh(z)] for wages, year 1992

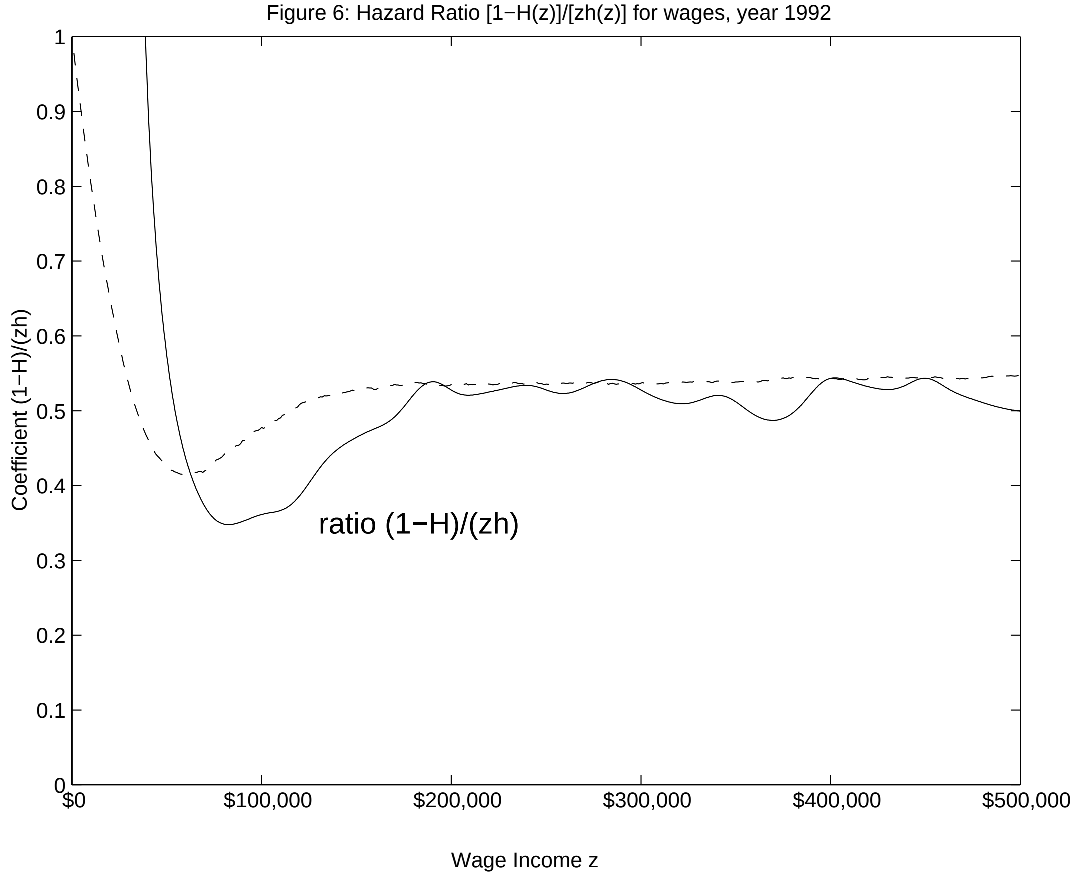

Figure 7: Optimal Rates, Utilitarian Criterion, Utility type I, ζ^c=0.25 and 0.5

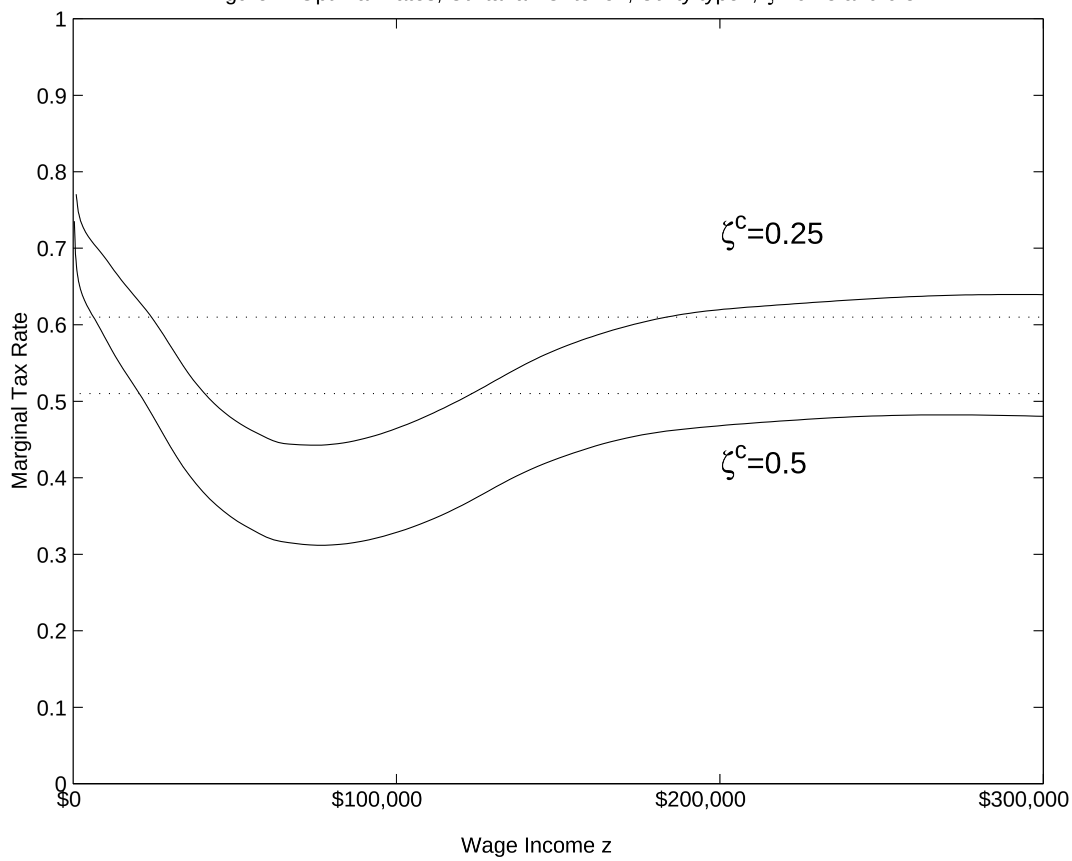

Figure 8: Optimal Rates, Utilitarian Criterion, Utility type II, ζ^c=0.25 and 0.5

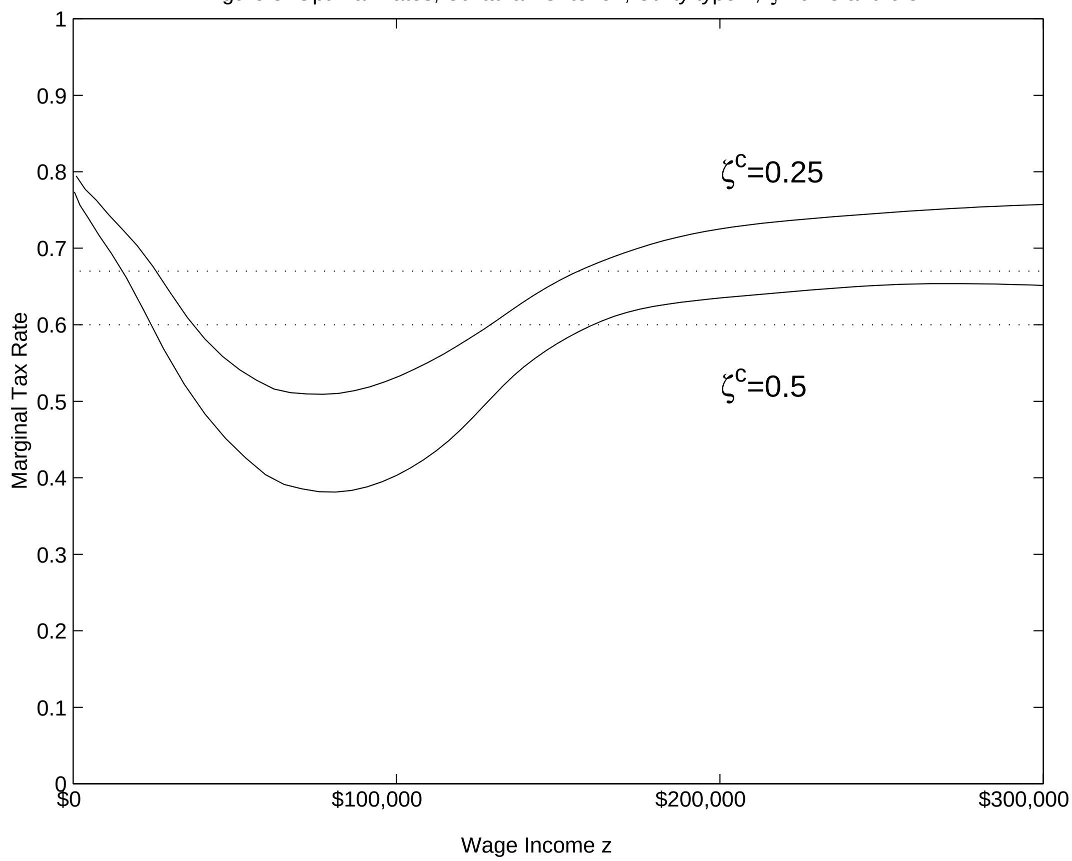

Figure 9: Optimal Rates, Rawlsian Criterion, Utility type I, ζ^c=0.25 and 0.5

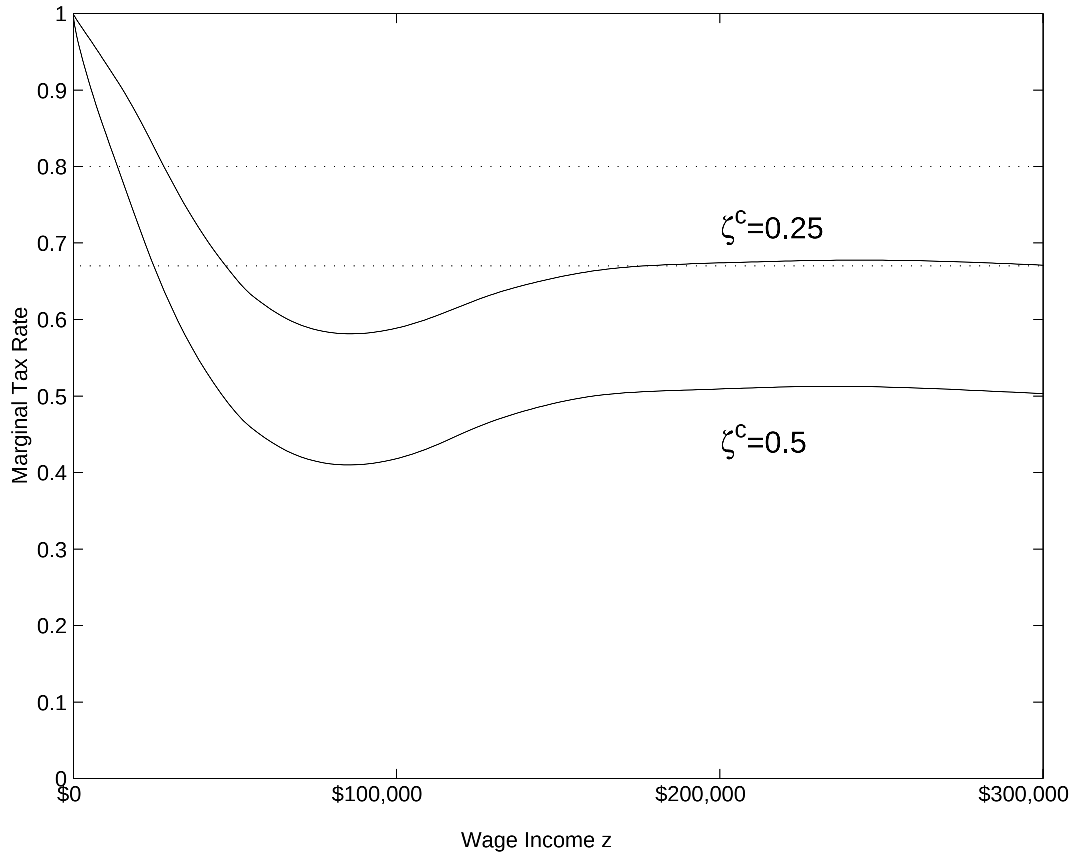

Figure 10: Optimal Rates, Rawlsian Criterion, Utility type II, ζ^c=0.25 and 0.5

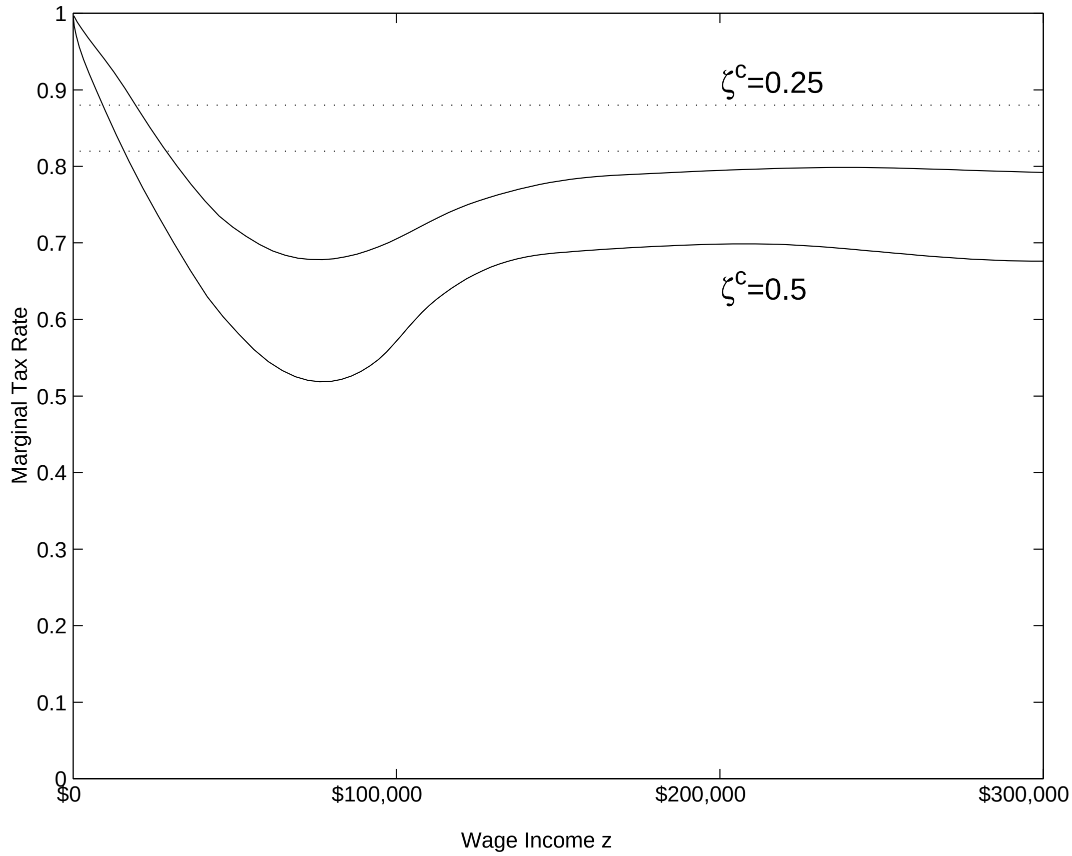

Table I: Asymptotic Marginal Rates (Optimal Rates for High Income Earners)

		Uncompensated Elasticity = 0			Uncompensated Elasticity = 0.2			Uncompensated Elasticity = 0.5	
		Compensated Elasticity			Compensated Elasticity			Compensated Elasticity	
		0.2	0.5	0.8	0.2	0.5	0.8	0.5	0.8
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: social marginal utility with infinite income (g) = 0									
Pareto	1.5	91	80	71	77	69	63	57	53
parameter	2	83	67	56	71	59	50	50	43
	2.5	77	57	45	67	51	42	44	37
Panel B: social marginal utility with infinite income (g) = 0.25									
Pareto	1.5	88	75	65	71	63	56	50	45
parameter	2	80	60	48	65	52	43	43	37
	2.5	71	50	38	60	44	32	38	31

Note: g is the ratio of social marginal utility with infinite income over marginal value of public funds. The Pareto parameter of the income distribution takes values 1.5, 2, 2.5. Optimal rates are computed according to formula (16)

Table II: Numerical Simulations for Optimal Tax Rates

	Utilitarian Criterion				Rawlsian Criterion			
	Compensated Elasticity				Compensated Elasticity			
	0.25		0.5		0.25		0.5	
	Non-linear	Linear	Non-linear	Linear	Non-linear	Linear	Non-linear	Linear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Utility Type I								
Optimal Average Rate	0.51	0.61	0.38	0.51	0.68	0.80	0.52	0.67
(Asymptotic Rate)	(0.68)		(0.51)		(0.68)		(0.51)	
Guaranteed Income Level	0.33	0.36	0.21	0.26	0.55	0.55	0.42	0.42
(linear over non-linear level)		(1.03)		(1.09)		(0.92)		(0.87)
Panel B: Utility Type II								
Optimal Average Rate	0.59	0.67	0.48	0.60	0.77	0.88	0.65	0.82
(Asymptotic Rate)	(0.81)		(0.69)		(0.81)		(0.69)	
Guaranteed Income Level	0.40	0.42	0.31	0.35	0.60	0.63	0.50	0.57
(linear over non-linear level)		(1.00)		(1.01)		(0.92)		(0.92)

Note: In the non-linear case, optimal rates are averaged with income weights; asymptotic rates are reported in parenthesis below average rates.
The guaranteed income level is expressed in percentage of average income. The ratio of the absolute guranteed level in the linear case over the absolute guarantee level in the non-linear case is reported in parenthesis.