



TAMPERE ECONOMIC WORKING PAPERS
NET SERIES

ON THE SHAPE OF OPTIMAL NON-LINEAR INCOME TAX SCHEDULE

Matti Tuomala

Working Paper 49
April 2006

<http://tampub.uta.fi/econet/wp49-2006.pdf>

DEPARTMENT OF ECONOMICS AND ACCOUNTING
FI-33014 UNIVERSITY OF TAMPERE, FINLAND

ISSN 1458-1191
ISBN 951-44-6634-9

On the shape of optimal non-linear income tax schedule

Matti Tuomala

2006-01-27

JEL.classification: H21, D31

Keywords: Optimal nonlinear income tax, distribution, numerical simulations

Abstract: This paper shows that the assumed distribution of earning abilities (wages) plays an important role in determining the optimal shape of non-linear income tax schedule. We replace the assumption of lognormal distribution used in several papers by the Champernowne distribution. Using numerical simulations we show that the U-shaped pattern of the marginal income tax rates can be obtained without assuming constant labour supply elasticity as in Diamond (1998) and Saez (2001). Our numerical results also suggest that it is either a sufficiently high inherent inequality or a combination of sufficiently high inherent inequality and sufficiently low revenue requirement that leads to a pattern of optimally increasing marginal tax rates. Furthermore, the paper also shows that the revenue requirement has a central role in determining the extent of redistribution and the level of the guaranteed minimum income.

1. Introduction

Optimal income tax schedules have been calculated under various assumptions. In simulations that followed Mirrlees (1971), the optimal income tax schedule typically exhibits an inverted U-shape. Results also turn out to be rather sensitive to particular assumptions. One of the key factors explaining the shape of optimal non-linear income tax schedule is the shape of the wage distribution among workers. As noted by Diamond (1998, page 93); “Presumably the relatively constant marginal tax rate in those simulations (Mirrlees,1971) would have had a different shape with a different assumed family of distributions”. One of our aim in this paper is just to explore how sensitive the shape of the non-linear tax schedule is to the chosen distribution.

There are, however, studies indicating that results can change significantly if either a larger standard deviation is assumed, or if the distribution of wages is Pareto rather than lognormal. Kanbur-Tuomala (1994) retain the assumption of a log-normal distribution, but simulate cases in which the standard deviation is greater than in earlier calculations. They show that increases in wage inequality can alter the qualitative pattern of optimal marginal tax rates. The optimal graduation can indeed be such that marginal tax rates increase for the majority of the population, but there continues to exist a significant income range at the top where marginal tax rates decline. Diamond (1998) in turn presents evidence that the wage distribution assumed can affect the shape of the optimal income tax schedule. He shows that if preferences are quasi-linear in consumption, an explicit formula for the marginal tax rate can be derived in terms of the distribution of wages, the elasticity of labour supply and the social welfare function¹. For this case, he finds that above some critical wage level, the pattern of marginal rates is U-shaped if the elasticity of labour supply is constant, the density of wages is single peaked and wages follows the Pareto distribution above the modal wage². Boadway et al (2000) conduct a similar analysis for the quasi-linear preferences in leisure. Saez (2001) in a more general treatment, using an elasticity approach, showed that the optimal income tax schedule would be U-shaped. As Diamond (1998) Saez (2001) assumes that the elasticity of labour supply is constant at all productivity levels. He simulates on US data a marginal tax rate schedule that slopes upwards at around fairly high income level.

¹ See also Atkinson (1990,1995).

² Dahan-Strawczynski (2000) show, however, that a rising marginal tax rates at high incomes presented by Diamond (1998) depend on the assumption of utility of consumption. Their simulations focus only on high levels of income, not the whole schedule.

Because one of the key factors explaining the shape of an optimal income tax schedule is the assumed family of distributions of earning abilities, it is an interest to look at other distributions than the lognormal and the Pareto distribution. As in Bevan (2005) we replace the lognormal distribution by the Champernowne distribution³. As commonly known the lognormal distribution fits reasonable well over a large part of income range but diverges markedly at the both tails. The Pareto distribution in turn fits well at the upper tail. Champernowne (1952) proposes a model in which individual incomes were assumed to follow a random walk in the logarithmic scale. Here we use the two parameter version of the Champernowne distribution. This distribution approaches asymptotically a form of Pareto distribution for large values of wages but it also has an interior maximum (see Figure 1).

In principle, the revenue requirement will affect the shape of the optimal tax schedule. There are very few analytical results (if any) in this area. The previous numerical calculations suggest that within a range of -10% of GDP to 10% of GDP the revenue requirement has just no effect on the shape of the marginal tax rate schedule, except for raising or lowering of the whole schedule (see Tuomala, 1984 and 1990). While going outside the range, at least in negative direction, is perhaps implausible in the context of the national economy, when redistributing between groups within an economy this is no longer the case. For example, the subsidy given to pensioners as group (as a fraction of their pre-transfer income) is probably quite high. Immonen et al (1999) show that in some cases it is optimal to transfer 50% or even 100% of pre-transfer group income to a group. They find that in these cases the optimal within group tax schedule shows increasing marginal tax rates. On the other hand in the previous numerical calculations revenue requirement has been implausible low. Therefore in this paper we study numerically how the higher revenue requirement will affect the shape of the tax schedule.

Other questions we explore in this paper are how sensitive is the shape of tax schedule to the interaction between the family of distributions of wages assumed and the utility function, how sensitive is the level of the lump sum transfer component of the tax system to the specification of the model, and what is the relationship between the lump sum transfer and the progressivity of tax schedule.

³ It is often (also in an earlier version of this paper) referred to as the Fisk distribution (see Fisk, 1961a,b).

The rest of the paper is organized as follows. Section 2 provides a summary of the original Mirrlees model and highlights the role of the underlying factors in determining the pattern of marginal tax rates. Section 3 outlines numerical solution and presents optimal asymptotic rates for a specific example. Section 4 provides numerical simulations with different specifications of the model and section 5 concludes.

2. The Mirrlees model

There is a continuum of individuals, each having the same preference ordering, which is represented by an additive utility function $u = U(x) - V(y)$ defined over consumption x and hours worked y , with $U_x > 0$ and $V_y < 0$ (subscripts indicating partial derivatives) and where $V(\cdot)$ is convex. Individuals differ only in the pre-tax wage n they can earn. There is a distribution of n on the interval (l, h) represented by the density function $f(n)$. Earnings are equal $z = ny$.

Suppose that the aim of policy can be expressed as maximizing the following social welfare criterion

$$S = \int_l^h W(u(n))f(n)dn, \quad (1)$$

where $W(\cdot)$ is an increasing and concave function of utility. The government cannot observe individuals' productivities, only the distribution $f(n)$, and thus is restricted to setting taxes and transfers as a function only of earnings, $T(z(n))$. The government maximizes S subject to the revenue constraint

$$\int_l^h T(z(n))f(n)dn = R \quad (2)$$

where R is the required revenue for essential public goods. In addition to the revenue constraint, the government faces incentive compatibility constraints. These in turn state that each n individual maximizes utility by choice of hour.

Totally differentiating utility with respect to n , and making use of workers utility maximization condition, we obtain the incentive compatibility constraints,

$$\frac{du}{dn} = -\frac{yV_y}{n} = g(u, y). \quad (3)$$

Since $T = ny - x$, we can think of government as choosing schedules $y(n)$ and $x(n)$. In fact it is easier to think of it choosing a pair of functions, $u(n)$ and $y(n)$, which maximize welfare index (1) subject to the incentive compatibility condition (3) and the revenue requirement (2). The first order conditions of this problem with respect to y gives

$$\lambda(n + \frac{V_y}{U_x})f(n) + \mu(n)(V_y + yV_{yy})/n = 0 \quad (4)$$

where $-g_y = (V_y + yV_{yy})$, λ is the multiplier on the revenue constraint and

$$\mu(n) = \int_l^n ((W'U_x - \lambda)(1/U_x))f(p)dp. \quad (5)$$

is the multiplier on the incentive compatibility constraint. This latter satisfies the transversality conditions

$$\mu(l) = \mu(h) = 0. \quad (6)$$

It is useful to rewrite marginal rates⁶ in terms of traditional labour supply elasticities. Now we can write marginal rates, $\tau(z) = T'(z)$;

⁴ The first order condition of individual's optimisation problem is only a necessary condition for the individual's choice to be optimal, but we assume here that it is sufficient as well. Assumptions that assure sufficiency are provided by Mirrlees (1976). Note also that while we here presume an internal solution for y , (3) remains valid even if individuals were bunched at $y=0$ since, for them, $du/dn=0$.

⁵ Integrating the first order condition of the problem with respect to u and using the transversality condition (6) we have (5).

⁶ See also Revesz (1989), Roberts (2000) and Saez (2001). Using formulas for E^u (=uncompensated wage elasticity) and E^c (=compensated wage elasticity) we can show that $-\frac{g_y}{V_y} = (1 + E^u) / E^c$.

Differentiating the FOC of the individual maximization, $U_x n(1 - \tau) + V_y = 0$, with respect to net wage, labour

supply and virtual income, b , we have after some manipulation elasticity formulas; $E^u = \frac{(V_y / y) - (V_y / U_x)^2 U_{xx}}{V_{yy} + (V_y / U_x)^2 U_{xx}}$,

(income effect parameter) $I = \frac{-(V_y / U_x)^2 U_{xx}}{V_{yy} + (V_y / U_x)^2 U_{xx}}$, and from the Slutsky equation $E^c = E^u - I$, then

$$E^c = \frac{(V_y / y)}{V_{yy} + (V_y / U_x)^2 U_{xx}}.$$

$$\frac{\tau}{1-\tau} = \frac{((1+E^u)/E^c)U_x}{\lambda n f(n)} \mu(n) \quad (7)$$

To get a better idea of the shape of the marginal tax rate schedule, following Atkinson (1995) and Diamond (1998), we multiply and divide (5) by $(1-F(n))$ to obtain:

$$\frac{\tau}{1-\tau} = \underbrace{\left[\frac{1+E^u}{E^c} \right]}_{A_n} \underbrace{\left[\frac{1-F(n)}{n f(n)} \right]}_{B_n} \underbrace{\left[\frac{U_x}{C_n} \int_l^n (W^u U_x - \lambda)(1/U_x) f(p) dp \right]}_{D_n} \quad (8)$$

The formula (8) is rather complicated. It does however provide several insights. There are four elements on the right hand side of (8) that determine optimum marginal tax rates.

(1) First, the term on the right hand side of (8), A_n , expressing labour supply responses in uncompensated and compensated elasticities, represents the standard efficiency effect, reflecting also conventional wisdom. It says that, other things equal, the marginal tax rate is decreasing in E^c and E^u . It is also important to note that, for a given compensated elasticity E^c , the decomposition into uncompensated and income effects matters. The higher are income effects (in absolute terms) relative to uncompensated effect, the higher is the marginal tax rate. What is also important here is that the elasticity may vary across population. This means that we need to know how the elasticity varies with the wage rate.

(2) The second term, B_n , tells that the shape of the wage distribution affects the optimal marginal tax rate at the wage level n through the ratio $\frac{1-F(n)}{n f(n)}$ (known as the inverse hazard ratio). When we increase the marginal tax rate at some n , we collect more revenue on more productive individuals, who are $1-F(n)$ in number. In other words an increase in marginal rate depends on the proportion of the population above n . The purpose of higher marginal rate is to increase the average tax rate higher up the

scale. Hence $1 - F(n)$ is in the numerator. And we distort only the behaviour of the marginal type, which explains why $f(n)$ in turn is in the denominator.

The marginal tax rate is higher when n is lower in the distribution ($\frac{1 - F(n)}{n}$ is decreasing in n) and when $nf(n)$, an indicator of the extent of earnings at the wage level n , is smaller. Hence, raising marginal rates on very low incomes, say z^* , raise substantial revenue because most of the taxpayers has income higher than this level. Moreover, higher marginal rates at the bottom are inframarginal for this large group. The high marginal rates act as a lump sum tax on higher earnings. Secondly, in the denominator f is not very large and n in turn is low, so there is little lost revenue. These considerations explain why several simulations produce high marginal rates at the bottom of the distribution. It has to remember that our story above is based on ceteris paribus assumption. Equation (8) also suggests that, other things equal, the marginal tax rate should be lower the denser the population at that point, i.e. higher $f(n)$. In other words the larger the fraction of the taxpayers paying more tax and the smaller is the group being distorted. On the other hand for the typical distribution the density weighted by n is likely to decline with n above some point suggesting a higher marginal tax rate on high earners. Figure 2 shows this for the lognormal and the Champernowne distribution. (The ranking of the distributions in Figures 1 and 2: The lognormal distribution has the highest peak, then the Champernowne distribution with $\theta=4$ (a shape or Pareto parameter), and the lowest peak with $\theta=2$)

Figure 1: f_c (Champernowne distribution) and f_l (lognormal distribution)

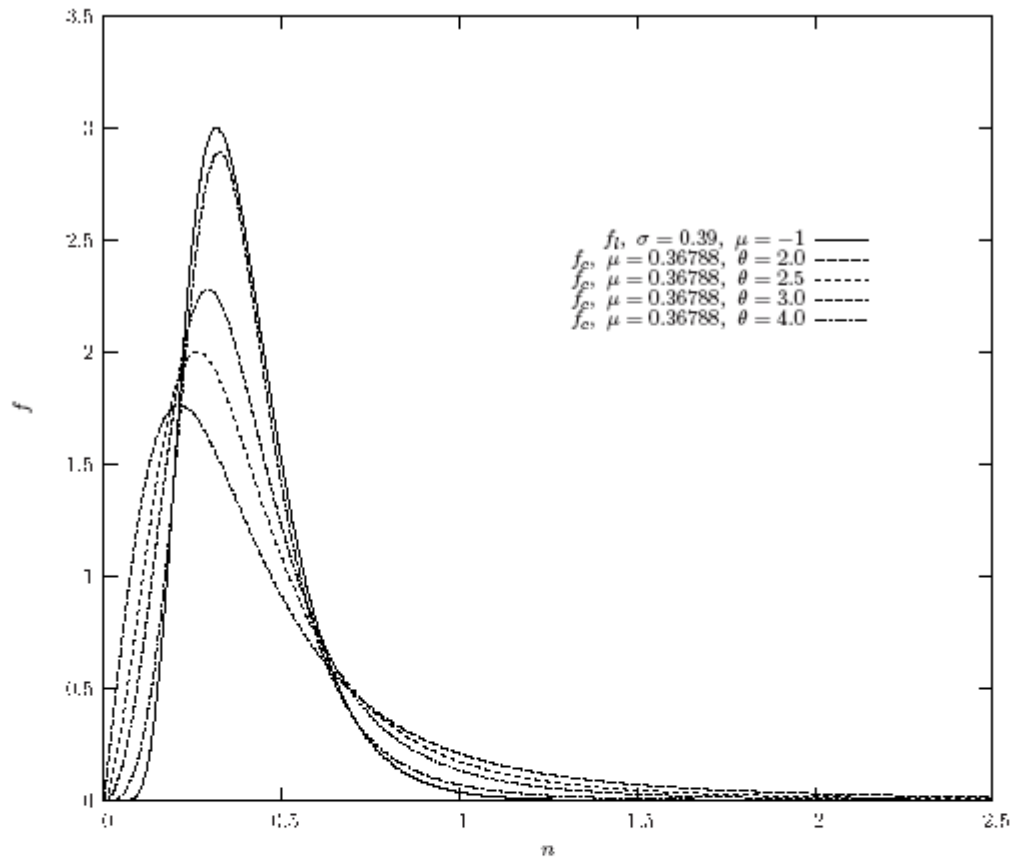
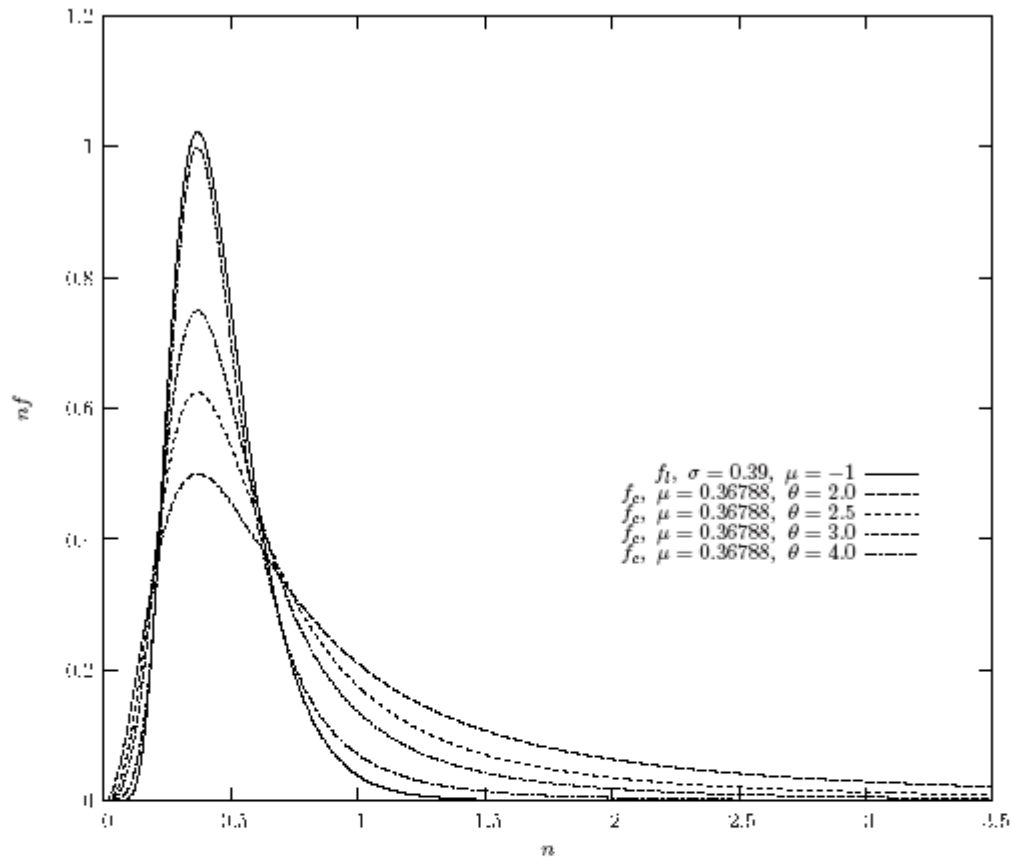
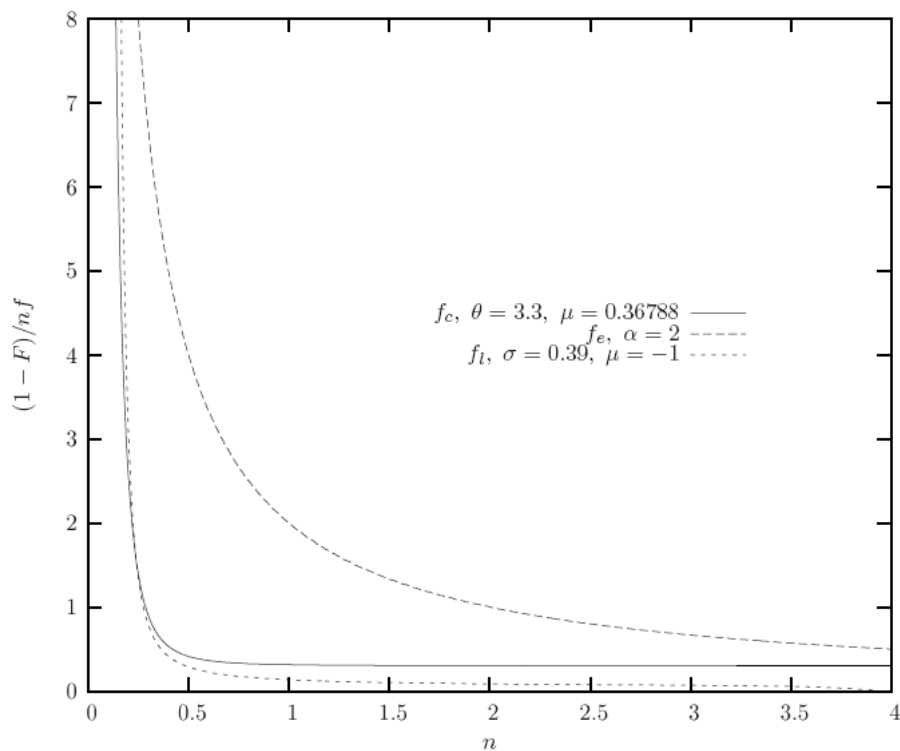


Figure 2 nf_c (Champernowne distribution) and nf_l (lognormal distribution)

In the case of lognormal distribution, then the ratio falls with n , tending in the limit to zero as n tends to infinity. In the case of the Champernowne distribution this ratio in turn tends to constant as n increases. The upper part of the Champernowne distribution follows the Pareto distribution, $\frac{1-F(n)}{nf(n)}$ is constant and is equal to $1/\theta$. Figure 3 shows this effect for different distributions (exponential, lognormal and Champernowne distribution),

Figure 3: f_c (Champernowne distribution), f_e (exponential distribution), f_l (lognormal distribution)



(3) The third term, C_n , reflects income effects and (4) the fourth term, D_n , in turn incorporates distributional concerns. In the fourth term, the integral term $\mu(n)$ measures the social welfare gain from slightly increasing the marginal tax rate at n and distributing as a poll subsidy to those below n the revenue raised from the consequent increase in average tax rates above n . From the transversality conditions we can deduce that $\mu(n)$ increases with n for low n (their social utility

of income, $W'U_x$, exceeds the marginal social cost of public funds, λ ,) and decreases with n for high n . The turning point depends on λ^7 . The lower is λ , the higher is the n at which the turning point occurs. Thus as the revenue requirement falls, and hence λ falls, the range over which $\mu(n)$ is increasing stretches further. It can be shown that $W'U_x$ is decreasing in n^* (n^* is the skill level at which $W'(u(n^*))U_x(x(n^*), y(n^*)) = \lambda$) so long as $W(u)$ is concave and leisure is normal (see proof in appendix A).

Since $\mu(n)$ affects the marginal tax rate positively, this means that the range over which the latter increases also stretches further - at least for this reason. In this sense, therefore, more tax revenue leads to a less progressive tax structure. The intuition, put crudely, is that the lower is the revenue requirement, the more the government can afford to support the poor by a generous poll subsidy, recouping at least part of this by a pattern of rising marginal tax rates on the better off.

As should be obvious from the above discussion, the exact pattern of this term in (8) follows as n rises depends on the social welfare function and the shape of the wage distribution. So the shape of the wage distribution is also important here. Moreover it is obvious in the integral term in (8) that the functional form of U_x has the important role in determining the shape of the schedule.

3. Numerical solution

Following Mirrlees (1971) we define

$$v = \frac{1 + V_y / nU_x}{V_y + yV_{yy}}, \quad (9)$$

and rewrite (4)

$$-n^2 f v + \int_l^n ((W' / \lambda) - (1/U_x)) f(n) dn = 0. \quad (10)$$

⁷ Note λ is itself a function of the overall distribution, since from (6) $\lambda = 1 / (\int_l^h (1/U_x) f(p) dp)$ (the social welfare function is utilitarian, i.e. $W^2=1$)

Differentiating (10) with respect to n and we have

$$v' = -\frac{v}{n} \left(2 + \frac{nf'}{f} \right) + \frac{1}{n^2} \left(\frac{W'}{\lambda} - \frac{1}{U_x} \right) \quad (11)$$

(11) and (3), two differential equations in u and v , provide the solution to the optimal income tax problem, together with the condition

$$n^2 f v = \int_l^n \left((W'/\lambda) - (1/U_x) \right) f(n) dn \quad (12)$$

and

$$\text{at } n=n_o^8; \quad n_o^2 f(n_o) v_o = \int_l^{n_o} \left((W'/\lambda) - (1/U_x) \right) f(n) dn \quad (13)$$

and the transversality condition $\mu(h) = 0$ and (12) require that

$$\lambda n^2 f v \rightarrow 0 (n \rightarrow \infty) . \quad (14)$$

The condition (14) guarantees an accurate value for n_o .

Before going to numerical results we analyze asymptotic marginal tax rates in the following case:

(i) The Champernowne distribution

There are a number of mathematical distribution functions extensively used in describing wage distributions such as lognormal, Pareto and Gamma. Empirical evidence is not conclusive about the quality of each in fitting actual distributions. Specifically, Pareto distributions are found to fit reasonable well at the upper tail of distributions, but the fit over the whole range of income turns out to be quite poor. As for the other functions such as lognormal and Gamma, while providing a good fit over a large part of the income range, they markedly differ at the upper tail. The explanation for this different performance seems to be that these functions are defined so as to reach a maximum in the interior of the interval definition, thereby giving a better fit over the values

⁸ n_o , largest n for which $y(n)=0$, may be in some cases rather large. In the interval $[0, n_o]$, $y=0$ and $x=x_0$ and then $u=U(x_0)-V(0)$ (*). (*) and (13) are needed for starting values in numerical solutions of (11) and (3) conditional on a trial value for n_o .

around the mode. However these functions have the drawback that their elasticity increases unboundly after the mode has been attained, thus contradicting the large evidence of a constant elasticity at the upper tail, which is precisely what characterizes Pareto distribution. To avoid this we adopt the Champernowne distribution. Here we use the two parameter version of the Champernowne distribution. The parameter μ is the median value and θ is a constant corresponding to Pareto's constant for high incomes. The Champernowne distribution approaches asymptotically a form of Pareto distribution for large values of n , but it also has an interior maximum. As the lognormal the Champernowne distribution exhibits the following features: asymmetry, a left humpback and long right-hand tail. It has a thicker upper tail than in the lognormal case.

The probability density function of the Champernowne distribution

$$f(n) = \theta \left(\frac{\mu^\theta n^{\theta-1}}{(\mu^\theta + n^\theta)^2} \right) \quad (15)$$

in which θ is a shape parameter and μ is a scale parameter. The cumulative distribution function is

$$F(n) = 1 - \frac{\mu^\theta}{(\mu^\theta + n^\theta)} \quad (16)$$

For the inverse hazard rate

$$\lim_{n \rightarrow \infty} \frac{1 - F(n)}{nf(n)} = \lim_{n \rightarrow \infty} \frac{\mu^\theta + n^\theta}{\theta n^\theta} \rightarrow \frac{1}{\theta}. \quad (17)$$

and in the elasticity form

$$\frac{nf'}{f} = \frac{\theta(\mu^\theta - n^\theta)}{(\mu^\theta + n^\theta)} - 1 \quad (18)$$

or

$$\frac{nf'}{f} = \frac{\theta \left(\frac{\mu^\theta}{n^\theta} - 1 \right)}{\left(\frac{\mu^\theta}{n^\theta} + 1 \right)} - 1 \quad (19)$$

and

$$\left(2 + \frac{nf'}{f} \right) = 1 - \frac{\theta \left(1 - \left(\frac{\mu}{n} \right)^\theta \right)}{\left(1 + \left(\frac{\mu}{n} \right)^\theta \right)} \quad (20)$$

(ii) The utility function is $u = \log x + \log(1 - y)$.

To find asymptotic marginal tax rates we write $s=1-y$ and put $t=\log n$. Now we can write

$v = s(s - \frac{x}{n})$. We use a constant absolute utility-inequality aversion form for the social welfare

function of the government is specified

$$(iii) \quad S(u) = -\frac{1}{\beta} e^{-\beta u} \quad (21)$$

where β measures the degree of inequality aversion in the social welfare function of the government (in the case of $\beta = 0$, we define $W = u$).

Now if $f(n)$ is the Champernowne distribution, then (11) becomes

$$\frac{dv}{dt} = -v \left(1 - \frac{\theta \left(1 - \left(\frac{\mu}{n} \right)^\theta \right)}{\left(1 + \left(\frac{\mu}{n} \right)^\theta \right)} \right) - s \left(1 - \frac{v}{s^2} \right) + \lambda e^{-t} \quad (22)$$

Since

$$\lim_{n \rightarrow \infty} \frac{nf'}{f} = -1 - \theta. \text{ i.e as in Pareto distribution}$$

and

$$\lim_{n \rightarrow 0} \frac{nf'}{f} = -1 + \theta$$

we can rewrite

$$\frac{dv}{dt} = -v(1 - \theta) - s \left(1 - \frac{v}{s^2} \right) + \lambda e^{-t} \quad (23)$$

Hence from (3)

$$\frac{du}{dt} = \frac{du}{dn} \frac{dn}{dt} = \frac{1-s}{s} \quad (24)$$

Using (24) we can write

$$\frac{ds}{dt} = \frac{v(1 + \theta) - 2s^2 + \lambda e^{-t}}{2s} \quad (25)$$

Denote $\frac{v}{s^2} = \tau$ i.e the marginal tax rate

It follows that

$$\frac{ds}{dt} = \frac{s}{2}(\tau(1+\theta) - 2) + \frac{\lambda e^{-t}}{2s} \quad (26)$$

and

$$\frac{d\tau}{dt} = \tau \left[\left(\theta - 1 - \frac{1}{s} \right) - \frac{1}{s} + \frac{\lambda e^{-t}}{s^2} \right] \quad (27)$$

For $\frac{d\tau}{dt} = 0$, it has to be $\tau = \frac{1}{(1+\theta)s}$ and for $\frac{ds}{dt} = 0$ in turn $\tau = \frac{2}{1+\theta}$.

As $n \rightarrow \infty$, $\tau \rightarrow \frac{2}{1+\theta}$.

If the Pareto exponent were 3, we would have an optimal tax rate of 50%. This is just the same result given by Mirrlees (1971,p.200).

We can give a more fully description of the solution in the case of the Pareto distribution. When the path to the singular solution starts from $s=1$, this implies that the marginal tax rates increase monotonically from $\tau = \frac{1}{(1+\theta)}$ to $\tau = \frac{2}{1+\theta}$. It is also important to note that these results are independent of the form of the government's maximand and of the net revenue requirement.

5. Results of numerical simulations

On the basis of the first order conditions it is possible to say relatively little about the general shape of the tax schedule. Therefore, following the lead of Mirrlees (1971), numerical calculations have proved useful in generating useful results⁹. We follow this route here. We assume $f(n)$ is the Champernowne distribution. In our basic case a scale parameter μ is as in Mirrlees (1971) with lognormal distribution; $e^{-1}=0.36788$ and $\theta = 3.3$.¹⁰ But the distribution of earnings is not the distribution of n -it is the distribution of z . Ideally we would like to use in numerical simulations empirical earnings distributions. This cannot apply directly because the distribution of z is affected by income taxation. This means that when we change utility function or its parameters, we also change the distribution of n so that resulting distribution of z (absent the tax) remains the same. Otherwise, we get an effect through the changes in utility functions, but also through a change in the distribution of z . To check this in our calculations we compare the inverse hazard ratios for the distribution of n (see figure 3) and the distribution of z . The inverse hazard ratio for the distribution of z is $\frac{(1-H(z))}{zh(z)}$ where $h(z)$ is the density of the distribution of z and $H(z)$ is the cumulative distribution of z . Figures B1a and b in appendix B present the graphs of the ratio for different specifications. It turns out that the ratios in different cases do not differ much from each others.

The utility function has the constant elasticity of substitution form

$$u(x, l) = [x^{-a} + (1 - y)^{-a}]^{-1/a} \quad (28)$$

where the elasticity of substitution between consumption and leisure $\varepsilon=1/(1+a)$.

Our calculations were carried out for the following special cases of (28);

$\varepsilon = 1$ (log-log or Cobb-Douglas case)

$$u = \log x + \log(1 - y) \quad \text{(case u1)} \quad (29)$$

⁹ Tuomala (1984, 1990) gives details of the computational procedure.

¹⁰ Bevan (2005) uses $\theta=3.37$ based on British income data (Royal Commission on the distribution of income and wealth, 1971)

and

$$\varepsilon = 0.5 \text{ (a=1)}$$

$$u = -\frac{1}{x} - \frac{1}{(1-y)}. \quad \text{(case u2)} \quad (30)$$

We use a constant absolute utility-inequality aversion form for the social welfare function of the government (21). We present simulations for $\beta = 0$ and $\alpha = 1$. The curvature in the utility of consumption modifies social marginal weights $W'U_x$ and makes the government preferences (implicit) more redistributive. Hence the overall curvature for (21) is $1 + \beta$. Overcoming possible philosophical problems we may take a view that β is an observable variable, not a social judgment (see Kaplow (2004)). $R=1-X/Z=1 - \int x(n)f(n)dn / \int z(n)f(n)dn$ is specified as a fraction of national income varying between $X/Z=1.1$ and $X/Z=0.8$. If $X/Z=1$, then taxation is purely redistributive. If $X/Z > 1$, then we have outside resources available (e.g foreign aid, the state owned firms make positive profits)

The results of the simulations are summarized below in Tables 1 and 2 and Figures 4a-7b. The Tables 1 and 2 give labour supply, y , gross income, z , net income, x (also x at the point n_0 , denoted by x_0 and F at the point n_0 , denoted by F_0)¹¹ and optimal average (ATR) and marginal tax rates (MTR) at various percentiles of the ability distribution, when $X/Z=0.9$ or $R=0.1$. Tables also provide the decile ratio (P90/P10) for net income and gross income. Unlike the scalar inequality measures the use of fractile measures such as the decile ratio allows us to consider changes in inequality at various different points in the distribution. Since marginal tax rates may be a poor indication of the redistribution powers of an optimal tax structure we measure the extent of redistribution, denoted by RD, as the proportional reduction between the decile ratio for market income, z , and the decile ratio for disposable income, x . To relate these results to empirical labour supply studies we give the values of the uncompensated elasticity, E'' and income effect parameter I.

¹¹ $y(n_0)=0$. The range of n was n_0 to 1.5, at which point the integrated value of $f(n)$ was more than 0.9999.

Table 1 (case u1)

$\beta=0$	$\theta=3.3$	$X/Z=0.9$					
F(n)	y	z	x	ATR%	MTR%	E ^u	I
0.10	0.32	0.06	0.08	-35	38	0.58	-0.5
0.50	0.43	0.16	0.15	-7	30	0.16	-0.5
0.90	0.46	0.33	0.27	19	32	0.09	-0.5
0.97	0.47	0.49	0.37	23	33	0.07	-0.5
0.99	0.47	0.69	0.50	27	33	0.05	-0.5
P90/P10		5.44	3.26				
RD(%)			40				
$n_0=0.08$	$x_0=0.046$	$F_0=0.01$					

Table 2 (case u2)

$\beta=0$	$\theta=3.3$	$X/Z=0.9$					
F(n)	y	z	x	ATR%	MTR%	E ^u	I
0.10	0.50	0.09	0.14	-50	58	0.18	-0.22
0.50	0.55	0.20	0.19	6	52	-0.003	-0.30
0.90	0.53	0.38	0.28	27	52	-0.09	-0.37
0.97	0.51	0.54	0.35	33	53	-0.13	-0.41
0.99	0.49	0.72	0.44	39	51	-0.18	-0.46
P90/P10		4.22	2.0				
RD(%)			52.6				
$n_0=0.02$	$x_0=0.1$	$F_0=0.0$					

Table 1 and Figure 4a and b show that replacing the lognormal distribution by the Champernowne distribution the shape of the income tax schedule differs remarkably from those computed in Mirrlees (1971). In the case u1 we have an U-shaped marginal tax rate structure and the marginal rates are rather high at the bottom of the distribution.¹² The marginal tax rate curves are decreasing from the bottom to middle incomes (around median). The marginal tax rate curves slope upward

¹² Bevan (2005) also computes an U-shaped pattern with the log-log utility function.

starting at around the range 60-70 percentile of wage distribution and then rise until income level at the 99th percentile. The U-shaped pattern is very close to the marginal rate structure commonly observed in many countries. This pattern has sometimes been referred to as a tax structure which sets the floor for poverty and the ceiling to riches on the egalitarian grounds. High marginal tax rates at lower income levels are due to the interaction between the various schemes for income support and income taxation at the lower end of the income scale. The U shape of tax schedules in Saez (2001) was a direct consequence of the U-shaped pattern of inverse hazard ratio. In our calculations the inverse hazard ratio is the L-shaped in all specifications considered, (see Figure B1a and B1b in appendix B). Hence, the shape tax schedule is determined by the interplay between all factors discussed in the context of formula (8). In the case u2 the marginal tax rates are much higher than in the case u1. (see Table 2 and Figures 5a-6b). In this case we can also find a slight U-shaped marginal tax rate curve. The marginal tax rates increase however for the lowest percentile and then fall and above middle incomes are roughly constant and finally rise again although not to the height at the rates applied to lower income individuals. In this case the marginal tax rate curves slope upward starting at around the range 70 – 80 percentile of the wage distribution, and then rise until income level at 97th percentile point. The point in which the marginal tax rate curve slopes upwards is much higher in Saez (2001), around 80000 dollars/year than in our cases. In the case u1 the elasticities are decreasing in incomes and in the case u2 the compensated elasticities decrease until the median income and then they remain constant.¹³ Our results show that it is possible to get an U-shaped pattern without assuming constant wage elasticity as in Diamond (1998) and Saez (2001).

Unfortunately there is little evidence regarding the relationship between labour supply elasticities and wage rates. Analyzing data from an earlier labour supply study, Sadka, Garfinkel and Moreland (1982) computed the compensated wage elasticity for each quintile of income distribution and find that it decreases as income increases. Røed and Ström (2002) (table 1 and 2) offer a review of the existing

¹³ Saez(2001,p.225) criticizes CES-specification because the income effects are unrealistically large, around -1. Emmanuel Saez has kindly clarified to me that his statement is not always correct and should have been qualified to that income effect is around -1 asymptotically in the case $\epsilon < 1$. Using (28) and denoting $I = n \frac{dy}{dR}$, where

$x = ny + R$ (R=the virtual income). We have $I = -\frac{1}{(1+n^{(\epsilon-1)})}$. Now we see that in the case $\epsilon=1$ (case u1) $I=-0.5$. When $\epsilon < 1$, then I converges to -1 as n goes to infinity. As seen in Table 2 $I=-0.46$ at the 99 percentile point of the n-distribution. It is very far from -1.

more recent evidence. They conclude that the limited evidence indicates that labour supply elasticities are declining with household income. High labour supply elasticities among low-wage workers is also confirmed by empirical evaluations of various in-work benefit schemes operating in the US, UK and some other countries, see Blundell (2000) for a recent review. By contrast, there is empirical evidence on the elasticity of taxable income that higher elasticities are among high income individuals. See eg. Gruber-Saez (2002).

As shown in Figures 4a-7b and Figures B2a and b in appendix B for different utility functions and X/Z-ratios (revenue requirement), especially in the u1 case, the marginal rate curves are clearly U-shaped and even more with lower revenue requirement (higher X/Z). In other words with lower revenue requirement (higher X/Z) the tax schedule is more progressive in the sense that the range of rising marginal rates increases. This is just what our discussion suggested in the context of the marginal tax formula (8). Some intuition for this might be developed as follows. First, it may be useful to describe the income tax schedule so that it consists of two elements; the guaranteed income; an individual with no income would get the lump sum subsidy or the guaranteed income $x(n_o) = x_o = 0 - T(z = 0) = -T(0)$. ($T(0)$ is negative if the government has redistributive goals) and the pattern of marginal tax rates $\tau(z)$. The latter element describes both how the guaranteed element is clawed back or taxed away and how tax burden increase with income¹⁴. Consider the case where revenue requirement is positive but there is high concern for the poor. In this case it is likely that there will be a high guaranteed income (see Table 3) but also high marginal tax rates on the low income people to claw back revenue. As the revenue requirement falls, and in fact as it becomes negative so that outside resources are available, the minimum income requirement for the poor can be met easily without having to claw back revenue with high marginal tax rates. Thus we should expect to see low marginal tax rates on the poor.

¹⁴ Mathematically; total tax burden at income z is $T(z) = T(0) + \int_0^z \tau(m)dm$.

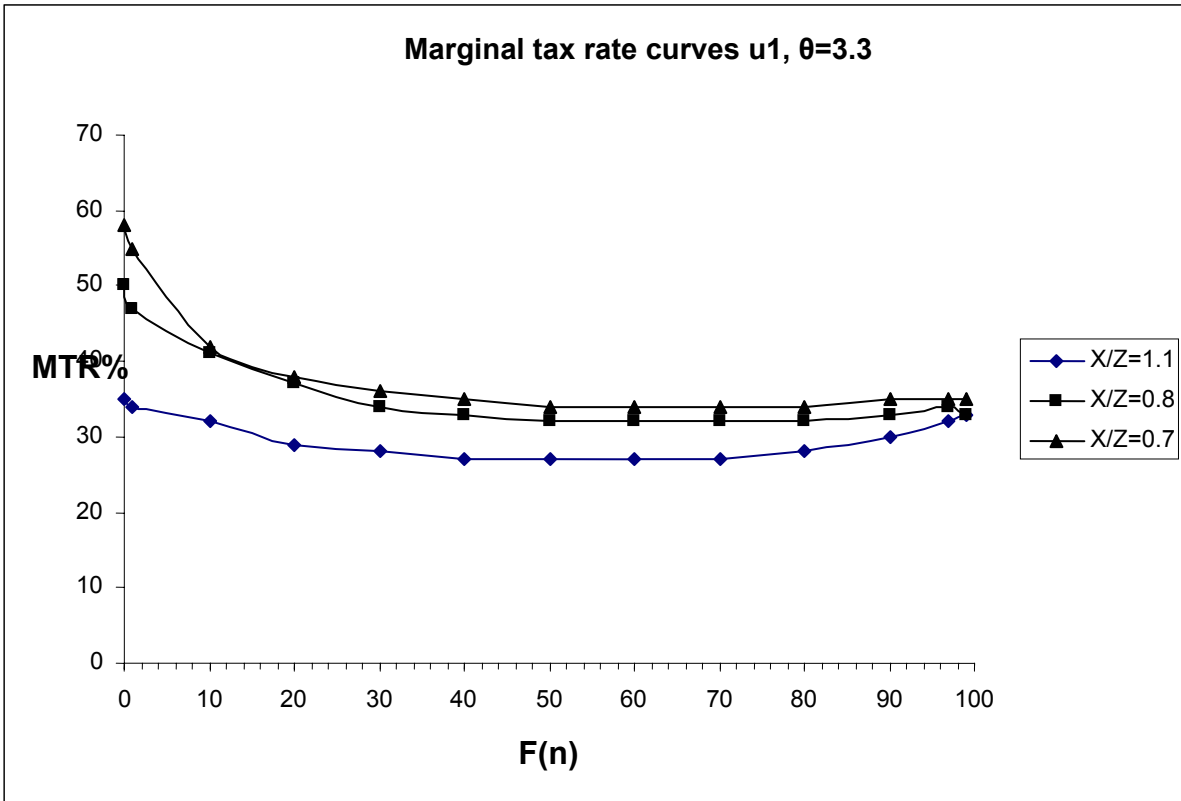


Figure 4a

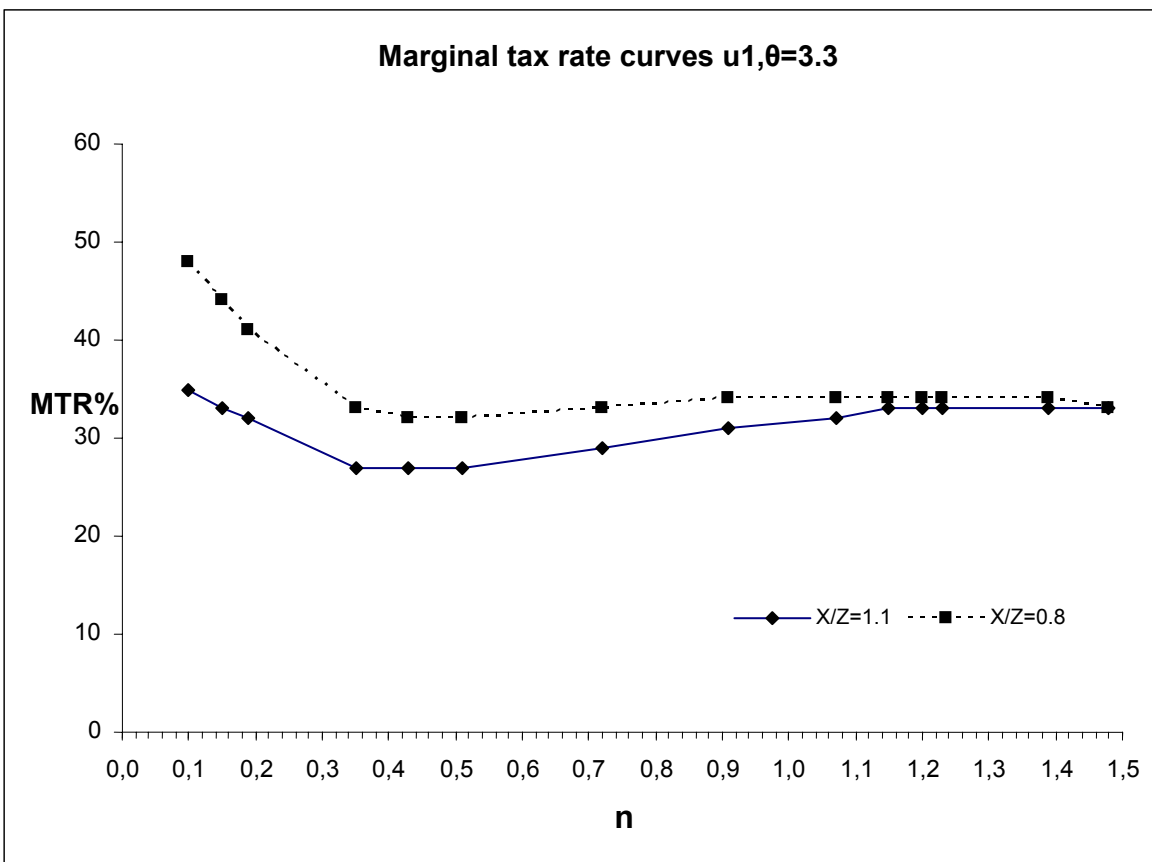


Figure 4b

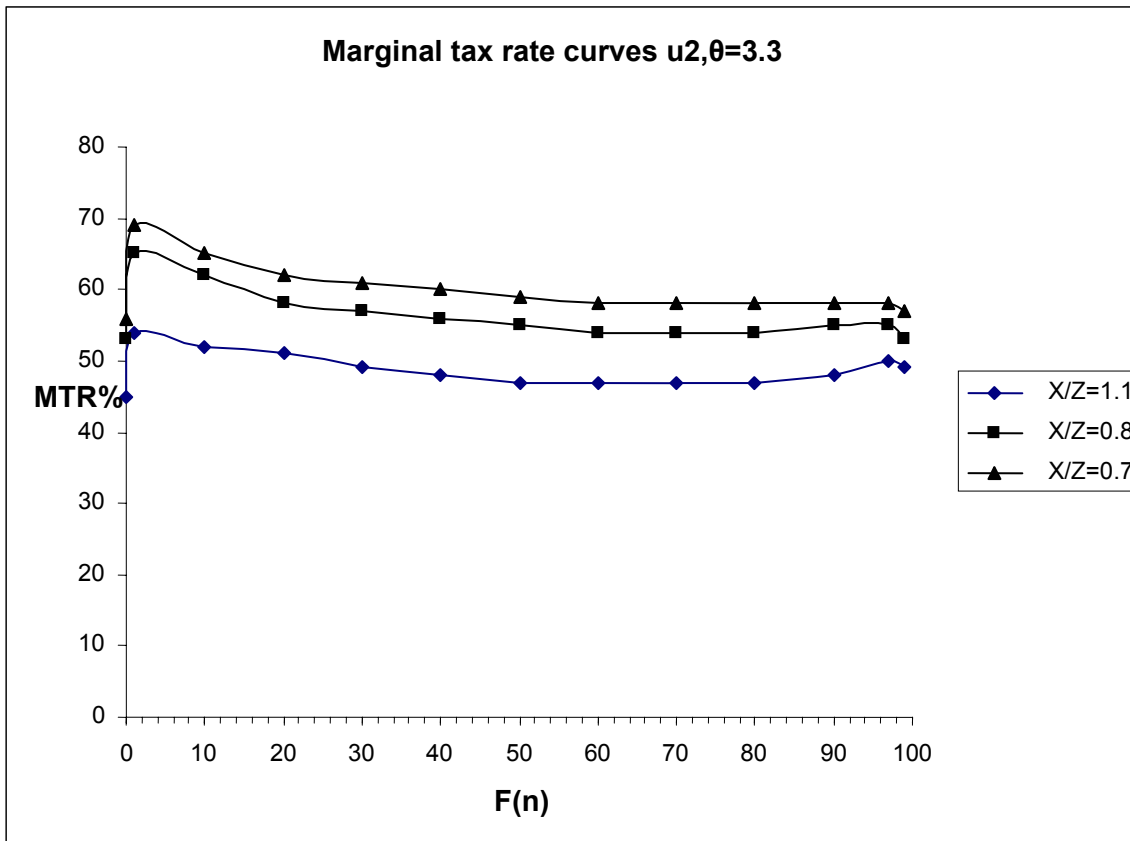


Figure 5a

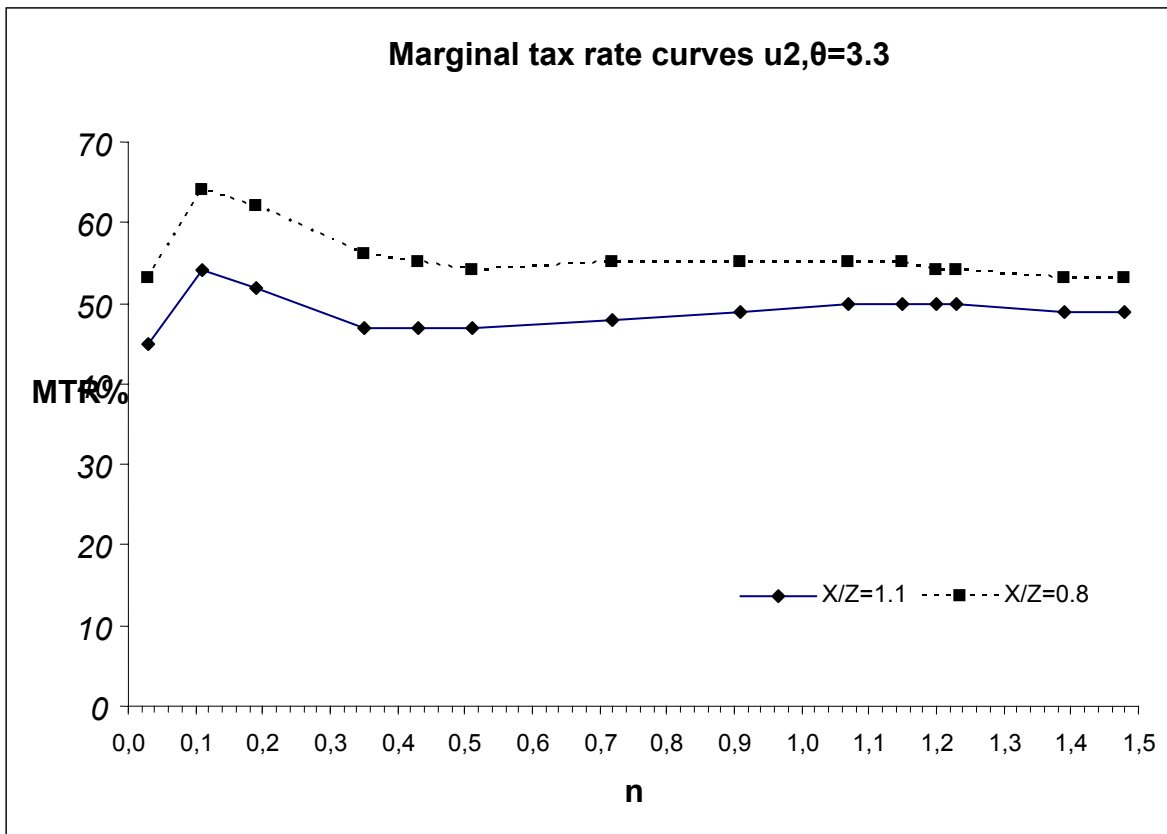


Figure 5b

To see how sensitive the shape of the tax schedule is to the choice of the parameter θ in the Champernowne distribution we computed solutions for $\theta=2.0$ and 2.5 in the case u2. Figures 6a and 6b depict these cases. When $\theta=2.0$ the marginal tax rate is increasing with income. When $\theta=2.5$, the marginal tax rates increase for the lowest decile and then the marginal tax rates remain constant. In the case u1 (see Figure 6c) with θ lower than 2.2 the marginal tax rate increase with income. These results reinforce the findings of Kanbur-Tuomala (1994) with lognormal distribution that when higher values of inherent inequality (here smaller θ) are used optimal marginal tax rates increase with the income over the majority of the population. Unlike in Kanbur-Tuomala (1994) with lognormal distribution we have here - practically speaking - a progressive marginal rate structure throughout.

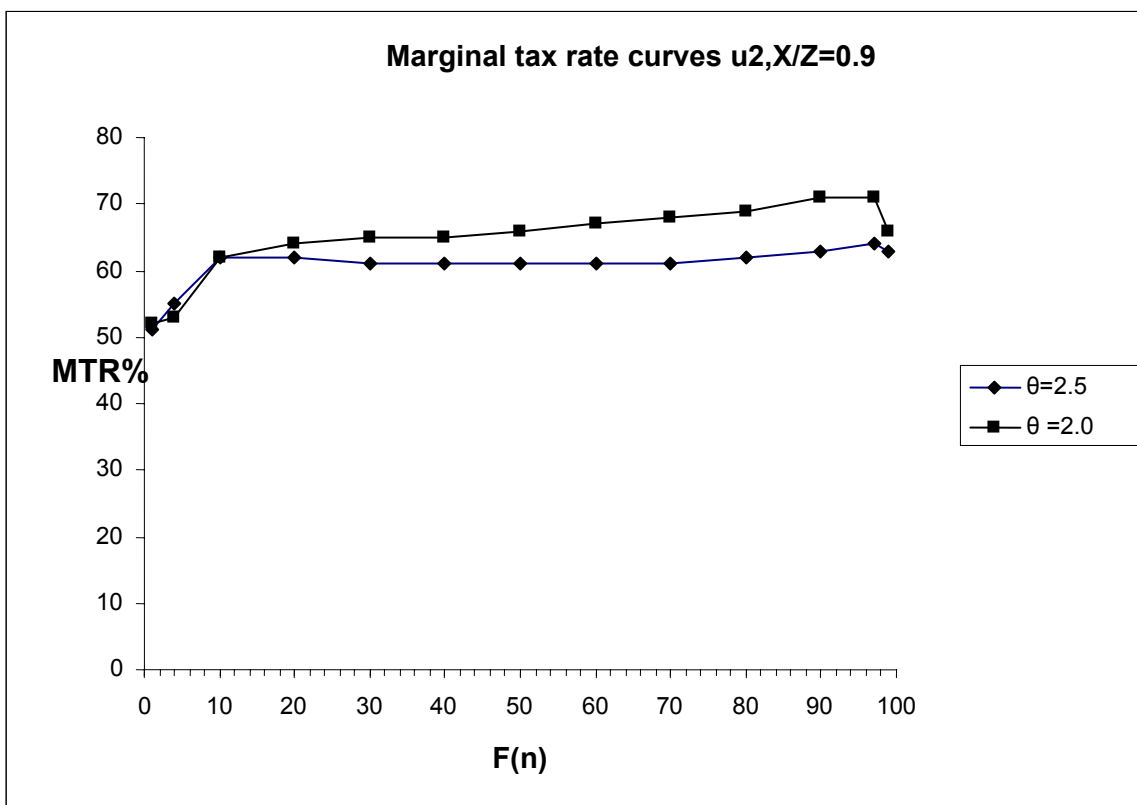


Figure 6a

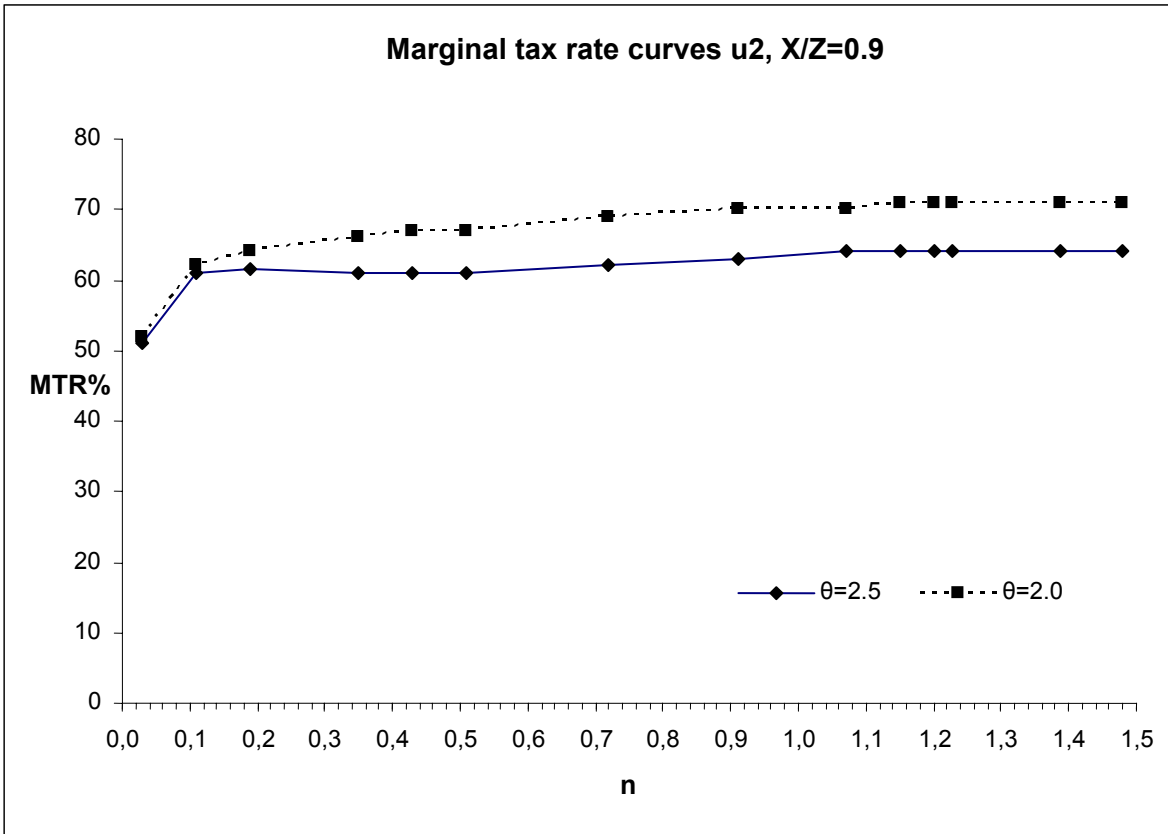


Figure 6b

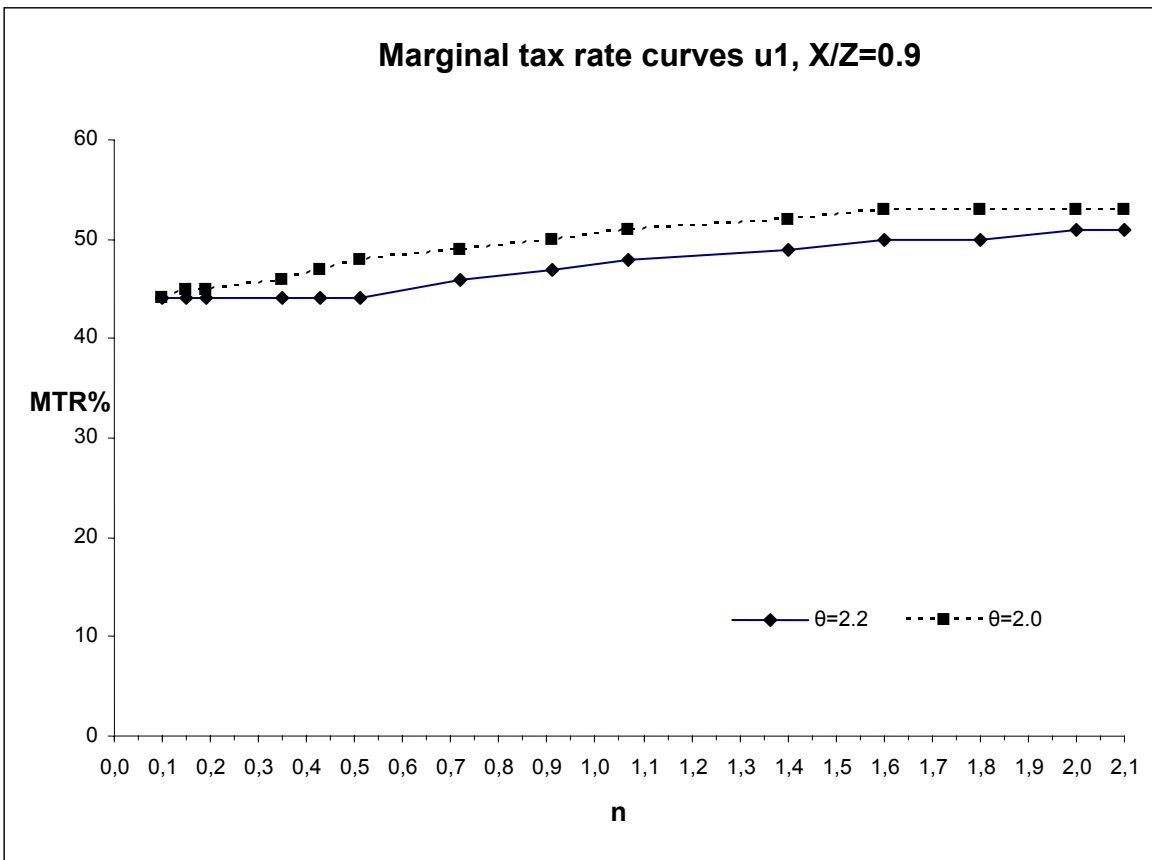


Figure 6c

What happens if higher inequality aversion is to be applied. Our numerical results (Figures 7a,b) seem to suggest that marginal tax rates tend to increase for all taxpayers with increasing inequality aversion. The shape of the tax schedule seems to change in a quite similar way as with increasing revenue requirement. Hence a high degree of inequality is akin to a high revenue requirement.

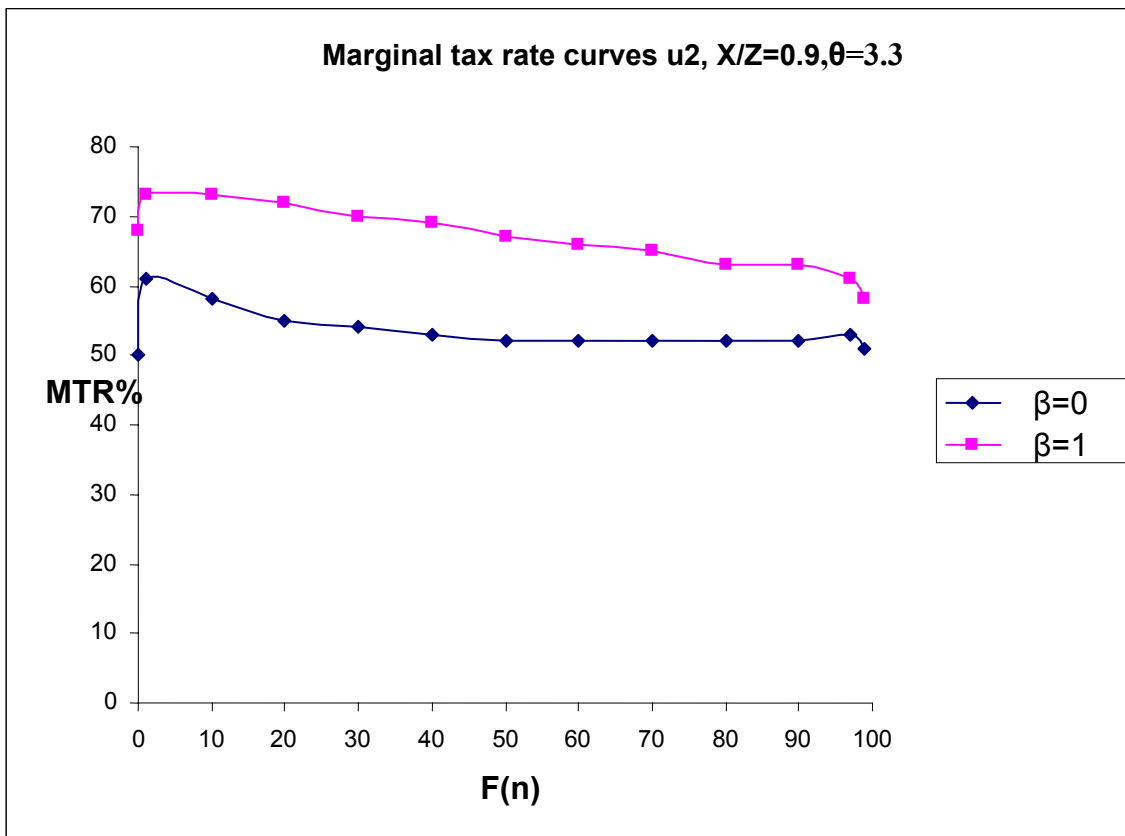


Figure 7a

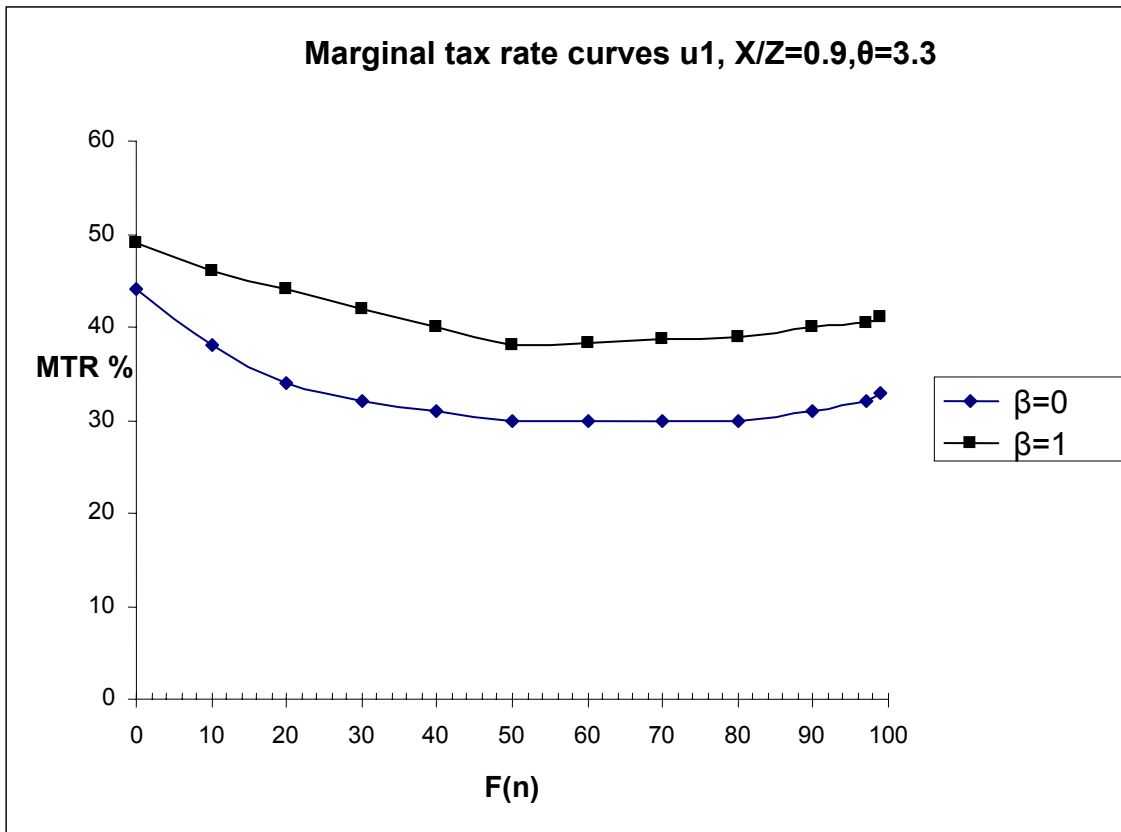


Figure 7b

It may also be some interest to compare the asymptotic rates to the optimal marginal tax rates. In the case u1 the asymptotic marginal rate must converge to $\frac{2}{1+\theta}$, which is 46.5 % in our specification.

At the income level of $F(n)=0.99$ optimal rates are far from their asymptotic rates (see Table 1). This also differs from the results in Saez (2001). The optimal rates he calculated are close to their asymptotic rates at the income level of \$200000 (up to this income level marginal tax rates increase)

The optimum is typically characterized by a certain fraction of individuals, at the bottom end, choosing not to work. Given that high marginal tax rates are optimal near the bottom, the finding that the percentage of those choosing not work at all is rather high might be unsurprising. It turns out to be so that things are not that straightforward. Namely in the case u2 (see Table 2) we have the highest marginal tax rates near the bottom, but the percentage of those choosing not to work is smaller than in other cases calculated.

We also explore how sensitive is the level of the guaranteed income or lump sum transfer component of the tax system to the specification of the model. And what is the relationship between the lump sum transfer and the progressivity of tax schedule. As we see from Table 2 and Figure 6 in the case u2 the marginal tax rates are much higher than in other cases considered and the guaranteed income x_0 is also much higher and rather substantial compared with other cases. The table 3 displays the ratio of the guaranteed income to the average gross income with different utility functions and revenue requirement, X/Z . This ratio is clearly higher in the u2 case than other cases. Also in this case the extent of redistribution in terms of our measure RD is much larger (50.5%) than in other cases.

Table 3. The ratio of the guaranteed income to the average gross income

X/Z	u1	u2
1.1	0.40	0.60
1.0	0.32	0.53
0.9	0.24	0.46
0.8	0.19	0.39
0.7	0.13	0.35

5. Conclusions

This paper shows that the assumed distribution of wages plays an important role in determining the optimal shape of non-linear income tax schedule. We replace the assumption of lognormal distribution used in several papers by the Champnowne distribution. Using numerical simulations we show that the U-shaped pattern of the marginal income tax rates can be obtained without assuming constant labour supply elasticity as in Diamond (1998) and Saez (2001). This paper also shows numerically that either a sufficiently high inherent inequality or a combination of sufficiently high inherent inequality and sufficiently low revenue requirement that leads to a pattern of optimally increasing marginal tax rates. Furthermore, it is also showed the central role of the revenue requirement in determining the shape of the schedule, the extent of redistribution and the level of the guaranteed minimum income.

References

- Aitchison, J. & Brown, J.A.C., 1957, *The lognormal distribution with special reference to its uses in economics*, Cambridge University Press,.
- Atkinson, A.B. and Stiglitz, J, 1980, *Lectures on Public Economics*, McGraw Hill,.
- Atkinson, A.B., 1990, Public economics and the economic public, *European Economic Review*, 34, 225-248, ..
- Atkinson, A.B., 1995, *Public economics in Action, The basic income/flat tax proposal*, The Lindahl Lectures, Oxford University Press,.
- Bevan, D., 2005, *On the shape of the optimal tax-transfer schedules under non-welfarist objectives*, mimeo, Oxford University.
- Blundell, R., , 2000, Work incentives and ‘in-work’ benefit reform: a review. *Oxford Review of Economic Policy*, 16, 1, 27-44.
- Boadway, R., Cuff, K and M. Marchand, 2000, Optimal income taxation with quasi-linear preferences revisited, *Journal of Public Economic Theory*,.
- Diamond, P., 1998. Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates, *American Economic Review*, 88(1), pp.83-95.
- Champernowne, D.G., 1952, The graduation of income distributions, *Econometrica* 20, 591-615.
- Dahan. M. and Strawczynski, M., 2001 The optimal non-linear income tax, *American Economic Review*. 90(3), pp.681-686,.
- Fisk, P.R., 1961a, The graduation of income distribution, *Econometrica*, 29, 171-185.
- Fisk, P.R., 1961b, Estimation of location and scale parameters in a truncated grouped sech-square distribution, *Journal of the American Statistical Association*, 56, 692-702.
- Ritva Immonen-Ravi Kanbur-Mick Keen-Matti Tuomala, 1998, Tagging and taxing: the optimal use of categorical and income information, *Economica*, vol 65, 179-92
- Kaplow, L, 2004, *Concavity of Utility, Concavity of Welfare and Redistribution of Income*”, Discussion paper, Harvard Law School.
- Kanbur, R. and Tuomala, M, 1994, Inherent inequality and the optimal graduation of marginal tax rates, *Scandinavian Journal of Economics* 96 (2), pp. 275-282.
- Mirrlees, J.A., 1971, An exploration in the theory of optimum income taxation, *Review of Economic Studies* 38, pp.175-208.
- Mirrlees, J.A., 1976, Optimal tax theory, a synthesis, *Journal of Public Economics* 6, 327-58.
- Revesz, J., 1989. The optimal taxation of labour income, *Public Finance* 44, pp.453-75.

- Roberts, K., 2000, A reconsideration of optimal income tax, in *Incentives, Organization and public Economics*, papers in Honour of Sir James Mirrlees, eds by P.Hammond and G.Myles, Oxford University Press.
- Röed, K and S.Ström , 2002, Progressive taxes and the labour market: Is the trade-off between equality and efficiency inevitable? *Journal of Economic Surveys* vol 16 no1.
- Sadka, E, Garfinkel.I and K. Moreland, 1982, Income testing and Social Welfare: An Optimal Tax – Transfer Model in *Income –Tasted Transfers programs: The case for and against* in Garfinkel (ed.)
- Saez, E., 2001, Using elasticities to derive optimal income tax rates, *Review of Economic Studies* 68, pp.205-229.
- Stern, N., 1986, On the specification of labour supply functions, in *Unemployment, search and labour supply*, edited by R. Blundell and I.Walker, Cambridge University Press.
- Tuomala,M., 1984, On the optimal income taxation: some further numerical results, *Journal of Public Economics* 23, 351-66.
- Tuomala, M, 1990, *Optimal income tax and redistribution*, Clarendon Press, Oxford.
- Tuomala, M, 2003, *On Optimal non-linear income taxation: numerical results revisited*, Tampere Economic Working Papers, Net series no 17, University of Tampere.

Appendix A

$W'(u(n^*))u_x(x, y)$ is decreasing in n^*

Proof: We prove this in case of a general utility function $u=u(x, y)$.

Differentiating $W'(u(n^*))u_x(x, y)$ we have

$$W''u_x u' + W'(u_{xx}x' + u_{xy}y') \quad (A1)$$

So long as W is concave the first term of (A1) is negative. Hence it suffices to show that $\psi = (u_{xx}x' + u_{xy}y') < 0$

Since $x(n) = ny(n) - T(ny(n))$, we have

$$x' = (1 - \tau)(ny' + y)$$

so that $\psi = (u_{xx}(1 - \tau)y + (u_{xx}(1 - \tau)n + u_{xy})y')$

Using the first order conditions of individual's utility maximization we have

$$\psi = (u_{xx}(1 - \tau)y - \left(\frac{u_y}{u_x}\right)(u_{xx} - u_{xy}\left(\frac{u_x}{u_y}\right))y')$$

Define leisure $l = 1 - y$ and $\zeta(x, l) = u(x, 1 - l)$. So $\zeta_l = -u_y$ and $\zeta_{lx} = -u_{yx}$.

Now

$$\psi = (u_{xx}(1 - \tau)y - \left(\frac{\zeta_l}{\zeta_x}\right)(\zeta_{xx} - \zeta_{xl}\left(\frac{\zeta_x}{\zeta_l}\right))y') < 0$$

if leisure is normal.

Above we assume that $y > 0$. For those n for which $y(n) = 0$, $W'u_x$ is constant.

Appendix B

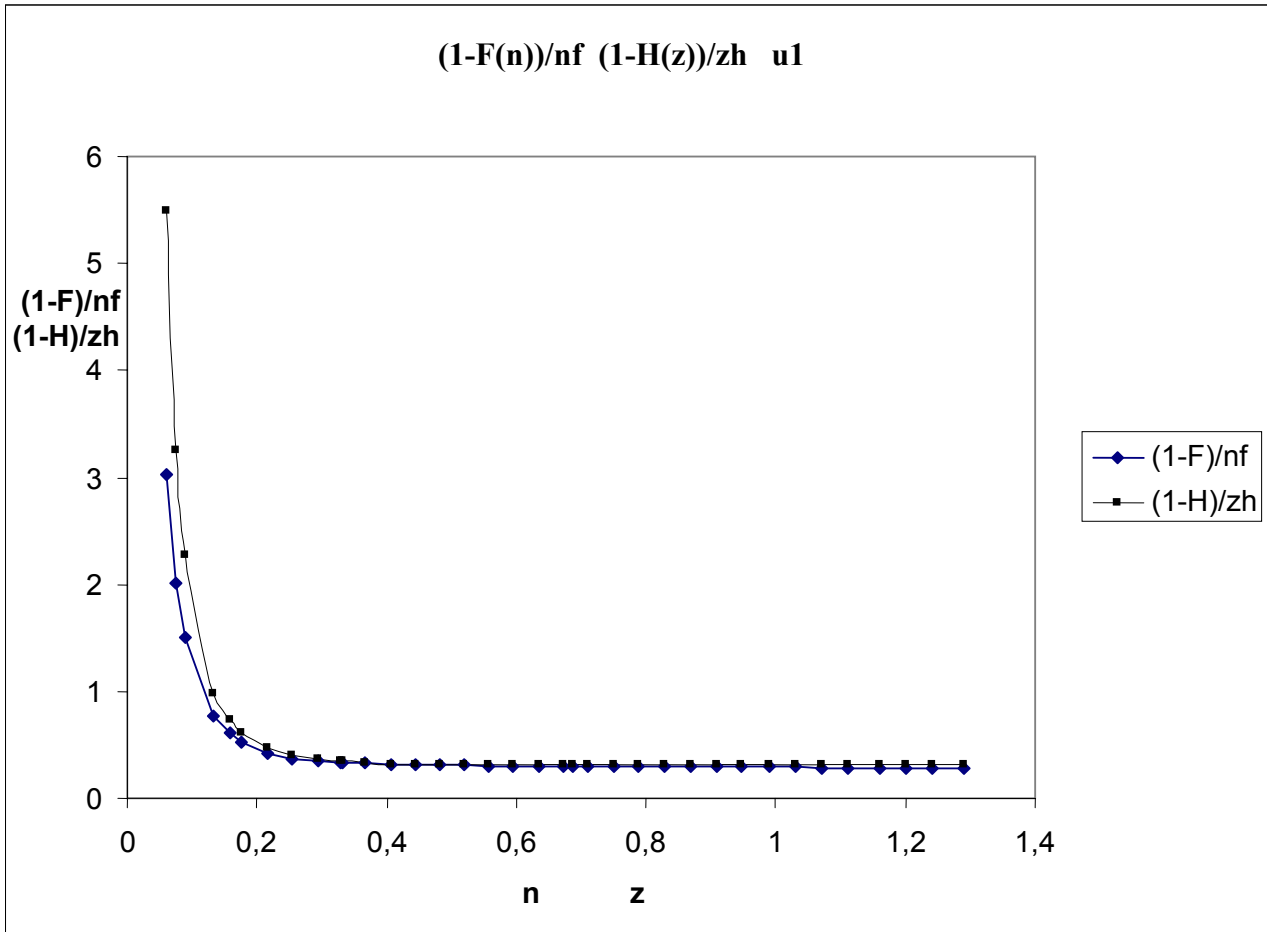
Fig B1a $(1-H(z))/zh(z)$ 

Figure B1b

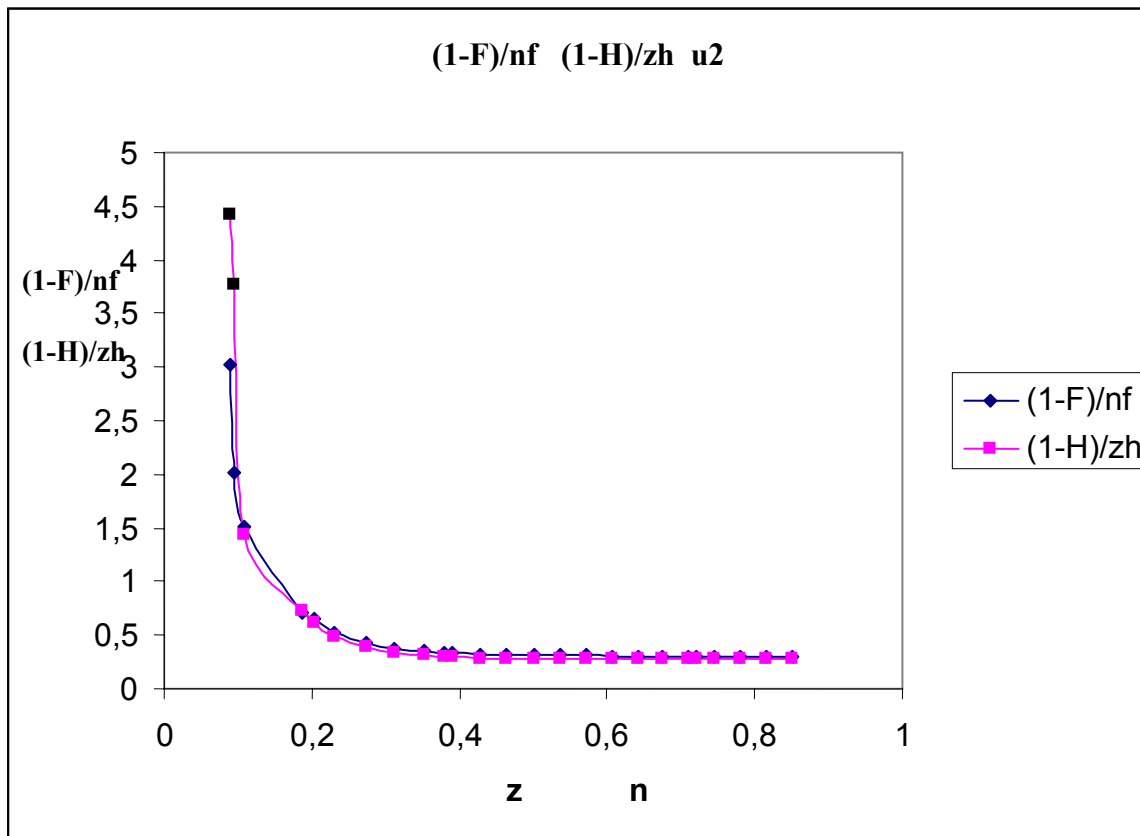


Figure B2a

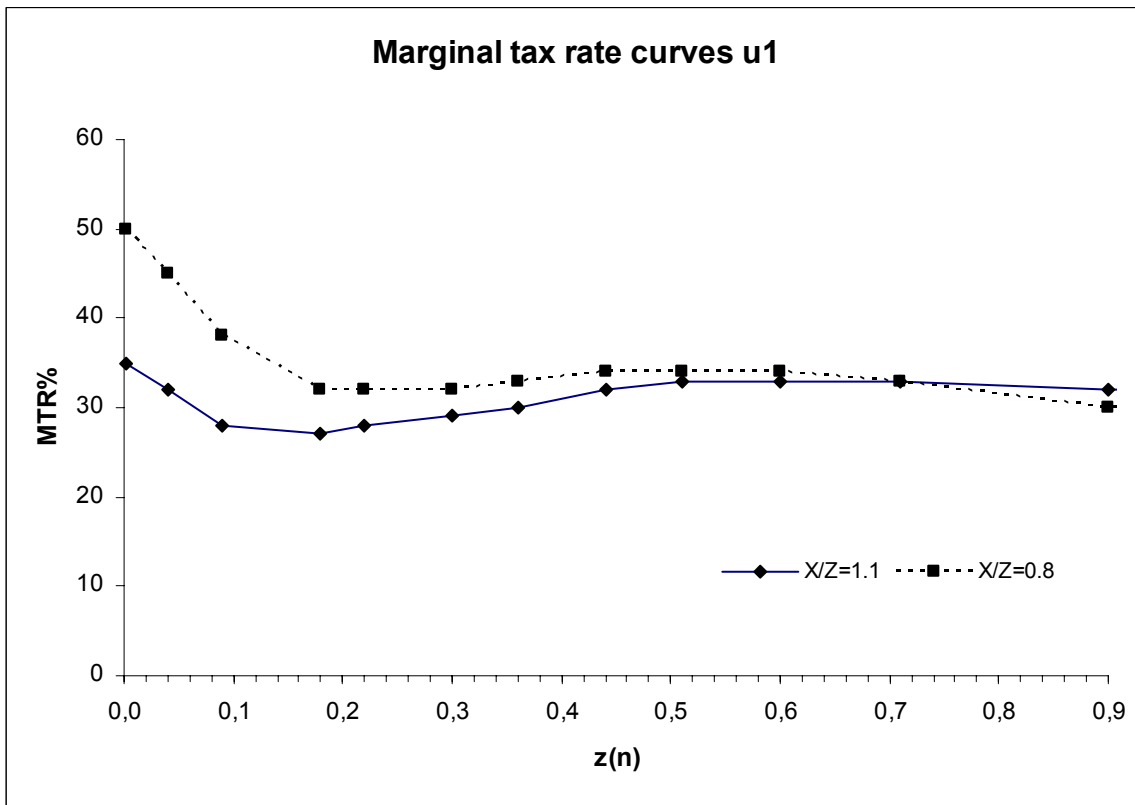


Figure B2b

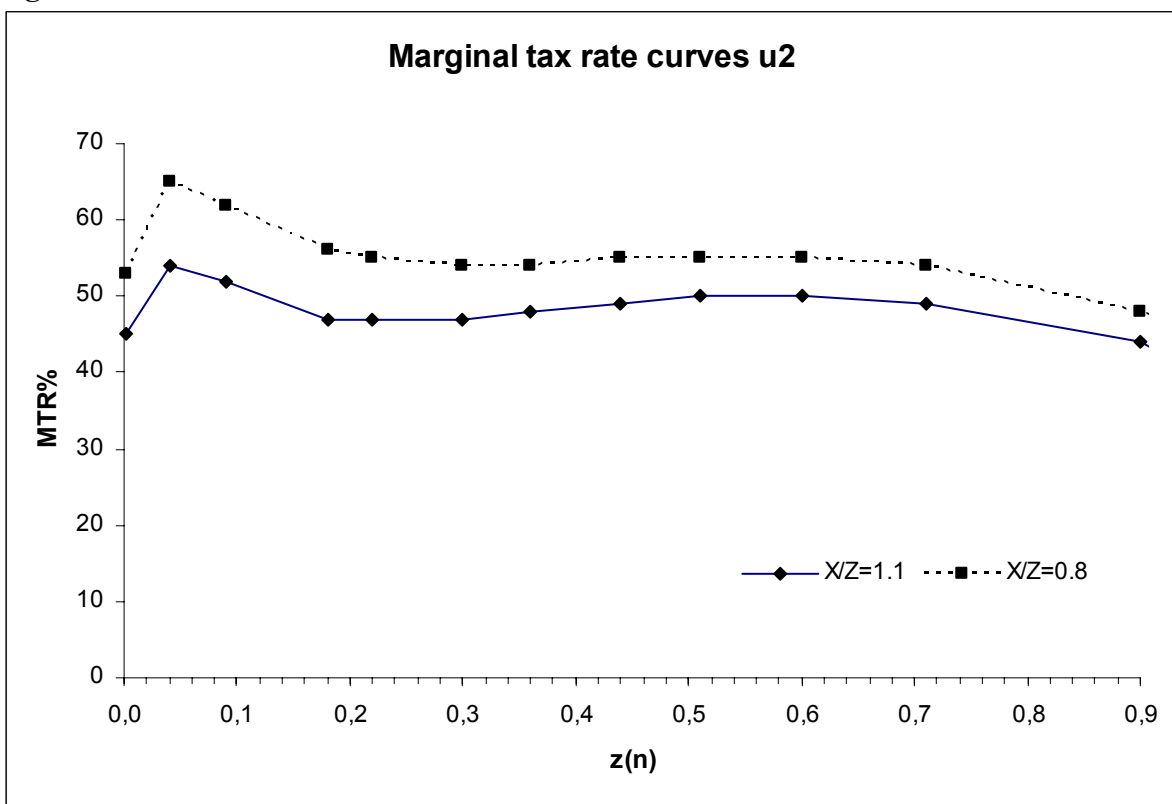


Figure B3a

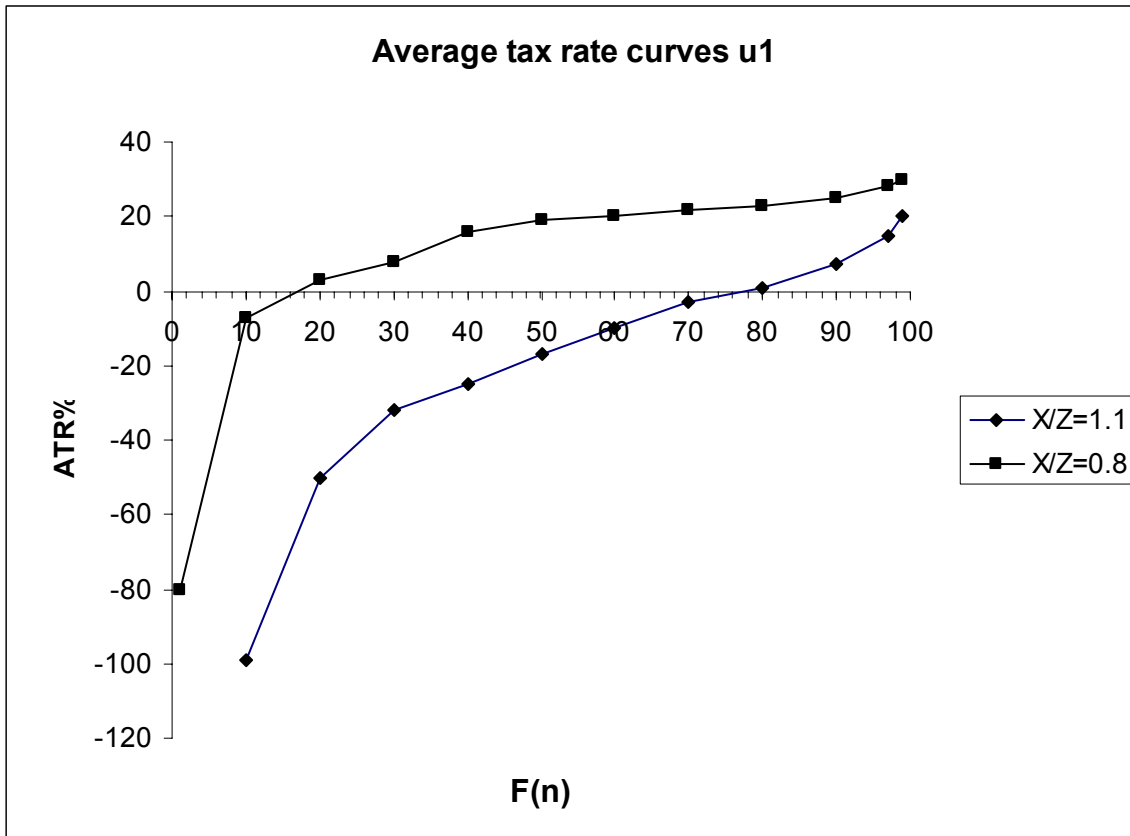


Figure B3b

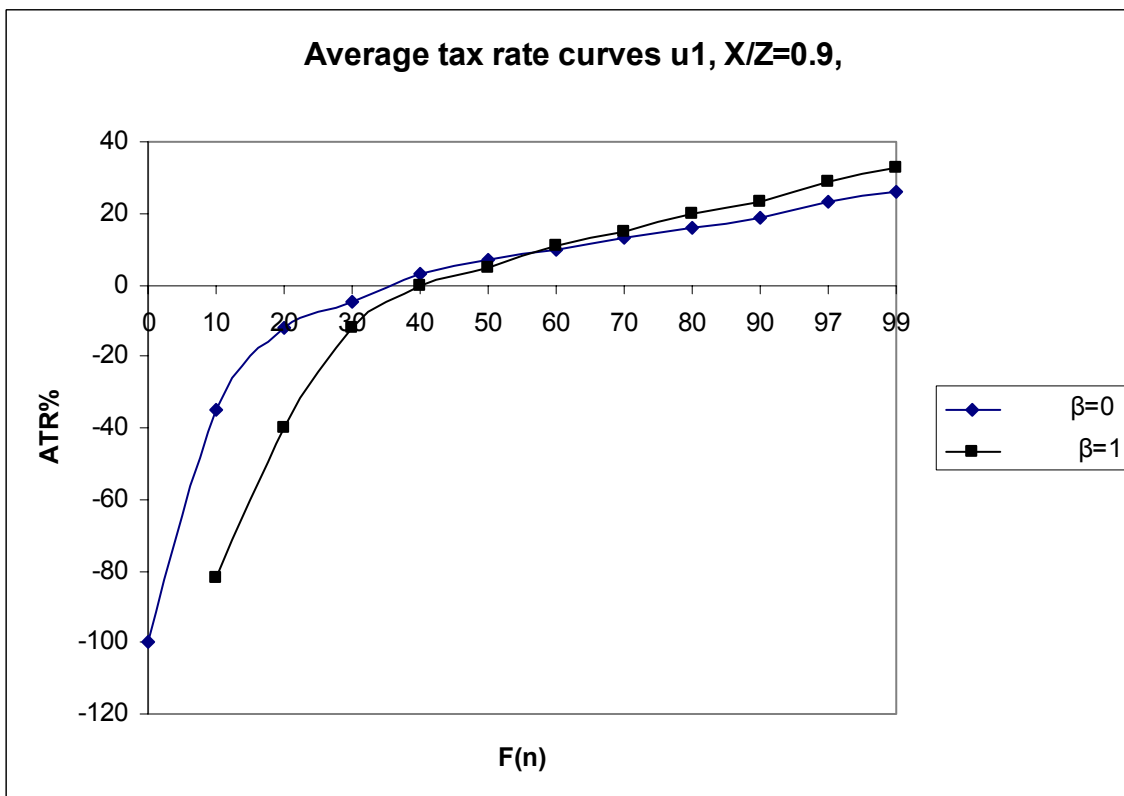


Figure B3c

