Mirrleesian optimal taxation: Theory and numerical solutions

Dushko Josheski[1]

Abstract

In this paper Mirrleesian optimal taxation will be reviewed. Models in optimal tax theory typically posit that the tax system should maximize a social welfare function subject to a government budget constraint, considering how individuals respond to taxes and transfers. James Mirrlees (1971) launched the second wave of optimal tax models by suggesting a way to formalize the planner's problem that deals explicitly with unobserved heterogeneity among taxpayers.There are static and dynamc versions of this model and we will review them or introduce them in this paper. Social welfare is larger when resources are more equally distributed, but redistributive taxes and transfers can negatively affect incentives to work and earn income in the first place. This creates the classical trade-off between equity and efficiency which is at the core of the optimal labor income tax problem.We will describe main theoretical findings in this literature as well as numerical examples with their policy implications.

Keywords: Optimal taxation, Mirrlees tax model, asymmetric information, non-linear tax rates, second-best analysis of taxes

1.Introduction

This paper will review topic from optimal Mirrleesian taxation. In the classical framework initiated by Mirrlees (1971), the theory studies the maximization of a utilitarian social welfare function by a benevolent planner who only observes the pretax labor income of agents whose wages differ, but whose preferences are identical. The other studies have relaxed the assumptions in order to take heterogeneity among agents into account. These studies include: Mirrlees (1976), Saez (2001), Choné and Laroque (2010), see Fleurbaey , Maniquet (2018). Mainly approach is based on asymmetric information. Public policies apply to the individuals on the basis of what the government knows about them. *Second* welfare theorem[2] states, that where a number of convexity and continuity assumptions are satisfied, an optimum is a competitive equilibrium once initial endowments have been suitably distributed. In general, complete information about the consumers for

[1] Goce Delchev University - Stip, North Macedonia, dusko.josevski@ugd.edu.mk

[2] Second fundamental theorem is giving conditions under which a Pareto optimal allocation can be supported as a price equilibrium with lump-sum transfers, i.e. Pareto optimal allocation as a market equilibrium can be achieved by using appropriate scheme of wealth distribution (wealth transfers) scheme (Mas-Colell, Whinston et al. 1995)

the transfers is required to make the distribution requires, so the question of feasible lump-sum transfers arises here. Usually the optimal tax systems combine flat marginal tax rate plus lump sum grants to all the individuals (so that the average tax rate rises with income even if the marginal does not), Mankiw NG, Weinzierl M, YaganD.(2009).Rigorous derivations of the optimal tax rates include:Atkinson,Stiglitz,(1980);Kaplow,(2008);Mirrlees(1976),Mirrlees(1986);Stiglitz,(1987);Tuomala,(1990). The choice of the optimal redistributive tax involves tradeoffs between three kinds of effects : equity effect (it changes the distribution of income) , the efficiency effect form reducing the incentives, the insurance effect from reducing the variance of individual income streams, Varian,H.R.(1980). Saez (2001) argued that "unbounded distributions are of much more interest than bounded distributions to address high income optimal tax rate problem". Saez (2001) investigated (four cases)[3] and the optimal tax rates are clearly U-shaped, see Diamond (1998) too. Saez,S.Stantcheva (2016),define social marginal welfare weight as a function of agents consumption, earnings, and a set of characteristics that affect social marginal welfare weight and a set of characteristics that affect utility. Piketty, Saez,Stantcheva(2014),derived optimal top tax rate formulas in a model where top earners respond to taxes through three channels: labor supply, tax avoidance, and compensation bargaining. Dynamic taxation most famous examples in the literature are: Diamond-Mirrlees (1978);Albanesi-Sleet(2006),Shimer-Werning(2008),Ales-Maziero(2009),Golosov-TroshkinTsyvinsky(2011).Sizeable literature in NDPF studies optimal taxation in dynamic settings,(Golosov,Kocherlakota,Tsyvinski(2003), Golosov,Tsyvinski, and Werning (2006), Kocherlakota (2010).Here we will derive optimal linear, non-linear tax rates for top earners and we will derive results in heterogenous preferences environment for dynamic taxation.

2.Mirrlees framework optimal top tax rate : derivation

The effect of small tax reform in MIrrless (1971) model is examined in Brewer, M., E. Saez, and A. Shephard (2010) ,where indirect utility function is given as :

$U(1-\tau,R)=max_w\big((1-\tau)w+R,z\big)$, where w represents the taxable income R is a virtual income intercept, and τ is an imposed income tax. Marshalian labor supply is w$=w(1-\tau,R)$, uncompensated elasticity of the supply is given as:$\varepsilon^u=\frac{(1-\tau)}{w}\frac{\partial w}{\partial(1-\tau)}$, income effect is $\eta=(1-\tau)\frac{\partial w}{\partial R}\leq 0$.Hicksian supply of labor is given as:$w^c\big((1-\tau,u)\big)$, this minimizes the cost in need to achieve slope $1-\tau$, compensated elasticity now is :$\varepsilon^c=\frac{(1-\tau)}{w}\frac{\partial w^c}{\partial(1-\tau)}>0$, Slutsky equation now becomes: $\frac{\partial w}{\partial(1-\tau)}=\frac{\partial w^c}{\partial(1-\tau)}+z\frac{\partial z}{\partial R}\Rightarrow\varepsilon^u=\varepsilon^c+\eta$, where η represents income effect :$\eta=(1-\tau)\frac{\partial w}{\partial R}\leq 0$.

With small tax reform taxes and revenue change i.e.:$dU=u_c\cdot[-wdt+dR]+dw[(1-\tau)u_c+u_z]=u_c\cdot[-zdt+dR]$.Change of taxes and its impact on the society is given as:$dU_i=-u_c dT(w_i)$. Envelope theorem here says :$U(\theta)=$

[3] Utilitarian criterion, utility type I and II and Rawlsian criterion, utility type I and II.

$\max_x F(x,\theta), s.t. c > G(x,\theta)$, and the preliminary result is $:U'(\theta) = \frac{\partial F}{\partial \theta}\left(x^*(\theta), \theta - \lambda^*(\theta)\frac{\partial G}{\partial \theta}x^*(\theta), \theta\right)$. Government is maximizing :

$$0 = \int G'(u^i)u_c^i \cdot \left[\left(W - w^i\right) - \frac{\tau}{d(1-\tau)}eW\right] \quad (1)$$

1. mechanical effect is given as:$dM = [w - w^*]d\tau$,
2. welfare effect is :$dW = -\bar{g}dM = -\bar{g}[w - w^*]$, and at last
3. the behavioral response is : $dB = -\frac{\tau}{1-\tau} \cdot e \cdot wd\tau$.

And let's denote that:

$$dM + dW + dB = d\tau\left[1 - \bar{g}[w - w^*] - e\frac{\tau}{1-\tau} \cdot w\right] \quad (2)$$

When the tax is optimal these three effects should equal zero i.e. $dM + dW + dB = 0$ given that:$\frac{\tau}{1-\tau} = \frac{(1-\bar{g})[w-w^*]}{e \cdot z}$, and we got $\tau = \frac{1-\bar{g}}{1-\bar{g}+a\cdot e}, a = \frac{w}{w-w^*}$,and $dM = d\tau[w - w^*] \ll dB = d\tau \cdot e\frac{\tau}{1-\tau} \cdot w, when\ w^* > w^T$, where w^T is a top earner income. Pareto distribution is given as:

$$1 - F(w) = \left(\frac{k}{w}\right)^a, f(w) = a \cdot \frac{k^a}{w^{1+a}} \quad (3)$$

a is a thickness parameter and top income distribution is measured as:

$$w(w^*) = \frac{\int_{z^*}^{\infty} sf(s)ds}{\int_{z^*}^{\infty} f(s)ds} = \frac{\int_{z^*}^{\infty} s^{-a}ds}{\int_{z^*}^{\infty} s^{-a-1}ds} = \frac{a}{(a-1)} \cdot w^* \quad (4)$$

Empirically $a \in [1.5,3]$, $\tau = \frac{1-\bar{g}}{1-\bar{g}+a\cdot e}$.General non-linear tax without income effects is given as:

$$\frac{T\prime(w_n)}{1-T\prime(w_n)} = \frac{1}{e}\left(\frac{\int_n^{\infty}(1-g_m)dF(m)}{w_n h(w)}\right) = \frac{1}{e}\left(\frac{1-H(w_n)}{w_n h(w_n)}\right) \cdot \left(1 - G((w_n)\right) \quad (5)$$

Where elasticity or efficiency $e = \left[\frac{1-\tau}{w}\right] \times \frac{dw}{d(1-\tau)}$. Where $G((w_n) = \frac{\int_n^{\infty} g_m dF(m)}{1-F(n)}$,and $g_m = G'(u_m)/\lambda$ this is welfare weight of type m.But non-linear tax witn income effect takes into account small tax reform where tax rates change from $d\tau$ to $[w^*, w^* + dw^*]$.Every tax payer with income $w > w^*$ pays additionaly $d\tau dw^*$ valued by $(1 - g(w))d\tau dw^*$.Mechanical effect is :

$$M = d\tau dw^* \int_{z^*}^{\infty}(1 - g(w))d\tau dw^* \quad (6)$$

Total income response is :$I = d\tau dw^* \int_{z^*}^{\infty}\left(-\eta_Z \frac{T'(w)}{1-T'(w)}(w)\right)h(w)dw$. Change at the taxpayers form the additional tax is :$dz = -\varepsilon_{(z)}^c \frac{T''dz}{1-T'} - \eta\frac{d\tau dw^*}{1-T'(w)} \Rightarrow -\eta\frac{d\tau dw^*}{1-T'(w)+z\varepsilon_{(w)}^c T\prime\prime(w)}$, if one sums up all effects can be obtained:

$$\frac{T'(w)}{1-T'(w)} = \frac{1}{\varepsilon_{(z)}^c}\left(\frac{1-H(w^*)}{z^*h(w^*)}\right) \times \left[\int_{z^*}^{\infty}(1 - g(w))\frac{h(w)}{1-H(w^*)}dz + \int_{z^*}^{\infty} -\eta\frac{T'(w)}{1-T'(w)}\frac{h^*(w)}{1-H(w^*)}dw\right] \quad (7)$$

With linear tax: $\frac{\dot{z}_n}{z_n} = \frac{1+\varepsilon^u_{(n)}}{n}$ and with non-linear tax:

$$\frac{\dot{w}_n}{w_n} = \frac{1+\varepsilon^u_{(n)}}{n} - \dot{w}_n \frac{T''(w_n)}{1-T''(w_n)} \varepsilon^c_{w(n)} \quad (8)$$

Optimal tax formula here if $dM + dW + dB = 0$ is given as : $\tau = \frac{1-\bar{g}}{1-\bar{g}+\alpha\cdot e}$; $\alpha = \frac{w}{w-w^*}$ where $\bar{g} = \frac{\int g_i \cdot w_i}{w \cdot \int g_i}$ and$g_i = G'(u')u^i_c$.

2.1 Formal derivation of optimal non-linear tax rates with no income effects

This point actually follows Mirrleees (1971) and Diamond (1998) , in deriving non-linear optimal tax rate with no-income effects.Utility function is quasi linear:

$$u(c,l) = c - v(l) \quad (9)$$

c is disposable income and the utility of suuply of labor $v(l)$ is increasing and convex in l. Earnings equal $w = nl$ where n represents innate ability. CDF of skills distribution is $F(n)$, it's PDF is $f(n)$and support range is $[0,\infty)$. Government cannot observe abilities instead it can set taxes asa function of labor income $c = w - \tau(w)$.Individual n chooses l_n to maximize :

$$\max(nl - \tau n(l) - v(l)) \quad (10)$$

When marginal tax rate τ is constant, the labor supply f-ction is given as: $l \rightarrow l\big(n(1-\tau)\big)$ and it is implicitly defined by the $n(1-\tau) = v'(l)$. And $\frac{dl}{d(n(1-\tau))} = \frac{1}{v''(l)}$, so the elasticity of the net-of-tax rate $1-\tau$ is:

$$e = \frac{\left(\frac{n(1-\tau)}{l}\right)dl}{d\big(n(1-\tau)\big)} = \frac{v'(l)}{lv''(l)} \quad (11)$$

As there are no income effects this elasticity is both the compensated and the uncompensated elasticity.The government maximizes SWF :

$$W = \int G(u_n) f(n)dn \;\; s.t. \int cnf(n)dn \leq \int nlnf(n)dn - E(\lambda) \quad (12)$$

u_n denotes utility, $w_n = nl_n$ denotes earnings, c_n denotes consumption or disposable income, and $c_n = u_n + v(l_n)$.By using the envelope theorem and the FOC for the individual, u_n satisfies following:

$$\frac{du_n}{dn} = \frac{lnv'(ln)}{n} \quad (13)$$

Now the Hamiltonian is given as:

$$\mathcal{H} = \big[G(u_n) + \lambda \cdot \big(nl_n - u_n - v(l_n)\big)\big] f(n) + \phi(n) \cdot \frac{lnv'(ln)}{n} \quad (14)$$

In previous $\phi(n)$ is the multiplier of the state variable. The FOC with respect to l is given as:

$$\lambda \cdot \big(n - v'(l_n)\big) + \frac{\phi(n)}{n} \cdot [v'(l_n) + l_n v''(l_n)] = 0 \quad (14)$$

FOC with respect to u is given as:

$$-\frac{d\phi(n)}{n} = [G'(u_n) - \lambda] \quad (15)$$

If integrated previous expression gives: $-\phi(n) = \int_n^\infty [\lambda - G'(u_m)] f(m)dm$ where the transversality condition $\phi(\infty) = 0$,and $\phi(0) = 0$, and $\lambda = \int_0^\infty G'(u_m) f(m)dm$ and social marginal welfare weights $\frac{G'(u_m)}{\lambda} = 1$. Using this equation for $\phi(n)$ and all previous $n - v'(ln) = n\tau'(w_n)$, and that

$$\frac{[v'(l_n) + l_n v''(l_n)]}{n} = \left[\frac{v'(l_n)}{n}\right]\left[1 + \frac{1}{e}\right] \quad (16)$$

We can rewrite FOC with respect to l_n as:

$$\frac{\tau'(w_n)}{1-\tau'(w_n)} = \left(1 + \frac{1}{e}\right) \cdot \left(\frac{\int_n^\infty (1-g_m) dF(m)}{nf(n)}\right) \quad (17)$$

In previous expression $g_m = \frac{G'(u_m)}{\lambda}$ which is the social welfare on individual m . The formula was derived in Diamond (1998) . If we denote $h(w_n)$ as density of earnings at w_n if the nonlinear tax system were replaced by linearized tax with marginal tax rate $\tau = \tau'(w_n)$ we would have that following equals $h(w_n)dw_n = f(n)dn$ and $f(n) = h(w_n)l_n(1 + e)$,henceforth $nf(n) = w_n h(w_n)(1 + e)$ and we can write previous equation as:

$$\frac{\tau'(w_n)}{1-\tau'(w_n)} = \frac{1}{e} \cdot \left(\frac{\int_n^\infty (1-g_m) dF(m)}{w_n h(w_n)}\right) = \frac{1}{e} \cdot \left(\frac{1-H(w_n)}{w_n h(w_n)}\right) \cdot \left(1 - G(w_n)\right) \quad (18)$$

In the previous expression $G(w_n) = \int_n^\infty \frac{dF(m)}{1-F(n)}$ is the average social welfare above w_n .If we change variables from $n \to w_n$, we have $G(w_n) = \int_{w_n}^\infty \frac{g_m dH(w_m)}{1-H(w_n)}$.The transversality condition implies $G(w_0 = 0) = 1$.

2.2 Optimal linear tax formula

First modern treatment of optimal linear tax was provided by Sheshinski (1972).Optimal linear tax formulae is given as:

$$\int_0^\infty \tau(w) f(n)dn = \int_0^\infty (w - \alpha - \beta w) f(n)dn = 0 \quad (19)$$

$f(n)$ is PDF of ability n, α is a tax parameter and is a lump-sum tax if $\alpha < 0$ and tax-subsidy if $\alpha > 0$ given to an individual with no income.$1 - \beta$ is a marginal tax rate i.e. $0 \le \beta \le 1$ so that marginal tax rate is non negative in the linear tax function which is $\tau(w) = -\alpha + (1 - \beta)w$, after tax consumption is $c(w) = w - \tau(w) = \alpha + \beta w$.Optimal labor supply is given as: $\ell = \hat{\ell}(\beta n, \alpha)$.If λ is the lowest elasticity of labor supply function and it is equal to $\lambda = \lim_n \inf \left[\frac{\beta}{\hat{\ell}} \frac{\partial \hat{\ell}}{\partial \beta}\right]$ so that $\frac{\beta}{\hat{\ell}} \frac{\partial \hat{\ell}}{\partial \beta} \ge \lambda$. Revenue maximizing linear tax rate is given as: $\frac{\tau^*}{1-\tau^*} = \frac{1}{e}$ or $\tau^* = \frac{1}{1+e}$ .Government FOC given $SWF = \int \omega_i G\left(u^i(1 - \tau)w^i + \tau w(1 - \tau) - E, w^i\right)\right) df(i)$ is :

$$0 = \frac{dSWF}{d\tau} = \int \omega_i G'(u_i) u_c^i \cdot \left((w - w^*) - \tau \frac{dw}{d(1-\tau)}\right) df(i) \quad (20)$$

Social marginal welfare weight g_i is given as: $g_i = \frac{\omega_i G'(u_i) u_c^i}{\int \omega_j G'(u_j) u_c^j df(j)}$.So that optimal linear tax formula is:

$$\tau = \frac{1-\bar{g}}{1-\bar{g}+e} \qquad (21)$$

where $\bar{g} = \frac{\int g_i \cdot w_i df(i)}{w}$.

2.3 Diamond ABC formula

Here in this paragraph a Diamond (1988) formula has been derived. Welfare weights are distributed with a CDF: $\Psi(n)$ and PDF : $\psi(n)$. The government maximization function is (objective function) is given as:

$$\int_{\underline{n}}^{\overline{n}} u(n)\psi(n)dn \qquad (22)$$

Now by assumption $\int_{\underline{n}}^{\overline{n}} u(n)\psi(n)dn = 1$, which implies that $\lambda = 1$,λ aggregates the social welfare weights across the entire economy.

$$\lambda = \int_{\underline{n}}^{\overline{n}} \Psi u(n)\psi(n)dn \qquad (23)$$

FOC can be found as previously, form the Hamiltonian $\mathcal{H} = \left[\Psi(u_n) + \lambda \cdot \left(nl_n - u_n - v(l_n)\right)\right]\psi(n) + \phi(n) \cdot \frac{lnv'(ln)}{n}$. In previous $\phi(n)$ is the multiplier of the state variable. The FOC with respect to l is given as:$\lambda \cdot \left(n - v'(l_n)\right) + \frac{\phi(n)}{n} \cdot [v'(l_n) + l_n v''(l_n)] = 0$.FOC with respect to u is given as:

$$-\frac{d\phi(n)}{n} = [\Psi(u_n) - \lambda] = -\phi'(n) - \lambda f(n) \qquad (24)$$

Or alternatively: $-\phi(n) = \int_n^{\bar{n}} \left(f(n) - \Psi(n)\right)dn = \Psi(n) - F(n)$

$$\frac{\tau'(w_n)}{1-\tau'(w_n)} = \left(\frac{1+e}{e}\right) \cdot \left(\frac{\psi(n)-F(n)}{nf(n)}\right) \qquad (25)$$

To write ABC formula we divide and multiply by $1 - F(n)$:

$$\frac{\tau'(w_n)}{1-\tau'(w_n)} = \underbrace{\left(\frac{1+e}{e}\right)}_{A(n)} \cdot \underbrace{\left(\frac{\psi(n)-F(n)}{1-F(n)}\right)}_{B(n)} \cdot \underbrace{\left(\frac{1-F(n)}{nf(n)}\right)}_{C(n)} \qquad (26)$$

Where $A(n) = \frac{1+e}{e}$ is the elasticity and efficiency argument, $B(n) = \frac{\psi(n)-F(n)}{1-F(n)}$ measures the desire for redistribution, $C(n) = \frac{1-F(n)}{nf(n)}$ measures the thickness on the right tail of distribution. In the Rawlsian case $\Psi(n) = 1$ previous formula will converge to:

$$\frac{\tau'(w_n)}{1-\tau'(w_n)} = \left(\frac{1+e}{e}\right) \cdot \left(\frac{1-F(n)}{nf(n)}\right) \qquad (27)$$

2.4 Formal derivation of optimal non-linear tax rates with income effects

Utility function takes form $\tilde{u}(c,l) = u(c) - v(l)$ where $u'(c) > 0; u''(c) \leq 0$.Elasticity of labor supply is :

$$\frac{v'(l)}{u'(c)} = \big(1 - \tau'(w)\big)n \tag{28}$$

The uncompensated response of labor supply is given as:

$$\frac{\partial l^u}{\partial(1-\tau'(w))n} = \frac{u'(c)+l\big(1-\tau'(w)\big)nu''(c)}{v''(l)-\big(1-\tau'(w)\big)^2 n^2\, u''(c)} \tag{29}$$

And uncompensated elasticity is implied:

$$\varepsilon^u = \frac{\frac{u'(c)}{l}+\frac{v'(l)^2}{u'(c)^2}u''(c)}{v''(l)-\frac{v'(l)^2}{u'(c)^2}u''(c)} \tag{30}$$

The response of labor to income changes is given as:

$$\frac{\partial l}{\partial y} = \frac{(1-\tau'(w)nu''(c)}{v''(l)-\big(1-\tau'(w)\big)^2 n^2 u''(c)} \tag{31}$$

By using the Slutsky equation we have:

$$\frac{\partial l^c}{\partial\big(1-\tau'(w)\big)n} = \frac{u'(c)+l\big(1-\tau'(w)\big)nu''(c)}{v''(l)-\big(1-\tau'(w)\big)^2 n^2 u''(c)} - \frac{l\big(1-\tau'(w)\big)nu''(c)}{v''(l)-\big(1-\tau'(w)\big)^2 n^2 u''(c)} = \frac{u'(c)}{v''(l)-\big(1-\tau'(w)\big)^2 n^2 u''(c)} \tag{32}$$

Henceforth :

$$\varepsilon^c = \frac{\frac{v'(l)}{l}}{v''(l)-\big(1-\tau'(w)\big)^2 n^2 u''(c)} \tag{33}$$

Here everything is as previous except now we cannot replace $c(n)$ in the resource constraint by using def. of indirect utility here we will define consumption as expenditure function $\tilde{c}(\tilde{u}(n), w(n), n)$.Previous resource constraint for this economy with no income effects was:

$$\int_{\underline{n}}^{\overline{n}} c(n)f(n)dn \geq \int_{\underline{n}}^{\overline{n}} w(n)f(n)dn - E \tag{34}$$

So this new function we will differentiate w.r.t. $\tilde{u}(n), w(n)$.Indirect utility is defined as :

$$\tilde{u}(n) = u\big(\tilde{c}(n)\big) - v\left(\frac{w^*(n)}{n}\right) \tag{35}$$

At optimum conditions that hold are:

$$\begin{gathered} d\tilde{u}(n) = u'\big(\tilde{c}(n)\big)d\tilde{c}(n) \\ 0 = u'\big(\tilde{c}(n)\big)d\tilde{c}(n) - \frac{1}{n}v'\left(\frac{w^*(n)}{n}\right)dw^*(n) \end{gathered} \tag{36}$$

If we rearrange we will get :

$$\begin{gathered} \frac{d\tilde{c}(n)}{d\tilde{u}(n)} = \frac{1}{u'\big(\tilde{c}(n)\big)} \\ \frac{d\tilde{c}(n)}{dw^*(n)} = \frac{v'\left(\frac{w^*(n)}{n}\right)}{nu'\big(\tilde{c}(n)\big)} \end{gathered} \tag{37}$$

Hamiltonian for this problem is given as:

$$\mathcal{H} = [G(u(n) + \lambda(w(n) - \tilde{c}(\tilde{u}(n), w(n), n)]f(n) + \phi(n)\frac{w(n)}{n^2}v'\left(\frac{w(n)}{n}\right) \tag{38}$$

FOC's are given a

$$\frac{\partial\mathcal{H}}{\partial w(n)} = \lambda\left[1 - \frac{v'\left(\frac{w(n)}{n}\right)}{nu'(c(n))}\right]f(n) + \frac{\phi(n)}{n^2}\left[v'\left(\frac{w(n)}{n}\right) + \frac{w(n)}{n}v''\left(\frac{w(n)}{n}\right)\right] = 0$$
$$\frac{\partial\mathcal{H}}{\partial u(n)} = \left[G'(u(m) - \frac{\lambda}{u'c(m)}\right]f(n)d = -\phi'(n) \quad (39)$$

For the multiplier $\phi'(n)$ the equilibrium value is given as:

$$\phi(n) = \int_{\underline{n}}^{\overline{n}}\left[G'(u(m) - \frac{\lambda}{u'c(m)}\right]f(m)dm \quad (40)$$

With the definition of the two elasticities we can write :

$$\left[v'\left(\frac{w(n)}{n}\right) + \frac{w(n)}{n}v''\left(\frac{w(n)}{n}\right)\right] = v'\left(\frac{w(n)}{n}\right)\left[1 + \frac{w(n)}{n}\frac{v''\left(\frac{w(n)}{n}\right)}{v'\left(\frac{w(n)}{n}\right)}\right] = v'\left(\frac{w(n)}{n}\right)\left(\frac{1+\varepsilon^u}{\varepsilon^c}\right) (41)$$

The optimal tax formula then will become :

$$\frac{\tau'(w_n)}{1-\tau'(w_n)} = \left(\frac{1+e^u}{e^c}\right)\cdot\left(\frac{\eta(n)}{nf(n)}\right) \quad (41)$$

Where $\eta(n) = \frac{u'(c(n)\phi(n))}{\lambda}$.

2.5 Pareto efficient taxes (due to Werning (2007))

If the tax system $T_0\ (w)$ in place is Pareto-optimal, it means that there exists no feasible adjustment in the tax schedule such that all individuals in the economy are weakly better off. We can characterize Pareto frontier such as:

$$\max \int_{\underline{n}}^{\overline{n}} u(n)\psi(n)dn \quad s.t.$$
$$u(c(n) - h\left(\frac{w(n)}{n}\right) \geq u(c(n') - h\left(\frac{w(n')}{n}\right),\forall n,n' \quad (42)$$
$$\int_{\underline{n}}^{\overline{n}}[w(n) - c(n)]f(n)dn \geq E$$

Definition: *Pareto efficient tax structures are those (given the admissible set of taxes and the required public revenue) which are such that no one can be better off without making someone worse off.*

Proposition 1: *A tax code fails to be constrained Pareto optimal if and only if there exists a feasible tax reform that (weakly) reduces taxes at all incomes*

Proof: (if) suppose we weakly reduce taxes all over the entire economy, then every individual is at least as well off. (only if) suppose there exists a Pareto improving feasible tax reform $T_1\ (w)$. Then we have:

$$u\big(w_1(n) - T_1\big(w_1(n)\big), w_1(n), n\big) \geq u\big(w_0(N) - T_0\big(w_0(n)\big), w_0(n), n\big) \geq u\big(w_1(n) - T_0\big(w_1(n)\big), w_1(n), n\big) \blacksquare \quad (43)$$

Here we are going to assess the Pareto efficiency of a tax schedule. Here first assumption is that elasticity of labor supply is zero. Now, let ε_w^* represents the compensated elasticity of labor supply with respect to real wage. Let the distribution of income generated by the current tax system be Pareto:

$$h(w) = k(w)^{-k-1}\underline{w}\,k\ for\ w \geq \underline{w}\ and\ k > 0 \quad (44)$$

and now let's suppose that there is linear flat tax :$t(w) = t + \tau(w)$ .Where τ represents marginal tax rate and intercept t.Here we assume that ε_w^* does not vary across individuals. This will be true in the case of this utility function[4]:$u(c,w,\theta) = c - w\theta^{\alpha}$.Now, starting from a general test for Pareto efficiency we will derive inequality for τ, ε_w^*, k. The starting point here is this inequality which states that marginal tax rate must be lower than 100% :

$$\frac{\tau(\theta)}{1-\tau(\theta)}\frac{\varepsilon_w^*}{\Phi}\left(-\frac{dlog\frac{\tau(\theta)}{1-\tau(\theta)}}{dlogw} - 1 - \frac{dlog(\varepsilon_w^*(w))}{dlogw} - \frac{dlog(h^*(w)}{dlogw} - \frac{\partial MRS}{\partial c}w\right) \leq 1 \quad (45)$$

The assumptions to use this inequality are as follows:

1. By quasi-linear utility preferences we have : $-\frac{\partial MRS}{\partial c} = 0$
2. A flat tax implies no convexity $t'' = 0$, a constant marginal tax rate $MTR = \tau(\theta) = \tau$ and also $\frac{dlog\frac{\tau(\theta)}{1-\tau(\theta)}}{dlogw} = 0$
3. Now, the logarithm of Pareto income density is given as:
$$\log(h \cdot (w)) = \log\, k - (k\, +\, 1)\log w\, +\, k\log \underline{w} \quad (46)$$

First of this log density with respect to income gives:

$$\frac{d\log(h^*(w)}{d\log w} = \frac{d(\log\, k-(k+1)\log w+k\log\, \underline{w})}{d\log\, w} = \frac{-(k+1)d\log\, w}{d\log w} = -(k+1) \quad (47)$$

So the first inequality in this part $\frac{\tau(\theta)}{1-\tau(\theta)}\frac{\varepsilon_w^*}{\Phi}\left(-\frac{dlog\frac{\tau(\theta)}{1-\tau(\theta)}}{dlogw} - 1 - \frac{dlog(\varepsilon_w^*(w))}{dlogw} - \frac{dlog(h^*(w)}{dlogw} - \frac{\partial MRS}{\partial c}w\right) \leq 1$ would become:

$$\frac{\tau(\theta)}{1-\tau(\theta)}\varepsilon_w^* k \leq 1 \quad (48)$$

The parameter k has been estimated by Saez (2001) to be of value 1.6[5]. The thicker the tail of the distribution, the smaller is a. Pareto distribution is given as PDF lower CDF and upper CDF [6].PDF (probability density function) :

$$f(x, x_m, \alpha) = \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}} \quad (49)$$

Lower cumulative distribution function (lower CDF):$P(x, x_m, \alpha) = \int_{x_m}^{x} f(x, x_m, \alpha)dx = 1 - \left(\frac{x^m}{x}\right)^{\alpha}$.Upper cumulative distribution function (upper CDF):$Q(x, x_m, \alpha) = \int_{x}^{\infty}(x, x_m, \alpha)dx = \left(\frac{x_m}{x}\right)^{\alpha}$

4. The compensated elasticity of labor supply with respect to real wage ε_w^* has been estimate approximately to be 0.5 see Gruber, Saez (2002).

5. So that $\frac{1}{\varepsilon_w^*} \in \left[\frac{1}{6}; \frac{10}{3}\right]$ or $\frac{1}{2*3} = \frac{1}{6}$ and $\frac{1}{0.2*1.5} = \frac{10}{3}$ which lies around a central value of $\frac{1}{0.5 \cdot 2.5} = 0.8$

And the second inequality from above now would become:

$$\frac{\tau(\theta)}{1-\tau(\theta)} \leq 0.8 \quad (50)$$

[4] $\boldsymbol{\theta}$ represents every individual's characteristics
[5] This value is approx..for US incomes above 0.3 m.
[6] This part is for readers that are not familiar with basic statistics

Gruber, Saez (2002) estimate that for the US taxpayer with incomes above 100K$ have elasticity around 0.57. And those <100K$ have elasticity around 0.2 or even less. Then the inequality will be affected in two ways:

1. $\varepsilon_w^*(\theta)$ will be higher for higher incomes
2. $\frac{dlog(\varepsilon_w^*(w))}{dlogw} > 0$

The inequality then becomes:
Relative to the average-income constant elasticity benchmark case the upper bound on the marginal tax ratio $\frac{1}{\varepsilon_w^*\left(k - \frac{dlog(\varepsilon_w^*(w))}{dlogw}\right)}$ is affected as follows for high and low earners:

- For high earners :
  1. is directly negative affected by the factor $\frac{1}{\varepsilon_w^*}$
  2. is positively affected by the factor $\frac{dlog(\varepsilon_w^*(w))}{dlogw}$
- for low earners:
  1. is directly positively affected by the factor $\frac{1}{\varepsilon_w^*}$
  2. is positively affected by the factor $\frac{dlog(\varepsilon_w^*(w))}{dlogw}$

Thus, in order to pass the efficiency test: (1)a higher maximal marginal tax rate for low-income earners is acceptable, (2) the effect on the maximal tax rate for high-income earners is theoretically ambiguous even if I suspect the direct negative effect to dominate because locally the logarithm of elasticity is relative stable compared to the parameter k and hence a lower maximal marginal tax rate for high- income earners is acceptable. This is very intuitive: if low-income earners are less elastic, we can tax them relative more.
Now, let's see how progressivity would affects tax schedule in question here. Convexity implies that $\tau'' > 0$. To keep things simple, we continue to assume that there is:

- quasi-linearity of preferences: $-\frac{\partial MRS}{\partial c} w = 0$
- a constant compensated elasticity of labor supply with respect to the real wage: $\frac{dlog(\varepsilon_w^*(w))}{dlogw} = 0$.Then the inequality becomes:

$$\frac{\tau(\theta)}{1-\tau(\theta)} \frac{\varepsilon_w^*}{\Phi} \left(- \frac{dlog\frac{\tau(\theta)}{1-\tau(\theta)}}{dlogw} - 1 - \frac{dlog(h^*(w)}{dlogw} \right) \leq 1 \qquad (51)$$

Given the convexity we also have because $\tau'(w) < 1 : \Phi(w) = 1 + we_w^*(w) \frac{\tau''(w)}{1-\tau'(w)} > 1$. Now we have that:

$$\frac{dlog(h^*(w)}{dlogw} = \frac{dlog(h^*(w)\Phi(w)^{-1}}{dlogw} \qquad (52)$$

We know from previously that $\Phi(w)$ increases in w or equivalently that $\Phi(w)^{-1}$ decreases in w and thus that the absolute value of the slope of the virtual density is higher than the real density $-\frac{dlog(h^*(w)}{dlogw} > -\frac{dlog(h(w)}{dlogw}$. Compared to the flat tax rate, the upper bound on the marginal tax ratio:

$$\tau'(\theta) = \frac{1}{\frac{\varepsilon_w^*}{\Phi}\left(- \frac{dlog\frac{\tau(\theta)}{1-\tau(\theta)}}{dlogw} - 1 - \frac{dlog(h^*(w)}{dlogw} \right)} \qquad (53)$$

And it is affected in three ways:

- positively by Φ
- positively by $-\frac{dlog\frac{\tau(\theta)}{1-\tau(\theta)}}{dlogw} < 0$
- negatively by the distinction between the virtual and the real density

We expect the positive effect to dominate and thus the upper bound on the marginal tax could then be higher. Werning (2008) proposed Pareto efficient income taxation with dual optimization problem in the original Mirrlees (1971) framework. Namely this model starts from the Mirrleesian framework with additively separable preferences like this:$u(c,y,\theta) = u(c) - \theta h(y)$.Where θ denotes heterogenous disutility from producing output y. Cardinality of preferences[7] is irrelevant and only ordinal preferences matter[8]. The expenditure function $e(v,y,\theta)$ is inverse from $u(\cdot,y,\theta)$,and $F(\theta)$ represents the distribution of θ in the population, and its PDF can be represented as $f(\theta)$. Some tax function is $t(y)$ and workers' utility $v(\theta)$ is maximized:$v(\theta) \equiv \max_y u(y - t(y), y, \theta)$ and $c(\theta) = e(v(\theta), y(\theta), \theta)$ is a consumption function dependent on workers' characteristics, $y(t) = y - t(y)$ and an allocation is resource feasible if :

$$\int \big(y(\theta) - c(\theta)\big) dF(\theta) + e \geq 0 \qquad (54)$$

Here e is an endowment. The allocation generated by some tax schedule is (constrained) Pareto efficient if there is no other tax schedule that induces a resource feasible allocation where nobody is worse off, and some workers are strictly better off. The marginal tax rate is :

$$\tau(\theta) = t'\big(y(\theta)\big) = 1 + \frac{u_y(c(\theta),y(\theta),\theta)}{u_c(c(\theta),y(\theta),\theta)} = 1 - \frac{\theta h'(y(\theta))}{u'^{(c(\theta))}} = 1 - e_y(v(\theta), y(\theta), \theta) (55)$$

There is above mentioned dual problem for the social planer.Here is introduced Pareto planning problem and sufficient and necessary conditions for optimality of the solution to the planers' problem[9].

$$\max_{\tilde{y},\tilde{v}} \int (\tilde{y}(\theta), -e(\tilde{v}(\theta),\theta) dF(\theta) \text{ s.t. } \tilde{v}(\theta) = \tilde{v}\big(\tilde{\theta}\ \big) - \int_{\theta}^{\bar{\theta}} u_\theta(e(\tilde{v}(z), \tilde{y}(z), z)\tilde{y}(z), z\ )dz \qquad (56)$$

In previous $\tilde{y}(\theta)$ is non-increasing $\tilde{v}\big(\tilde{\theta}\ \big) \geq v(\theta)$. The objective is to maximize aggregate net
resources, output minus consumption. FOC necessary to be verified in order allocation to be Pareto efficient. Lagrangian for the FOC's is:

$$\mathcal{L} = \int (\tilde{y}(\theta), -e(\tilde{v}(\theta),\theta) dF(\theta) + \int \Big(\tilde{v}(\theta) - \tilde{v}(\bar{\theta}) + \int_{\theta}^{\bar{\theta}} u_\theta(e(\tilde{v}(z), \tilde{y}(z), z)\tilde{y}(z), z\ )dz\Big) d\mu(\theta) \qquad (57)$$

Integrating second term by parts we have:

[7] In economics, a cardinal utility function or scale is a utility index that preserves preference orderings uniquely up to positive affine transformations, see Ellsberg (1954)

[8] In economics, an ordinal utility function is a function representing the preferences of an agent on an ordinal scale. Ordinal utility theory claims that it is only meaningful to ask which option is better than the other, but it is meaningless to ask how much better it is or how good

[9] A Pareto improvement would always be possible: if another allocation provided the same utility but increased net resources, then these resources can be used to construct another allocation that increases utility for some workers and is resource feasible.

$$\mathcal{L} = \int (\tilde{y}(\theta), -e(\tilde{v}(\theta), \theta) dF(\theta) - \tilde{v}(\bar{\theta})\mu(\bar{\theta}) + \mu(\underline{\theta})\tilde{v}(\underline{\theta}) + \int \tilde{v}(\theta) d\mu + \int \mu(\theta) u_\theta\, (\tilde{v}(\theta), \tilde{y}(\theta), \theta) d\theta \qquad (58)$$

About the efficiency conditions, the FOC for $\tilde{y}(\theta)$ evaluated at $(y(\theta), v(\theta))$ gives:

$$(1 - e_y(v(\theta), y(\theta), \theta) f(\theta) = -\mu(U_{\theta_c}(e(v(\theta), y(\theta), \theta) e_v(v(\theta), y(\theta), \theta) + u_{\theta_y}\big(e(v(\theta), y(\theta), \theta)\big) \qquad (59)$$

Implying $\mu(\theta) = \tau(\theta)\frac{f(\theta)}{h'(y(\theta))}$ .The FOC for $v(\bar{\theta})$ is $\mu(\bar{\theta}) \geq 0$, or if θ is bounded away from zero the FOC for $v(\underline{\theta})$ gives $\mu(\bar{\theta}) \leq 0$. And so : $\tau(\bar{\theta}) \geq 0$ and $\bar{\tau}(\underline{\theta}) \leq 0$. For interior θ, the FOC with respect to $\tilde{v}(\theta)$ evaluated at $(y(\theta), v(\theta))$ gives:

$$\dot{\mu}(\theta) =\leq e_v(v(\theta), y(\theta), \theta) f(\theta) \qquad (60)$$

By differentiation equation gives: $\dot{\mu}(\theta) = \mu(\theta)\left(\frac{\tau'(\theta)}{\tau(\theta)} + \frac{f'(\theta)}{f(\theta)} - \frac{h''(y(\theta))}{h'(y(\theta))} y'(\theta)\right)$

Substituting $\mu(\theta) = \tau(\theta)\frac{f(\theta)}{h'(y(\theta))}$ and $\dot{\mu}(\theta) = \mu(\theta)\left(\frac{\tau'(\theta)}{\tau(\theta)} + \frac{f'(\theta)}{f(\theta)} - \frac{h''(y(\theta))}{h'(y(\theta))} y'(\theta)\right)$ into the $\dot{\mu}(\theta) =\leq e_v(v(\theta), y(\theta), \theta) f(\theta)$ we get :

$$\tau(\theta)\left(\frac{d\log\tau(\theta)}{d\log(\theta)} + \frac{d\log f(\theta)}{d\log\theta} - \frac{d\log h'(y(\theta))}{d\log\theta}\right) \leq 1 - \tau(\theta) \qquad (61)$$

The integral form of this efficiency condition is given as:

$$\frac{\tau'(\theta) f(\theta)}{h' y(\theta)} + \int_\theta^{\bar{\theta}} \frac{1}{u'(c(\tilde{\theta}))} f(\tilde{\theta}) d\tilde{\theta} \leq 0 \qquad (62)$$

Proposition 1 : Given the utility function $u(c, y, \theta)$ and a density of skills $f(\theta)$, a differentiable tax function $t(y)$ inducing an allocation .$(c(\theta), y(\theta))$ is *Pareto efficient* if and only if condition$\frac{\tau'(\theta) f(\theta)}{h' y(\theta)} + \int_\theta^{\bar{\theta}} \frac{1}{u'(c(\tilde{\theta}))} f(\tilde{\theta}) d\tilde{\theta} \leq 0$ holds, where $\tau(\theta) = t'(y(\theta))$.

Proof: Now, we define $\tilde{h}\,(\theta) = h(\tilde{y}(\theta))$ and we will write the planning problem as:

$$\max_{\underline{\tilde{v}}, \tilde{h}} \int \left(h^{-1}\left(\tilde{h}(\theta)\right) - u^{-1}\left(\underline{\tilde{v}} - \int_{\underline{\theta}}^{\theta} \tilde{h}(z) dz + \theta \tilde{h}(\theta)\right)\right) dF(\theta) \quad (63)$$

Subject to : $\underline{\tilde{v}} - \int_{\underline{\theta}}^{\theta} \tilde{h}(z) dz - v(\theta) \geq 0$

And $\tilde{h}(\theta) \in ni(\Theta)$, where $ni(\Theta)$ is the set of non-increasing real-valued functions over Θ .This is a convex optimization problem the objective to be maximized is concave and the constraints are linear (convex). Now, $ni(\Theta)$ is a closed convex cone, closed under multiplication by positive scalars in the linear space of bounded functions $\mathcal{B}(\Theta)$ endowed with the supremum norm. Previous constraint $\underline{\tilde{v}} - \int_{\underline{\theta}}^{\theta} \tilde{h}(z) dz - v(\theta) \geq 0$ can be expressed as: $G(\tilde{h}) \in P$, where the mapping $G: ni(\Theta) \to c(\Theta)$ is convex, and P is the positive cone of the $c(\Theta)$.Previous constraint $\underline{\tilde{v}} - \int_{\underline{\theta}}^{\theta} \tilde{h}(z) dz - v(\theta) \geq 0$ allows for an interior point $\forall \underline{\tilde{v}} >$

$v(\underline{\theta}); \tilde{h}(\theta) = h(\theta) = h(\tilde{y}(\theta))$. All the conditions required in Luenberger (1969) are met and maximizing Lagrangian is sufficient and necessary for optimality. The Lagrangian here is:

$$\mathcal{L} = \int \left(h^{-1}(\tilde{h}(\theta) - u^{-1}\left(\underline{v} - \int_{\underline{\theta}}^{\theta} h(z)dz + \theta\tilde{h}(\theta)\right)\right) dF(\theta) + \int \left(\underline{v} - \int_{\underline{\theta}}^{\theta} \tilde{h}(z)dz - v(\theta)\right) d\lambda(\theta) \quad (64)$$

For some nondecreasing function $\lambda(\theta)$, the multiplier on the inequality $\underline{\tilde{v}} - \int_{\underline{\theta}}^{\theta} \tilde{h}(z)dz - v(\theta) \geq 0$, is normalized so that $\lambda(\bar{\theta}) = 0$. Fréchet derivative[10] is given by the following:

$$\partial\mathcal{L}(h; \Delta_{\underline{v}}; \Delta_{\tilde{h}}) = \int \left((h^{-1})' \left(h(\theta)\right)\Delta_{\tilde{h}}(\theta) - (u^{-1})'(u(\theta)) \left(\Delta_{\underline{v}}(\theta) + \theta\Delta_{\tilde{h}}(\theta)\right)\right) + \int \Delta_{\tilde{v}}(\theta) d\lambda(\theta) \quad (65)$$

Where in previous: $\Delta_{\tilde{v}}(\theta) = \Delta_{\underline{v}} - \int_{\underline{\theta}}^{\theta} \Delta_{\tilde{h}}(z)dz$.Where z is the function of earnings. Now by substituting for $\Delta_{\tilde{v}}(\theta)$ and by integration by parts we get:

$$\partial\mathcal{L}(h; \Delta_{\underline{v}}; \Delta_{\tilde{h}}) = \int \left((h^{-1})' \left(h(\theta)\right) - (u^{-1})'(u(\theta)\theta) f(\theta)\Delta_{\tilde{h}}(\theta)d\theta\right) + \int \left(\int_{\underline{\theta}}^{\theta} (u^{-1})'(u(z)) f(z)dz\right) \Delta_{\tilde{h}}(\theta)d\theta + \int \lambda(\theta)\Delta_{\tilde{h}}(\theta)d\theta - \Delta_{\underline{v}}(\lambda(\underline{\theta}) + \int (u^{-1})'(u(\theta) f(\theta)d\theta) \quad (65)$$

By collecting the terms we get :

$$\partial\mathcal{L}(h; \Delta_{\underline{v}}; \Delta_{\tilde{h}}) = \int \mathcal{A}(\theta)\Delta_{\tilde{h}}(\theta)d\theta = \Delta_{\tilde{h}}(\underline{\theta}) \int_{\underline{\theta}}^{\theta} A(z)dz + \int \int_{\underline{\theta}}^{\theta} A(z)dz \, d\Delta_{\tilde{h}}(\theta) \quad (66)$$

Where: $\mathcal{A}(\theta) = \left(\left((h^{-1})' \left(h(\theta)\right) - (u^{-1})'(u(\theta)\theta) f(\theta)\right)\right) + \int_{\underline{\theta}}^{\theta} (u^{-1})'(u(z)) f(\theta)dz + \lambda(\theta)$ and $\mathcal{L}(h; \Delta_{\underline{v}}; \Delta_{\tilde{h}})$ is convex, and the necessary and sufficient conditions for $\tilde{h}(\theta) \in ni(\Theta)$ to be maximized are :

$$\partial\mathcal{L}(h; \Delta_{\underline{v}}; \Delta_{\tilde{h}}) \geq 0 \,; \forall \Delta_{\tilde{h}} \in ni(\Theta); \partial\mathcal{L}(h; \underline{v}; h) = 0 \quad (67)$$

Lemma 1. (optimality and FOC's to allow for Gateaux differentials[11] instead of Frechet derivatives Amador, Werning, and Angeletos (2006)) .Let f be a concave functional on P a convex cone in X.Take $x_0 \in P$ and define $h\ (x0) \equiv \{h : h = x - x0 \ and\ x \in P\}$.Then, $\exists \delta f\ (x_0, h)$ for $h \in h\ (x_0)$. Assume that $\exists \delta f\ (x_0, \alpha_1 h_1 + \alpha_2 h_2)$ for $h_1, h_2 \in h\ (x_0)$, and $\delta f\ (x_0, \alpha_1 h_1 + \alpha_2 h_2) = \alpha_1 \delta f\ (x_0, h_1) + \alpha_2 \delta f\ (x_0, h_2)$ for $\alpha_1, \alpha_2 \in R$. A necessary condition for $x_0 \in P$ to maximize f is that: $\delta f\ (x_0, x) \leq 0\ \forall\ x \in P\,; \delta f\ (x_0, x_0) = 0$. Thus, we obtain that a necessary and sufficient condition for the Lagrangian to be maximized at (u_0, w_0) over Φ and that is:

$$\partial\mathcal{L}\left(\underline{w_0}; u_0; \underline{w_0}; u_0 \middle| \Lambda_0\right) = 0 \quad (68)$$

$$\partial\mathcal{L}\left(\underline{w_0}; u_0; h_{\underline{w}}, h_u \middle| \Lambda_0\right) \leq 0 \,; \forall (h_{\underline{w}}, h_u) \in \Phi \quad (69)$$

[10] It is commonly used to generalize the derivative of a real-valued function of a single real variable to the case of a vector-valued function of multiple real variables, and to define the functional derivative used widely in the calculus of variations.

[11] Gateaux differential or Gateaux derivative is a generalization of the concept of directional derivative in differential calculus. Like the Fréchet derivative on a Banach space, the Gateaux differential is often used to formalize the functional derivative

Since $\Delta_{\underline{v}} \lesseqgtr 0$ we obtain that: $\lambda(\underline{\theta}) + \int (u^{-1})'(u(\theta)\theta) f(\theta) = 0$
Because $\Delta_{\tilde{h}}(\theta) \lesseqgtr 0$ and $\Delta_{\tilde{h}} > 0$ it follows that we must have :
$\int_{\underline{\theta}}^{\theta} \mathcal{A}(z)dz = 0$; $\int_{\underline{\theta}}^{\theta} \mathcal{A}(z)dz \leq 0$
From $\partial\mathcal{L}(h; \underline{v}; h) = 0$, if the original $h(\theta) = h(y(\theta))$ strictly increasing near in neighborhood it follows that:

$$\int_{\underline{\theta}}^{\theta} \mathcal{A}(z)dz \leq 0 \Rightarrow \mathcal{A}(\theta) = 0 \qquad (70)$$

In addition we must have $\lambda(\theta)$, and by using the fact $h^{-1}(\tilde{h}(\theta) - (u^{-1})' \left(\underline{v} - u(\theta)\right)\theta = \tau(\theta)/h'(y(\theta))$ and that $(u^{-1})'(u(\theta)) = e_v(v(\theta), y(\theta), \theta)$ we obtain that :

$$-\lambda(\theta) = \frac{\tau(\theta) f(\theta)}{h'(y(\theta))} + \int_{\theta}^{\overline{\theta}} e_v(v(z), y(z), z) f(z)dz \qquad (71)$$

And previous expression is decreasing, by differentiation of this expressions and setting $-\lambda'(\theta) \leq 0$ gives $\tau(\theta)\left(\frac{d\log\tau(\theta)}{d\log(\theta)} + \frac{d\log f(\theta)}{d\log\theta} - \frac{d\log h'(y(\theta))}{d\log\theta}\right) \leq 1 - \tau(\theta)$ ■.
Now, if $F(\theta(y)) = 1 - G(y)$ which implies that $\frac{d\log\theta(y)}{d\log y} = -\varepsilon_{\theta,y}$ and $\theta'(y) < 0$. Where $\varepsilon_{\theta,y}$ is the elasticity of $\theta(y)$ with respect to y .

$$\varepsilon_{\theta,y} \equiv \left|\frac{y\theta'(y)}{\theta(y)}\right| = -\frac{d\log(1-t'(y))}{d\log y} - \frac{d\log u'(y-t'(y))}{d\log y} + \frac{d\log h'(y)}{d\log y} \qquad (72)$$

And $f(\theta(y)) = -\frac{g(y)}{\theta'(y)}$

$$-\frac{d\log f(\theta(y))}{d\log\theta}\varepsilon_{\theta,y} = \frac{d\log g(y)}{d\log y} - \frac{d\log -\theta'(y)}{d\log y} + 1 - \varepsilon_{\theta,y} - \frac{d\log\varepsilon_{\theta,y}}{d\log y} \qquad (73)$$

And multiplying $\tau(\theta)\left(\frac{d\log\tau(\theta)}{d\log(\theta)} + \frac{d\log f(\theta)}{d\log\theta} - \frac{d\log h'(y(\theta))}{d\log\theta}\right) \leq 1 - \tau(\theta)$ by $\varepsilon_{\theta,y}$ and by substituting this last expression:

$$-\frac{d\log(1-t'(y))}{d\log y} - \frac{t'(y)}{(1-t'(y))}\left(\frac{d\log g(y)}{d\log y} - \frac{d\log h'(y(\theta))}{d\log\theta} + 1 - \varepsilon_{\theta,y} - \frac{d\log\varepsilon_{\theta,y}}{d\log y}\right) \leq \varepsilon_{\theta,y} \qquad (74)$$

By rearrangement this gives:

$$t'(y)\left(-\frac{d\log(1-t'(y))}{d\log y} + \frac{d\log h'(y(\theta))}{d\log\theta} - \frac{d\log g(y)}{d\log y} - 1 + \frac{d\log\varepsilon_{\theta,y}}{d\log y}\right) \leq$$
$$-2\frac{d\log(1-t'(y))}{d\log y} + \frac{d\log h'(y(\theta))}{d\log\theta} - \frac{d\log u'(y-t'(y))}{d\log y} \qquad (75)$$

Now one extension flat tax rate. We are assuming power utility function given as: $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$ and $h(y) = \alpha y^{\eta}$ and we are supposing top tax rate:

$$\bar{\tau} \equiv \lim_{\theta\to 0}\tau(\theta) = \lim_{y\to\infty} t'(y) < 1 \qquad (76)$$

$$\lim_{y\to\infty}\frac{d\log(1-t'(y))}{d\log y} = 0\,;\ \frac{d\log\varepsilon_{\theta,y}}{d\log y} = 0 \qquad (77)$$

For high income consumption becomes proportional to income:

$$\lim_{y\to\infty}\frac{d\log(1-t'(y))}{d\log y} = -\sigma \text{ and } \lim_{y\to\infty}\frac{d\log h'(y(\theta))}{d\log\theta} = \eta - 1 \qquad (78)$$

Now by substituting these expressions in $t'(y)\left(-\frac{d\log\left(1-t'(y)\right)}{d\log y}+\frac{d\log h'(y(\theta))}{d\log\theta}-\frac{d\log g(y)}{d\log y}-1+\frac{d\log\varepsilon_{\theta,y}}{d\log y}\right)\leq -2\frac{d\log\left(1-t'(y)\right)}{d\log y}+\frac{d\log h'(y(\theta))}{d\log\theta}-\frac{d\log u'\left(y-t'(y)\right)}{d\log y}$ gives:

$$\bar{\tau}\leq\frac{\sigma+\eta-1}{\varphi+\eta-2} \tag{79}$$

Where $\varphi=-\lim_{y\to\infty}\frac{d\log g(y)}{d\log y}$, the value $\varphi-1>0$ to ensure that income has finite mean, and it is called asymptotic Pareto distribution parameter. The Pareto distribution had a density that is a power function $g(y)=\mathcal{A}y^{-(\varphi)}$, so that these holds: $\frac{d\log g(y)}{d\log y}=-\varphi$ .In $\bar{\tau}\leq\frac{\sigma+\eta-1}{\varphi+\eta-2}$ if $\varphi\approx 3$ as per Saez (2001) , then $\sigma<2$ and σ cannot be interpreted as risk aversion but as control variable[12] for controlling the income and substitution effects for labor.Now in a case of flat tax $t(y)=\bar{\tau}(y)$ for a flat tax rate following result is yielded:

$$\tau(y)=\frac{\sigma+\eta-1}{-\frac{d\log g(y)}{d\log y}++\eta-2} \tag{80}$$

Now if we assume and transfers t_0 and that $t(y)=\bar{\tau}(y)-t_0$ where $t_0>0$ we get :

$$-\frac{d\log u'(y-t(y))}{d\log y}=-\sigma\frac{1-t'(y)}{1-\left(\frac{t(y)}{y}\right)}=\sigma\frac{1-\bar{\tau}}{1-\bar{\tau}+\frac{t_0}{y}}\leq\sigma \tag{81}$$

Which goes $-\frac{d\log u'(y-t(y))}{d\log y}\in(0,\sigma)$ for $y\in(0,\infty)$. So that $\frac{d\log\varepsilon_{\theta,y}}{d\log y}\geq 0$, aditionaly:

$$\frac{dlog}{d\log y}\left(-\frac{d\log u'(y-t(y))}{d\log y}\right)=\frac{dlog}{d\log y}\left(\frac{1-\bar{\tau}}{1-\bar{\tau}+\frac{t_0}{y}}\right)=\frac{\frac{t_0}{y}}{1-\bar{\tau}+t_0/y}\leq 1 \tag{82}$$

Which implies that :

$$\frac{d\log\varepsilon_{\theta,y}}{d\log y}\leq\frac{\sigma}{\sigma+\eta-1} \tag{83}$$

And sufficient condition for $t'(y)\left(-\frac{d\log\left(1-t'(y)\right)}{d\log y}+\frac{d\log h'(y(\theta))}{d\log\theta}-\frac{d\log g(y)}{d\log y}-1+\frac{d\log\varepsilon_{\theta,y}}{d\log y}\right)\leq -2\frac{d\log\left(1-t'(y)\right)}{d\log y}+\frac{d\log h'(y(\theta))}{d\log\theta}-\frac{d\log u'\left(y-t'(y)\right)}{d\log y}$ to hold is :

$$\bar{\tau}<\frac{\eta-1}{-\frac{d\log g(y)}{d\log y}+\eta-2+\frac{\sigma}{\sigma+\eta-1}}<\frac{\eta-1}{-\frac{d\log g(y)}{d\log y}+\eta-1} \tag{84}$$

3.Golosov et al.(2016) framework: heterogenous preferences

This economy is described by $t+1$ periods denoted by $t=0,1,\ldots,t+1$.Agents preferences are described by a time separable utility function over consumption c_t and labor l_t , and discount factor $\beta\in(0,1)$, and expectation operator in period $t=1$, E_0 and utility function $u:\mathbb{R}_+^2\to\mathbb{R}$. Where ; $E_0\sum_{t=0}^{t+1}\beta^t(c_t,l_t)$.In period $t=0$

[12] A control variable (or scientific constant) in scientific experimentation is an experimental element which is constant and unchanged throughout the course of the investigation. Control variables could strongly influence experimental results, were they not held constant during the experiment in order to test the relative relationship of the dependent and independent variables. The control variables themselves are not of primary interest to the experimenter.

agent skills are θ_0 and the distribution of those skills is $F(\theta_0)$.In period $t+1\,; t \geq 1$ skills follow Markov process $F_t(\theta_t|\theta_{t-1})$,where θ_{t-1} represents skill realization, and PDF is $f_t(\theta_t|\theta_{t-1})$ .People retire at period $\hat{t}$ in which case $F_t(0|\theta)=1$ $\forall t, \wedge \forall t \geq \hat{t}$.

Assumption 1. $\forall t \geq \hat{t}$,pdf is differentiable with $f_t' \equiv \frac{\partial f_t}{\partial \theta}$ and $f_{2,t}' \equiv \frac{\partial f_t}{\partial \theta_{t-1}}$,where $\forall \theta_{t-1}$, where $\psi(\theta|\theta_{t-1}) = \frac{\theta_{t-1}\int_\theta^\infty \frac{\partial f_t}{\partial \theta_{t-1}}(x|\theta_{t-1})dx}{\theta f_t(\theta|\theta_{t-1})}$, is bounded one sided $|_\theta^\infty$ $\forall \theta$ and this limit is finite :$\lim_{\theta\to\infty} \frac{1-F_t(\theta|\theta_{(t-1)})}{\theta f_t(\theta|\theta_{t-1})}$.

If previous process is AR(1) then ψ is equal to autocorrelation of the shock process $\forall \theta$. Skills are non-negative $\theta_t \in \Theta = \mathbb{R}^+, \forall t$. Agent types are also persistent like in Hellwig (2021) :

$$\Theta(\theta|\theta_{t-1}) = \frac{\frac{\partial f_t(\theta_t|\theta_{t-1})}{\partial \theta_{t-1}}}{f_t(\theta|\theta_{t-1})} \tag{85}$$

Where $\frac{\partial f_t(\theta_t|\theta_{t-1})}{\partial \theta_{t-1}} = -\rho \frac{\partial f_t(\theta_t|\theta_{t-1})}{\partial \theta_t}$,when $\rho = 0$, θ_t is i.i.d. and when $\rho = 1$ θ_t is random walk with persistence.

Assumption 2. Single crossing condition strictly decreasing:$\frac{u_{c\theta}}{u_c} - \frac{u_{y\theta}}{u_y} > 0$

Where y are the earnings of the agent. Social planer evaluates welfare by Pareto weights $\alpha: \Theta \to \mathbb{R}_+$ .Then α is normalized to 1 $\int_0^\infty \alpha(\theta)dF_0(\theta) = 1$ Social welfare is given by:

$$SWF = \int_0^\infty \alpha(\theta)(E_0 \textstyle\sum_{t=0}^{t+1} \beta^t (c_t, l_t))dF_0(\theta) \tag{86}$$

Assumption 3. u is continuous and twice differentiable in both arg. and satisfies $u_c > 0; u_l < 0; u_{cc} \leq 0; u_{ll} \leq 0$, and $\frac{\partial}{\partial \theta}\frac{u_y\left(c,\frac{y}{\theta}\right)}{u_c\left(c,\frac{y}{\theta}\right)}$ .There the optimal allocation solve mechanism design problem as in Golosov, Kocherlakota, Tsyvinski (2003):

$$\max_{c_{t(\theta_t)}, y_{t(\theta_t)}; \theta_t \in \Theta_t; t \in (0,\hat{t})} \int_0^\infty \alpha(\theta)\left(E_0\left(\textstyle\sum_{t=0}^{t+1} \beta^t \left(c_t(\theta_t), \frac{y_t(\theta_t)}{\theta_t}\right)|\theta_t\right)\right)dF_0(\theta) \tag{87}$$

s.t. IC (incentive compatibility) constraint :

$$E_0\left(\textstyle\sum_{t=0}^{t+1} \beta^t u\left(c_t(\theta_t), \frac{y_t(\theta_t)}{\theta_t}\right)|\theta_t\right) \geq$$
$$E_0\left(\textstyle\sum_{t=0}^{t+1} \beta^t u\left(c_t\left(\sigma^t(\theta_t)\right), \frac{y_t(\sigma^t((\theta_t)))}{\theta_t}\right)|\theta_t\right), \forall \sigma\hat{t} \in \textstyle\sum^t, \sigma^t \in \sigma^{\hat{t}}, \theta \in \Theta \tag{88}$$

and feasibility constraint:

$$\int_0^\infty E_0\left\{\textstyle\sum_{t=0}^{\hat{t}} R^{-t} c_t(\theta_t)|\theta_t\right\}dF_0(\theta) \leq \int_0^\infty E_0\left\{\textstyle\sum_{t=0}^{\hat{t}} R^{-t} y_t(\theta_t)|\theta_t\right\}dF_0(\theta) \tag{89}$$

Now, $\omega(\hat{\theta}, \theta)$ is state variable following Fernandes,Phelan (2000). Dynamic generalization of Envelope condition of Mirrlees (1971) and Milgrom and Segal (2002) , Kapicka (2013), Williams (2011), Pavan, Segal and Toikka (2014).So now we have:

$$\begin{cases} u\left(c_t(\theta_t), \frac{y_t(\theta_t)}{\theta_t}\right) + \beta\omega_{t+1}(\theta_{t-1}\,|\theta_t) \geq u\left(c_t(\theta_{t-1}, \hat{\theta}), \frac{y_t(\theta_{t-1},\hat{\theta})}{\theta_t}\right) + \beta\omega_{t+1}(\theta_{t-1}, \hat{\theta}|\theta_t), \forall \hat{\theta}, \theta \in \Theta, \forall t, \\ \omega_{t+1}(\theta_{t-1}, \hat{\theta}|\theta_t) = E_t\left\{\sum_{s=t+1}^{\hat{t}} \beta^{s-t-1} u\,(c_s, (\hat{\theta}_s), \frac{y_s(\hat{\theta}_s)}{\theta_s}|\theta_t\right\} \end{cases}$$

(90)

First and second derivative of utility are: $w(\theta) = \omega(\theta|\theta)$ and $w_2(\theta) = \omega_2(\theta|\theta)$.The value function takes form of:

$$\begin{cases} V_t\left(\widehat{w},\widehat{w}_2,\underline{\theta}\right)=\min\limits_{c,y,w,w_2}\int_0^\infty\left(c(\theta)-y(\theta)+\frac{1}{R}V_{t+1}(w(\theta),w_2(\theta),\theta)\right)f_t(\theta|\underline{\theta})\,d\theta, s.t. \\ \dot{u}(\theta)=u_\theta\left(c(\theta),\frac{y(\theta)}{\theta}\right)+\beta w_2(\theta), \widehat{w}=\int_0^\infty u(\theta)f_t\left(\theta|\underline{\theta}\right)d\theta\,, \widehat{w}_2=\int_0^\infty u(\theta)f_{2,t}\left(\theta|\underline{\theta}\right)d\theta \\ u(\theta)=u\left(c(\theta),\frac{y(\theta)}{\theta}\right)+\beta w(\theta) \end{cases} \quad (91)$$

Labor $\left(1-\tau_t^y(\theta_t)\right)$ and savings distortions $(1-\tau_t^s(\theta_t))$ are defined as:

$$1-\tau_t^y(\theta_t)\equiv\frac{-u_l\left(c_t(\theta_t),\frac{y_t(\theta_t)}{\theta_t}\right)}{\theta_t u_c\left(c_t(\theta_t),\frac{y_t(\theta_t)}{\theta_t}\right)};\ 1-\tau_t^s(\theta_t)\equiv\frac{1}{\beta R}\frac{u_c\left(c_t(\theta_t),\frac{y_t(\theta_t)}{\theta_t}\right)}{E_t\left\{u_c\left(c_{t+1}(\theta_{t+1}),\frac{y_{t+1}(\theta_{t+1})}{\theta_{t+1}}\right)\right\}} \quad (92)$$

In the case of separable preferences, let $\varepsilon_t(\theta)\equiv\frac{u_{ll,t}(\theta)l_t(\theta)}{u_{l,t}(\theta)}$ is the inverse of Frisch elasticity of labor[13] , and $\sigma_t(\theta)\equiv-\frac{u_{cc,t}(\theta)c_t(\theta)}{u_{c,t}(\theta)}$ represents the intertemporal elasticity of substitution. Preferences are isoelastic: $u(c,l)=\frac{c^{1-\sigma}-1}{1-\sigma}-\frac{l^{1+\varepsilon}}{1+\varepsilon}$.Optimal tax rate here is:

$$\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta_t)}=(1+\varepsilon)\frac{1-F_0(\theta)}{\theta f_0(\theta)}\int_0^\infty\exp\left(\int_0^x\sigma_t(\tilde{x})\frac{\dot{c}(\tilde{x})}{c_t(\tilde{x})}d\tilde{x}\right)\left(1-\lambda_{1,t}\bar{\alpha}_t(x)u_{c,t}(x)\right)+\beta R\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta_t)}\frac{A_t(\theta)}{A_{t-1}}\frac{u_{c,t}(\theta)}{u_{c,t-1}}\psi_t(\theta)\,, t>0 \quad (93)$$

In previous expression: $A_t(\theta)=(1+\varepsilon)$; $B_t(\theta)=\frac{1-F_0(\theta)}{\theta f_0(\theta)}$; $C_t(\theta)=\left(\int_0^x\sigma_t(\tilde{x})\frac{\dot{c}(\tilde{x})}{c_t(\tilde{x})}d\tilde{x}\right)\left(1-\lambda_{1,t}\bar{\alpha}_t(x)u_{c,t}(x)\right)$; $D_t(\theta)=\frac{A_t(\theta)}{A_{t-1}}\frac{u_{c,t}(\theta)}{u_{c,t-1}}\psi_t(\theta)$ where also : $\lambda_{1,t}=\int_0^\infty\frac{f_t(x)}{u_{c,t}(x)}dx$; $\bar{\alpha}_t(\theta)=\alpha(\theta)\ if\ t=0; \bar{\alpha}_t(\theta)=1\ if\ t>0$ .In a case when $\sigma=0$ and $t=0$ previous optimal labor tax becomes:

$$\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta_t)}=(1+\varepsilon)\frac{1-F_0(\theta)}{\theta f_0(\theta)}\int_0^\infty\left(1-\alpha(x)\right)\frac{f_0(x)dx}{1-F_0(\theta)} \quad (94)$$

And if $t>0$ then previous intratemporal components will be equal to zero ($A_t(\theta)=B_t(\theta)=C_t(\theta)=0$ and optimal marginal tax rate will be qual to intertemporal component

$$\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta_t)}=\beta R\rho\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta_t)} \quad (95)$$

In the case of nonseparable preferences between labor and consumption almost all principles as in the case with separable preferences hold,$\gamma_t(\theta)\equiv\frac{u_{c,l,t}(\theta)l_t(\theta)}{u_{c,t}(\theta)}$ represents the degree of complementarity between consumption and labor, and the MPC from after-tax income on the right upper tail of the distribution $\bar{x}=\lim\limits_{\theta\to\infty}\frac{c_t(\theta)}{(1-\tau_t^y(\theta)y_t(\theta)}$.Labor distortions are :

[13] The Frisch elasticity measures the relative change of working hours to 1% increase in real wage given the marginal utility of wealth $\boldsymbol{\lambda}$.In the steady state benchmark model is given as: $\frac{\frac{dh}{h}}{dw/w}=\frac{1-h}{h}\left(\frac{1-\eta}{\eta}\boldsymbol{\theta}-\mathbf{1}\right)^{-1}$

$$\begin{cases} A_t(\theta) = (1 + \varepsilon(\theta) - \gamma_t(\theta)) \\ \mathrm{C_t(\theta)} = \int_\theta^\infty \exp\left(\int_\theta^x \left[\sigma_t(\tilde{x}) \frac{\dot{c}(\tilde{x})}{c_t(\tilde{x})} - \gamma_t(\tilde{x}) \frac{\dot{y}_t(x)}{y_t(\tilde{x})}\right] d\tilde{x}\ \right)\left(1 - \lambda_{1,t}\bar{\alpha}_t(x)u_{c,t}(x)\right)\frac{f_t(x)dx}{1-F_t(\theta)} \\ D_t(\theta) = \frac{A_t(\theta)}{A_{t-1}}\frac{u_{c,t}(\theta)}{u_{c,t-1}}\frac{\theta_{t-1}\ \int_\theta^\infty \exp(-\int_\theta^x \gamma_t(\tilde{x})\frac{d\tilde{x}}{\tilde{x}})\ f_{2,t}*x)dx}{\theta f_t(\theta)} \end{cases} \quad (96)$$

Now about the income and substitution effects, let $\varepsilon_t^u(\theta), \varepsilon_t^c(\theta)$ be the compensated and uncompensated elasticities and the income effect is $\eta_t(\theta) = \varepsilon_t^u(\theta) - \varepsilon_t^c(\theta)$,now we can rewrite labor distortions $A_t(\theta), C_t(\theta)$:

$$\begin{cases} A_t(\theta) = \frac{1+\varepsilon_t^u(\theta)}{\varepsilon_t^c(\theta)} \\ \mathrm{C_t(\theta)} = \int_\theta^\infty \exp\left(g_t;(x;\theta)\right)\left(1 - \lambda_{1,t}\ \bar{\alpha}_t(x)u_{c,t}(x)\right)\frac{f_t(x)dx}{1-F_t(\theta)} \end{cases} \quad (97)$$

$g_t = \int_\theta^x \left\{\frac{-\eta_t(\tilde{x})}{\varepsilon_t^c(\tilde{x})}\frac{\dot{y}_t}{y_t}\tilde{x} - \sigma_t(\tilde{x})\ \frac{\left(1-\tau_t^y(\tilde{x})\right)\dot{y}_t - \dot{c}_t}{c_t}\tilde{x}\right\} d\tilde{x}$, $A_t(\theta), C_t(\theta)$ are similar in their dependence on $\varepsilon_t^u(\theta), \varepsilon_t^c(\theta)$ as in Saez (2001). Preferences here are given as in Greenwood, Hercowitz,, Huffman (1988): $u(c,l) = \frac{1}{1-v}\left(c - \frac{1}{1+\varepsilon}\, l^{1+\frac{1}{\varepsilon}}\right)$.Labor distortions here are given as:

$$\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta_t)} \sim \left[a\frac{1}{1+\frac{1}{\varepsilon}} - \varepsilon\frac{-\bar{\sigma}(1-\bar{x})}{\bar{x}}\right]^{-1} ;\theta \to \infty \quad (98)$$

3.1 Dynamic Mirrlees taxation: two period example

Government computes allocations subject to IC constraints and then implicit taxes are inferred from the resulting wedges between marginal rates of substitution (MRS) and marginal rates of transformation (MRT). Assumption of the model here are:

1. Workers are heterogenous plus random
2. The government does not observe individual skills, but it knows the distribution of skills *apriori*
3. There are no *apriori* restrictions on fiscal policy *e.g. lump-sum taxes are available -possible
4. Government can commit
5. Preferences are separable between consumption and leisure (government should be able to observe marginal utility of consumption)
6. There is no aggregate uncertainty

Without aggregate uncertainty perfect consumption insurance is possible (everybody gets the same consumption). However, if government cannot observe the skills.Assumptions here are:

1. ∃ continuum of workers who live in 2 period and the maximization problem is

$$\max\ E(u(c_1) + v(n_1) + \beta[u(c_2) + v(n_2)] \quad (99)$$

2. Skills production is :$y = \theta \cdot n$

y represents observable output,θ are skills, n is effort/labor. Furthermore: θ_i is only observed by the agent i at the beginning of period,$\Pi_1(i)$ represents period 1 distribution of skills, and here $\Pi_2(j|i)$ is the conditional distribution of skills 2. Government maximization problem is given as:

$$\max_{\substack{c_1(i),c_2(i)\\y_1(i),y_2(i)}} \sum_i \left\{u(c_1, l_{i,j} + v\left(\frac{y_1(i)}{\theta_1(i)}\right) + \beta \sum_j \left[u(c_2, l_{i,j}) + v\left(\frac{y_2(i)}{\theta_2(i)}\right)\right]\right\} \Pi_2(j|i)\Pi_1(i) \quad \text{s.t.}$$

1) Resource constraint :

$$\sum_i \left\{[c_i, l_{i,j} + \frac{1}{R}\sum_j c_2, l_{ij}\Pi_2(j|i)]\,\Pi_1\,(i)\right\} + G_1 + \frac{1}{R}G_2 \le \sum_i \left[y_1(i) + \frac{1}{R}\sum_j y_2(i,j)\Pi_2(j|i)\right]\Pi_1(i) + Rk_1 \qquad (100)$$

2) Incentive compatibility constraints are given below:

$$u(c_1 l_{i,j}) + v\left(\frac{y_i l_{i,j}}{\theta l_{i,j}}\right) + \beta \sum_i \left[u(c_2, l_{i,j}) + v\left(\frac{y_2(i,j)}{\theta_2(i,j)}\right)\right]\Pi_2(j|i) \ge u\left(c_1 l(i_r) + v\left(\frac{y_i l(i_r)}{\theta(i)}\right) + \beta \sum_j \left(u(c_2(i_r, j_r)) + v\left(\frac{y_2(i_r,j_r)}{\theta_2(i,j)}\right)\Pi_2(j|i)\right)\right. \qquad (101)$$

3. Revelation principle: Government asks what your skill is and allocates consumption plus labor contingent on your answer. So now here we have i_r-which denotes first-period skills report (which depends on realized i) and j_r-which represents the 2nd period skills report (which depends on realized j). Characterization of optimum

Let's consider the following simple variational argument:

1) Fix a 1st period realization i and a hypothetical optimum $c_1^*(i), c_2^*(i)$.
2) Increase 2nd period utility uniformly across 2nd period realizations :$u(\tilde{c}_2(i,j;\Delta) \equiv u(c_2^*(i,j)) + \Delta$
3) Hold total utility constant by decreasing 1st period utility by $\beta\Delta$: $u(\tilde{c}_1(i,j,\Delta)) = u(c_1^*(i)) - \beta\Delta$
4) Note that this variation does not affect IC constraint and only the resource constraint is potentially affected.
5) Therefore, for $c_1^*(i); c_2^*(i)$ to be optimal, $\Delta = 0$ must minimize resources expended on the allocation.

One can express the resource costs of the perturbed allocation as follows:

$$\tilde{c}_i(i;\Delta) + R^{-1}\sum_j \tilde{c_2}\,(i,j,\Delta)\Pi(j|i) = u^{-1}(u(c_1(i) - \beta\Delta) + R^{-1}\sum_j u^{-1}(u(c_2(i,j) + \Delta)\Pi(j|i)$$

FOC evaluated at $\Delta = 0$ is as follows:

$$\frac{1}{u'(c_1(i))} = \frac{1}{\beta R}\sum_j \frac{1}{u'(c_2(i,j))}\Pi_2(j|i) \qquad (102)$$

Previous equation is inverse Euler equation, $\frac{d}{dx}f^{-1}(x) = \frac{1}{f'(x)}$. We outline three cases as follows:

1) Skills observable ⇒ $u'(c_1) = \beta R u'(c_2)$
2) Skills unobservable ⇒ $u'(c_1) = \beta R u'(c_2)$ but not random constant overtimes

3) Skills observable plus random: $\frac{1}{u'(c_1)} = \frac{1}{\beta R} E\left[\frac{1}{u'(c_2)}\right] > \frac{1}{\beta R}\frac{1}{Eu'(c_2)} \Rightarrow$ $u'(c_1(i)) < \beta R\, E[u'(c_2(i,j))] \Rightarrow \tau_k > 0$

Previous is Jensen's inequality. Intuition here is that savings affects incentive to work, so government needs to discourage savings to prevent the flowing deviation by highly-skilled: 1) save more today; 2) work less tomorrow. Some other features of optimal fiscal policy are:

1) On average wealth taxes across individuals are zero ex-ante
2) However, they depend on future labor income-if labor income is below average, your capital tax is positive. If your labor income is above average, then your capital tax is negative.
3) So this tax or this fiscal policy might be regressive for incentive reasons

The fact that the capital tax varies in this regressive way makes investment risky and creates a positive risk premium[14]. This explains how $\tau_k > 0$

4. Numerical solution of linear and nonlinear top-earners marginal tax rates

Here we are utilizing this equation : $\tau = \frac{1-\bar{g}}{1-\bar{g}+e}$. The first column of the table follows realistic scenario with elasticity of range $e = 0.25$, as in Saez et al., (2012) and Chetty, (2012) , and Piketty, Saez (2013) .The second column is with estimates in range $e = 0.5$ which is high range elasticity scenario and a third scenario is $e = 1$ which is well above estimates in the current literature.

Table 1 Linear optimal tax rates per Piketty, Saez (2013)

	$e = 0.25$		$e = 0.5$		$e = 1$	
	$\bar{g}$	τ	$\bar{g}$	τ	$\bar{g}$	τ
Rawlsian revenue maximizing rate	0	0.8	0	0.67	0	0.50
Utilitarian CRRA=1 $u_c = \frac{1}{c}$	0.61	0.61	0.54	0.48	0.44	0.36
Median voter I $\frac{w_{median}}{w_{average}}$	0.7	0.55	0.7	0.38	0.7	0.23
Median voter II $\frac{w_{median}}{w_{average}}$	0.75	0.50	0.75	0.33	0.75	0.20
very low tax country 10%	0.97	0.1	0.94	0.1	0.88	0.1
low tax country 35%	0.87	0.35	0.807	0.35	0.46	0.35
high tax country 50%	0.75	0.5	0.5	0.5	0	0.5

Source: Author's calculation

The first row of table 1 is Rawlsian criterion with $\bar{g} = 0$. The second row is utilitarian criterion with coefficient of risk aversion (CRRA) equal to one and social marginal welfare weights are proportional to $u_c = \frac{1}{c}$ where $c = (1-\tau)w + R$ where R is disposable income. Chetty (2006) proved and showed that $CRRA = 1$ is consistent with empirical labor supply behavior and that is a reasonable benchmark.

[14] The risk premium is the rate of return on an investment over and above the risk-free or guaranteed rate of return. To calculate risk premium, investors must first calculate the estimated return and the risk-free rate of return.

First scenario with $e = 0.25$ shows that revenue maximizing tax rate is 80% which is higher even for the countries with highest marginal tax rate which is around 50%. The optimal tax rate under Utilitarian criterion is 61%. The optimal tax rate for median earner is 55% or 38% under $e = 0.5$ and 36% under $e = 1$. In the examples with very low tax country one can see that a tax rate of 10% is optimal in a situation where $g = 0.97$ i.e. in a country with very low redistributive tastes. A tax rate of 50% would be optimal in a country with $\bar{g} = 0.75$. A high elasticity estimate $e = 0.5$ would generate tax rate of 67% above current rates in every country. The median voter tax rate in such a situation would be 38%, Utilitarian criterion generate tax rate of 48% in this situation. In the unrealistically high elasticity scenario $e = 1$ the revenue maximizing tax rate is 50% which is about the current rate in countries with highest $\frac{Tax}{GDP}$ ratios.

Example 1 with non-linear taxes

Here we are using this exact formula for calculation: : $\bar{\tau} = \frac{1-\bar{g}}{1-\bar{g}+\bar{\varepsilon}^{u}+\bar{\varepsilon}^{c}(\alpha-1)}$,and we get table that consists of three global columns with supposed elasticities (uncompensated) $\varepsilon_u \in (0,0.2,0.5)$ and supposed compensated elasticities $\varepsilon_c \in (0.2,0.5,0.8)$.

Table 2 Non-linear income taxes under different uncompensated and compensated elasticities

	$\varepsilon_u = 0$			$\varepsilon_u = 0.2$			$\varepsilon_u = 0.5$		
$\varepsilon_c =$	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
				$\bar{g} = 0$					
a=1.5	0.91	0.80	0.71	0.77	0.69	0.63	0.63	0.57	0.53
a=2	0.83	0.67	0.56	0.71	0.59	0.50	0.59	0.50	0.43
a=2.5	0.77	0.57	0.45	0.67	0.51	0.42	0.56	0.44	0.37
				$\bar{g} = 0.25$					
a=1.5	0.88	0.75	0.65	0.71	0.63	0.56	0.56	0.50	0.45
a=2	0.79	0.60	0.48	0.65	0.52	0.96	0.52	0.43	0.37
a=2.5	0.71	0.50	0.38	0.60	0.44	0.35	0.48	0.38	0.31
				$\bar{g} = 0.5$					
a=1.5	0.83	0.67	0.56	0.63	0.53	0.45	0.45	0.40	0.36
a=2	0.71	0.50	0.38	0.56	0.42	0.33	0.42	0.33	0.28
a=2.5	0.63	0.40	0.29	0.50	0.34	0.26	0.38	0.29	0.23
				$\bar{g} = 0.75$					
a=1.5	0.71	0.50	0.38	0.45	0.36	0.29	0.29	0.25	0.22
a=2	0.56	0.33	0.24	0.38	0.26	0.20	0.26	0.20	0.16
a=2.5	0.45	0.25	0.17	0.33	0.21	0.15	0.24	0.17	0.13

Source: Author's calculation

Another example with non-linear U-shaped taxes as per Diamond (1998).The formulae that we are using here is :$\tau' = \frac{(e^{-1}+1)(1-g)}{[a+(e^{-1}+1)(1-g)]}$

Table 3 Non-linear income tax rates as per Diamond (1998) and authors own calculations

	$g = 0$			$g = 0.25$			$g = 0.5$			$g = 0.975$		
α=	0.5	1.5	5	0.5	1.5	5	0.5	1.5	5	0.5	1.5	5
e												
0.2	0.92	0.8	0.55	0.90	0.75	0.47	0.86	0.67	0.38	0.23	0.09	0.03
0.5	0.86	0.67	0.38	0.82	0.60	0.31	0.75	0.50	0.23	0.13	0.05	0.01
0.75	0.82	0.61	0.32	0.78	0.54	0.26	0.70	0.44	0.19	0.10	0.03	0.01
1	0.80	0.57	0.29	0.75	0.50	0.23	0.67	0.40	0.17	0.09	0.03	0.01
1.5	0.77	0.53	0.25	0.71	0.45	0.20	0.63	0.36	0.32	0.08	0.03	0.01
2	0.75	0.50	0.23	0.69	0.43	0.18	0.60	0.33	0.13	0.07	0.02	0.01

Source: Author's calculation

Form previous table one can see that highest non-linear income taxes are generated with high tastes for redistribution where $g = 0$ and Pareto shape parameter $\alpha = 0.5$ and with labor elasticity $e = 0.2$.Gnerated tax rates are $\tau \in (0.92,0.88,0.55)$ for Pareto shape parameters $\alpha \in (0.5,1.5,5)$.For the same elasticities and Pareto shape parameters but with very low almost non-existent redistributive tastes generated low tax rates are: $\tau \in (0.23,0.09,0.03)$ respectively. On a very high (unrealistically high) labor elasticities generated are tending to zero $\tau \to 0$ .Next original Mirrlees (1971) paper main result has been simulated.

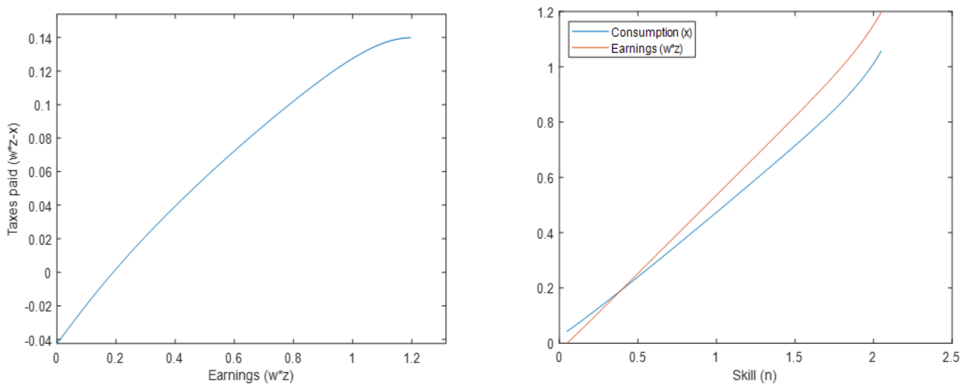

Figure 1 Mirrleesian taxation

Table 4 FOC's for the Mirrlees model

iteration	Func-count	f(x)	Norm of step	First-order optimality
0	3	1.37E-01		
1	6	9.01E-04	0.000224	0.00276
2	9	2.13E-04	2.97E-01	0.000677
3	12	5.02E-08	4.86E-01	9.93E-06
4	15	2.87E-14	6.74E-03	7.50E-09

Table 5 skills, consumption and earnings for the Mirrlees model

F(n)-skills	x-cons.	y-income	x(1-y)	z-earnings
0	0.0424	0	0.0424	0
0.1	0.116	0.3894	0.0708	0.0869

0.5	0.18	0.4382	0.1011	0.1612
0.9	0.2888	0.4686	0.1535	0.2842
0.99	0.4315	0.4841	0.2226	0.4412

Table 6 average and marginal tax rates for Mirrlees model

z-earnings	x-consumption	average tax rate	marginal tax rate
0	0.0424	-Inf	0.2147
0.05	0.0847	-0.54	0.2336
0.1	0.1271	-0.1558	0.2223
0.2	0.214	0.0273	0.1993
0.3	0.3031	0.0817	0.1824
0.4	0.3937	0.1052	0.1698
0.5	0.4856	0.1171	0.1599

Optimal mirrleesian taxation is flat for a long range of top incomes >1.

5. Conclusion

Optimal tax rates as this paper shows depend on redistributive tastes of the supposedly benevolent social planers. The marginal social welfare weight on a given individual measures the value that society puts on providing an additional dollar of consumption to this individual.As the numerical solutions in the non-linear optimal tax rates showed that high tax rates are obtained when there unrealistically low uncompensated and compensated elasticities, also the shape parameter of Pareto distribution must be lower. For high tax countries e.g. countries with highest tax burden around 50% the area that provides such high tax rates is where compensated elasticity is between 0.2 and 0.5 and uncompensated elasticity and unrealistically high compensated elasticity between 0.5 and 0.8 but medium redistributive tastes $\bar{g} = 0.5$. Or alternatively, if uncompensated elasticity is high $\varepsilon_u = 0.5$ than also the taste for redistribution must be high e.g. $\bar{g} \in (0{,}0.25)$.For low tax countries the area where those taxes are provided is in high Pareto distribution parameter and very low taste for redistribution. These are very loose results and are conditioned by themselves and their combinations. In turn there is not straightforward solution to the optimal linear or non-linear labor income tax problem. Pareto efficient tax rates differ from those proposed by Mirrlees (1971). In the dynamic Mirrlees approach, when it comes to the result for capital, capital is taxed to provide more efficient labor supply incentives when there is imperfect information (private distributions of ability unknown to other parties) and as a part of optimal insurance scheme against stochastic earning abilities. Intuition here is that savings affects incentive to work, so government needs to discourage savings to prevent the flowing deviation by highly skilled: 1) save more today; 2) work less tomorrow. That was the second model we reviewed and from there some optimal fiscal policy features are:1) On average wealth taxes across individuals are zero ex-ante ;2) However, they depend on future labor income-if labor income is below average, your capital tax is positive. If your labor income is above average, then your capital tax is negative. 3) So, this tax or this fiscal policy might be regressive for incentive reasons. So in general about dynamic Mirrlees approach it can be concluded that: this approach assumes that agents' abilities to earn income are heterogeneous, stochastic, and private information. Tax instruments ex ante are unrestricted. The model solves for the optimal allocations using dynamic mechanism design (subject only to incentive compatibility constraints) and then considers how to implement these allocations using decentralized tax systems, see also Stantcheva (2020).

References

1. Albanesi, S., Sleet, C. (2006). Dynamic Optimal Taxation with Private Information. The Review of Economic Studies, 73(1), 1–30. http://www.jstor.org/stable/3700615
2. Ales,L.Maziero,P.(2008). Accounting for private information. Federal Reserve Bank of Minneapolis, working paper
3. Amador, Manuel, Iván Werning, and George-Marios Angeletos, (2006). Commitment vs. Flexibility. Econometrica. 74 (2): 365–396
4. Atkinson, A.B. and Stiglitz,J. (1980).Lectures on Public Economics, Chap 14-4 New York: McGraw Hill, 1980. (web)
5. Brewer, M., E. Saez, and A. Shephard (2010).Means Testing and Tax Rates on Earnings,in The Mirrlees Review: Reforming the Tax System for the 21st Century, Oxford,University Press, 2010. (web)
6. Chetty, R. (2006). A new method of estimating risk aversion. American Economic Review, 96(5), 1821–1834
7. Chetty,R. (2012). Bounds on elasticities with optimization frictions:A synthesis of micro and macro evidence on labor supply. Econometrica, 80(3), 969–1018
8. Choné, Philippe, and Guy Laroque. (2010). Negative Marginal Tax Rates and Heterogeneity. American Economic Review 100 (5): 2532–47.
9. Diamond, P. (1998). Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates. American Economic Review, 88, 83–95.
10. Diamond,P.,Helms,J. and Mirrlees (1978).Optimal taxation in a stochastic economy, A Cobb-Douglas example, M.I.T. Working Paper no. 217
11. Fernandes, Ana and Phelan, Christopher, (2000),.A Recursive Formulation for Repeated Agency with History Dependence, Journal of Economic Theory, 91, issue 2, p. 223-247.
12. Fleurbaey,Marc, Maniquet François (2018). Optimal Income Taxation Theory and Principles of Fairness. Journal of Economic Literature 2018, 56(3), 1029–1079
13. Golosov, M., Kocherlakota, N., & Tsyvinski, A. (2003). Optimal Indirect and Capital Taxation. The Review of Economic Studies, 70(3), 569–587. http://www.jstor.org/stable/3648601
14. Golosov, M., Troshkin, M.,Tsyvinski.(2016). Redistribution and Social Insurance. The American Economic Review, 106(2), 2016,359–386. http://www.jstor.org/stable/43821455

15. Golosov, Mikhail, Aleh Tsyvinski, and Ivan Werning. (2006). New dynamic public finance: A users guide. NBER Macroeconomics Annual 21:317-363.
16. Golosov,M-Troshkin,M,-Tsyvinsky,A.(2011). Optimal Dynamic Taxes. NBER
17. Greenwood, J., Hercowitz,Z. , Huffman,G.W.:Investment, Capacity Utilization, and the Real Business Cycle. American Economic Review 78 (3),1988, 402–17
18. Gruber, J.Saez,E.(2002). The Elasticity of Taxable Income: Evidence and Implications, Journal of Public Economics 84(1):1-32
19. Hellwig, C.(2021).Static and Dynamic Mirrleesian Taxation with Non-Separable Preferences: A Unified Approach (June 1, 2021). CEPR Discussion Paper No. DP16254, Available at SSRN: https://ssrn.com/abstract=3886723
20. Kapička,(2013). Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach, The Review of Economic Studies, Volume 80, Issue 3, July 2013, Pages 1027–1054
21. Kaplow, L. (2008).The theory of taxation and public economics. Princeton: Princeton University Press.
22. Kocherlakota, Narayana, (2010). The New Dynamic Public Finance, 1 ed., Princeton University Press.
23. Mankiw NG, Weinzierl M, Yagan D.(2009),Optimal Taxation in Theory and Practice. Journal of Economic Perspectives. 2009;23 (4) :147-174.
24. Mas-Colell, A., et al. (1995). Microeconomic Theory, Oxford University Press.
25. Milgrom, P. Segal, I.(2002). Envelope Theorems for Arbitrary Choice Sets. EconometricaVolume70, Issue2 March 2002 Pages 583-601
26. Mirrlees, J. A. (1971). An exploration in the theory of optimal income taxation. Review of Economic Studies,38, 175–208.
27. Mirrlees, J. A. (1976). Optimal tax theory:A synthesis. Journal of Public Economics, 6, 327–358.
28. Mirrlees, J. A. (1986). The theory of optimal taxation. In K. J. Arrow, M. D. Intriligator (Eds.), Handbook of mathematical economics.Vol. 3 (pp. 1197–1249). Amsterdam: North-Holland.
29. Pavan, Alessandro, Ilya Segal, and Juuso Toikka.(2014). Dynamic mechanism design: a Myersonian approach. Econometrica 82, no. 2 (2014): 601–53. http://www.jstor.org/stable/24029270.
30. Piketty, T., Saez,E. (2013), Chapter 7 - Optimal Labor Income Taxation, Elsevier,Volume 5, Pages 391-474, ISSN 1573-4420, ISBN 9780444537591
31. Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva.(2014). Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities. American Economic Journal: Economic Policy 6, no. 1 (February 2014): 230–71.
32. Saez, E. ,S. Stantcheva (2016). Generalized social marginal welfare weights for optimal tax theory. The American Economic Review, 106(1), pp.24-45.

33. Saez, E., Slemrod, J., Giertz, S. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. Journal of Economic Literature, 50(1), 3–50
34. Saez,E.(2001).Using elasticities to derive optimal income tax rates, The Review of Economic Studies, 68(1), pp.205- 229.
35. Sheshinski, Eytan. (1972). The Optimal Linear Income-Tax. Rev. Econ. Studies 39 (3): 297–302
36. Shimer, R.,Werning, I. (2008). Liquidity and Insurance for the Unemployed. The American Economic Review, 98(5), 1922–1942. http://www.jstor.org/stable/29730157
37. Stantcheva,S. (2020). Dynamic Taxation. Annual Review of Economics Vol. 12: 2020, pp.801-831
38. Varian,H.R.(1980).Redistributive taxation as social insurance. Journal of public Economics,14(1), pp.49-68.
39. Werning,I.(2007).Pareto efficient income taxation. NBER Public Economics meeting
40. Williams, N. (2011). Persistent private information. Econometrica, 79(4), 1233–1275. http://www.jstor.org/stable/41237860