DUSHKO JOSHESKI*
TATJANA BOSHKOV*

336.221-022.233:336.225.53]:303.725.36
(Original scientific paper)

MIRRLEESIAN OPTIMAL TAXATION: REVIEW OF THE STATIC AND DYNAMIC FRAMEWORK

Abstract: Modern tax theory has been subject of investigation of this paper. Namely static taxation theory with issues such as: income effects, inverse elasticity rule, and linear and non-linear taxa formulas. The issue of heterogeneity is present in the dynamic taxation models and inverse Euler equation where savings affects incentive to work, so government needs to discourage savings to prevent the flowing deviation by highly skilled to save more today and work less tomorrow. Other important implication for fiscal policy is that if your labor income is below average, your capital tax is positive. Numerical results at the end have confirmed results from theoretical models in optimal taxation outlined in this paper.

Keywords: Optimal taxation, non-linear tax rates, inverse Euler equation, Mirrlees taxation

JEL: H20, H21

* Associate Professor, Ph.D., Goce Delchev University – Stip, Republic of North Macedonia, dusko.josevski@ugd.edu.mk

* Associate, Ph.D., Goce Delchev University – Stip, Republic of North Macedonia, tatjana.dzaleva@ugd.edu.mk

Introduction

Modern tax theory relies heavily on Mirrlees (1971) idea. This equation is the staple in the public finance field: Mirrlees (1976), Saez (2001), Choné, Laroque (2010), Fleurbaey, Maniquet (2018). Atkinson, Stiglitz, (1980); Kaplow, (2008); Mirrlees (1976), Mirrlees (1986); Stiglitz, (1987); Tuomala, (1990) include rigorous derivations of tax formulas. 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)[1] 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-Troshkin Tsyvinsky (2011). Sizeable literature in NDPF studies optimal taxation in dynamic settings, (Golosov, Kocherlakota, Tsyvinski (2003), Golosov, Tsyvinski, and Werning (2006), Kocherlakota (2010). This paper will review the basis of the optimal tax theory and will show the derivation of the optimal tax formula in linear-and non-linear cases and in dynamic Mirrlees model too.

1.ATKINSON-STIGLITZ THEOREM: COMMODITY TAXATION AS SUPPLEMENTARY TO LABOR TAXATION

The question here is whether governments can increase social welfare by adding differentiated commodity taxation $\tau = (\tau_1, \dots, \tau_k)$ in addition to nonlinear tax on earnings w. Atkinson and Stiglitz (1976) theorem:

Theorem: Commodity taxes cannot increase social welfare if utility functions are weakly separable in consumption goods versus leisure and the subutility of consumption goods is the same across individuals, i.e.,$u_i(c_1, \dots, c_k, w) = u_i(v(c_1, \dots, c_k), w)$ with the subutility function

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

$v(c_1, \ldots, c_k)$ homogenous across individuals. Laroque (2005) and Kaplow (2006) have provided intuitive proof of this theorem.

1.1. Inverse elasticity rule

In Ramsey (1927), utility function is given of the following type: $u = f(p_1, p_2, p_3, \ldots\ldots, w)$, $p_1, p_2, p_3, \ldots\ldots$ are the prices and w is an income. This result is known as Roy's identity, Roy (1947)[2], it is :$\frac{\partial u}{\partial p_i} = -f_i \frac{\partial u}{\partial w}$. With the horizontal demand curves, price of the producers is fixed, change in the goods price is only equal to the change in taxes. Then, $dp_1 = d\tau_1 > 0$, $dp_2 = d\tau_2 < 0$. Change in taxes must satisfy the following equation:

$$dU = \frac{\partial U}{\partial p_i} d\tau_1 + \frac{\partial U}{\partial p_2} d\tau_2 = 0, \frac{d\tau_2}{d\tau_1} = -\frac{F_1}{F_2}, \tag{1}$$

change in the revenues caused by the change in taxes is: $\frac{\partial(\tau_1 f_1)}{\partial t_1} = F_1 + \frac{\tau_1 df}{dp_1} = F\left(1 + \frac{\tau_1 dF_1 p_1}{p_1 dp_1 F_1}\right) = F_1\left(1 - \frac{\tau_1}{p_1}\varepsilon_u^1\right)$, where ε_u^1 represents the compensated elasticity of the demand for good 1. Change of revenues of good 2 is: $\frac{\partial(t_2 F_2)}{\partial t_2} = F_2\left(1 - \frac{\tau_2}{p_2}\varepsilon_u^2\right)$. This identity must hold: $\frac{t_2}{p_2}\varepsilon_u^2 - \frac{t_1}{p_1}\varepsilon_u^1 = 0$, for the linear demand curve results is: $\frac{t}{p} = \frac{kQ}{bp} = \frac{k}{\varepsilon_u^d}$. This conclusion is supported by the findings of Feldstein (1978). Ramsey's model was used in life cycle models, for best reference see Atkinson, A.B. and Stiglitz, J. (1976), Atkinson, A.B. and A. Sandmo (1980), Atkinson, A.B. and Stiglitz,J. (1980).

2.OPTIMAL LINEAR TAX FORMULAE

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

$$\int_0^\infty \tau(w) f(n) dn = \int_0^\infty (w - \alpha - \beta w) f(n) dn = 0 \tag{2}$$

$f(n)$ 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. $0 \leq \beta \leq 1$ so that marginal tax rate is non negative

[2] The lemma relates the ordinary (Marshallian) demand function to the derivatives of the indirect utility function.

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 $\frac{\beta}{\hat{\ell}} \frac{\partial \hat{\ell}}{\partial \beta} \geq \lambda$. 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 :

$$0 = \frac{dSWF}{d\tau} = \int \omega_i G'(u_i) u_c^i \cdot \left((w - w^*) - \tau \frac{dw}{d(1-\tau)}\right) df(i) \tag{3}$$

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:

$$\tau = \frac{1-\bar{g}}{1-\bar{g}+e} \tag{4}$$

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

3.OPTIMAL NON-LINEAR TAX FORMULAE: DERIVATION

Utility function is quasi linear: $u(c,l) = c - v(l)$. c is disposable income and the utility of supply of labor $v(l)$ is increasing and convex in l. Earnings equal $w = nl$ where n represents innate ability. CDF of skills distribution is $F(n)$, its PDF is $f(n)$and support range is $[0, \infty)$. The government cannot observe abilities instead it can set taxes as a function of labor income $c = w - \tau(w)$.Individual n chooses l_n to maximize: $\max(nl - \tau n(l) - v(l))$. When marginal tax rate τ is constant, the labor supply f-ction is given as: $l \rightarrow l(n(1-\tau))$ 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:

$$e = \frac{\left(\frac{n(1-\tau)}{l}\right) dl}{d(n(1-\tau))} = \frac{v'(l)}{l v''(l)} \tag{5}$$

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

$$SWF = \int G(u_n) f(n) dn \;\; s.t. \int cnf(n) dn \leq \int nlnf(n) dn - E(\lambda) \tag{6}$$

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: $\frac{du_n}{dn} = \frac{lnv'(ln)}{n}$. Now the Hamiltonian is given as:

$$\mathcal{H} = \left[G(u_n) + \lambda \cdot \left(nl_n - u_n - v(l_n)\right)\right]f(n) + \phi(n) \cdot \frac{lnv'(ln)}{n} \quad (7)$$

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

$$\lambda \cdot \left(n - v'(l_n)\right) + \frac{\phi(n)}{n} \cdot [v'(l_n) + l_n v''(l_n)] = 0 \quad (8)$$

FOC with respect to u is given as:$-\frac{d\phi(n)}{n} = [G'(u_n) - \lambda]$ .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

$$\frac{[v'(l_n) + l_n v''(l_n)]}{n} = \left[\frac{v'(l_n)}{n}\right]\left[1 + \frac{1}{e}\right] \quad (9)$$

FOC with respect to l_n is:

$$\frac{\tau'(w_n)}{1-\tau'(w_n)} = \left(1 + \frac{1}{e}\right) \cdot \left(\frac{\int_n^\infty (1-g_m)dF(m)}{nf(n)}\right) \quad (10)$$

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

$$\frac{\tau'(w_n)}{1-\tau'(w_n)} = \frac{1}{e} \cdot \left(\frac{\int_n^\infty (1-g_m)dF(m)}{w_n h(w_n)}\right) = \frac{1}{e} \cdot \left(\frac{1-H(w_n)}{w_n h(w_n)}\right) \cdot \left(1 - G(w_n)\right) \quad (11)$$

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

4.GOLOSOV ET AL. (2016) FRAMEWORK: HETEROGENOUS PREFERENCES

This economy is described by $t+1$ periods denoted by $t=0,1,..,t+1$.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 $u\colon \mathbb{R}^2_+ \rightarrow \mathbb{R}$. Where; $E_0 \sum_{t=0}^{t+1} \beta^t(c_t, l_t)$. In period $t=0$ 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 skills 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: $\lim_{\theta \to \infty} \frac{1-F_t\left(\theta|\theta_{(t-1)}\right)}{\theta f_t(\theta|\theta_{t-1})}$.

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) :

$$\Theta(\theta|\theta_{t-1}) = \frac{\frac{\partial f_t(\theta_t|\theta_{t-1})}{\partial \theta_{t-1}}}{f_t(\theta|\theta_{t-1})} \tag{12}$$

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:

$$\frac{u_{c\theta}}{u_c} - \frac{u_{y\theta}}{u_y} > 0 \tag{13}$$

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

$$SWF = \int_0^{\infty} \alpha(\theta)\left(E_0 \sum_{t=0}^{t+1} \beta^t(c_t, l_t)\right) dF_0(\theta) \tag{14}$$

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):

$$\max_{c_{t(\theta_t)}, y_{t(\theta_t)}; \theta_t \in \Theta_t; t \in (0,\hat{t})} \int_0^\infty \alpha(\theta) \left(E_0 \left(\sum_{t=0}^{t+1} \beta^t \left(c_t(\theta_t), \frac{y_t(\theta_t)}{\theta_t} \right) | \theta_t \right) \right) dF_0(\theta) \quad (15)$$

s.t. IC (incentive compatibility) constraint:

$$E_0 \left(\sum_{t=0}^{t+1} \beta^t u \left(c_t(\theta_t), \frac{y_t(\theta_t)}{\theta_t} \right) | \theta_t \right) \geq$$

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

feasibility constraint:

$$\int_0^\infty E_0 \{ \sum_{t=0}^{\hat{t}} R^{-t} c_t(\theta_t) | \theta_t \} dF_0(\theta) \leq \int_0^\infty E_0 \{ \sum_{t=0}^{\hat{t}} R^{-t} y_t(\theta_t) | \theta_t \} dF_0(\theta) \quad (16)$$

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:

$$\begin{cases} u\left(c_t(\theta_t), \frac{y_t(\theta_t)}{\theta_t} \right) + \beta\omega_{t+1}(\theta_{t-1} | \theta_t) \geq u\left(c_t(\theta_{t-1}, \hat{\theta}), \frac{y_t(\theta_{t-1}, \hat{\theta})}{\theta_t} \right) + \beta\omega_{t+1}(\theta_{t-1}, \hat{\theta} | \theta_t), \forall\hat{\theta}, \theta \in \Theta, \forall t, \\ \omega_{t+1}(\theta_{t-1}, \hat{\theta} | \theta_t) = E_t \left\{ \sum_{s=t+1}^{\hat{t}} \beta^{s-t-1} u\ (c_s, (\hat{\theta}_s), \frac{y_s(\hat{\theta}_s)}{\theta_s} | \theta_t \right\} \end{cases} \quad (17)$$

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

$$\begin{cases} V_t(\hat{w}, \hat{w}_2, \underline{\theta}) = \min_{c,y,w,w_2} \int_0^\infty \left(c(\theta) - y(\theta) + \frac{1}{R} V_{t+1}(w(\theta), w_2(\theta), \theta) \right) f_t(\theta|\underline{\theta})\, d\theta, s.t. \\ \dot{u}(\theta) = u_\theta \left(c(\theta), \frac{y(\theta)}{\theta} \right) + \beta w_2(\theta), \hat{w} = \int_0^\infty u(\theta) f_t(\theta|\underline{\theta}) d\theta\, , \hat{w}_2 = \int_0^\infty u(\theta) f_{2,t}(\theta|\underline{\theta}) d\theta \\ u(\theta) = u\left(c(\theta), \frac{y(\theta)}{\theta} \right) + \beta w(\theta) \end{cases} \quad (18)$$

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

$$1 - \tau_t^y(\theta_t) \equiv \frac{-u_l\left(c_t(\theta_t), \frac{y_t(\theta_t)}{\theta_t}\right)}{\theta_t u_c\left(c_t(\theta_t), \frac{y_t(\theta_t)}{\theta_t}\right)};\ 1 - \tau_t^s(\theta_t) \equiv \frac{1}{\beta R} \frac{u_c\left(c_t(\theta_t), \frac{y_t(\theta_t)}{\theta_t}\right)}{E_t\left\{ u_c\left(c_{t+1}(\theta_{t+1}), \frac{y_{t+1}(\theta_{t+1})}{\theta_{t+1}}\right)\right\}} \quad (19)$$

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[3], 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: $u(c,l) = \frac{c^{1-\sigma}-1}{1-\sigma} - \frac{l^{1+\varepsilon}}{1+\varepsilon}$. The optimal tax rate here is:

$$\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta_t)} = (1+\varepsilon)\frac{1-F_0(\theta)}{\theta f_0(\theta)}\int_0^\infty \exp\left(\int_0^x \sigma_t(\tilde{x})\frac{\dot{c}(\tilde{x})}{c_t(\tilde{x})}d\tilde{x}\right)\Big(1 - \lambda_{1,t}\bar{\alpha}_t(x)u_{c,t}(x)\Big) + \beta R\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta_t)}\frac{A_t(\theta)}{A_{t-1}}\frac{u_{c,t}(\theta)}{u_{c,t-1}}\psi_t(\theta)\,, t > 0 \tag{20}$$

In the previous expression: $A_t(\theta) = (1+\varepsilon)$; $B_t(\theta) = \frac{1-F_0(\theta)}{\theta f_0(\theta)}$; $C_t(\theta) = \left(\int_0^x \sigma_t(\tilde{x})\frac{\dot{c}(\tilde{x})}{c_t(\tilde{x})}d\tilde{x}\right)\Big(1-\lambda_{1,t}\bar{\alpha}_t(x)u_{c,t}(x)\Big)$; $D_t(\theta) = \frac{A_t(\theta)}{A_{t-1}}\frac{u_{c,t}(\theta)}{u_{c,t-1}}\psi_t(\theta)$ 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:

$$\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta_t)} = (1+\varepsilon)\frac{1-F_0(\theta)}{\theta f_0(\theta)}\int_0^\infty \big(1-\alpha(x)\big)\frac{f_0(x)dx}{1-F_0(\theta)} \tag{21}$$

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 equal to intertemporal component

$$\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta_t)} = \beta R\rho\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta_t)} \tag{22}$$

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 $\bar{x} = \lim_{\theta\to\infty}\frac{c_t(\theta)}{(1-\tau_t^y(\theta) y_t(\theta)}$. Labor distortions are :

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

$$\begin{cases} A_t(\theta) = \left(1+\varepsilon(\theta)-\gamma_t(\theta)\right) \\ C_t(\theta) = \int_\theta^\infty \exp\left(\int_\theta^x \left[\sigma_t(\tilde{x})\frac{\dot{c}(\tilde{x})}{c_t(\tilde{x})} - \gamma_t(\tilde{x})\frac{\dot{y}_t(x)}{y_t(\tilde{x})}\right] d\tilde{x}\right)\left(1-\lambda_{1,t}\overline{\alpha}_t(x)u_{c,t}(x)\right)\frac{f_t(x)dx}{1-F_t(\theta)} \\ D_t(\theta) = \frac{A_t(\theta)}{A_{t-1}}\frac{u_{c,t}(\theta)}{u_{c,t-1}}\frac{\theta_{t-1}\int_\theta^\infty \exp(-\int_\theta^x \gamma_t(\tilde{x})\frac{d\tilde{x}}{\tilde{x}})f_{2,t}{*}x)dx}{\theta f_t(\theta)} \end{cases} \quad (23)$$

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 $A_t(\theta), C_t(\theta)$:

$$\begin{cases} A_t(\theta) = \frac{1+\varepsilon_t^u(\theta)}{\varepsilon_t^c(\theta)} \\ C_t(\theta) = \int_\theta^\infty \exp\left(g_t;(x;\theta)\right)\left(1-\lambda_{1,t}\,\overline{\alpha}_t(x)u_{c,t}(x)\right)\frac{f_t(x)dx}{1-F_t(\theta)} \end{cases} \quad (24)$$

$g_t = \int_\theta^x \left\{\frac{-\eta_t(\tilde{x})}{\varepsilon_t^c(\tilde{x})}\frac{\dot{y}_t}{y_t}\tilde{x} - \sigma_t(\tilde{x})\,\frac{\left(1-\tau_t^y(\tilde{x})\right)\dot{y}_t-\dot{c}_t}{c_t}\tilde{x}\right\}d\tilde{x}$, $A_t(\theta), C_t(\theta)$ 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:

$$\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta_t)} \sim \left[a\frac{1}{1+\frac{1}{\varepsilon}} - \varepsilon\frac{-\overline{\sigma}(1-\bar{x})}{\bar{x}}\right]^{-1};\theta \to \infty \quad (25)$$

4.1 Dynamic Mirrlees taxation: two period example

The 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). The assumptions 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. The 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 a perfect consumption insurance is possible (everybody gets the same consumption). However, if government cannot observe the skills. The assumptions here are:

1. ∃ continuum of workers who live in the 2nd period and the maximization problem is
2. max $E(u(c_1)+v(n_1)+\beta[u(c_2)+v(n_2)]$
3. The 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:

$$\max_{\substack{c_1(i),c_2(i)\\ y_1(i),y_2(i)}} \sum_i \left\{u(c_1,l_{i,j}+v\left(\frac{y_1(i)}{\theta_1(i)}\right)+\beta\sum_j\left[u(c_2,l_{i,j})+v\left(\frac{y_2(i)}{\theta_2(i)}\right)\right]\right\}\Pi_2(j|i)\Pi_1(i) \quad (26) \text{ s.t.}$$

1) **Resource constraint** :

$$\sum_i\left\{[c_i,l_{i,j}+\frac{1}{R}\sum_j c_2,l_{ij}\Pi_2(j|i)]\,\Pi_1\,(i)\right\}+G_1+\frac{1}{R}G_2\leq\sum_i\left[y_1(i)+\frac{1}{R}\sum_j y_2(i,j)\Pi_2(j|i)\right]\Pi_1(i)+Rk_1 \quad (27)$$

2) **Incentive compatibility constraints** are given below:

$$u(c_1l_{i,j})+v\left(\frac{y_il_{i,j}}{\theta l_{i,j}}\right)+\beta\sum_i\left[u(c_2,l_{i,j})+v\left(\frac{y_2(i,j)}{\theta_2(i,j)}\right)\right]\Pi_2(j|i)\geq u\left(c_1l(i_r)+v\left(\frac{y_il(i_r)}{\theta(i)}\right)+\beta\sum_j\,(u(c_2(i_r,j_r))+v\left(\frac{y_2(i_r,j_r)}{\theta_2(i,j)}\right)\Pi_2(j|i)\right) \quad (28)$$

4. **Revelation principle**: The 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: $u(\tilde{c}_2(i,j;\Delta) \equiv u\big(c_2^*(i,j)\big) + \Delta$
3) Hold total utility constant by decreasing 1st period utility by $\beta\Delta$: $u\big(\tilde{c}_1(i,j,\Delta)\big) = u\big(c_1^*(i)\big) - \beta\Delta$
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:

$$\tilde{c}_i(i;\Delta) + R^{-1}\sum_j \tilde{c_2}\,(i,j,\Delta)\Pi(j|i) = u^{-1}(u(c_1(i) - \beta\Delta) + R^{-1}\sum_j u^{-1}(u(c_2(i,j) + \Delta)\Pi(j|i) \quad (29)$$

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

$$\frac{1}{u'(c_1(i))} = \frac{1}{\beta R}\sum_j \frac{1}{u'(c_2(i,j))}\Pi_2(j|i) \quad (30)$$

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 $\Rightarrow u'(c_1) = \beta R u'(c_2)$
2) Skills unobservable $\Rightarrow u'(c_1) = \beta R u'(c_2)$ but not random constant overtimes
3) Skills observable plus random: $\frac{1}{u'(c_1)} = \frac{1}{\beta R} E\left[\frac{1}{u'(c_2)}\right] > \frac{1}{\beta R}\frac{1}{Eu'(c_2)} \Rightarrow$ $u'\big(c_1(i)\big) < \beta R\, E\big[u'\big(c_2(i,j)\big)\big] \Rightarrow \tau_k > 0$

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[4]. This explains how $\tau_k > 0$

5.NUMERICAL SOLUTIONS TO MIRRLEES STATIC MODEL: GRAPHIC AND TABULAR

In this first example we are using non-linear tax formula : $\bar{\tau} = \frac{1-\bar{g}}{1-\bar{g}+\bar{\varepsilon}^{u}+\bar{\varepsilon}^{c}(\alpha-1)}$,.Table 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 1 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

[4] 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.

				$\bar{g} = 0.5$					
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

Highest optimal tax rates are obtained where there are low uncompensated and compensated utility as well as low Pareto parameter.

Table 2 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. Chetty (2006) proved and showed that $CRRA = 1$ is consistent with empirical labor supply behavior and that is a reasonable benchmark.MATLAB example settings are:

$$\begin{cases} u = \alpha \log x + \log(1 - y) \\ G(u) = -\frac{1}{\beta} e^{-\beta u} \\ f(n) = \frac{1}{n} \exp\left[-\frac{(\log n + 1)^2}{2}\right] \end{cases} \tag{31}$$

Skills are assumed to be lognormally distributed with the average $\bar{n} = \frac{1}{\sqrt{e}} = 0.607$. So now, the equations:

$$\begin{cases} \frac{dv}{dn} = -\frac{v}{n}\left(2 + \frac{nf'}{f}\right) - \frac{1}{n^2 u_1} + \frac{\lambda G'}{n^2} \\ \frac{du}{dn} = -\frac{y u_2}{n} \end{cases} \tag{32}$$

Would become:

$$\begin{cases} \frac{dv}{dn} = -\frac{v \log n}{n} - \frac{x}{\alpha n^2} + \frac{\lambda}{n^2}\, e^{-\beta u} \\ \frac{du}{dn} = \frac{y}{n(1-y)} \end{cases} \tag{33}$$

Figure 1 Mirrleesian taxation

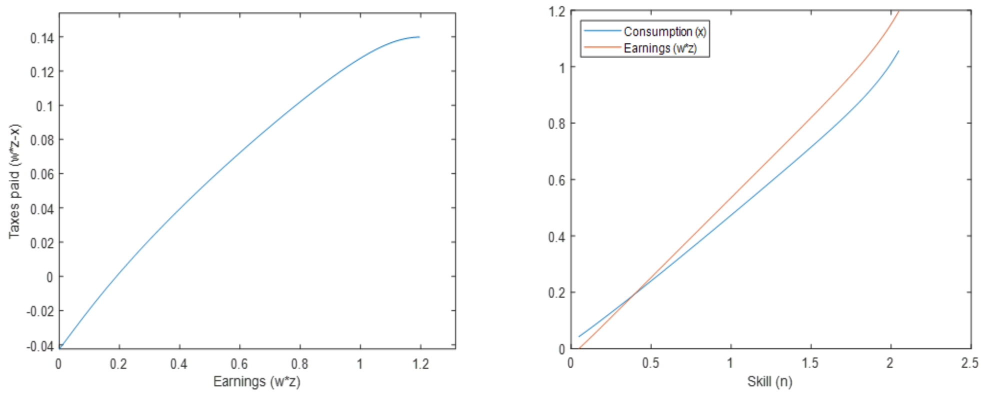

Table 3 FOC's for the Mirrlees model

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 4 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 5 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

The optimal Mirrleesian taxation is flat for a long range of top incomes >1.

Conclusion

In static models with Utilitarian SWF there is not substantial evidence for progressive taxation optimal tax rates and as this paper shows they depend on redistributive tastes of the supposedly benevolent social planers.

The numerical solutions in the non-linear optimal tax rates showed that high tax rates are obtained when there are unrealistically low uncompensated and compensated elasticities, also the shape parameter of Pareto distribution must be lower. For high tax countries, with 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 $\bar{g}=0.5$. If uncompensated elasticity is high $\varepsilon_u=0.5$ then 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. In the dynamic Mirrlees approach, 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. Savings affects incentive to work, so government needs to discourage savings to prevent the flowing deviation by highly skilled. Dynamic Mirrlees 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. Atkinson, A.B. and Sandmo A. : Welfare Implications of the Taxation of Savings, Economic Journal, Vol. 90, 1980, pp.529-49. (web)
2. Atkinson, A.B. and Stiglitz,J: The design of tax structure: Direct versus indirect taxation,Journal of Public Economics, Vol. 6, 1976, pp.55-75. (web)
3. Atkinson, A.B. and Stiglitz,J. : Lectures on Public Economics, Chap 14-4 New York: McGraw Hill, 1980. (web)
4. Chetty, R. A new method of estimating risk aversion. American Economic Review, 96(5), 2006, pp.1821–1834
5. Choné, P., Laroque.G.: Negative Marginal Tax Rates and Heterogeneity. American Economic Review 100 (5), 2010), pp.2532–47.

6. Diamond, P: Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates. American Economic Review, 88, 1998, pp.83–95.
7. Farhi E, Werning I : Insurance and taxation over the life cycle. Rev. Econ. Stud. 80, 2013), pp.596–635
8. Fernandes, A, Phelan, C: A Recursive Formulation for Repeated Agency with History Dependence. Journal of Economic Theory 91 (2), 2000, pp. 223–47
9. Fleurbaey,M., Maniquet F : Income Taxation Theory and Principles of Fairness. Journal of Economic Literature 2018, 56(3), 1029–1079
10. Golosov, M., Troshkin, M.,Tsyvinski, A. :Redistribution and Social Insurance. The American Economic Review, 106(2), 2016,359–386. http://www.jstor.org/stable/43821455
11. Golosov, M., Kocherlakota,N. and Tsyvinski,A: Optimal Indirect and Capital Taxation. Review of Economic Studies 70 (3): 2003,569–87
12. Greenwood, J., Hercowitz,Z. , Huffman,G.W.:Investment, Capacity Utilization, and the Real Business Cycle. American Economic Review 78 (3),1988, 402–17
13. Hellwig, C.: 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
14. Kapička: 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
15. Kaplow, L. The theory of taxation and public economics. Princeton: Princeton University Press. 2008
16. Kaplow,L. On the undesirability of commodity taxation even when income taxation is not optimal. Journal of Public Economics, 90(6–7), 2006, pp.1235–50.
17. Laroque, G. R. : Indirect taxation is superfluous under separability and taste homogeneity:A simple proof. Economics Letters, 87(1), 2005, pp.141–144.
18. Milgrom, P. Segal, I.: Envelope Theorems for Arbitrary Choice Sets. EconometricaVolume70, Issue2 March 2002 Pages 583-601
19. Mirrlees, J. A. : An exploration in the theory of optimal income taxation. Review of Economic Studies,38, 1971, pp.175–208

20. Mirrlees, J. A. : Optimal tax theory:A synthesis. Journal of Public Economics, 6, 1976, Pp.327–358
21. Mirrlees, J. A. 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. 1986
22. Pavan, A., Ilya S., and Juuso T. : Dynamic mechanism design: a Myersonian approach. Econometrica 82, no. 2 (2014): pp.601–53. http://www.jstor.org/stable/24029270.
23. Piketty, T., Saez,E. Chapter 7 - Optimal Labor Income Taxation, Elsevier,Volume 5, Pages 391-474, ISSN 1573-4420, ISBN 9780444537591, 2013
24. Ramsey, F. A Contribution to the Theory of Taxation. Economic Journal. 37: 1927, pp.47 -61.
25. Roy, R: La Distribution du Revenu Entre Les Divers Biens. Econometrica. 15 (3): .1947, 205–225. JSTOR 1905479.
26. Saez,E:Using elasticities to derive optimal income tax rates, The Review of Economic Studies, 68(1),2001,pp.205- 229.
27. Sheshinski, E. : The optimal linear income tax. Review of Economic Studies, 39(3), 1972,297–302.
28. Stiglitz,J.:Pareto efficient and optimal taxation and the new new welfare economics. InA. J.Auerbach, & M. Feldstein (Eds.), Handbook of Public Economics.Vol. 2, (pp. 991–1042). Amsterdam: North-Holland, 1987
29. Tuomala, M. Optimal income tax and redistribution. Oxford: Clarendon Press, 1990
30. Williams, N. Persistent private information. Econometrica, 79(4), 2011, pp.1233–1275. http://www.jstor.org/stable/41237860
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. Diamond, P. and Mirrlees, J., :A model of social insurance with variable retirement, Journal of Public Economics, 10, issue 3, 1978, pp. 295-336.
33. Albanesi, S., Sleet, C. : Dynamic Optimal Taxation with Private Information. The Review of Economic Studies,73(1), 2006, pp.1–30.
34. Shimer, R., & Werning, I. :Liquidity and Insurance for the Unemployed. The American Economic Review, 98(5),2008,1922–1942.

35. Ales,L.Maziero,P. : Accounting for private information. Federal Reserve Bank of Minneapolis, working paper , 2008
36. Golosov, M., Kocherlakota, N., & Tsyvinski, A. : nOptimal Indirect and Capital Taxation. The Review of Economic Studies, 70(3), 2003, 569–587.
37. Golosov, M-Troshkin, M,-Tsyvinsky, A. : Optimal Dynamic Taxes. NBER, 2011)
38. Golosov, M., Tsyvinski, A. Werning.I : New dynamic public finance: A users guide. NBER Macroeconomics Annual 21: 2006, pp.317-363.
39. Kocherlakota, N., : The New Dynamic Public Finance, 1 ed., Princeton University Press. 2010
40. Stantcheva,S. » Dynamic Taxation. Annual Review of Economics Vol. 12: 2020, pp.801-831