# Optimal taxation in theory

**Dushko Josheski** *, **Tatjana Boshkov***

University Goce Delcev-Shtip , R.North Macedonia

Working paper

## Abstract

In this paper optimal income taxation theories are subject of investigation following the classic paper in public finance by Mirrlees (1971), than the models of Sadka (1976), Seade,(1977), Akerlof (1978),Stiglitz (1982), Diamond (1998), and Saez (2001) , Piketty-Saez-Stantcheva (2014), all related to the classic paper by Mirrlees (1971). The problem is to maximize integral over population of the social evaluation of individual utility, that depends on individual consumption and labor. This paper first posed the problem of asymmetric information since the basic idea of the paper is that a first-best redistribution scheme is based on innate ability, and the information about ability is known to the individual, the government observes instead earnings. Mirrlees (1971), provides analytical solutions for the second-best efficient tax system in presence of such an adverse selection. Untill late 1990s, Mirrlees results not closely connected to empirical tax studies and little impact on tax policy recommendations Since late 1990s, Diamond (1998), Saez (2001) have connected Mirrlees model to practical tax policy / empirical tax studies. Mankiw, Weinzierl, and Yagan (2009) provide MATLAB code for analyzing the Mirrlees model MTR and wages, they are using log-normal and Pareto distributions. Later we look up to theory for optimal commodity sales taxes Ramsey (1927),using Ramsey rule utilized in Feldstein (1978) also , Diamond-Mirrlees (1971a), Diamond-Mirrlees (1971b) propose alternative to Ramsey proposition.

Key words: Optimal taxes, public finance, optimal minimum wage, asymmetric information

## Introduction

Mirrlees (1986), elaborates that a good way of governing is to agree upon objectives, than to discover what is possible and to optimize. The central element of the theory of optimal taxation is information. Public policies apply to the individuals on the basis of what the government knows about them. Second welfare theorem 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 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, Yagan D.(2009)[1]. 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

* Assistant professor, UGD-Shtip, R.North Macedonia, dusko.josevski@ugd.edu.mk
* Assistant professor, UGD-Shtip, R.North Macedonia, tdzaleva@gmail.com
[1] A key determinant of the optimal tax schedule (tax bracket) is the shape of the ability of the distribution.

income streams, Varian,H.R.(1980).In his model Varian (1980) derives optimal linear and nonlinear tax schedule. He uses Von Neumann-Morgenstern utility function(VNM decision utility, or decision preferences) [2] , with declining absolute risk aversion, see Kreps (1988). Varian (1980), concentrates especially on the problem of social insurance that previously was treated by Diamond, Mirrlees (1978) ,where in their model were emphasized the insurance-incentive aspects involved the retirement decision .Diamond, Helms and Mirrlees (1978),analyze the presence of uncertainty in the analysis of optimal taxation, with Cobb-Douglas utility function, with elasticity of substitution between labor and leisure <1 s that backward bending labor supply curve can be observed. Two period model with uncertainty showed how stochastic economies differ from the economies without uncertainty, since these second best insurance/redistribution programs differ in the outcomes from the first best resut economies without government intervention. In general if income contains random component then a system of redistributive taxation would contribute in the reduction of the variance in the after-tax income. In general Varian (1980) finds for linear and non-linear optimal tax, that if the consumption values are bounded, the optimal tax will always exist and would be a continuous function of observed income. Also in this model marginal tax are positive and the optimal tax will be increasing in contrast to the findings of Mirrlees (1971). In early contribution Ramsey (1927) , supposed that the planner must raise tax revenue only through imposition of tax on commodities only. In his model taxes should be imposed in inverse proportion to the representative customer's elasticity of demand for the good, so that commodities with more inelastic demand are taxed more heavily. But form the standpoint of public economics, goal is to derive the best tax system. In perfect economy with absent of any market imperfection (externality), if the economy is described by the representative agent, that consumer is going to pay the entire bill of the government, so that the lump-sum tax is the optimal tax. Governments in real world however cannot observe individual ability. Mirrlees (1971) , in the basic version of the model allowed individuals to differ in their innate ability. The planer can observe income, but the planner cannot observe ability or effort. By recognizing unobserved heterogeneity, diminishing marginal utility of consumption, and incentive effects, the Mirrlees approach formalizes the classical tradeoff between efficiency and equity. In this framework the optimal tax problem is a problem of imperfect information between taxpayers and the social planner. Saez (2001) argued that "unbounded distributions are of much more interest than bounded distributions to address high income optimal tax rate problem". In all of the cases that Saez (2001) investigated (four cases)[3] the optimal tax rates are clearly U-shaped. This paper by using the elasticity estimates from the literature, the formula for the asymptotic top rates suggests that the marginal rates for the labor income should not be lower than 50% and they could be as much as high as 80%.This paper used methodology proposed by Diamond (1998). Diamond and Mirrlees (1971a) and Diamond, Mirrlees (1971b) , are proposing alternative in Ramsey proposition by allowing the social planer to considers a numerous tax systems. In the first paper Diamond and Mirrlees (1971a), they prove how some market imperfections eg. capital market imperfections (consumers can lend but not borrow), the market situation will alter the optimal tax structure. Diamond and Mirrlees (1971b), are proposing tax rules for single good economy (changes on demand due to the tax structure differ from proportionality with larger than average percentage fall in the demand for goods with large income derivatives (elasticities)) ,in three-good economy the tax rate is proportionately greater for a good with smaller cross elasticity of compensated demand with the price of labor, in many commodities economy , households with low social marginal income utility predominate among the purchasers of the commodity, that

[2] This theorem serves as a basis of the expected utility theory. This theory actually represents maximizing the expected value of some function defined over the potential outcomes at some specified point in the future
[3] Utilitarian criterion, utility type I and II and Rawlsian criterion, utility type I and II.

commodity should be taxed more heavily, and vice versa, this taxation increases total welfare. It also shows that at optimum , the social marginal weighted utility changes in taxation are proportional to the changes in total tax revenue (income and commodity tax revenue), calculated at fixed prices, with consumer behavior corresponding to the price change. This study thus is not suggesting that commodity taxation is superior to income taxation. Also, this paper proves that the presence of commodity taxes implies the desirability of aggregate production efficiency even if the production is not Pareto optimal i.e. results is second best. In the second best setting however aggregate production efficiency over the whole economy may not be desirable, because distortionary taxes on transactions of individuals and firms will be needed to redistribute the real income or finance the production of public goods so that second best optimum will be reached (Second fundamental welfare theorem), Hammond (2000). Diamond and Mirrlees (1971a), continue to point out that there should not be taxation on intermediate goods such as capital held by the producers, also see Judd (1999).The general result Judd (1999) finds is that optimal tax on capital should be zero except for the initial period. Judd (1985), also found a zero optimal long-run capital income tax rate for steady states of the general competitive equilibrium and heterogeneous infinitely-lived agents with nonseparable preferences. But the famous Atkinson,Stiglitz (1976) results(result on the role of indirect taxation with an optimal nonlinear income tax) states that commodity taxes are not useful under these assumptions about the utility function: weak separability of function, and homogeneity across individuals in sub-utility of consumption. Proof of this theorem can be found in Laroque, G.(2005) , and Kaplow, L.(2006).The Atkinson-Stiglitz result is obtained by embedding the Ramsey model within Mirrlees model.Also zero-top tax rate suggest important task for the policy makers to identify the shape of the high-end of the ability distribution (they cannot observe the effort and ability in direct way but they can observe income). Tuomala (1990), confirms that marginal tax rate decreases as income increases except at income levels within a bottom decile. Ordover, J., Phelps,E. (1979), provided that if consumption have weakly separable utility functions and government has instruments that allow it to fix the capital stock on the socially optimal level, then the optimal tax rate on capital is zero, Salanie (2003). Chamley (1986), results on zero capital income tax states: " When the consumption decisions in a given period have only a negligible effect on the structure of preferences for periods in the distant future, then the second-best tax rate on capital income tends to zero in the long run". But these are (Ramsey capital income tax )two period models if more periods are included than the optimal tax formula would be more complex, as in Auerbach, Kotlikoff (1987a), and Auerbach, Kotlikoff (1987b). But what about estate and gift taxes and property taxes? Modigliani,F.,Brumberg ,R.H. (1954), Modigliani, F. (1966),Modigliani(1976) , Modigliani, F. (1986), Modigliani(1988) , view states that life cycle wealth accounts for the bulk of wealth (in US). Kotlikoff and Summers (1981) challenged this old view[4]. Here key problem is that the definition of life-cycle vs. inherited wealth is not conceptually clean. Previous Kotlikoff-Summers controversy consisted in the fact that estimates of the share of inherited wealth in aggregate wealth for Modgiliani (1986), Modigliani (1988) definition was 20% as low, and for Kotlikoff and Summers (1981) was as 80% high (data were the same). Piketty, T., Postel-Vinay,G.,Rosenthal,J.L,(2014), give better definition that the individual wealth is a sum of individual earnings minus expenditures (accrued amount ) multiplied by compound interest rate. Feldstein (1978), showed that elimination of tax on capital income is only optimal only when the structure of preferences satisfy certain separability condition. And for the capital taxation to be optimal it must be that uncompensated elasticity of savings (elasticity of the Marshalian demand for savings) is zero, even when the compensated elasticity of consumption of old population (Hicksian demand for consumption) is high (he reported result of -0.75). Now, if the

[4] Why is this important? ...taxation of capital income and estates, Role of pay-as-you-go vs. funded retirement programs.

labor and consumption are equivalent for the individuals, but savings pattern are different, results is that individuals will save more with consumption tax, than with labor tax. In OLG closed economy capital stock is due to life-time savings. The full neutrality result implies extra savings of young is equal to the consumption of old capital stock plus new government deficit (no change in capital stock)[5].In equilibrium where endowment is zero at equilibrium, and Hicksian demand for consumption is infinite i.e. compensated elasticity of consumption when old is infinite. But according to Saez,Stantcheva (2016a), because individuals derive utility from wealth, micro foundations for this wealth in the utility function are : bequest motives, entrepreneurship, or services from wealth it means that steady-state features finite finite supply elasticities of capital to capital tax rates. And because there is bi-heterogeneity of the agent's income and capital, Atkinson-Stiglitz zero-tax result does not apply herein. The optimal tax rate on inheritance (bequest in utility) case is zero, when the elasticity of bequest is infinite nesting the zero tax result. However, when in the model are imputed bequests, inequality is bi-dimensional and earnings are no longer the unique determinant of lifetime resources. That means that here A-S zero-tax result fails, see Piketty, T. , Saez,E., (2013), Farhi and Werning (2010).Also, Stiglitz, J.(1982) , showed that when leisure and goods are separable, differential taxation of commodities cannot be used as a basis of separation between the two and therefore is sub-optimal, Saez (2002). Commodity taxation is desirable when government is using social weights that are correlated with the consumption patterns and are conditional on income, or when the consumption patterns are related to the intrinsic earning ability and leisure choices[6]. Saez, E. ,S. Stantcheva (2016b),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. Chari and Kehoe(1999), besides developing stronger zero-optimal capital income tax rate than Chamley (1986), are developing Barro's (1979) result on tax smoothing, where in deterministic concept, optimal tax rates are constant, while in stochastic economy with incomplete markets tax rates follow a random pattern generated by a martingale process[7] .And the tax smoothing hypothesis requires tax rate to be changed (altered) only when some unpredicted shock occurs. This means that there should be no predictable changes in tax rates in times without shocks. The optimal capital tax formula is a function of social marginal welfare weights, that are product of Pareto weight and the utility of consumption of individuals, and this weights are normalized across the population to one. Optimal linear income and linear capital tax are inversely related to the elasticity, the revenue maximizing tax rates are calculated when weights on capital and labor are zero. The non-linear capital and labor taxes are dependent on the average welfare weight of capital income higher than the product of rate of return of capital and capital stock itself, and average welfare weight higher than the individual earnings. Pareto weights here proportional to net rate of return of capital and density of taxed labor income, and probability density function of tax system which is linearized at points of net tax return (substitution effects, no income effects) and earnings. Auerbach, A. (2009), Kaplow(1994), propose equivalence of consumption taxes and labor taxes: a linear consumption at some inclusive rate, is equivalent to a labor tax income combined with the initial wealth. In this setting consumption tax is equal to labor tax if there is no initial wealth and differences in wealth arise only from wealth preferences.

[5] Aggregate interest rate should equal to interest rate for the government debt.

[6] And if in the presence of optimal income taxation whether if a small commodity tax can be replicated by a small income change, and when this is not a case commodity taxation allows government to expand its own taxation power and therefore it is desirable.

[7] Martingale is a sequence of random variables (i.e., a stochastic process) for which, at a particular time, the conditional expectation of the next value in the sequence, given all prior values, is equal to the present value.

## Optimal taxation models: Mirrless (1971)

In the Mirrlees (1971) model, all individuals have same utility function which depends positively on consumption, and negatively on labor supply ,which can be denoted as $u(c, l)$.Let's suppose the utility function g the agents in the economy Mirrlees (1971) model:

*Equation 1*

$$\tilde{U}(c, l) = c - \frac{l^2}{2}$$

Where $y = \theta l$ и $\theta$ represents the level of skils of the worker. Now his social welfare function SWF is :$SWF(v) = \log(v)$.Now lets find the distribution of skills when $T(y) = 0.3$ which is Pareto with $h(y) = ky^{-k-1}y^k$[8].Equation for the distribution of skills is $f(\theta) = h\big(y(\theta)\big)y'(\theta)$,from the quasi-linear utility functions : $U(c, y, \theta) = c - \frac{1}{2}\left(\frac{y}{\theta}\right)^2$.And the tax function $T(y) = \tau y$, individual with skill level $\theta$ solves :

*Equation 2*

$$\max_y (1-\tau)y - \frac{1}{2}\left(\frac{y}{\theta}\right)^2$$

FOC is given as :$(1-\tau) - \frac{y}{\theta^2} = 0$, which implies that $y = (1-\tau)\theta^2$ and $f(\theta) = h\big(y(\theta)\big)y'(\theta) = k(\theta)^{-k-1}y^k 2(1-\tau)\theta = k\big((1-\tau)\theta^2\big)^{-k-1}y^k 2(1-\tau)\theta = 2k(1-\tau)^{-k}\theta^{-2k-1}y^k = 2k\theta^{-2k-1}\theta_l^{2k}$ .By integration one could get :$F(\theta) = \int_{\theta_l}^{\theta} f(\theta)d\theta = \int_{\theta_l}^{\theta} 2k(1-\tau)^{-k}\theta^{-2k-1}y^k d\theta = [-(1-\tau)^{-k}\theta^{-2k}y^k]_{\theta_l}^{\theta} = (1-\tau)^{-k}\theta_l^{-2k-1}\left((1-\tau)\theta_l^2\right)^k - (1-\tau)^{-k}\theta^{-2k}y^k = 1 - \theta^{-2k}\theta_l^{2k}$.Now we can solve for numerical optimum. Let's use $y = 2$ and $k = 4$ and truncate the distrubition[9] at the top $x$ percentile for some small $x$.

In this case : $\max_{v(\theta), u(\theta)} \int_{\theta_l}^{\theta^h} W[v(\theta)]f(\theta)d\theta$.Subject to :

$$\int_{\theta_l}^{\theta^h} \big(y(\theta) - e(v(\theta), y(\theta), \theta)\big)f(\theta)d\theta \geq 0; v'(\theta) = u_\theta\,[e(v(\theta), y(\theta), \theta)]$$

$y(\theta)$ is non decreasing function. Hamiltonian is formed as :$H = W\,[v(\theta)]f(\theta) + \lambda\,(y(\theta) - e(v(\theta), y(\theta), \theta))f(\theta) + \eta\,(\theta)U_\theta\,[e(v(\theta), y(\theta), \theta), y(\theta), \theta]$.

Standard conditions are as:

1. $\frac{\partial H}{\partial y} = 0 \Rightarrow \lambda f\big(1 - e_y\big) + \eta\big[u_{\theta c}e_y + u_{\theta y}\big] = 0$
2. $\frac{\partial H}{\partial v} == \eta' \Rightarrow W'f - \lambda e_v f + \eta u_{\theta c} e_v = -\eta'$

Transfersality conditions : $\eta(\theta_l) = \eta\big(\theta^h\big) = 0$ .From $W = \log(v)$ и $u(c, y, \theta) = c - \left(\frac{1}{2}\right)\left(\frac{y}{\theta}\right)^2$ will get the following derivatives :$u_\theta = \frac{y^2}{\theta^3}; u_{\theta c} = 0;\ u\theta_y = \frac{2y}{3}\ ;\ W' = \frac{1}{v}$.Let us remember that

[8] This is a density of earnings function , dependent on the skills of workers

[9] In statistics truncated distribution is a conditional distribution that comes as a result of the restriction of the domain of some other distribution or probability .

$v = u(e(v,y,\theta),y,\theta)$,we have $1 = u_c e_v$ и $0 = u_c e_y + u_y$,therefore :$e_v = \frac{1}{u_c}; e_y = -\frac{u_y}{u_c} = \frac{y}{\theta^2}$. If we substitute in the optimality and control equations about the state variables one can get :

*Equation 3*

$$\lambda f\ \left(1 - \frac{y}{\theta^2}\right) + \eta\left[\frac{2y}{\theta^3}\right] = 0 \text{ и } \frac{f}{v} - \lambda f = -\eta'$$

If we solve in the first equation for $y(\theta)$ we get : $y(\theta) = \frac{\lambda f\,(\theta)\theta^3}{\lambda f\,(\theta)\theta - 2\eta(\theta)}$. With the equation $\eta'\eta'(\theta) = \left(\lambda - \frac{1}{v(\theta)}\right)f(\theta)$. If we substitute for $y(\theta)$ in the constraint :$v'(\theta) = u_\theta[e(v(\theta),y(\theta),\theta),y,\theta] = \frac{y^2}{\theta^3} = \left(\frac{\lambda f\,(\theta)}{\lambda f\,(\theta)\theta - 2\eta(\theta)}\right)^2\theta^3$.In Saez (2001) optimal tax formula is given as :

*Equation 4*

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

In the previous expression $\tau$ are taxes, $\bar{g}$ is the ratio of social marginal utility for the top bracket taxpayers to the marginal value of public funds for the government, which depends on the social welfare function [10]. Utility in the social welfare function provides a guideline for the government for achieving optimal distribution of income, Tresch, R. W. (2008) . In the Saez (2001) optimal tax formula also : $\bar{\varepsilon}^u = \int_{\bar{w}}^{\infty} \varepsilon_w^u w h(w)dw/w_m$ , where Marshallian demand for labor is given as :$w = w(1-\tau,R)$ where $R$ is the non-labour income, and $w$ are earnings(wages)[11]. And compensated elasticity of earnings is : $\bar{\varepsilon}^c = \frac{1-\tau}{w}\left(\frac{\partial w}{\partial(1-\tau)}\right)|_u$ .Those two are related by the Slutsky equation : $\varepsilon^c = \varepsilon^u - \eta$, when there are no behavioral responses there is only meachnical effect denote by M and $M = [w_m - \bar{w}\,]d\tau$ , where $w_m - \bar{w}$ represents the earnings of the agent above medium population earnings. Behavioral responses are equal to : $dw = -\frac{\partial w}{\partial(1-\tau)}d\tau + \frac{\partial w}{\partial R}dR = -\left(\varepsilon^u w - \frac{(1-\tau)\partial w}{\partial R}\right)\left(\frac{d\tau}{1-\tau}\right)$ ,or the total behavioral response $\beta = -\left(\varepsilon^u w\frac{(1-\tau)\partial w}{\partial R}\right)\left(\frac{\tau d\tau}{1-\tau}\right)$.Saez(2001) result for high income earners is given as : $\frac{\tau}{1-\tau} = \frac{(1-\bar{g})\,(w_m/(\bar{w}-1)}{\frac{\bar{\varepsilon}^u w_m}{\bar{w} - \int_{\bar{w}}^{\infty} \eta_w h(w)dw}}$.In the Mirrlees(1971) model government , maximizes[12] :$SWF = \int_0^\infty G(u_w)f(w)dw$. In the previous expression $G(u_w)$ represents the concave utility function[13].The constraint here is given as: $\int_0^\infty G(u_w)f(w)dw \leqq \int_0^\infty w_l f(w)dw - E$ ,where $E$ are government expenditures. Now, about Pareto distributions it is well known fact that :$\frac{ratio\ average}{threshold} = constant$ .Now if we denote the average wage $w^*(w) > w$, and if $w$ is a threshold, then $w^*(w)$ can be expressed as :$w^*(w) = \int_{w_m > w} wf(w)dw/$

---

[10] Social welfare function can be :$SWF = \int U^i\,di$-Utilitarian or Benthamite, $SWF = min_i U^i$-Rawlsian $SWF = \int U^i di \rightarrow G(U) = \frac{U^{1-\gamma}}{1-\gamma}$ if $\gamma = 0$ function is utilitarian , Rawlsian if $\gamma = \infty$. With Pareto weights: $SWF = \int \mu_i U^i\,di$ where $\mu_i$ is exogenous.

[11] Income effects are captured through $\eta = (1-\tau)\partial w/\partial R$ ,average income effects are :$\bar{\eta} = \int_{\bar{w}}^{\infty} \eta_w h(w)dw$

[12] Here we make assumption that wages =skill level

[13] Now, for a concave function $f: (a,b) \rightarrow R$ is continuous in $IntA$. This function $f: (a,b) \rightarrow R$ is concave in the interval $(a,b)$ , if for every $x_1, x_2 \in (a,b), a \in (0,1)$, it follows $f(ax_1 + (1-a)x_2) < af(x_1) + (1-a)f(x_2)$.

$\int_{w_m>w} f(w)dw = \int_{w_m>w} dw/w^a / \int_{w_m>w} dw/w^{1+a} = \frac{aw}{a-1}$ .In the previous expression $a$ represents the shape parameter of the Pareto distribution. And $a = \frac{b}{b-1}$ i.e. $\frac{w^*(w)}{w} = b$ .About the Pareto distribution PDF of this distribution is given as :$1 - F(w) = \left(\frac{k}{w}\right)^a$ ,and CDF of the function is given as $f(w) = \frac{ak^a}{w^{1+a}}$,[14]that is $\lim_{w\to\infty} \frac{\left(\frac{k}{w}\right)^a}{w\cdot\left(\frac{ak^a}{w^{1+a}}\right)}$ by applying $\lim_{x\to a}[c \cdot f(x)] = c \cdot \lim_{x\to a} f(x) \Rightarrow$ $\frac{1}{ak^a} \cdot \lim_{w\to\infty} \frac{\left(\frac{k}{w}\right)^a}{w\cdot\left(\frac{ak^a}{w^{1+a}}\right)} = \frac{1}{ak^a} \cdot \lim_{w\to\infty} (k^a) = \frac{1}{ak^a} \cdot k^a = \frac{1}{a}$ .hence the formula of marginal income for top earners $\tau^* = \frac{1}{1+a\cdot\varepsilon}$.This is the tax rate that maximizes government revenues. Now, $\frac{1-F(w)}{wf}$ represents the ratio of people with wages above $w$ ,which is the mass of people paying more tax, and on the right the people affected by the adverse incentive effects. The social optimum means U shaped pattern of marginal tax rates. Diamond (1998) , gives such an example in his ABC tax model. There salaries are distributed $w(w_l, w_h)$.The government objective function than becomes :$\int_{w_l}^{w_h} u(w)\psi(w)dw$ , in the previous expression $\psi(w)$ represents the distribution function. Now if $\int_{w_l}^{w_h} u(w)\psi(w)dw = 1$ , it means that $\lambda = 1$ [15] , or $\lambda = \int_{w_l}^{w_h} G'\big(u(w)\big)f(w)dw = 1$ ,and $\mu(w)$denotes the multiplier on the incentive constraint of type w, and is equal to $-\mu(w) = \int_{w_l}^{w_h}\big(f(w) - \psi(w)\big)\, dw = \Psi(\mathrm{w}) - \mathrm{F}(\mathrm{w})$,here $\Psi(\mathrm{w})$ is a CDF of the function,and $\mathrm{F}(\mathrm{w})$ is a distribution of skills. The optimal tax government formula with Rawlsian government [16]would be :

*Equation 5*

$$\frac{T'\big(w(h)\big)}{1-T'\big(w(h)\big)} = \left(\frac{1+\varepsilon}{\varepsilon}\right)\frac{1-F(w)}{wf(w)} \quad \text{or} \quad \frac{T'\big(w(h)\big)}{1-T'\big(w(h)\big)} = \left(\frac{1+\varepsilon}{\varepsilon}\right)\frac{\psi(w)-F(w)}{wf(w)}$$

Now if we divide and multiply by $1 - F(w)$ we get :$\frac{T'\big(w(h)\big)}{1-T'\big(w(h)\big)} = \left(\frac{1+\varepsilon}{\varepsilon}\right)\frac{\Psi(w)-F(w)}{1-F(w)}\frac{1-F(w)}{wf(w)}$ .In the previous formula $\left(\frac{1+\varepsilon}{\varepsilon}\right) = A(w)$ , elasticity and efficiency argument, $\frac{\Psi(w)-F(w)}{1-F(w)} = B(w)$, measures the desire for redistribution :if the sum of weights $\psi(w)f(w)$ is below $w$ is relative high to the weights above , the government will like to tax more, this part $\frac{1-F(w)}{wf(w)} = C(w)$ measures the density of the right tail of the distribution and higher density will be associated with higher taxes. In Piketty, T., Saez.E., and Stantcheva,S.(2014), it is well defined aggregate elasticity of income as:

$\varepsilon = \frac{1-\tau}{z}\frac{dz}{d(1-\tau)}$ , where $z$ is taxable income and $z = y - x$, where $y$ is the real income, and $x$ is sheltered income [17],taxable income s used in the calculation for Pareto parameter $a = \frac{z}{z-\bar{z}}$. Tax avoidance elasticity component is given as $\varepsilon_1 = \frac{1-\tau}{z}\frac{dx}{1-\tau}$ , and $\varepsilon_2 = \frac{1-\tau}{z}\frac{dy}{1-\tau}$ is the real labor supply elasticity. Now, when government raises slightly $\tau \to d\tau$ there is: mechanical effect from the increase in taxes i.e. $dM = (z - z^*)d\tau$ , welfare effect $dW = -\bar{g}dM = -\bar{g}(z - z^*)d\tau$ , where social marginal weight for individual is :$g_i = G'\big(u^i\big)u_c^i/\lambda$ , where $\lambda$ is a multiplier of government

[14] $\frac{\left(\frac{k}{w}\right)^a}{w\cdot\left(\frac{ak^a}{w^{1+a}}\right)} = \frac{\frac{k^a}{w^a}}{w\cdot\left(\frac{ak^a}{w^{1+a}}\right)} = \frac{\frac{k^a}{w^a}}{w\cdot\frac{ak^a}{w\cdot w^a}} = \frac{1}{a}$

[15] This is an expression for the marginal value of public funds to the government

[16] The social welfare function that uses as its measure of social welfare the utility of the worst-off member of society. The following argument can be used to motivate the Rawlsian social welfare function.

[17] Investments or investment accounts that provide favorable tax treatment , or activities and transactions that lower taxable income.

constraint which is $\int \tau(wl)f(w)dw \geq E$ ,average income in economy is $\bar{z} = \int zh(z)dz$, $h(z)$ is a density $\frac{\partial L}{\partial \tau(z)} = \lambda$, $c = \bar{z} - E$, while $\bar{g}$ social marginal weight for top earners is given as: $\bar{g} = \frac{\int g_i \cdot z_i}{z^* \cdot \int g_i}$ , where $z - z^*$ is the mechanical redistribution effect. And, the third effect is behavioral response of the top earners: $dB = -\frac{\tau}{1-\tau} \cdot \frac{1-\tau}{z} \cdot \frac{dz}{d(1-\tau)} \cdot z \cdot d\tau \; = -\frac{\tau}{1-\tau} \cdot \varepsilon \cdot z \cdot d\tau$ . From here it can be derived Diamond (1998) optimal tax formula :$\frac{\tau(z)}{1-\tau(z)} = \frac{1}{\varepsilon(z)} \cdot \left[\frac{1-F(z)}{z \cdot F(z)}\right] \cdot [1 - G(z)]$, this is distribution shape parameter $\frac{1-F(z)}{z \cdot F(z)}$ , $G(z)$ are social marginal welfare weights For numerical solutions of the Mirrless model (1971) , one can look up to Brewer, M., E. Saez, and A. Shephard (2010) ,

*Equation 6*

$$\frac{\tau' z(h)}{1 - \tau' z(h)} = \left(1 + \frac{1}{\varepsilon}\right) \frac{1}{hf(h)} \int_h^\infty \left(1 - \frac{G' u(h)}{\lambda}\right) f(h) dh$$

Where $\lambda = \int G'(u)dh$.Few general conclusions about marginal tax rates in the literature appear:1. $\tau(z(h)) \geq 0$ , a in Mirrlees (1971), 2. $\tau(z(highest\ h)) = 0$ ($f(h)$bounded above ,Sadka, (1976), 3. $\tau(z(lowest\ h)) = 0$ ($h$: $y(h) > 0$),Seade,(1977). Akerlof (1978),introduces optimal redistribution system with and without tagging, in Mirrless(1971) –Fair (1971) model (those are separate papers), by allowing tax administration to treat different labeled (identified, tagged)[18] groups of people who are in need ,taxpayers (as opposed to beneficiaries) are denied the benefit of the transfer, and lump sum transfer is made to the tagged people. The utility function here is: $u = \frac{1}{2} * \max u[(q_D - \tau_D) - \delta + u(q_E + t_E)] + 1/2\, u(q_E + t_E)$ In previous expression $q_D, q_E$ represents the output of the skilled and unskilled workers in difficult and easy jobs respectively. And, $\tau_D$ is a tax on income in difficult jobs, while $t_E$ transfer of income in easy jobs, $\delta$ reflects distaste of workers for difficult jobs. From negative tax[19] formula, $T = -a\bar{y} + \tau y$; $\tau = \beta a + g$ , where $\tau$ is marginal tax rate ,$a$ is the fraction of per capita income$\bar{y}$ received by a person with zero gross income,$\beta$ are tagged group of poor people from population, $\beta a$ is a fraction of population that receives minimum support. Skilled workers are assumed to work in difficult jobs, so their utility is: $\max u[(q_D - \tau_D) - \delta; u(q_E + t_E)]$,and $u(q_D - \tau_D) - \delta \geq u(q_E + t_E)$,and then tax collections per skilled worker equal transfers i.e. $\tau_D = t_E$, but if the skilled workers work in the easy jobs than the result would be zero $t_E = 0$ , if $u(q_D - \tau_D) - \delta < u(q_E + t_E)$. In the tax equal budget constraint: $u(q_D - \tau_D^*) - \delta < u(q_E + t_E^*)$, or $\tau_D^* = t_E^*$ are optimal tax –cum-transfer rates. And now utility function becomes:

$$u^{tag*} = \frac{1}{2} * \max u[(q_D - \tau_D) - \delta + u(q_E + t_E)] + 1/2(1 - \beta)\; u(q_E + t_E) + 1/2\; \beta u(q_E + t)$$

Subject to :$\tau_D = (1 - \beta)t_E + \beta t$ , or transfers to easy jobs, if $u(q_D - \tau_D) - \delta \geq u(q_E + t_E)$ and $(2 - \beta)\tau_E + \beta t = 0$ , if $u(q_D - \tau_D) - \delta < u(q_E + t_E)$.Generalization of Mirrless and Fair problem includes administrative costs for grouping (tagging ) people in different groups $c(\Lambda)$, utility is to be maximized over different types of people $x$: $u = \int u_x f(x)dx$,where $f(x)$ denotes the distribution of people of type $x$,the group to which agent belongs is $\sigma$ so $u_x = (y - \tau, x, \sigma)$ , constrant for the utility maximization involves administrative costs also : $\int_x \tau_\sigma(y(x), \sigma(x)) f(x)dx + c(\Lambda) = 0$, $\sigma(x)$ is a group to which individual belongs to. And other constraint: $u[w(x,\sigma)\ell(x,\sigma) - \tau\_\sigma(w\,(x,\sigma)\ell(x,\sigma)), x, \sigma]$ . Where $w(x,\sigma)$ is a wage of a person with characteristics $x$ , that belongs

[18] A system of tagging permits relatively high welfare payments with relatively low marginal rates of taxation.
[19] In economics, a negative income tax is a welfare system within an income tax where people earning below a certain amount receive supplemental pay from the government instead of paying taxes to the government.

to a group $\sigma$, and $\ell(x,\sigma)$ is a labor input , and this is generalization of Mirrless and Fair models with tagging.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 :$U(1-\tau,R) = max_z\big((1-\tau)z + R, z\big)$ ,where $z$ represents the taxable income $R$ is a virtual income intercept, and $\tau$ is an imposed income tax. Marshalian labor supply is :$z = z(1-\tau,R)$, uncompensated elasticity of the supply is given as:$\varepsilon^u = \frac{(1-\tau)}{z}\frac{\partial z}{\partial(1-\tau)}$ , income effect is $\eta = (1-\tau)\frac{\partial z}{\partial R} \leq 0$.Hicksian supply of labor is given as:$z^c\big((1-\tau,u)\big)$, this minimizes the cost in need to achieve slope $1-\tau$ , compensated elasticity now is : $\varepsilon^c = \frac{(1-\tau)}{z}\frac{\partial z^c}{\partial(1-\tau)} > 0$, Slutsky equation now becomes: $\frac{\partial z}{\partial(1-\tau)} = \frac{\partial z^c}{\partial(1-\tau)} + z\frac{\partial z}{\partial R} \Rightarrow \varepsilon^u = \varepsilon^c + \eta$, where $\eta$ represents income effect :$\eta = (1-\tau)\frac{\partial z}{\partial R} \leq 0$ .With small tax reform taxes and revenue change i.e.:$dU = u_c \cdot [-zdt + dR] + dz[(1-\tau)u_c + u_z] = u_c \cdot [-zdt + dR]$.Change of taxes and its impact on the society is given as:$dU_i = -u_c dT(z_i)$. Envelope theorem here says :$U(\theta) = \max_x F(x,\theta), s.t. c > G(x,\theta)$ , and the preliminary result is :$U'(\theta) = \frac{\partial F}{\partial \theta}\Big(x^*(\theta),\theta - \lambda^*(\theta)\frac{\partial G}{\partial \theta}x^*(\theta),\theta\Big)$. Government is maximizing :$0 = \int G'(u^i)u_c^i \cdot \Big[\big(Z - z^i\big) - \frac{\tau}{d(1-\tau)}eZ\Big]$, mechanical effect is given as:$dM = [z - z^*]d\tau$, welfare effect is :$dW = -\bar{g}dM = -\bar{g}[z - z^*]$, and at last the behavioral response is :$dB = -\frac{\tau}{1-\tau}\cdot e \cdot zd\tau$. And lets denote that:

*Equation 7*

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

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

*Equation 8*

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

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

*Equation 9*

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

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:

*Equation 10*

$$\frac{T'(z_n)}{1-T'(z_n)} = \frac{1}{e}\left(\frac{\int_n^{\infty}(1-g_m)dF(m)}{z_n h(z_n)}\right) = \frac{1}{e}\left(\frac{1-H(z_n)}{z_n h(z_n)}\right)\cdot\big(1 - G((z_n)\big)$$

Where $G((z_n) = \frac{\int_n^{\infty} g_m dF(m)}{1-F(n)}$ ,and $g_m = G'(u_m)/p$ 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 $[z^*, z^* + dz^*]$.Every tax payer with income $z$ above $z^*$ pays additionaly $d\tau dz^*$ valued by $(1 - g(z))d\tau dz^*$.Mechanical effect is :

*Equation 11*

$$M = d\tau dz^* \int_{z^*}^{\infty} (1 - g(z)) d\tau dz^*$$

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

*Equation 12*

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

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

*Equation 13*

$$\frac{\dot{z}_n}{z_n} = \frac{1+\varepsilon_{(n)}^{u}}{n} - \dot{z}_n \frac{T''(z_n)}{1-T''(z_n)} \varepsilon_{z(n)}^{c}$$

This model was later augmented with the migrations by Mirrless (1982).Migrations are of importance for the top incomes(brain drain). In the model earnings are fixed $c$ , and $p(c|z)$ represents the number of residents earning $z$, while $c = z - T(z)$ represents the disposable income. Now , one small tax reform $dT(z)$ ,for those earning income $z$. Mechanical effect of net-welfare is :$M + W = (1 - g(z))P(c|z)dT$ .Migration equal taxes average or total:$M + W = \frac{\partial P(c|z)}{\partial C} \frac{z-T(z)}{P(c|z)}$.Cost of imposing taxes are :$B = -\frac{T(z)}{z-T(z)}$.Optimal tax applies when : $M + W + B = 0$.And the formula for the optimal tax with migrations becomes :

*Equation 14*

$$\frac{T(z)}{z-T(z)} \frac{1}{\eta_m(z)} \cdot (1 - g(z))$$

$\eta_m(z)$ (elasticity of taxable top income) depends on the size of jurisdiction; it's large for the cities, and its small (zero) for the world, redistribution is easier in larger jurisdictions. Formula for maximizing the revenues from top incomes is :

*Equation 15*

$$\tau = \frac{1}{1 + a \cdot e + \bar{\eta}^m}$$

Where $\bar{\eta}^m$ is elasticity of the top earners towards disposable income.

## Ramsey model (1927) in theory

In Ramsey (1927), utility function is of type: $U = f(p_1, p_2, p_3, \dots\dots, Y)$, $p_1, p_2, p_3, \dots\dots$ are prices and $Y$ is income. Standard result is known as Roy's identity , Roy (1947)[20], is : $\frac{\partial U}{\partial p_i} = -Q_i \frac{\partial U}{\partial Y}$. With the horizontal demand curves, price of the producers is fixed, change in the goods price is only equal to the change in taxes. Than, $dp_1 = dt_1 > 0$, $dp_2 = dt_2 < 0$.Change in taxes must satisfy the following equation: $dU = \frac{\partial U}{\partial p_i} d\tau_1 + \frac{\partial U}{\partial p_2} d\tau_2 = 0$,and $\frac{d\tau_2}{d\tau_1} = -\frac{Q_1}{Q_2}$, , change in the revenues caused by the change in taxes is : $\frac{\partial(\tau_1 Q_1)}{\partial t_1} = Q_1 + \frac{\tau_1 dQ}{dp_1} = Q_1\left(1 + \frac{\tau_1 dQ_1 p_1}{p_1 dp_1 Q_1}\right) = Q_1\left(1 - \frac{\tau_1}{p_1}\varepsilon_u^1\right)$, where $\varepsilon_u^1$ represents the compensated elasticity of the demand for good 1. Change of revenues as a result of change of taxes on good 2 is: $\frac{\partial(t_2 Q_2)}{\partial t_2} = Q_2\left(1 - \frac{\tau_2}{p_2}\varepsilon_u^2\right)$. Total change in revenues is given as:

*Equation 16*

$$\frac{dR}{dt_1} = Q_1\left(1 - \frac{\tau_1}{p_1}\varepsilon_u^1\right) + \frac{d\tau_2}{d\tau_1} Q_2\left(1 - \frac{\tau_2}{p_2}\varepsilon_u^2\right) = Q_1\left[\left(1 - \frac{\tau_1}{p_1}\varepsilon_u^1\right) - \left(1 - \frac{\tau_2}{p_2}\varepsilon_u^2\right)\right] = Q_1\left[\frac{\tau_2}{p_2}\varepsilon_u^2 - \frac{\tau_1}{p_1}\varepsilon_u^1\right]$$

With the optimal tax structure, this identity must holds: $\frac{t_2}{p_2}\varepsilon_u^2 - \frac{t_1}{p_1}\varepsilon_u^1 = 0$, 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), "when lump-sum taxation is not available (or, equivalently, when a tax on leisure is impossible), all other commodities should be taxed at differential rates (positive and negative) that depend on their relative demand elasticities and cross elasticities".Ramsey model was found useful 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). Here the problem of utility maximization is given as: $u\,(q, w(1-\tau_L)) = \max_{c_1,c_2,l}\ u(c_1, c_2, l)$ s.t. $c_1 + c_2/(1 + r(1-\tau_K)) = wl(1-\tau_L)$, where $q = \frac{1}{1+r(1-\tau_K)}$; $p = \frac{1}{1+r}$ , are the prices after taxation respectively on $c_2$ .Optimal tax rates could be obtained by solving standard Ramsey problem: $\max_{\tau_L,,\tau_K} u\left(q, w(1-\tau_L)\right)$ subject to $wl\tau_L + (q-p)c_2 \geq g\,(\lambda)max$, where $g$ is exogenous tax requirement. When the compensated elasticity of the supply is : $\frac{r\tau K}{1+r} = (\sigma_{L2} - \sigma_{22}) = \frac{\tau L}{1-\tau_L}(\sigma_{LL} - \sigma_{2L})$, where following applies also: $\sigma_{22} = \left(\frac{q}{c_2}\right)\partial c_2^c/\partial q < 0$ ; $\sigma_{L2} = \left(\frac{q}{l}\right)\partial l^c/\partial q$; $\sigma_{2L} = \left(\frac{w(1-\tau_L)}{c_2}\right)\partial c_2^c/\partial(w(1-\tau_L))$ . Optimal tax formula can be simplified : $\frac{r\tau_K}{1+r}\sigma_{22} = \frac{\tau_L}{1-\tau_L}\sigma_{LL}$. Inverse elasticity rule says if : $\sigma_{LL} << |\sigma_{22}|$ than $\tau_k$ will be of relative small size to $\tau_L$ .Feldstein (1978) makes famous theoretical argument why $\sigma_{22}$ can be large even if $\varepsilon_u = \left(\frac{q}{s}\right)\partial s/\partial q$ [uncompensated savings elasticity] is zero:Budget $c_1 + qc_2 = w(1-\tau L)l + y$, Slutsky equation when endowment is zero is : $\frac{\partial c_2^c}{\partial q} = \frac{\partial c_2}{\partial q} + \frac{c_2 \partial c_2}{\partial y} \Rightarrow \sigma_{22} = \varepsilon_{2q}^u + \frac{q\partial c_2}{\partial y c_2} = \frac{s}{q}$ , $\varepsilon_{2q}^u = \left(\frac{q}{c_2}\right)\partial c_2/\partial q = \varepsilon_{sq}^u - 1 \Rightarrow \sigma_{22} = \varepsilon_{sq}^u - 1 + q\partial c_2/\partial y c_1 + qc_2 = w(1-\tau_L)l + y \Rightarrow \partial c_1/\partial y + q\partial c_2/\partial y = w(1-\tau_L)\partial l/\partial y + 1 \approx 1 \sigma_{22} \cong \varepsilon_{sq}^u - \partial c_1/\partial y \cong -\partial c_1/\partial y \leq -0.75$.

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

## Optimal minimum wage with no and with taxes and transfers

In the model with constant returns to scale there is no profit at equilibrium, and following Lee,D., Saez,E. (2012) ,optimal wage of agent $i$ equals marginal productivity i.e. $w_i = \frac{\partial F(h_l,h_h)}{\partial h_l}$, where $h_l, h_h$ represent the low skilled and high skilled workers respectively. Because there are zero profits at equilibrium here because of the constant returns to scale that means that:$\Pi = F(h_l, h_h) - w_l h_l - w_h h_h = 0$,consumption equal to $c_i = w_i - \tau_i$,agents are heterogeneous and the costs for their efforts are given as : $\theta = (\theta_l, \theta_h)$ ,where $(\theta_l, \theta_h)$ are the costs of low and high skilled workers, low skills and high skills also means different occupations. This also implies that:$\frac{w_l}{w_h} = F_h(1, h_l, h_h)/F_1(1, h_l, h_h)$,low skilled labor demand elasticity is given as :$\varepsilon_l = D_l(w_l) \cdot -\left(\frac{w_l}{h_l}\right)$, resource constraint in the economy[21] is given as : $c_0 h_o + c_l h_l + c_h l_h \leq w h_l + w_h l_h$.Social welfare function in this case is given as : $SWF = (1 - h_l - h_h)G(c_0) + \int_{\Theta_l} G(c_l - \theta_l)\, dH(\theta) + \int_{\Theta_h} G(c_h - \theta_h)\, dH(\theta)$ [22].Since there are no income effects $\lambda = g_o h_0 + g_l h_l + g_h h_h = 1$ ,where $\lambda$ is the marginal value of public funds, and marginal weights are defined as $g_0 = G'(c_0)/\lambda$ and $g_i = \int_{\Theta_i} G'(c_i - \theta_i) dH(\theta)/\lambda(h_i)$. The concavity of the social welfare function implies that $g_0 > g_l > g_h$.If there are no taxes and transfers then minimal wage will equal $\overline{w} = w_l^* + dw_l$ , where $dw_l$ are transfer from other factors to minimal workers. If the minimal wage increases we are facing changes :$dw_l, dw_h, dh_l, dh_h$.And then :$\mathrm{d}\Pi = \sum_i \left(\frac{\partial w_i}{\partial h_i}\right) dh_i - w_l dh_l - w_h h_h = w_l dh_l + w_h h_h = 0$. FOC of the previous expression is given as :

*Equation 17*

$$\frac{dSWF}{d\overline{w}} = \left(-\frac{dh_l}{w_l} - \frac{dh_h}{w_h}\right) G(0) + \frac{dh_l}{d\overline{w}} G(0) + \frac{dh_h}{d\overline{w}} G(0) + \int_{\Theta_l} G'(c_l - \theta_l)\, dH(\theta) + \frac{dw_h}{d\overline{w}} \int_{\Theta_h} G'(c_h - \theta_h)\, dH(\theta) \text{ or, } \frac{dSWF}{d\overline{w}} = \lambda h_l (g_l - g_h) > 0.$$

Now if government can introduce taxes and transfer at optimum we have $g_l = 1$ full redistribution to low skilled workers, and $g_o h_0 + g_l h_l + g_h h_h = 1$. If the consumption changes with wages than $\Delta c_i = c_i - c_0$ and the resource constraint is given as : $h_l(w_l - \Delta c_l) + h_h(w_h - \Delta c_h) \geq c_0$.Lagrangian for the previous function is given as : $\mathcal{L} = SW + \lambda\, [h_l \cdot (w1_l - \lambda c_l) + h_h \cdot (w_h - \lambda c_h) - c_0]$.Skills ratio doesn't change with taxes and so $\frac{h_l}{h_h} = \rho\left(\frac{w_l}{w_h}\right)$, and $\frac{d\mathcal{L}}{dc_0} = \lambda(g_o h_0 + g_l h_l + g_h h_h - 1)$ ,at the optimum $g_l = 1$ . And $\frac{d\mathcal{L}}{dc_1} = \int_{\Theta_l} G(c_0 + \Delta c_1 - \theta_l)\, dH(\theta) - \lambda h_l = \lambda(g_l - 1)h_l$. Now the optimal tax rate when in a model there are extensive labour supply responses is:

$T_0^* = \frac{1}{1 + \frac{h_l e_0}{h_l + h_h}}$ , where $e_0$ represents the elasticity of transition from unemployed to low skill , and $T_0^*$ is the first tax at state 0, marginal tax rate at state 1 is $T_1^* = \frac{1}{1 + \frac{h_l e_1}{h_l + h_h}}$. And $e_1$ is the elasticity of transition from low skilled to high skilled employee. Transitional taxes from unemployed to low skilled and from low skilled to high skilled are given as :$T_0 = 1 - \frac{y_l - y_0}{w_h}$ and $T_1 = 1 - \frac{y_l - y_h}{w_h - w_l}$. $y_l - y_0$ is the after tax income gap, same applies for $y_l - y_h$. In the Stiglitz (1982) model of endogenous wages high skill and low skill workers have same amount of working hours $l_i$. The production

[21] $c_0 h_o$ are consumption and skills of unemployed agent

[22] $H(\theta) = H(\theta_l, \theta_h)$

function is of constant returns of scale also. Wages here are equal to the marginal product of labour :$w_i = \frac{\partial F(l_l,l_h)}{\partial l_i}$, the resource constraint in the economy is given as:$\sum_i c_i \leq F(l_l, l_h)$.Production function follows Mirrlees (1971) assumptions and it linear :$F(l_l, l_h) = \theta_l l_l + \theta_h l_h$,where $\theta$ represents the ability of an agent $i$ and wages are $w_i$. Linear welfare weights are $\psi_l, \psi_h$ and if $\psi_l > \psi_h$ in equilibrium .the following Lagrangian applies here:

*Equation 18*

$$\mathcal{L} = \psi_l u(c_l, l_l) + \psi_h u(c_h, l_h) + \lambda\left(F(l_l, l_h) - \sum_i c_i\right) + \mu\left(u(c_h, l_h) - u\left(c_l, \frac{w_l l_l}{w_h}\right)\right) + \sum_i \eta_i (w_i - F_i)(l_l, l_h)$$

$\lambda$ is the marginal value of public funds, and $\mu$ is value of relaxing the incentive constraint for type H,and how they are related :$(\psi_h + \mu)u_c(c_h, l_h) = \lambda$,or $(\psi_h + \mu)u_l(c_h, l_h) = -\lambda F_h(l_h, l_l) + \sum_i \eta_i F_{i_h}(l_l, l_h)$.Marginal taxes or labor wedge[23] is given as:

*Equation 19*

$$T'(w_h) = 1 + \frac{u_l(c_h, l_h)}{u_c(c_h, l_h) w_h} \text{ or } T'(w_h) = 1 + \frac{-\lambda F_h(l_h, l_l) + \sum_i \eta_i F_{i_h}(l_l, l_h)}{\lambda F_h(l_h, l_l)} = \frac{\sum_i \eta_i F_{i_h}(l_l, l_h)}{\lambda F_h(l_h, l_l)}$$

First order conditions are given as:

$$-\mu u_l\left(c_l, \frac{w_l l_l}{w_h}\right)\frac{l_l}{w_h} + \eta_l = 0 \text{ and } \mu u_l\left(c_l, \frac{w_l l_l}{w_h}\right)\frac{w_l l_l}{w_h^2} + \eta_h = 0$$

Because $\{u_l < 0; \eta_l < 0; \eta_h > 0\} \Rightarrow F_l > 0; F_h < 0,;$ $\sum_i \eta_i F_{i_h}(l_l, l_h) < 0$ , this means that the marginal tax rate $T'(w_h) < 0$,which says top earners are subsidized at the margin because their labor raises the wages of lower earners.

## Numerical examples of optimal taxation (review)

In this section we are trying to explain previous theory by following numerical examples or some using real historical data from developed countries. In the first example Mankiw NG, Weinzierl M, Yagan D.(2009) , they are using log-normal [0,1] or Pareto normal distribution with CDF :

*Equation 20*

$$F(x) = \frac{\int_a^x \frac{1}{z\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{lnz - \mu}{\sigma}\right)^2\right] dz}{\int_a^b \frac{1}{z\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{lnz - \mu}{\sigma}\right)^2\right] dz}$$

And PDF :

$$f(x) = \frac{\frac{1}{x\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{lnx - \mu}{\sigma}\right)^2\right]}{\int_a^b \frac{1}{z\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{lnz - \mu}{\sigma}\right)^2\right] dz}$$

And the Lognormal-Pareto density is derived from :

[23] Average labor wedge measures the extent to which imposed taxes on labor income discourage employment.

*Equation 21*

$$f(x) = \frac{\alpha\theta^{\alpha}}{(x-b)^{\alpha+1}}; x \gg \theta + b$$

Where $\alpha > 0, \theta + b > 0, \int_{\alpha}^{\infty} f(x)dx = 0; f_1(\theta + b) = f_2(\theta + b)$ and $f'_1(\theta + b) = f'_2(\theta + b)$, see Teodorescu,S.(2010). Truncated Lognormal-Pareto density is given as:

*Equation 22*

$$f_{LP}^{c}(x) = \begin{cases} \frac{1}{[1+\Phi(k)]\sqrt{2\pi}(x-\alpha)\sigma} \exp\left[-\frac{1}{2}\left(\frac{\ln(x-\alpha)-\mu}{\sigma}\right)^2\right]; \alpha < x \leq \theta + b \\ \frac{\alpha\theta^{\alpha}}{[1+\Phi(k)](x-b)^{\alpha+1}}; \theta + b \leq x < \infty \end{cases}$$

Where $\alpha = \frac{\theta}{\theta+b-\alpha}\left[1 + \frac{\ln(\theta+b-\alpha)-\mu}{\sigma^2}\right] - 1$ and $k = \frac{\ln(\theta+b-\alpha)-\mu}{\sigma}$, where $\sigma > 0$ , $\mu \in \mathbb{R}$. Some of the parameters used in the model are explained below :

## Description of the model 1

Utility parameters are : $\gamma = 1.5; \alpha = 2.55; \sigma = 3$. Utility function is given as:

*Equation 23*

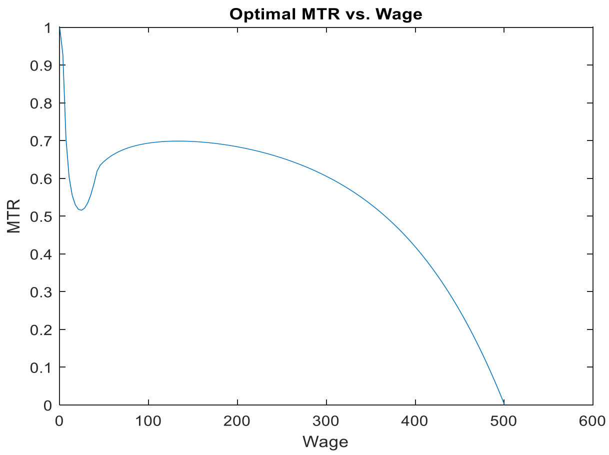
$$u = \frac{c^{1-\gamma} - 1}{1-\gamma} - \frac{\alpha^{\sigma}}{\sigma}$$

Adjusting for marginal tax rates half-way schedule is [24]:

*Equation 24*

$$tax\_marg = \frac{(tax\_marg + tax\_marg\_alternative)}{2}$$

*Figure 1 MTR schedule and distribution of wages*




From the previous graph one can see that for the top earners marginal tax rate equal zero, as is predicted in Mirrless (1971). Next , as representative or seminal paper in the numerical methods

[24] This part contains part of the MATAB code used for computation in Mankiw NG, Weinzierl M, Yagan D.(2009).

used to calculate optimal taxation models is this one based on the [Aiyagari model (1994)] and [Aiyagari model (1995)] with infinitely-lived households, labour endowment shocks l(t) following an AR(1) process, where HH's solve both the consumption-savings problem and labour supply problem. HH's can save via capital K at interest rate r (endogenously determined) but cannot borrow. HH's have convex preferences and the government imposes labour and capital income tax as well as consumption tax. The government uses tax revenues to finance an exogenous level of G (government spending) and a representative firm maximizes their profits given aggregate capital K aggregate labour N, the wage rate w and depreciation delta. Individuals are subject to exogenous income shocks. These shocks are not fully insurable because of the lack of a complete set of Arrow-Debreu contingent claims, Arrow , K., (1953). Incomplete markets case is when at date 0 there is trade on $K \leq S$ assets, i.e. number of Arrow-Debreu securities is less or equal than the states of nature[25].

## Description of the model 2

Parameters of the benchmark model are :

$z = 1$;- total factor productivity ;$\alpha = 0.4$; - production function parameter (share of production due to capital) ;$\delta = 0.08$; - proportion of capital saved today for the next period, $\beta$ = 0.96; - discount factor ;$\rho = 0.90$; - parameters of labor endowment shock process l(t), and $\rho$ being the autocorrelation coefficient for the AR(1) process: $log\big(l(t+1)\big) = \rho * log(l(t)) + \epsilon(t)$,where $\sigma = 0.20$; $\epsilon(t)$ is normally distributed with mean zero and standard deviation $\sigma$. $\tau_{y_{bench}} = 0.3$; labour and capital income tax rate for benchmark , $\tau_{c_{bench}}$ = 0.075; consumption tax rate for benchmark. $\lambda = 2$; % utility function parameter for HH preferences $\mu = 0.10$; parameter used for determining equilibrium interest rate and NA = 400; is the number of intervals in A grid-space, for assets (analogous to K). and NL = 5; is number of "l" states, for labour efficiency endowment (analgous to Z).Initial value function is V_benchmark(1:NL,1:NA) = 0. Intial guess for the interest rate is : dist_r = 1;$r = \frac{1}{\beta} - 1 - 0.001$; $r = 0.0379$.Results for this section are $K_{bench} = 7.4287$[26];$N_{bench} = 1.1838$[27]; $G_{bench} = 1.1192$[28]; $Y_{bench}$ =2.4679; $aggregate_{c_{bench}} = 7.4287$[29];$aggregate_{v_{bench}} = -33.0068$[30];$\tau_{c_{bench}} = 0.0750$[31]. A value function is often denoted v() or V(). Its value is the present discounted value, in consumption or utility terms, of the choice represented by its arguments. The classic example, from [Stokey, N., Lucar,R.. and Prescott ,E.(1989)], is:

*Equation 25*

$$v(k) = maxk' \{ u(k,k') + \beta v(k') \}$$

where k is current capital, k' is the choice of capital for the next (discrete time) period, u(k, k') is the utility from the consumption implied by k and k', $\beta$ is the period-to-period discount factor.In the reformed economy new values of some of the parameters are : $\tau_{y_{reform}} = 0$; here we set the labour

[25] When security is sold, when $s$ state occurs, money is transferred in a way determined by the securities, and the allocation of commodities occurs at market in a usual way, without further risk bearing.
[26] Benchmark capital
[27] Benchmark labor
[28] Governments balanced budget
[29] Aggregate benchmark consumption
[30] Initial guess for value function
[31] Consumption tax benchmark value

and capital income tax rate for the reform economy as 0.And $\tau_{c_{reform}} \;\; = \;\; 0.1507;$ % here we set the consumption tax for the reform economy according to the definition:

*Equation 26*

$$\tau_{c_{reform}} = \frac{G_{bench}}{aggregate_{c_{bench}}} = \frac{1.1192}{7.4287} \approx 0{,}150658930903119$$

In the reform economy expected results are: $K_{reform} \;\; = 9.1932; N_{reform} \; = 1.1838; \; G_{reform} \; = 1.1192; Y_{reform} \; = 2.6875; aggregate_{c_{reform}} \; = 9.1932; aggregate_{v_{reform}} \; = -26.5039; \tau_{c_{reform}} \; = \;\; 0.1217.$

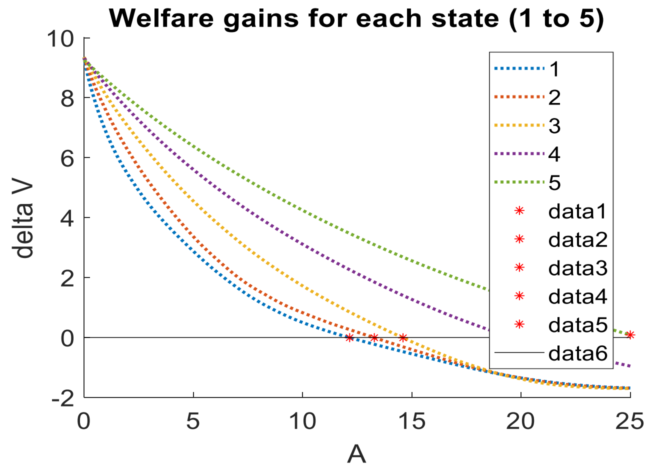
Equilibrium interest rate and wage rate for the benchmark economy: r=0.0529 ; w=1.2508.Equilibrium interest rate and wage rate for the reform economy: $r = \;\; 0.0370 \;\;\; w = 1.3617$.Next, equations about the changes between benchmark and reform economies are given;

1. $\Delta K \; = \; \big(K_{reform} \; - \; K_{bench}\big)/K_{bench} \; ) .* \; 100;$
2. $\Delta Y \; = \; ((Y_{reform} \; - \; Y_{bench}) \; / \; Y_{bench} \; ) .* \; 100;$
3. $\Delta aggC \; = \; ((aggregate_{c_{reform}} \; - \; aggregate_{c_{bench}}) \; / aggregate_{c_{bench}}) \; * \; 100;$
4. $\Delta aggV \; = \; abs\left(\frac{aggregate_{v_{reform}} - aggregate_{v_{bench}}}{(aggregate_{v_{bench}})}\right) \; * \; 100;$
5. $\Delta r \; = \; ((computed_{r_{reform}} \; - \; computed_{r_{bench}}) \; / \; (computed_{r_{bench}}) \; * \; 100;$
6. $\Delta w \; = \; \left(\frac{(computed_{w_{reform}} - computed_{w_{bench}})}{computed_{w_{bench}}}\right) \; * \; 100;$

*Table 1 Changes between benchmark and reform economies*

| K | Y | aggr C | aggr V | r | w |
|---|---|---|---|---|---|
| 237.521 | 88.983 | 237.521 | 197.019 | -300.839 | 88.724 |

*Figure 2 Welfare gains of each state*




On the previous graphs are presented welfare gains from each of the five states, for which transitional probabilities were calculated, these are the number of "l" states, for labor efficiency endowment.

*Table 2 Transition probabilities matrix (benchmark model)*

| | | | | |
|---|---|---|---|---|
| 0.8491 | 0.1509 | 0.0000 | 0.0000 | 0.0000 |
| 0.0195 | 0.8962 | 0.0843 | 0.0000 | 0.0000 |
| 0.0000 | 0.0427 | 0.9147 | 0.0427 | 0.0000 |
| 0.0000 | 0.0000 | 0.0843 | 0.8962 | 0.0195 |
| 0.0000 | 0.0000 | 0.0000 | 0.1509 | 0.8491 |

*Table 3 Transition probabilities matrix (reform model)*

| | | | | |
|---|---|---|---|---|
| 0.8491 | 0.1509 | 0.0000 | 0.0000 | 0.0001 |
| 0.0195 | 0.8962 | 0.0843 | 0.0000 | 0.0000 |
| 0.0000 | 0.0427 | 0.9147 | 0.0427 | 0.0000 |
| 0.0000 | 0.0000 | 0.0843 | 0.8962 | 0.0195 |
| 0.0000 | 0.0000 | 0.0000 | 0.1509 | 0.8491 |

Recall that the value function describes the best possible value of the objective, as a function of the state $x$.

*Figure 3 Value functions (benchmark economy)*

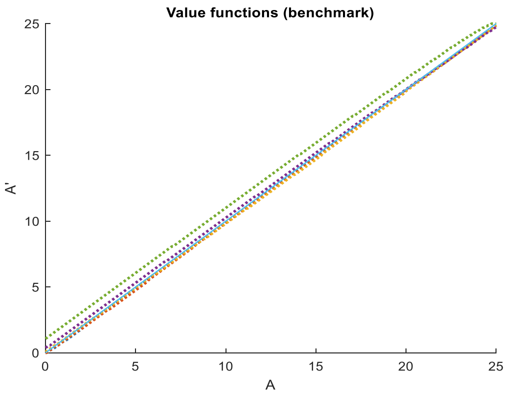
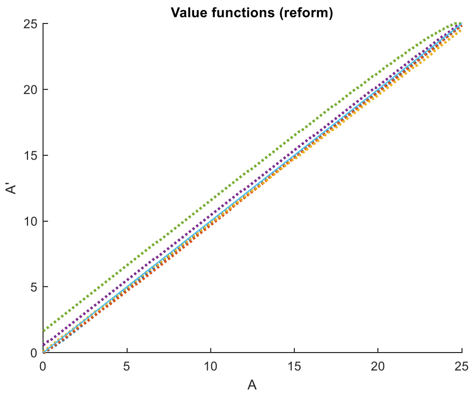



*Figure 4 Value functions (reform economy)*



## Description of model 3

This is dynamic Ramsey taxation model[32] and the model parameters are :

$\sigma = 2; \beta = 0.99; \theta = 0.38; \delta = 0.08; \alpha = 0.30; A = 10;$ . In this model utility of the representative agent with preferences is :

*Equation 27*

$$u(c,l) = \frac{1}{1-\sigma}\left[c^{\theta}(1-l)^{1-\theta}\right]^{1-\sigma}$$

The discount factor is $\beta$ . technology is Cobb-Douglas with :

*Equation 28*

$$F(k,l) = Ak^{\alpha}l^{1-\alpha}$$

Steady state values of the variables of interest are given as:

1. $R = \frac{1}{\beta}$
2. $r = \frac{R-1}{1-k} + \delta$
3. $\frac{k}{l} = \left(\frac{r}{\alpha A}\right)^{\frac{1}{\alpha-1}}$
4. $w = A(1-\alpha)\left(\frac{k}{l}\right)^{\alpha}$
5. $\frac{c}{l} = AF\left(\frac{k}{l},1\right) - \delta\left(\frac{k}{l}\right) = A\left(\frac{k}{l}\right)^{\alpha} - \delta\left(\frac{k}{l}\right)$
6. $l = \frac{\theta w}{(1-\theta)\left(\frac{c}{l}\right)+\theta w}$
7. $k = \frac{k}{l}l$
8. $c = \frac{c}{l}l$
9. $y = Ak^{\alpha}l^{1-\alpha}$

Evolution of capital around steady-state shows that $k = I + (1-\delta)k$ or that $\delta = \frac{1}{k}$.In 2007 in US economy investment accounted of 3.8 trillions US/dollars , capital was 47.9 trillions/US dollars . This implies that $\delta = 0.079$. Now:

*Equation 29*

$$\beta = R^{-1} = \left(1 + (1-k)(r-\delta)\right)^{-1} = (1 + (1-k)\left(a\left(\frac{k}{y}\right)^{-1} - \delta\right)^{-1}$$

By assumption $\alpha = 0.33$,and for the labor tax we have :

*Equation 30*

$$\frac{(1-\alpha)(1-\tau_l)y}{c} = \frac{1-\theta}{\theta}\frac{l}{1-l}$$

[32] MATLAB code written by Florian Scheuer, 2007

If $l = 0.31$ and labor income tax is $\tau_l = 0.28$, consumption in 2007 were 9.8 trillions us/dollars leads to that $\theta = 0.39$, and now we know :

*Equation 31*

$$A = \frac{y}{k^{\alpha} l^{1-\alpha}} = 8.6$$

Steady state value of capital is given as:

*Equation 32*

$$\tilde{V}\big(k(k)\big) = \frac{1}{(1-\beta)(1-\delta)}\left[c(k)^{\theta}\big(1 - l(k)\big)^{1-\theta}\right]^{1-\sigma}$$

We can compare this value $\tilde{V}\big(k(k)\big)$ with the one where value of the capital tax is zero $\tilde{V}\big(k(0)\big)$.

*Equation 33*

$$\frac{1}{(1-\beta)(1-\delta)}\left[1 + \lambda(k)c(k)^{\theta}\big(1 - l(k)\big)^{1-\theta}\right]^{1-\sigma} = \tilde{V}\big(k(0)\big)$$

As soon as the capital income tax is zero, by the Bellman equation :

*Equation 34*

$$V(k) = \max_{c,lk'} u(c,l) + \beta V\,(k')$$

Subject to constraints:

*Equation 35*

$$c + k' = F(k,l) + (1-\delta)k$$

Marginal utilities for the consumption and labor are given as:

*Equation 36*

$$u_c\,(c,l) = \left[c^{\theta}\ (1-\ l)^{1-\theta}\right]^{-\sigma} \theta c^{\,\theta-1}\,(1-\ l)^{1-\theta}\ ;u_l\,(c,l) = -\left[c^{\theta}\,(1-\ l)^{1-\theta}\right]^{\sigma} c^{\theta}\,(1 - \theta)(1-\ l)^{-\theta}$$

Intertemporal optimality condition is given as:

*Equation 37*

$$w\left[c^{\theta}(1-l)^{1-\theta}\right]^{-\sigma}\theta c^{\theta-1}(1-l)^{1-\theta} = \left[c^{\theta}(1-l)^{1-\theta}\right]^{-\sigma}c^{\theta}(1-\theta)(1-l)^{-\theta} \Rightarrow w\theta c^{\theta-1}(1-l)^{1-\theta} = c^{\theta}(1-\theta)(1-l)^{-\theta} \Rightarrow w\theta(1-l) = c(1-\theta) \Rightarrow w\theta(1-l) = \frac{c}{l}l(1-\theta) \Rightarrow l = \frac{\theta w}{(1-\theta)\left(\frac{c}{l}\right) + \theta w}$$ [33]

[33] $\beta^t u(c_t, l_t) = \lambda q_t; \beta^{t+1}u(c_{t+1}, l_{t+1}) = \lambda q_{t+1};\ \beta^t u_l(c_t, l_t) = -\lambda q_t w_t$

Next some of the previous calculations are calculated in MATLAB :

*Figure 5  Steady state value of consumption, investment and output as a function of grid of the capital tax  kappa=0*

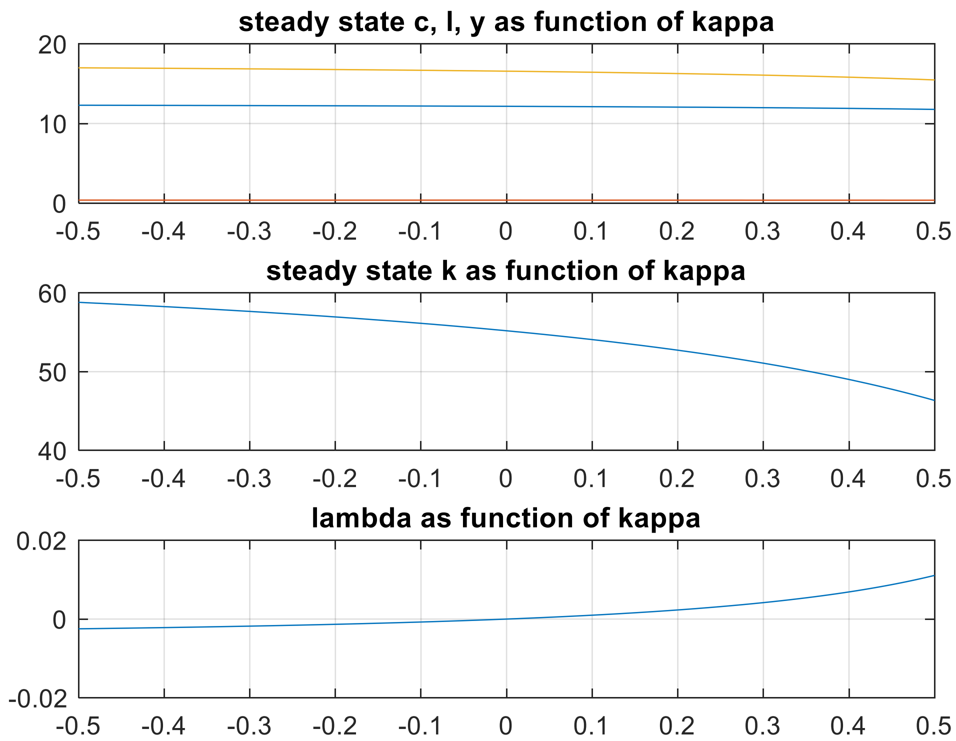
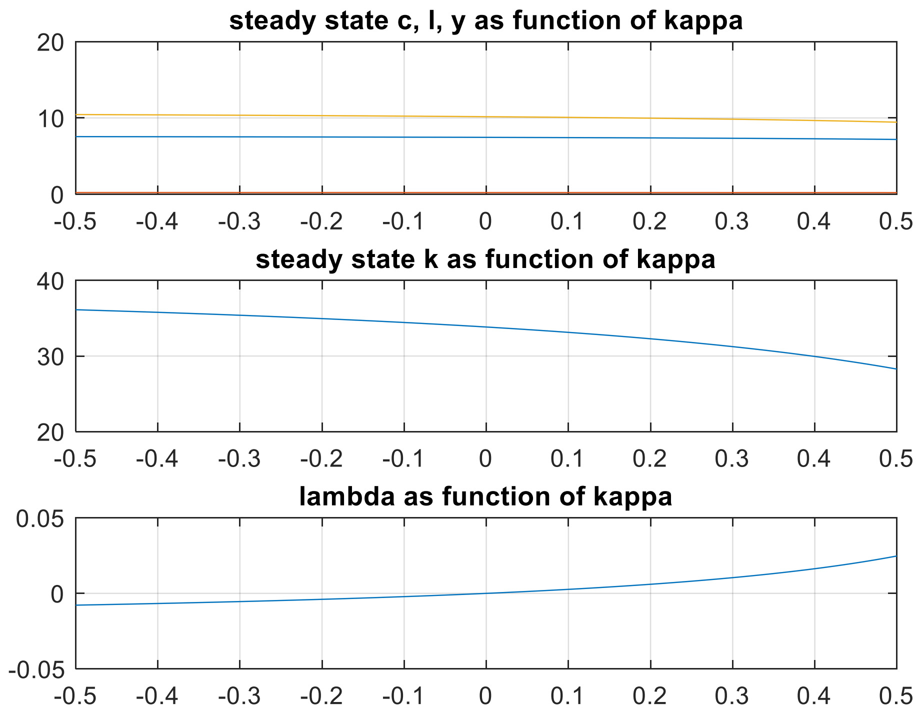



*Figure 6 Steady state value of consumption, investment and output as a function of grid of the capital tax  kappa=0.5*



## Conclusion/s

The theory of optimal taxation represents the study of designing and implementing a tax that maximizes a social welfare function subject to some economic constraints. This paper made attempt to review the past and the current literature on the optimal tax theory, empirical and theoretical. The developments of the tax theory have improved the tax policies in the past. For instance, worldwide trends towards reduction of capital income taxation in enacted by law tax rates. But also, we have seen that there are justifications for taxation of capital. Weak separability is the first reason that is bad when it comes to choose between present and future consumption. Though there is a lack of empirical proofs of this statement. The second reason being that if agents receive inheritance that is not taxed by the tax on bequest, then it may be optimal to tax the capital income. Though taxation of bequests (from which consumer derives utility until his death) by the Hicks-Leontief theorem Then another justification for capital taxation in the economy is the sub-optimality of the capital accumulation in the economy. In such a case when economies suffer from too little capital and when agents have a Cobb-Douglas utility function it will be useful to tax the capital and to transfer the revenue to the young population. Also, the last result can be achieved by imposing lump-sum tax to the old and subsiding the capital. So there is no a strong argument that support taxing capital same as income. But as in Stiglitz (1985) ,model of capital taxation were a skill premium (skill premium is defined as ratio of the wages skilled versus unskilled workers) is equal to the relative productivity. Ig this relative wage depends on the capital intensity (capital intensity is the amount of fixed capital in relation to other factors of production, especially labor i.e. capital to efficient labor ratio), and there is a good proof in the empirical literature that confirm that unskilled labor is a good substitutable to capital than skilled labor. Increase in a capital intensity increases relative wage of skilled versus unskilled labor, so that productivity as a function of capital is a decreasing function. Results from the paper also provide rationale for distortions (upward and downward) in the savings behavior in a simple two period model where high-skilled and low-skilled have different non-observable time preferences beyond their earning capacity. In the comparisons between benchmark(one with labor tax and consumption tax) and reform economy (no income tax and no capital tax only consumption tax) : benchmark capital stock is lower in the benchmark economy, benchmark labor is the same (is not affected by the taxation), Reform output is higher than the benchmark output, aggregate benchmark consumption is lower than the reform economy consumption, initial guess for value function is higher in the reform economy ,in the reform economy consumption tax benchmark value is higher than the benchmark economy. In the model of dynamic taxation: Ramsey taxation model, as it can be spotted from the results there is not much difference from the different consumption tax rate. These results prove that consumption tax is more optimal for the economy than income or capital tax. Though these results cannot be generalized for the actual economies. That is to say, confirmation of Diamond and Mirrlees (1971b),result, that tax system can be designed to minimize distortions and disincentives, and to eliminate production inefficiencies. Actually, this paper opened topic of research on tax mechanism design and minimization of tax incidence (tax burden), Arrow, K. et al. (2001). These results furthermore can be empirically tested in some well-defined models.

## References


1. Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," Quarterly Journal of Economics, 109(3), 659–684
2. Aiyagari, S. R. (1995): "Optimal Capital Income Taxation with Incomplete Markets,Borrowing Constraints, and Constant Discounting," Journal of Political Economy, 103(6),1158–75.
3. Akerlof, G. (1978)"The Economics of Tagging as Applied to the Optimal Income Tax,Welfare Programs, and Manpower Planning", American Economic Review, Vol. 68, 1978, 8-19. (web)
4. Arrow, K. (1953), "The Role of Securities in the Optimal Allocation of Risk Bearing", Review of Economic Studies, 1964, 31, 91-96.
5. Arrow, K. J., Bernheim, B. D., Feldstein,M.S., McFadden,D.,L., Poterba,J.M., Solow,R.,M., (2011). "100 Years of the American Economic Review: The Top 20 Articles." American Economic Review, 101 (1): 1-8. DOI: 10.1257/aer.101.1.1
6. Atkinson, A.B. and Stiglitz,J. (1976)"The design of tax structure: Direct versus indirect taxation",Journal of Public Economics, Vol. 6, 1976, 55-75. (web)
7. Atkinson, A.B. and A. Sandmo (1980) "Welfare Implications of the Taxation of Savings", Economic Journal, Vol. 90, 1980, 529-49. (web)
8. Atkinson, A.B. and Stiglitz,J. (1980)"Lectures on Public Economics", Chap 14-4 New York: McGraw Hill, 1980. (web)
9. Auerbach, A. (2009). "The choice between income and consumption taxes: A primer". Cambridge: Harvard University Press.
10. Auerbach, A., Kotlikoff,L. (1987a)."Evaluating Fiscal Policy with a Dynamic Simulation Model",American Economic Review, May 1987, 49?55
11. Auerbach,A.,J.,Kotlikoff,L.J.,(1987b), "Dynamic fiscal policy", Cambridge University Press
12. Barro,R.,J.(1979),"On the determination of public debt", Journal of political economy , 87:940 971
13. 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)
14. Chamley, C. (1986). "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives". Econometrica. 54 (3): 607–622. doi:10.2307/1911310.
15. Chari, V.V. ,Kehoe, P.J., (1999)." Optimal fiscal and monetary policy". Handbook of macroeconomics, 1, pp.1671-1745.
16. Cooley, T. (1995), "Frontiers of Business Cycle Research", Princeton University Press,Princeton NJ.
17. Diamond, P.(1998), "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", American Economic Review, 88, 83–95.
18. Diamond, P.,A.,Mirrlees , J.A.(1978)."A model of social insurance with variable retirement", Journal of public economics,10, pp.295-336
19. Diamond,P.,Helms,J. and Mirrlees (1978) "Optimal taxation in a stochastic economy, A Cobb-Douglas example", M.I.T. Working Paper no. 217
20. Fair,R.C. (1971), "The Optimal Distribution of Income,"Quart. J. Econ., Nov. 1971, 85, 55779.
21. Farhi, E., Werning, I. (2010): "Progressive Estate Taxation," Quarterly Journal of Economics,125 (2), 635–673. [1852,1853,1864-1866]
22. Feldstein, M.(1978), "The Welfare Cost of Capital Income Taxation", Journal of Political Economy, Vol. 86, No. 2, Part 2: Research in Taxation (Apr., 1978), pp. S29-S51

23. Hammond, P.(2000)." Reassessing the Diamond/Mirrlees Efficiency Theorem",Department of Economics, Stanford University, CA 94305-6072, U.S.A
24. Judd, K.L. (1985) "Redistributive taxation in a simple perfect foresight model". Journal of Public Economics 28, 59–83
25. Judd, K.L. (1999) "Optimal Taxation and Spending in General Competitive Growth Models"Journal of Public Economics 71: 1–26
26. Kaplow, L. (1994). "Taxation and risk taking: A general equilibrium perspective". National Tax Association 47 (4), 789–798.
27. Kaplow, L.(2006). "On the undesirability of commodity taxation even when income taxation is not optimal", Journal of Public Economics, Vol.90, 2006, 1235-1260. (web)
28. Kotlikoff, L. ,Summers,L. (1981). "The Role of Intergenerational Transfers in Aggregate Capital Accumulation", Journal of Political Economy, Vol. 89, 1981, 706-732.
29. Kreps, David M.(1988)," Notes on the Theory of Choice". Westview Press ,chapters 2 and 5.
30. Laroque, G.(2005)."Indirect Taxation is Superfluous under Separability and Taste Homogeneity: A Simple Proof", Economic Letters, Vol. 87, 2005, 141-144.
31. Lee,D., Saez,E. (2012). "Optimal minimum wage policy in competitive labor markets," Journal of Public Economics, vol 96(9-10), pages 739-749.
32. Mankiw NG, Weinzierl M, Yagan D.(2009)," Optimal Taxation in Theory and Practice". Journal of Economic Perspectives. 2009;23 (4) :147-174.
33. Mirrlees, J. A. (1971). "An Exploration in the Theory of Optimum Income Taxation". The Review of Economic Studies. 38 (2): 175–208. doi:10.2307/2296779. JSTOR 2296779
34. Mirrlees, J.A. (1982), "Migration and Optimal Income Taxes", Journal of Public Economics,Vol. 18, 1982, 319-341. (web)
35. Mirrlees, J.A., (1976). "Optimal tax theory: a synthesis". Journal of Public Economics 6, 327–358.
36. Mirrlees,J.A., Diamond,P.(1971b). "Optimal Taxation and Public Production II: Production Efficiency". American Economic Review. 61: 261-278.
37. Mirrlees,J.A.,(1986), "The theory of optimal taxation", Chapter 24 Handbook of Mathematical Economics, edited by K. J. Arrow and M. Intriligator, eds. (Amsterdam, North-Holland, 1986)
38. Mirrlees,J.A.,Diamond,P.(1971a). "Optimal Taxation and Public Production I: Production Efficiency". American Economic Review. 61: 8–27.
39. Modigliani, F. (1966). "The Life Cycle Hypothesis of Saving, the Demand for Wealth and the Supply of Capital". Social Research. 33 (2): 160–217. JSTOR 40969831.
40. Modigliani, F. (1976)."Life-cycle, individual thrift, and the wealth of nations," American Economic Review, 76(3), 297–313.
41. Modigliani, F. (1986). "Life cycle, individual thrift and the wealth of nations". American Economic Review 76 (3), 297–313
42. Modigliani, F. (1988)."The Role of Intergenerational Transfers and Lifecyle Savings in the Accumulation of Wealth", Journal of Economic Perspectives, Vol. 2, 1988, 15-40.
43. Modigliani,F.,Brumberg,R.H.(1954), "Utility analysis and the consumption function: an interpretation of cross-section data," in Kenneth K. Kurihara, ed., PostKeynesian Economics, New Brunswick, NJ. Rutgers University Press. Pp 388–436.
44. Ordover, J., Phelps,E. (1979). "The concept of optimal taxation in the overlapping generations model of capital and wealth". Journal of Public Economics 12: 1–26.
45. Piketty, T. , Saez,E., (2013) "A Theory of Optimal Inheritance Taxation", Econometrica, 81(5), 2013, 1851-1886.
46. Piketty, T., Postel-Vinay,G.,Rosenthal,J.L,(2014). "Inherited vs. Self-Made Wealth: Theory and Evidence from a Rentier Society (1872-1927)," Explorations in Economic History

47. Piketty,T., Saez,E., Stantcheva,S.( 2014). "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," American Economic Journal: Economic Policy, American Economic Association, vol. 6(1), pages 230-71
48. Ramsey, F. (1927). "A Contribution to the Theory of Taxation". Economic Journal. 37: 47 61. doi:10.2307/2222721.
49. Roy, R.(1947). "La Distribution du Revenu Entre Les Divers Biens". Econometrica. 15 (3): 205–225. JSTOR 1905479.
50. Sadka,E.(1976)," On Income Distribution, Incentive Effects and Optimal Income Taxation",The Review of Economic Studies, Vol. 43, No. 2, (Jun., 1976), pp. 261-267
51. Saez, E. (2001). "Using Elasticities to Derive Optimal Income Tax Rates". Review of Economic Studies. 68: 105–229. doi:10.1111/1467-937x.00166.
52. Saez, E. ,S. Stantcheva (2016a). "A Simpler Theory of Optimal Capital Taxation", NBER Working Paper 22664, 2016.
53. Saez, E. ,S. Stantcheva (2016b). "Generalized social marginal welfare weights for optimal tax theory". The American Economic Review, 106(1), pp.24-45.
54. Saez, E., (2002). "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes". Journal of Public Economics, 83(2), pp.217- 230.
55. Saez,E.(2001)."Using elasticities to derive optimal income tax rates", The Review of Economic Studies, 68(1), pp.205- 229.
56. Salanie,B.(2003)."The Economic of Taxation". The MIT Press Cambridge, Massachusetts London, England
57. Seade, J. K., (1977), "On the shape of optimal tax schedules", Journal of Public Economics, 7, issue 2, p. 203-235.
58. Stiglitz, J. (1985). "Inequality and capital taxation". IMSSS Technical Report 457, Stanford University.
59. Stiglitz, J.(1982) "Self-Selection and Pareto Efficient Taxation," Journal of Public Economics, 1982,17, 213–240.
60. Stokey, N., Lucar,R.. and Prescott ,E.(1989). Recursive Methods in Economic Dynamics. Harvard University Press. ISBN 0-674-75096-9.
61. Teodorescu,S.(2010), "On the truncated composite lognormal-Pareto model", MATH. REPORTS 12(62), 1 (2010), 71–84
62. Tresch, R. W. (2008). "Public Sector Economics". 175 Fifth Avenue, New York, NY 10010: PALGRAVE MACMILLAN. p. 67. ISBN 978-0-230-52223-7.
63. Tuomala, M.(1990). "Optimal Income Tax and Redistribution". New York: Oxford University Press.
64. Varian,H.R.(1980)."Redistributive taxation as social insurance". Journal of public Economics,14(1), pp.49-68.