UDK 336.226.111:336.225
DOI https://www.doi.org/10.46763/YBFTBL2221030j

DYNAMIC OPTIMAL TAXATION MIRRLEES' APPROACH: A REVIEW

Dushko Josheski[1], Tatjana Boshkov[2], Mico Apostolov[3]

[1]Assistant professor, UGD-Shtip, R. North Macedonia, dusko.josevski@ugd.edu.mk
[2]Associate professor, Goce Delcev University of Stip,email: tatjana.dzaleva@ugd.edu.mk
[3] Professor Goce Delcev University of Stip,email: mico.apostolov@ugd.edu.mk

Abstract
This paper reviews recent advances in dynamic capital taxation considering the dynamic Mirrlees approach. Dynamic taxes are not restricted ex ante and are set for redistribution and insurance considerations. Capital is taxed in order to improve incentives to work. On average wealth taxes across individuals are zero ex-ante. 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. The fact that the capital tax varies in this regressive way makes investment risky and creates a positive risk premium, which explains how it is possible to have a positive intertemporal wedge/tax. Even though taxes are zero ex-ante.
Keywords: dynamic Mirrlees model, dynamic taxes, inverse Euler equation, labor taxes, capital taxes, heterogenous ability

Introduction

When we think about the optimal taxation the obvious questions that we are raising are: whether taxes depend on age, or whether taxes depend on past history, or whether the optimal tax system should be progressive or regressive. The standard theory of optimal taxation posits that a tax system [28]should be chosen to maximize a social welfare function subject to a set of constraints, see Mankiw, Wenzierl,Yagan (2009).This paper will review some selected advances in the theory of dynamic taxation. The approach we will investigate is by papers based on Mirrlees (1971).This approach by the static Mirrlees(1971) supposes the idea that agents' productive abilities are private information and are heterogenous, and lets productive ability to evolve in time, see Stancheva (2020).Static taxation is studied in well known distinguished papers by: Mirrlees (1971), Diamond (1998), Saez(2001), Werning (2007).Dynamic taxation most famous example in the literature are: Diamond-Mirrlees (1978); Albanesi-Sleet (2006), Shimer-Werning (2008), Ales-Maziero (2009), Golosov-Troshkin-Tsyvinsky (2011).A sizeable literature in NDPF(New Dynamic Public Finance) studies optimal taxation in dynamic settings, see also(Golosov, Kocherlakota, and Tsyvinski (2003), Golosov,Tsyvinski, and Werning (2006), Kocherlakota (2010). The government's inability to condition taxes directly on skills ends up implying that it has to treat agents as being privately informed about their productivities. It follows that the optimal tax problem in the NDPF is isomorphic to a dynamic contracting problem between a risk-neutral principal and a risk-averse agent who is privately informed about productivities, see Kocherlakota (2010).There exists a large literature on the dynamic principal agent models , which includes the works by: Rogerson (1985), Spear,Srivastava (1987), Green (1987), and Atkeson ,Lucas (1995). The analysis of the static Mirrlees problems (Mirrlees (1971), Atkinson,Stiglitz (1976), Tuomala (1990)) also points out that if the planner is more redistributive than utilitarian planner, the tax policy is substantially different from linear, and nonlinear taxes may yield large welfare gains. Models of optimization in these papers from NPDF literature and developing IC constraints follow methods developed by: Fernandes-Phelan (2000), Werning (2002), Abraham-Pavoni (2008), Kapicka (2013), Williams (2011), Pavan-Segal-Toikka (2014).In the static MIrrlees model developed by Diamond (1998) the optimal labor distortions depended on three parameters: the shape of the income distribution, the redistributionary objectives of the government, and labor elasticity. In the dynamic taxation model three differences are introduced: "(i)the use of dynamic incentives adds a force lowering labor wedges; (ii) conditional rather than unconditional distributions of skills are key determinants of wedges; (iii) persistence of shocks acts as a more redistributionary motive for the planner". See Golosov-Troshkin-Tsyvinsky (2011). One of the main findings from the dynamic Mirrlees literature is that:" taxes will be optimally smoothed over the life cycle, and that they will be featuring a persistent component that depends on last period's taxes and a drift term that captures the insurance motive", see Stantcheva (2020). Savings are typically discouraged at the optimum relative to the free-savings case because higher levels of assets and lower work effort are

[28] Optimal tax system should not be confused with Pareto efficient tax system. One definition we can adopt for the Pareto efficient tax system is given is Stiglitz (2018) as follows: Pareto efficient tax structures are those (given the admissible set of taxes and the required public revenue) which are such that no one can be better off without making someone worse off. What does "optimal" means? – "ex ante Pareto optimal tax systems, in which the government's redistributional motives are based on attributes other than skills themselves" , see Kocherlakota (2010).

complements. This is the inverse Euler logic that arises when labor effort is not observed and needs to be incentivized. This result is different from that by Chamley(1986)–Judd(1985) zero capital income tax result. So, this paper will review models developed in Stantcheva (2020), Golosov et al. (2003), and Farhi, Werning (2013). There is additional 2 period simple Mirrlees model explained.

The dynamic Mirrlees approach:I (per Stantcheva (2020))

Economy consists of agents who live T year , they work $l_t \geq 0$ hours per period t at wage rate w_t their income is $y_t = w_t l_t$, disutility of labor is $\phi_t(l_t)$ which is strictly increasing and convex. The wage rate $w_t = \theta_t$ is equal to ability θ_t , or the notation is $w_t(\theta_t)$, gross interest rate from physical capital accumulation is RInvestments in the physical capital are called savings. Ability of agents is heterogenous θ_t and the distribution of ability is $f^1(\theta_1)$, earning ability evolves according to Markov process[29] with time varying transition function $f^t(\theta_t|\theta_{t-1})$ and the support of the function is $\Theta \equiv [\underline{\theta}, \overline{\theta}]$. Agents per period utility separable in consumption and labor is:

equation 1

$$\tilde{u}_t(c_t, y_t, \theta_t) = u_t(c_t) - \phi_t\left(\frac{y_t}{\theta_t}\right)$$

u_t is increasing twice differentiable and concave. Now, θ^t is the history of ability shocks up to period t and Θ^t the set of possible histories at t, and $P(\theta^t)$ the probability of a history θ^t ; so that now we have : $P(\theta^t) = f^t(\theta_t|\theta_{t-1}) \dots f^2(\theta_2|\theta_1) f^1(\theta_1)$ .An allocation $x_t|\theta^t = x_t = \{x_t(\theta^t)\}_{\Theta_t} = \{c(\theta^t), y(\theta^t)\}_{\Theta_t}$.Utility of lifetime allocation discounted by the discount factor β is given by:

equation 2

$$U(c(\theta_t), y\,(\theta^t)) = \sum_{t=1}^{T} \int \beta^{t-1} \left[u_t\big(c(\theta^t)\big) - \phi_t\left(\frac{y(\theta^t)}{\theta_t}\right)\right] P(\theta^t) d\theta^t$$

Where in previous $d\theta^t \equiv d\theta_t, \dots, d\theta_1$.The planning problem is set up as follows. In every period the planer can observe the agents' outputy_t and consumption c_t, but ability θ_t is not observable and neither is labor supply $l_t = \frac{y_t}{\theta_t}$. So, if an agent produces low output, the planner does not know whether it was labor effort or ability that was low. This problem uses FOA (first order approach) that replaces the (infinite) set of incentive compatibility constraints with agents' envelope conditions, the program is made recursive, using as state variables the promised utility and its gradient. Now let's imagine a direct revelation mechanism[30] where in each period agents report their ability θ_t, where reporting strategy is $r_t = \{r_t(\theta^t)\}_{t=1}^{T}$., $\mathcal{R}$ represents the set of all reporting strategies $r^t = \{r_1(\theta_1), \dots, r_t(\theta^t)\}$ which is the history from this reporting strategy r_t .Continuation value after the history is denoted by $\omega^r(\theta^t)$ is the solution to:

equation 3

$$\omega^r(\theta^t) = u_t\big(c(r^t(\theta^t)\big) - \phi_t\left(\frac{y(r^t(\theta^t)}{\theta_t}\right) + \beta \int \omega^r(\theta^{t+1}) f^{t+1}(\theta_{t+1}|\theta_t) d\theta_{t+1}$$

The continuation value under truthful revelation is given as:

equation 4

$$\omega(\theta^t) = u_t\big(c(\theta^t)\big)\phi_t\left(\frac{y\big(r^t(\theta^t)\big)}{\theta_t}\right) + \beta \int \omega\,(\theta^{t+1}) f^{t+1}(\theta_{t+1}|\theta_t) d\theta_{t+1}$$

Incentive compatibility as per Stantcheva (2020) means imposing that truth-telling yields at least weakly higher continuation utility than any other reporting strategy; that is, we have incentive compatibility:

inequality 1

$$IC: \omega(\theta_1) \geq \omega^r(\theta_1); \forall \theta_1, \forall r$$

x^{ic} denotes the set of incentive compatible allocations ,to solve this FOA the set of assumptions used by Farhi, Werning (2013) and Stantcheva (2017) are being used.Lets consider a history θ^t and a deviation strategy $\tilde{r}_t$,

[29] A random process whose future probabilities are determined by its most recent values.A stochastic process $x(t)$ is called Markov if for every n and $t_1 < t_2 < \dots . . < t_n$ we have: $P\big(x(t_n)\big) \leq x_n \big| x\big(t_{n-1}, \dots, x(t_1)\big) = P\big(x(t_n) \leq x_n \big| x(t_{n-1})\big) \equiv P\big(x(t_n)\big) \leq x_n \big| x(t); \forall t \leq t_{n-1}\big) = P\big(x(t_n) \leq x_n | x(t_{n-1})\big)$, see Papoulis (1984)

[30] A direct revelation mechanism is one where each agent is asked to reveal its individual preferences, in which case $M = \Theta$ and $f = g$ where $\Theta = \Pi_{i \in \{1,2,..I\}}$, and agents are drawn by some known distribution and I are agents $i \in 1,2, \dots n$

under which the agents report thruthfully until period $r(\tilde{t}_s\,(\theta^s) = \theta_s, \forall s \leq t-1)$ but report $\tilde{r}_t(\theta^t) = \theta^t \neq \theta_t$ in period t . Under this assumption continuation utility is the solution to:

equation 5

$$\omega^{\tilde{r}}(\theta^t) = u_t(c(\theta^{t-1},\theta') - \phi_t\left(\frac{y(\theta^{t-1},\theta'}{\theta_t}\right) + \beta\int \omega^{\tilde{r}}(\theta^{t-1},\theta',\theta_{t+1})\, f^t(\theta_{t+1}|\theta_t)d\theta_{t+1}$$

IC in $IC: \omega(\theta_1) \geq \omega^r(\theta_1); \forall\theta_1, \forall r$ implies that after all θ^t , the temporal IC holds and we obtain:

equation 6

$$\omega(\theta^t) = \max_{\theta'} \omega^{\tilde{r}}(\theta^t)$$

Envelope theorem condition for every agent neccesary for IC see Milgrom Segal (2002):

equation 7

$$\dot{\omega}(\theta^t) := \frac{\partial\omega(\theta^t)}{\partial\theta_t} = \frac{w_{\theta,t}}{w_t} l(\theta^t)\phi_{l,t}\big(l(\theta^t)\big) + \beta\int \omega\big(\theta^{(t+1)}\big)\frac{\partial f^t(\theta_{t+1}|\theta_t)}{\partial\theta_t}d\theta_{t+1}$$

The relaxing planing problem denoted by $\mathcal{P}^{FOA}$ replaces IC by by the envelope condition and it is given by:

equation 8

$$\mathcal{P}^{FOA}: \min_{\{c,y\}} \Pi\big(\{c,y\}: \underline{U}(\theta)_\Theta\big) = \left[\sum_{t=1}^{T}\left(\frac{1}{R}\right)^{t-1} \int_\theta^t \left(c(\theta^t) - y(\theta^t)\mathcal{P}(\theta^t)d\theta^t\,\right)\right]$$

inequality 2

$$U(\{c,y\};\theta) \geq \underline{U}(\theta), y(\theta^t) \geq 0, c(\theta^t) \geq 0. \{c,y\} \in x^{FOA}$$

Now one simple expanation of revelation principle:

Theorem 1 Revelation principle Myerson,1981
Suppose that ψ was a Bayes-Nash equilibrium of the indirect mechanism Γ. Then there exists a direct mechanism that is payoff-equivalent and where truthful revelation is an equilibrium.
Proof: $\exists\psi$ and these strategies are equivalent in direct and indirect mechanisms. Direct revelation mechanism is the one where agent reports his preferences truthfully and hence $M = \Pi_{i\in 1,2,\ldots n} M_i$ (messages)agents type of profiles are :$\theta \in \{1,2,\ldots,n\}$ and $\theta \in \Theta$,the social choice function is $f: \Theta \to X$ where outcome $x \in X$ and in a message space there is mapping $g: M \to X$. Let's notice that if bidder(player)i with type θ deviates and reports his other type θ' that that agent earns $E_{\theta_{-i}} v_i\big(\psi_i(\theta_i')\psi_{-i}(\theta_{-i})\big) = E_{\theta_{-i}} v_i\big(\psi', \psi_{-i}(\theta_{-i})\big)$ for some $\psi'$and we know that (form above said):
equation 9

$$E_{\theta_{-i}} v_i\big(\psi_i(\theta_i), \psi_{-i}(\theta_{-i})\big) \geq E_{\theta_{-i}} v_i\big(\psi', \psi_{-i}(\theta_{-i})\big)$$

So, this last expression is not profitable ■.
Definition Incentive compatibility (Bayesian Incentive compatibility (BIC))
A social choice function $f: \Theta_1 \times \Theta_2 \ldots\ldots \times \Theta_n \to X$ is said to be incentive compatible (IC) or truthfully implementable if the Bayesian game (is a game in which the players have incomplete information about the other players) induce by the direct revelation mechanism(is one where each agent is asked to report his individual preferences, in which case $M = \Theta$ and $f = g$) or $\mathcal{D} = \big(\Theta_{i\in N}, f(\cdot)\big)$ has a pure strategy equilibrium(Bayesian-Nash equilibrium) $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_n^*(\cdot))$ where $s_i^*(\Theta_i) = \Theta_i$ and $\forall\theta_i \in \Theta$ and $\forall i \in N$.
Individual rationality (IR) axiom
First we define as in Myerson (1991), *two-person bargaining problem , to consist of a pair (F, v) where F is a convex subset of $\mathbf{R}^2$,$v = (v_1, v_2)$ is a vector in $\mathbf{R}^2$and the set $F \cap \{(x_1,x_2)|x_1 \geq v_1; x_2 \geq v_2\}$ is non-empty and bounded. Where F is a set of feasible payoff allocations and v represents the disagreement point. F is a convex means that the players are assumed that will agree on their jointly randomized strategies so that utility allocations $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are feasible and $0 \leq \theta \leq 1$ so that following expected utility allocation applies $\theta x + (1-\theta)y$ .Two-players strategic game form is given as: $\Gamma = \{(1,2), C_1, C_2 mu_1, u_2\}$ where C_1, C_2 are used to denote the pure players strategies set.*
Now, since we defined some terms that we were using previously we can go to writing the planing problem recursively with promised utility as a state variable. So the continuation value in period t for the agent is what has been promised by the social planner (government) in period $t-1$,now $v(\theta^t)$ is the expected future continuation utility:

equation 10

$$v(\theta^t) \equiv \int \omega\,(\theta^{t+1}) f^{t+1}(\theta_{t+1}|\theta_t) d\theta_{t+1}$$

And continuation utility $\omega(\theta^t)$ can be rewritten as:

equation 11

$$\omega(\theta^t) = u_t\big(c(\theta^t)\big) - \phi_t\left(\frac{y(\theta^t)}{\theta_t}\right) + \beta \int \omega\,(\theta^{t+1}) f^{t+1}(\theta_{t+1}|\theta_t) d\theta_{t+1}$$

Future marginal rent is defined as:

equation 12

$$\Delta(\theta^t) \equiv \int \omega(\theta^{t+1}) \frac{\partial f^t(\theta_{t+1}|\theta_t)}{\partial \theta_t} d\theta_{t+1}$$

So now the envelope condition can be rewritten as:

equation 13

$$\omega(\theta^t) = \frac{w_{\theta,t}}{w_t} l(\theta') \phi_{l,t}\big(l(\theta^t)\big) + \beta \int \omega(\theta^{t+1}) \frac{\partial f^t(\theta_{t+1}|\theta_t)}{\partial \theta_t} d\theta_{t+1}$$

The expected continuation costs of the planner at time t for the states $v_{t-1}, \Delta_{t-1}, \theta_{t-1}$ is :

equation 14

$$\mathcal{k}(v_{t-1}, \Delta_{t-1}, \theta_{t-1}, t) = \min \left[\sum_{t=1}^{T} \left(\frac{1}{R}\right)^{\tau-t} \int \big(c_t(\theta^t) - y_t(\theta^\tau)\big) \mathcal{P}(\theta^{t-1}\,) d\theta^{\tau-t} \right]$$

Where we have $d\theta^{\tau-t} = d\theta_t d\theta_{\tau-1}, \dots, d\theta_t$ and $\mathcal{P}(\theta^{\tau-t}) = f^t(\theta_t|\theta_{\tau-1}) \dots f^t(\theta_t|\theta_{t-1})$.

equation 15

$$\mathcal{k}(v, \Delta, \theta\_, t) = \min \int \left(c_t(\theta^t) - w_t(\theta) l(\theta) + \frac{1}{R} \mathcal{k}(v(\theta), \Delta(\theta), \theta, t+1) \right) f^t(\theta_t|\theta\_) d\theta$$

s.t.

$$\omega(\theta) = u_t\big(c(\theta)\big) - \phi_t\big(l(\theta)\big) + \beta v(\theta)$$

$$\dot{\omega}(\theta) = \frac{w_{\theta,t}}{w_t} l(\theta) \phi_{l,t}((\theta)) + \beta \Delta(\theta)$$

$$v = \int \omega(\theta) f^t(\theta|\theta\_) d\theta$$

$$\Delta = \int \omega(\theta)\ \frac{\partial f^t(\theta|\theta\_)}{\partial \theta\_} d\theta$$

Tax wedge[31] on labor $\tau_L(\theta^t)$ and the intertemporal wedge on savings (capital) $\tau_k(\theta^t)$ are given as:

equation 16

$$\tau_L(\theta^t) = 1 - \frac{\phi_{l,t}(l_t)}{w_t u_t'(c_t)}\ ; \tau_k(\theta^t) = 1 - \frac{1}{R\beta} \frac{u_t'(c_t)}{E_t\big(u_t'(c_{t+1})\big)}$$

Proposition1. at the optimum the labor wedge is equal to:

equation 17

$$\frac{\tau_{l,t}(\theta^t)}{1 - \tau_l(\theta^t)} = \frac{\mu(\theta^t) u_t'\big(c(\theta^t)\big)}{f^t(\theta_t|\theta_{t-1})} \frac{\varepsilon_{w,\theta,t}}{\theta_t} \frac{1 + \varepsilon_t^u}{\varepsilon_t^c}$$

Where $\mu(\theta^t) = \eta(\theta^t) + k(\theta^t)$, and where

[31] Tax wedge is defined as the ratio between the amount of taxes paid by an average single worker (a single person at 100% of average earnings) without children and the corresponding total labor cost for the employer. The average tax wedge measures the extent to which tax on labor income discourages employment. This indicator is measured in percentage of labor cost (OECD def,)

equation 18

$$\eta(\theta^t) = \frac{\tau_{l,t-1}(\theta^{t-1})}{1-\tau_{l,t-1}(\theta^{t-1})}\left[\frac{R\beta}{u'_{t-1}\left(c(\theta^{t-1})\right)}\frac{\varepsilon^c_{t-1}}{1+\varepsilon^u_{t-1}}\frac{\theta_{t-1}}{\varepsilon_{w,\theta,t-1}}\int_{\theta_t}^{\bar{\theta}}\frac{\partial f(\theta_s|\theta_{t-1})}{\partial\theta_{t-1}}d\theta_s\right]$$

And

equation 19

$$k(\theta^t) = \int_{\theta_t}^{\bar{\theta}}(1-g_s)\ \frac{1}{u'_t(c(\theta^{t-1},\theta_s)}f(\theta_s|\theta_{t-1})\ d\theta_s$$

Where : $g_s = u'_t(c(\theta^{t-1},\theta_s)\lambda_{t-1}$ and $\lambda_{t-1} = \int_{\underline{\theta}}^{\bar{\theta}}\frac{1}{u'_t(c(\theta^{t-1},\theta_s)}\ f(\theta_m|\theta_{t-1})d\theta_m$, where $k(\theta^t)$ captures the insurance motive and g_s represents the social marginal weight of an agent of type θ_s, measuring the social value of one more dollar transferred to that individual and $1/\lambda_{t-1}$ represents the social costs of public funds at time t.The redistributive term $\eta(\theta^t)$ can be written recursively in terms of the previous period's labor wedge weighted by a measure of ability persistence.In the first period heterogeneity in θ_1 leads to :
equation 20

$$\mu(\theta_1) = \int_{\theta_1}^{\bar{\theta}}\frac{1}{u'_t(c(\theta^{t-1},\theta_s)}\left(1-\lambda_0(\theta_s)u'_1\left(c_1(\theta_s)\right)\right)f(\theta s)$$

Where $\lambda_0(\theta_s)$ is the multiplier scaled by $f(\theta s)$ on type θ_s target utility. And we linear utility one can obtain that:

equation 21

$$1 = \int_{\underline{\theta}}^{\bar{\theta}}\lambda_0(\theta_s)f(\theta s)$$

Special autoregresive case with persistence p can be written as follows:

equation 22

$$\log(\theta) = plog(\theta_{t-1}) + \psi_t$$

Where PDF od ψ_t is given as: $f^{\psi}(\psi|\theta_{t-1})$ and first moment is given as: $E(\psi|\theta_{t-1}) = 0$.Evolution of labor wedge is given as:

equation 23

$$E_{t-1}\left(\frac{\tau_{l,t}}{1-\tau_l}\frac{\varepsilon^c_t}{1+\varepsilon^u_t}\frac{1+\varepsilon^u_{t-1}}{\varepsilon^c_{t-1}}\left(\frac{1}{R\beta}\frac{u'_{t-1}}{(u'_t)}\right)\right) = \varepsilon_{w,\theta,t-1}\frac{1+\varepsilon^u_{t-1}}{\varepsilon^c_{t-1}}cov\left(\frac{1}{R\beta}\frac{u'_{t-1}}{(u'_t)},\log(\theta_t)\right) + p\frac{\tau_{l,t-1}}{1-\tau_{lt-1}}$$

Now, dynamic Mirrlees' approach Mirrlees(1971) , is said to have strong implications of how savings (capital) should be taxed. At the optimum the inverse Euler holds and:

equation 24

$$\frac{R\beta}{u'_t * c(\theta^t))} = \int_{\underline{\theta}}^{\bar{\theta}}\frac{1}{u'_{t+1}\left(c(\theta^{t+1})\right)}f^{t+1}(\theta_{t+1}|\theta_t)d\theta_{t+1}$$

The inverse marginal utility in the current period is equal to the expected inverse marginal utility in the next period. By the concavity of marginal utility and Jenssen's inequality[32], it is case that:
equation 25

$$u'\left(c(\theta^t)\right) < \beta R\int_{\underline{\theta}}^{\bar{\theta}}u'_{t+1}\left(c(\theta^{t+1})\right)f^{t+1}(\theta_{t+1}|\theta_t)d\theta_{9t} + 1)$$

Now if we try to understand inverse Euler equation by cost of providing utility , we can pick a history θ^t and leave all allocations unchanged, except a node θ^t where we perturb utility by providing $\beta \times \Delta$ less utility in period t and Δ more utility for $\forall\theta_{t+1}$ after history θ^t , that is the perturbed utilities:

[32] About Jensen inequality: if $p_1, \ldots, p_n$ are positive numbers which sum to 1, and if f is real continuous function which is concave then the inequality gives: $f(\sum_{i=1}^n p_i x_i) \geq \sum_{i=1}^n p_i f(x_i)$, if $p_i = \frac{1}{n}$ in the concave case then:$\ln\left(\frac{1}{n}\sum_{i=1}^n x_i\right) \geq \frac{1}{n}\sum_{i=1}^n \ln x_i$, see Jensen (1906).

equation 26

$$u_t\big(c(\theta_t)\big) - \beta\Delta;\ u_t\ +\ 1(c(\theta_t, \theta_{t+1})\ +\Delta$$

Costs of providing utility must be minimized at $\Delta = 0$ and that costs are:

equation 27

$$c(u_t - \beta\Delta) + \frac{1}{R}\int_{\underline{\theta}}^{\overline{\theta}} \big(c(u_{t+1}(\theta^t, \theta_{t+1}) + \Delta)\big) f^t(\theta_{t+1}|\theta^t) d\theta_{t+1}$$

In the special case in which there is uncertainty in the first period but not thereafter, it is recovered the result by Atkinson, Stiglitz (1976) that capital should not be taxed.Atkinson and Stiglitz (1976) demonstrated the following theorem known as Atkinson, Stiglitz theorem:

Theorem: Commodity taxes cannot increase social welfare if utility functions are weakly separable in consumption goods versus leisure and the sub utility 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 sub utility function $v(c_1, \dots, c_k)$ homogenous across individuals.

Laroque (2005) and Kaplow (2006) have provided intuitive proof of this theorem as follows:

Proof: A tax system $(\tau(\cdot), t)$that includes both nonlinear income tax and a vector of commodity taxes can be replaces by a pure income tax $(\bar{\tau}(\cdot), t = 0)$ .This tax system keeps all individual utilities constant and raises at least as much tax revenue. Let $v(p + t, \gamma) = \max_c v(c_1, .. c_k)$ s.t. $(p + t) \cdot c \leq \gamma$ be the indirect utility of consumption goods which is common to al individuals. Now if we consider replacing $(\tau(\cdot), t)$ this tax system with $(\bar{\tau}(\cdot), t = 0)$ where $\bar{\tau}(w)$ is defined such that $v\big(p + t, w - \tau(w)\big) = v\big(p, w - \bar{\tau}(w)\big)$. Here $\bar{\tau}(w)$ naturally exists a $v(p, \gamma)$ is strictly increasing in γ .Which on turn implies that $u_i(v(p + t, w - \tau(w), w) = u_i(v(p + t, w - \bar{\tau}(w), w)$, $\forall w$ .So the utility and labor supply for $\forall i$ are unchanged .Attaining utility of consumption $v\big(p, w - \bar{\tau}(w)\big)$ at price p costs at least $w - \bar{\tau}(w)$.Now, let c_i be the consumer choice of individual i under the initial tax system $(\tau(\cdot), t)$.Individual i attains utility $v\big(p, w - \bar{\tau}(w)\big) = v\big(p, w - \bar{\tau}(w)\big)$ when choosing c_i. And, now $p \cdot c_i \geq w - \bar{\tau}(w)$ and we have that $\bar{\tau}(w) \geq \tau(w) + t \cdot c_i$ i.e. the government collects more taxes with $(\bar{\tau}(\cdot), t = 0)$ ∎

Dynamic Mirrlees approach to taxation: I (2 period example)

This models endogenizes the tax structure so that is not a flat tax. Government does not need to increase taxes just to raise revenue but also needs to provide incentives, since it does not observe and individuals' skill level. Government also wants to provide insurance against skills shocks (e.g. disability). The key issue here is resolving the trade-off between incentives and insurance. 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). Assumption of the model here are:

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

Without aggregate uncertainty perfect consumption insurance is possible (everybody gets the same consumption). Here the additional feature is the absent informational frictions. However, if government cannot observe the skills, then highly able will pretend to be disabled. Now will go to 2 period example section. Assumptions here are:

1. $\exists$ continuum of workers who live in 2 period and the maximization problem is

equation 28

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

2. Skills production is:

equation 29

$$y = \theta \cdot n$$

Where y represents the output, which is observable, θ represents the skills, n represents 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.

3. Resource constraint is given as:

inequality 3

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

Now we will outline the government maximization problem.

4. Government maximization problem is given as:

$$\max_{\substack{c_1(i),c_2(i)\\ y_1(i),y_2(i)}} \sum_I \left\{ u(c_1 l_{i,j}) + v\left(\frac{y_1(i)}{\theta_1(i)}\right) + \beta \sum_j \left[u(c_2 l_{i,j}) + v\left(\frac{y_2(i,j)}{\theta_2(i,j)}\right) \right] \right\} \Pi_2(j|i)\Pi_1(I)$$

s.t.

1) **Resource constraint** :

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

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

inequality 4

$$u(c_1 l_{i,j}) + v\left(\frac{y_i l_{i,j}}{\theta l_{i,j}}\right) + \beta \sum_i \left[u(c_2, l_{i,j}) + v\left(\frac{y_2(i,j)}{\theta_2(i,j)}\right) \right] \Pi_2(j|i)$$

$$\geq u\left(c_1 l(i_r) + v\left(\frac{y_i l(i_r)}{\theta(i)}\right) + \beta \sum_j \ (u(c_2(i_r, j_r)) + v\left(\frac{y_2(i_r, j_r)}{\theta_2(i,j)}\right) \Pi_2(j|i) \right)$$

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

5. 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

equation 30

$$u(\tilde{c}_2(i,j;\Delta) \equiv u\big(c_2^*(i,j)\big) + \Delta$$

3) Hold total utility constant by decreasing 1st period utility by $\beta\Delta$

equation 31

$$u\big(\tilde{c}_1(i,j,\Delta)\big) = u\big(c_1^*(i)\big) - \beta\Delta$$

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:

equation 32

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

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

equation 33

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

Previous equation is inverse Euler equation. Here only to note that: $\frac{d}{dx}f^{-1}(x) = \frac{1}{f'(x)}$. Here 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[33]. This explains how it is possible to have a positive intertemporal wedge/tax. Even though taxes are zero ex-ante.

Notes on Inverse Euler equation and Savings Distortions and Mirrlees model (Golosov et al. (2003)

Here we review a two period hazard model with savings distortions in a Mirrleesian setup namely the paper by Golosov, Kocherlakota,Tsyvisnki (2003). First two period moral hazard model and properties of efficient allocations in such models as in Rogerson (1985). The model is dynamic in two periods, effort in first period is e_0 consumption in both periods is c_0, c_1, stochastic output in second period is $y_1 = \theta_1$ with PDF density $f(\theta_1|e_0)$. Utility is separable and given as:

equation 34

$$u(c_0) - h(e) + \beta \int u\big(c_1(\theta_1)\big) f(\theta_1|e_0)$$

IC incentive compatibility requires:

inequality 5

$$(c_0) - h(e) + \beta \int u\big(c_1(\theta_1)\big) f(\theta_1|e_0) \geq (c_0) - h(e') + \beta \int u\big(c_1(\theta_1)\big) f(\theta_1|e')$$

Rewrite in terms of utility assignments: $u_t = u(c_t)$ ....

inequality 6

$$u_0 - h(e) + \beta \int u_1(\theta_1) f(\theta_1|e) \geq u_0 - h(e') + \beta \int u_1(\theta_1) f(\theta_1|e')$$

Planners' problem here is:

equation 35

$$\min\left\{c(u_0) + \frac{1}{R}\int [c(u_1(\theta_1) - y_1(\theta_1)]\, f(\theta_1|e_0)\right\}$$

equation 36

$$u_0 - h(e) + \beta \int u_1(\theta_1) f(\theta_1|e) = v_0$$

$$u_0 - h(e) + \beta \int u_1(\theta_1) f(\theta_1|e) \geq u_0 - h(e') + \beta \int u_1(\theta_1) f(\theta_1|e')$$

Now we introduce savings distortions, In previous $q = R^{-1} = \frac{1}{R}$ and now if $R = q^{-1}$ then we would have the standard Euler equation in the form:

equation 37

$$u'(c_0) = \beta R \int u'\big(c_1(\theta_1)\big) f(\theta_1|e_0)$$

Now it will be shown that previous does not hold, since there is distortion in savings. First fix e_0 and consider variations in consumption / utility:

[33] 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.

equation 38

$$\hat{u} = u_0 - \beta\Delta$$
$$\hat{u}(\theta_1) = u_1(\theta_1) + \Delta$$

There is no effect on utility or incentive constraint since:

equation 39

$$u_0 - h(e') + \beta \int u_1(\theta_1) f(\theta_1|e') = \hat{u}_0 - h(e') + \beta \int \hat{u}_1(\theta_1) f(\theta_1|e')$$

This applies for $\forall e'$, and we need to solve:

equation 40

$$\min_{\hat{u}_0, \hat{u}_1(\cdot), \Delta} \left\{ c(\hat{u}_0) + q \int c\big(\hat{u}_1(\theta_1)\big) f(\theta_1|e_0) \right\}$$

And: $\begin{matrix} \hat{u} = u_0 - \beta\Delta \\ \hat{u}(\theta_1) = u_1(\theta_1) + \Delta \end{matrix}$. And by substituting we get:

equation 41

$$\min_{\Delta} \left\{ c(u_0 - \beta\Delta) + \frac{1}{R} \int c\big(u_1(\theta_1)\big) f(\theta_1|e_0) \right\}$$

FOC s given as:

equation 42

$$c'(u_0 - \beta\Delta)\beta = q \int c'(u_1(\theta_1) + \Delta) f(\theta_1|e_0)$$

This condition is sufficient and necessary for an interior and we can use this to solve for Δ.First, if original allocation was optimal then $\Delta = 0$ .And by using $c = u^{-1}$ we can solve:

equation 43

$$\frac{1}{u'(c_0)} = \frac{1}{\beta R} \int \frac{1}{u'\big(c_1(\theta_1)\big)} f(\theta_1|e_0)$$

Previous is inverse Euler equation. Since $1/x$ is convex we apply Jensen inequality, if $VAR[c_1(\theta_1)] > 0$ then:

equation 44

$$u'(c_0) < \beta R \int u'\big(c_1(\theta_1)\big) f(\theta_1|e_0)$$

Where agents are "savings constrained". Wedge is given as:

equation 45

$$u'c_0) = \beta(1-\tau)R \int u'\big(c_1(\theta_1)\big) f(\theta_1|e_0)$$

And $\tau \geq 0$ actualy here $\tau_K > 0$. Or as in Auerbach (2013):

equation 46

$$u'(c_t) = (\beta r_{t+1}) E_t[u'(c_{t+1})]$$

It follows thet there must be positive wedge between the social return on capital R and the private return r- which is a capital income tax. Mirrlees model by Golosov, et al. (2003) is presented with a work time at $t = 1$

equation 47

$$u(c_0) + \beta \int [u(c_1(\theta_1) - h(y_1, \theta_1] f(\theta_1)$$

Here same optimality conditions apply: Inverse Euler equation, and previous is true also for a mixed model of moral hazard and adverse selection where effort affects distribution of θ . In general horizon and welfare utility function is given as:

equation 48

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t [u(c_t) - h(y_t, \theta_t]$$

Where in previous θ_t is in general stochastic process and private information., agent rewrite in terms of u_t and incentive constraint IC is given as:

equation 49

$$\mathbb{E}\sum_{t=0}^{\infty}\beta^t[u(\theta^t)-h(y(\theta^t),\theta_t]\geq \mathbb{E}\sum_{t=0}^{\infty}\beta^t\big[u(\sigma^t(\theta^t)-h\big(y\big(\sigma^t(\theta^t)\big),\theta_t\big]$$

Planners' costs are as previous:

equation 50

$$\mathbb{E}\sum_{t=0}^{\infty}q^t\big[c\big(u(\theta^t)\big)-y(\theta^t)\big]$$

Variations are as before:

equation 51

$$\hat{u}(\theta^t)=u\ (\theta^t)+\Delta(\theta^{t-1})-\beta\Delta(\theta^t)$$

And a "No Ponzi condition" implies:

equation 52

$$\lim\beta^t\mathbb{E}\Delta\big(\sigma^t(\theta^t)\big)=0$$

Preserve utility and IC incentive compatibility:

equation 53

$$\mathbb{E}\sum_{t=0}^{\infty}\beta^t u\big(\sigma^t(\theta^t)\big)=\mathbb{E}\sum_{t=0}^{\infty}\beta^t\hat{u}\big(\sigma^t(\theta^t)\big)$$

And hence to minimize costs:

equation 54

$$\mathbb{E}\sum_{t=0}^{\infty}q^t[c(uu(\theta^t)+\Delta(\theta^{t-1})-\beta\Delta(\theta^t)]$$

Theorem 1 (proof and set up by Golosov et al.(2003)) Let $u(c,l)=u(c)-v(l)$ suppose (c^*,y^*,K^*) solve $P_1(K_1^*)$ and that $\exists t<T$ and scalars $c^+\geq c_t^*;c_{t+1}^*,K_{t+1}^*\geq c_+>0$,then:

$$\frac{\beta(1-\delta+F_K(K_{t+1}^*,\int y_{t+1}^*d\mu))}{u'(c_t^*)}=E_t\left\{\frac{1}{u'(c_{t+1}^*)}\right\}$$

Previous its says that given any positive measure Borel[34] set B in Θ^t the average of $\frac{u'(c_t^*)}{u'(c_{t+1}^*)}$ across the agents who have a period t history in B is equal to $\beta(1-\delta+F_K(K_{t+1}^*)$. Second, it says that given that agent knows his period t history lies in B the agents expectations $\frac{u'(c_t^*)}{u'(c_{t+1}^*)}$ is equal to $\beta(1-\delta+F_K(K_{t+1}^*)$.Allocation in society is defined $(c,y,K)=(c_t,y_t,K_{t+1})_{t=1}^T$as:

equation 55

$$K_{t+1}\in R$$
$$c_t:\Theta^T\to R_+^j$$
$$(c_t,y_t)is\ \theta^t$$

Proof: minimization problem is as follows:

equation 56

$\min_{\eta_t,\varepsilon_t,\zeta_t}[\zeta_t+\int\eta_t d\mu]$ s.t.

$$\int\varepsilon_{t+1}d\mu=F(K_{t+1}^*+\zeta_t,\int y_{t+1}^*d\mu-F\ (K_{t+1}^*\int y_{t+1}^*d\mu)+(1-\delta)\zeta_t$$
$$u(c_t^*+\eta_t)+\beta u(c_{t+1}^*+\varepsilon_{t+1})=u(c_t^*)+\beta u(c_{t+1}^*)$$
$$c_t^*+\eta_t\geq 0,c_{t+1}^*+\varepsilon_{t+1}\geq 0,K_{t+1}^*+\zeta_t\geq 0$$

---

[34] Borel sets are the sets that can be constructed from open or closed sets by repeatedly taking countable unions and intersections. Formally, the class B of Borel sets in Euclidean R^n is the smallest collection of sets that includes the open and closed sets such that if $E,E_1,E_2,\ldots$ are in B, then so are union $\cup_{i=1}^{\infty}E_i$, $\cap_{i=1}^{\infty}E_i$, and $\mathbb{R}^n\backslash E$, where $F\backslash E$ is a set difference (Croft et al. 1991).

$$\eta_t \in L_t^\infty, \varepsilon_{t+1} \in L_{t+1}^\infty, \zeta_t \in R$$

The objective in previous min problem is to minimize resource use in period t.We claim here that previous minimization problem is solved for $(\eta_t, \varepsilon_{t+1}, \zeta_t) = 0$,and $\exists B \in \Theta^T$ such that $\mu(B) = 1$.And:

equation 57

$$u\left(c_t^*(\theta^T)\right) + \eta_t(\theta^T)) + \beta u\left(c_{t+1}^*(\theta^T)\right) + \varepsilon_{t+1}(\theta^T)) = u\left(c_t^*(\theta^T)\right) + \eta_t(\theta^T)) + \beta u\left(c_{t+1}^*(\theta^T)\right) \forall \theta\text{^}T \in B$$

Now we define (c', K') so that $c' = c^*$; $K' = K^*$ except that:

equation 58

$$c_t'(\theta^T) = c_t^*(\theta^T) + \eta_t(\theta^T)\ \forall\ \theta^T \in B$$
$$c_{t+1}'(\theta^T) = c_{t+1}^*(\theta^T) + \varepsilon_{t+1}(\theta^T)\ \forall\ \theta^T \in B$$
$$K_{t+1}' = K_{t+1}^* + \zeta_t$$

Here it is claimed that (c', y^*, K') is incentive feasible delivers the same value of the planner's objective as (c^*, y^*, K^*) and uses fewer resources. This allocation (c', y^*, K') is feasible if or because:

equation 59

$$\int c_t' d\mu + K_{t+1}' = \int c_t^* d\mu + K_{t+1}^* + \zeta_t + \int \eta_t d\mu < \int c_t^* d\mu + K_{t+1}^*$$

Next this allocation (c', y^*, K') will be shown to be IC. By construction:

equation 60

$$u\left(c_t'(\theta^T)\right) + \beta u\left(c_{t+1}'(\theta^T)\right) = u\left(c_t^*(\theta^T)\right) + \beta u\left(c_{t+1}^*(\theta^T)\right) \forall \theta^T$$

Then we know that $\forall \sigma \in \sum$ and for $\forall \theta^T$

equation 61

$$\sum_{s=1}^{T} \beta^{s-1} u\left(c_s'\left(\sigma(\theta^T)\right)\right)$$
$$= \sum_{s=1}^{t-1} \beta^{s-1} u\left(c_s^*\left(\sigma(\theta^T)\right)\right) + \beta^{t-1}\left[u(c_t'\left(\sigma(\theta^T)\right) + \beta u\left(c_{t+1}'\left(\sigma(\theta^T)\right)\right)\right] + \sum_{s=t+2}^{T} \beta^{s-1} u\left(c_s^*\left(\sigma(\theta^T)\right)\right)$$
$$= \sum_{s=1}^{t-1} \beta^{s-1} u\left(c_s^*\left(\sigma(\theta^T)\right)\right) + \beta^{t-1}\left[u\left(c_t^*\left(\sigma(\theta^T)\right)\right)\right] + \sum_{s=t+2}^{T} \beta^{s-1} u\left(c_2^*\left(\sigma(\theta^T)\right)\right) = \sum_{s=1}^{T} \beta^{s-1} u\left(c_s^*\left(\sigma(\theta^T)\right)\right)$$

It follows that (c', σ^*) for any σ agent get the same utility from c' as from c^*

equation 62

$$\int_{t=1}^{T} \beta^{t-1}\left[u(c_t') - v\left(\frac{y_t^*}{\theta_t}\right) d\mu\right] = \int_{t=1}^{T} \beta^{t-1}[u(c_t'(\sigma) - v(y_t^*(\sigma)/\theta_t)]\, d\mu$$

Lagrangian is given as;

equation 63

$$\mathcal{L}(\zeta_t, \eta_t, \varepsilon_{t+1}) = \zeta_t + \int \eta_t d\mu + \lambda_t^*\left[\int \varepsilon_{t+1} d\mu - (1-\delta)\zeta_t - F\ (K_{t+1}^* + \zeta_t, Y_{t+1}^*\right] - \left(\zeta_{t+1}^*, u(c_t^* + \eta_t) + \beta u(c_{t+1}^* + \varepsilon_{t+1})\right)$$

It follows that:

equation 64

$$\int \frac{\eta_t'}{u'(c_t^*) d\mu} = (z_{t+1}^*, \eta_t') \forall \eta_t' \in L_t^\infty$$
$$\beta^{-1}\lambda_t^* \int \frac{\varepsilon_{t+1}'}{u'(c_{t+1}^*) d\mu} = (z_{t+1}^*, \varepsilon_{t+1}'\ ) \forall \varepsilon_{t+1}' \in L_{t+1}^\infty$$
$$\lambda_t^* = (1 - \delta + F_k(K_{t+1}^*, Y_{t+1}^*)^{-1}$$

And so:

equation 65

$$\beta^{-1}[1-\delta+F_K(K_{t+1}^*,Y_{t+1}^*]^{-1}\int\frac{\eta'}{u'(c_{t+1}^*)d\mu}=\int\frac{\eta_t'}{u'(c_t^*)d\mu}\quad\forall\eta_t'\in L_t^\infty$$

Now, let $\eta_t' = 1_A u'(c_t^*)$ where A is an arbitrary element of Ω_t . ■ ||

Insurance and Taxation Over the Life Cycle by Farhi, Werning (2013)

Preferences and technology in this model are given as:

equation 66

$$u(c,y)=\mathbb{E}_0\sum_{t=1}^{T}\beta^{t-1}u^t(c_t,y_t,\theta_t)$$

Costs are presented as: $\mathbb{E}_0\sum_{t=1}^{T}R^{-(t-1)}(\,c_t-y_t)$ and life cycle is presented by work phase: $t<T_E$ and utility is $u^t(c,y,\theta)=\tilde{u}\left(c,\frac{y}{\theta}\right)$, and retirement phase is : $T_E<t\leq T$ and utility $u^t(c;y;\theta)=\begin{cases}\tilde{u}(c,0) & y=0\\ =\infty & y>0\end{cases}$.About uncertainty and information: θ_t is private information, and Markov process with support $\left[\underline{\theta_t}(\theta_{t-1}),\bar{\theta}_t(\theta_{t-1})\,\right]$ with PDF differential: $f^t(\theta_t|\theta_{t-1})$.Planers problem is:

equation 67

$$K_0(v)=\min_{c,y}\mathbb{E}\sum_{t=1}^{T}R^{-(t-1)}(c_T-y_t)$$

s.t.

$$u(c,y)\geq v\,;u(c,y)\geq u(c^\sigma,y^\sigma)\;\forall\sigma\in\Sigma$$

Continuation utility is:

equation 68

$$w(\theta^t)=u(c(\theta^t),y(\theta^t),\theta_t)+\beta v(\theta^t)$$
$$v(\theta^t)=\int w(\theta^{t+1})f^{t+1}(\theta_{t+1}|\theta_t)d\theta_{t+1}$$

Dynamic generalization of Envelope condition of Mirrlees (1971) and Milgrom and Segal (2002) , Kapicka (2013), Williams (2011), Pavan, Segal and Toikka (2014) . Necessary conditions for IC are:

equation 69

$$\frac{\partial}{\partial_t}w(\theta^t)=u_\theta\big(c(\theta^t)\big),y(\theta^t);\theta^t)+\beta\Delta(\theta^t)$$
$$\Delta(\theta^t)\equiv\int w(\theta^{t+1})f_{\theta_t}^{t+1}(\theta_{t+1}|\theta_t)d\theta_{t+1}$$

About wedges . Now intertemporal wedge is given as:

equation 70

$$1=\beta R\left(1-\tau_{K,t-1}\right)\mathbb{E}_{t-1}\left[\frac{\hat{u}_c^t(c_t,y_T,\theta_t)}{\hat{u}^{t-1}(c_{t-1}y_{t-1};\theta_{t-1}}\right]$$

Labor wedge is given as:

equation 71

$$1=\left(1-\tau_{L,t}\right)\frac{\hat{u}_c^t(c_t,y_t,\theta_t)}{-\hat{u}_y^t(c_t,y_t,\theta_t)}$$

Assumption here is separable utility: $u^t(c,y,\theta)=\hat{u}^t(c)-\hat{h}^t(y,\theta)$ and as proposition that inverse Euler holds:

equation 72

$$\frac{1}{\hat{u}^{t-1\prime}(c_{t-1})}=\frac{1}{\beta R}\mathbb{E}_{t-1}\left[\frac{1}{\hat{u}^{t\prime}(c_t)}\right]$$

So that intertemporal wedged is $\tau_{K,t-1}\geq 0$. Assumption is also isoelastic disutility of work:

equation 73

$$\hat{h}^t(y,\theta) = \left(\frac{k}{\alpha}\right)\left(\frac{y}{\theta}\right)^{\alpha}$$

Also productivity is AR(1) process :

equation 74

$$\log(\theta_t) = \rho \log(\theta_{t-1}) + \bar{\theta}_t + \varepsilon_t$$

And furthermore proposition is:

equation 75

$$\mathbb{E}_{t-1}\left[\frac{\tau_{L,t}}{1-\tau_{L,t}}\frac{1}{\beta R}\frac{\hat{u}^{t-1'}(c_{t-1})}{\hat{u}^{t'}(c_t)}\right] = \rho\ \frac{\tau_{L,t-1}}{1-\tau_{L,t-1}} + \alpha Cov_{t-1}\left(\log(\theta_t), \frac{1}{\beta R}\ \frac{\hat{u}^{t-1'}(c_{t-1})}{\hat{u}^{t'}(c_t)}\right)$$

Previous is labor wedge simple formula where :LHS is risk-adjusted conditional expectations $\frac{\tau_{L,t}}{1-\tau_{L,t}}$ and RHS (1) $\frac{\tau_{L,t}}{1-\tau_{L,t}}$ inherits mean reversion of θ .RHS(2) s a positive drift of $\frac{\tau_{L,t}}{1-\tau_{L,t}}$, there is a benefit of added insurance in the form of : $Cov_{t-1}\left(\log(\theta_t), \frac{1}{\beta R}\ \frac{\hat{u}^{t-1'}(c_{t-1})}{\hat{u}^{t'}(c_t)}\right)$, and incentive costs increase with Frisch elasticity $\frac{1}{\alpha-1}$.General stochastic process is given as:

equation 76

$$\phi_t^{log}\ (\theta_{t-1}) \equiv \int \log(\theta_t) f^t(\theta_t|\theta_{t-1}) d\theta_t$$

And proposition for labor wedge in stochastic case is:

equation 77

$$\mathbb{E}_{t-1}\left[\frac{\tau_{L,t}}{1-\tau_{L,t}}\frac{1}{\beta R}\frac{\hat{u}^{t-1'}(c_{t-1})}{\hat{u}^{t'}(c_t)}\right] = \theta_{t-1}\frac{d\phi_t^{log}}{d\theta_{t-1}}\ \frac{\tau_{L,t-1}}{1-\tau_{L,t-1}} + \alpha Cov_{t-1}\left(\log(\theta_t), \frac{1}{\beta R}\ \frac{\hat{u}^{t-1'}(c_{t-1})}{\hat{u}^{t'}(c_t)}\right)$$

Only difference in necessary IC conditions is :

equation 78

$$\Delta(\theta^t) \equiv \int w(\theta^{t+1}) f_{\theta_t}^{t+1}(\theta_{t+1}|\theta_t) d\theta_{t+1} + \frac{d\bar{\theta}_{t+1}}{d\theta_t} w(\bar{\theta}_{t+1}) f^{t+1}(\bar{\theta}_{t+1}|\theta_t) - \frac{d\underline{\theta}_{t+1}}{d\theta_t} w(\underline{\theta}_{t+1}) f^{t+1}(\underline{\theta}_{t+1}|\theta_t)$$

Labor wedges at top and bottom are given as follows:

equation 79

$$\frac{\bar{\tau}_{L,t}}{1-\bar{\tau}_{L,t}} = \frac{\tau_{L,t}}{1-\tau_{L,t}}\beta R \frac{\hat{u}^{t'}}{\hat{u}^{t-1'}}\frac{d\log\bar{\theta}_t}{d\log\theta_{t-1}}$$
$$\frac{\underline{\tau}_{L,t}}{1-\underline{\tau}_{L,t}} = \frac{\tau_{L,t}}{1-\tau_{L,t}}\beta R \frac{\hat{u}^{t'}}{\hat{u}^{t-1'}}\frac{dlog\ \underline{\theta}_t}{d\log\theta_{t-1}}$$

Previous generalizes Mirrlees (1971) and for a fixed support:

equation 80

$$\tau_L(\theta^{t-1}, \bar{\theta}_t) = \tau_L(\theta^{t-1}, \underline{\theta}_t) = 0\ ; \theta_t = \varepsilon\theta_{t-1}\ \varepsilon_t \in [\underline{\varepsilon}; \bar{\varepsilon}]$$
$$\tau_L(\theta^{t-1}, \bar{\theta}_t) \leq \tau_L(\theta^{t-1}) \leq \tau_L(\theta^{t-1}, \underline{\theta}_t)$$

Now productivity follows a Brownian diffusion:

equation 81

$$d\log\theta_t = \hat{\mu}_t^{\log}(\theta_t) d\theta_t + \hat{\sigma}_t dW_t$$

Dynamics of the model is given as:

equation 82

$$\frac{d\lambda_t}{\lambda_t} = \sigma_{\lambda,t}\hat{\sigma}_t dW_t\ ;\ d\gamma_t = \left[-\theta_t\sigma_{\lambda,t}\hat{\sigma}_t^2 + \left(\hat{\mu}_t + \theta_t\frac{d\hat{\mu}_t^{\log}}{d\theta}\right)\gamma_t\right] dt + \gamma_t\hat{\sigma}_t dW_t$$

Allocation and wedges in this economy are given as:

equation 83

$$\frac{1}{\hat{u}^{t\prime}(c_t)} = \lambda_t; \frac{1}{\hat{u}^{t\prime}(c_t)} - \frac{\theta_t}{h^{t\prime}\left(\frac{y_t}{\theta_t}\right)} = -\alpha\frac{\gamma_t}{\theta_t}$$

$\frac{\tau_{L,t}}{1-\tau_{L,t}} = -\alpha\frac{\gamma_t}{\lambda_t}\frac{1}{\theta_t};\ \tau_{K,t} = \sigma_{\lambda,t}^2\hat{\sigma}_t^2$

Dual variables are : $\lambda_t \equiv K_v(v_t, \Delta_t, \theta_t, t)$ and $\gamma_t \equiv K_\Delta(v_t, \Delta_t, \theta_t, t)$, and sufficient control is : $\sigma_\lambda(\lambda_t, \gamma_t, \theta_t, t)$. Labor wedge now by simple formula is given as:

equation 84

$$d\left(\frac{\tau_{L,t}}{1-\tau_{L,t}}\right) = \left[\frac{\tau_{L,t}}{1-\tau_{L,t}}\theta_t\frac{d\hat{\mu}_t^{\log}}{d\theta} + \alpha\sigma_{\lambda,t}\hat{\sigma}_t^2 + \frac{\tau_{L,t}}{1-\tau_{L,t}}\sigma_{\lambda,t}\hat{\sigma}_t^2\right]dt - \frac{\tau_{L,t}}{1-\tau_{L,t}}\sigma_{\lambda,t}\hat{\sigma}_t dW_t$$

Intuition behind labor wedge is:

equation 85

$$\lambda_t\frac{\tau_{L,t}}{1-\tau_{L,t}} = \frac{1}{\hat{u}^{t\prime}(c_t)} - \frac{\theta_t}{\hat{h}^{t\prime}\left(\frac{y_t}{\theta_t}\right)}$$

At first best : $\frac{1}{\hat{u}^{t\prime}(c_t)} = \frac{\theta_t}{\hat{h}^{t\prime}\left(\frac{y_t}{\theta_t}\right)}$ and $\frac{1}{\hat{u}^{t\prime}(c_t)}$ tracks $\frac{\theta_t}{\hat{h}^{t\prime}\left(\frac{y_t}{\theta_t}\right)}$at second best. General preferences are given with inverse Euler which requires separability,here we define :

equation 86

$$\eta_t = \frac{\partial \log |MRS_t|}{\partial \log \theta_t}$$

$$|MRS_t| = -\frac{u_y^t}{u_c^t}$$

Generalization of previous is given as:

equation 87

$$d\left(\frac{\tau_{L,t}}{1-\tau_{L,t}}\frac{1}{\eta_t}\frac{1}{u_c^t}\right) = \left[\lambda_t\sigma_{\lambda,t}\hat{\sigma}_t^2 + \frac{\tau_{L,t}}{1-\tau_{L,t}}\frac{1}{\eta_t}\frac{1}{u_c^t}\ \theta_t\frac{d\hat{\mu}_t^{\log}}{d\theta}\right]dt$$

In previous expression $\frac{\tau_{L,t}}{1-\tau_{L,t}}\frac{1}{\eta_t}$ represents discouragement. In one particular interesting case where: $\tilde{u}^t\left(\hat{u}^t(c) = \frac{k_t}{\alpha_t}\left(\frac{y}{\theta}\right)_t^\alpha\right)$ in this case $\eta_t = \alpha_t$ and

equation 88

$$d\left(\frac{\tau_{L,t}}{1-\tau_{L,t}}\right) = \left[\alpha_t\lambda_t\sigma_{\lambda,t}\hat{\sigma}_t^2 + \frac{\tau_{L,t}}{1-\tau_{L,t}}\frac{1}{u_c^t}\left(\theta_t\frac{d\hat{\mu}_t^{\log}}{d\theta} + \frac{1}{\alpha_t}\frac{d\alpha_t}{dt}\right)\right]dt + \frac{\tau_{L,t}}{1-\tau_{L,t}}\frac{1}{u_c^t}d\ (u_c^t)$$

Conclusion

This paper tried to review one part of the literature on dynamic optimal taxation. Namely Mirrleesian approach to dynamic taxation and models that were developed in this way for the last two decades were outlined. In the dynamic Mirrlees approach, when it comes to the result for capital, capital is taxed to provide more efficient labor supply incentives when there is imperfect information (private distributions of ability unknown to other parties) and as a part of optimal insurance scheme against stochastic earning abilities. Intuition here is that savings affects incentive to work, so government needs to discourage savings to prevent the flowing deviation by highly skilled: 1) save more today; 2) work less tomorrow. That was the second model we reviewed and from there some optimal fiscal policy features are:1) On average wealth taxes across individuals are zero ex-ante ;2) However, they depend on future labor income-if labor income is below average, your capital tax is positive. If your labor income is above average, then your capital tax is negative. 3) So, this tax or this fiscal policy might be regressive for incentive reasons.So in general about dynamic Mirrlees approach it can be concluded that: this approach assumes that agents' abilities to earn income are heterogeneous, stochastic, and private information. Tax instruments ex ante are unrestricted. The model solves for the optimal allocations using dynamic mechanism design (subject only to incentive compatibility constraints) and then considers how to implement these allocations using decentralized tax systems, see also Stantcheva (2020). Furthermore, the dynamic Mirrlees approach needs a general theory of approximation of the optimal, often complicated, policies. The tax system by countries is used as a means of social insurance. Friedman (1962) , proposed negative income tax in which individuals with low enough negative income will owe negative tax and collect transfers. In this model constant marginal tax rate is combined with lump-sum transfers, which results in a linear function with negative intercept, Sheshinski (1972).Later Mirrlees (1971) refined this idea by allowing non-linear taxes, see also Farhi, Werning (2012). Atkinson and Stiglitz (1976), provided a negative answer for capital taxation as a provider of social insurance they have left that role for labor income taxation only. But later literature on the issue was adding uncertainty which would make difference. So when individuals do face uncertainty, the optimal insurance arrangement calls for taxing capital positively: constrained efficient allocations satisfy an inverse Euler equation, as opposed to the agent's standard intertemporal Euler equation, implying a positive capital tax. See, Diamond; Mirrlees (1978); Rogerson (1985); Ligon (1998); Golosov, Kocherlakota,Tsyvinski (2003).This models some of which were discussed in the paper will provide authors argument for further research on the topic and investigation in the issues they raise.

References

1. Abraham, Arpad and Pavoni, Nicola, (2008). Efficient Allocations with Moral Hazard and Hidden Borrowing and Lending: A Recursive Formulation, Review of Economic Dynamics, 11, issue 4, p. 781-803.
2. Albanesi, S., Sleet, C. (2006). Dynamic Optimal Taxation with Private Information. The Review of Economic Studies, 73(1), 1–30. http://www.jstor.org/stable/3700615
3. Ales,L.Maziero,P.(2008). Accounting for private information. Federal Reserve Bank of Minneapolis, working paper
4. Atkeson, A., Lucas, R. E. Jr. (1995). Efficiency and equality in a simple model of efficient unemployment insurance. Journal of Economic Theory 66:64–88.
5. 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)
6. Auerbach,A.(2013). Capital Income Taxation, Corporate Taxation, Wealth Transfer Taxes and Consumption Tax Reforms. This paper has been prepared for the conference, "The Empirical Foundations of SupplySide Economics," at the Becker Friedman Insitutute, University of Chicago, September 27, 2013.
7. Chamley, C. (1986). Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. Econometrica. 54 (3): 607–622. doi:10.2307/1911310.
8. Croft, H. T.; Falconer, K. J.; and Guy, R. K.(1991). Unsolved Problems in Geometry. New York: Springer-Verlag, p. 3, 1991.
9. Diamond, P. (1998). Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates. American Economic Review, 88, 83–95.
10. Diamond, Peter and Mirrlees, James, (1978).A model of social insurance with variable retirement, Journal of Public Economics, 10, issue 3, p. 295-336.
11. Farhi E, Werning I. (2012). Capital Taxation: Quantitative Explorations of the Inverse Euler Equation. Journal of Political Economy, 120, 398 - 445.
12. Farhi E, Werning I. (2013). Insurance and taxation over the life cycle. Rev. Econ. Stud. 80:596–635
13. Farhi, E., Werning, I. (2013). Insurance and Taxation over the Life Cycle. The Review of Economic Studies, 80(2 (283)), 596–635. http://www.jstor.org/stable/43551497
14. Fernandes, Ana and Phelan, Christopher, (2000),.A Recursive Formulation for Repeated Agency with History Dependence, Journal of Economic Theory, 91, issue 2, p. 223-247.
15. Friedman, Milton. (1962). Capitalism and Freedom. Chicago: Univ. Chicago Press.
16. Golosov, M., Kocherlakota, N., & Tsyvinski, A. (2003). Optimal Indirect and Capital Taxation. The Review of Economic Studies, 70(3), 569–587. http://www.jstor.org/stable/3648601
17. Golosov, Mikhail, Aleh Tsyvinski, and Ivan Werning. (2006). New dynamic public finance: A users guide. NBER Macroeconomics Annual 21:317-363.
18. Golosov,M-Troshkin,M,-Tsyvinsky,A.(2011). Optimal Dynamic Taxes. NBER
19. Green, E. (1987). Lending and the smoothing of uninsurable income. In Contractual Arrangements for Intertemporal Trade (ed. E. C. Prescott and N. Wallace). University of Minnesota Press.
20. Jensen, J. L. W. V.(1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta Math. 30, 175-193, 1906.
21. Judd, K.L. (1985). Redistributive taxation in a simple perfect foresight model. Journal of Public Economics 28, 59–83
22. Kapička,(2013). Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach, The Review of Economic Studies, Volume 80, Issue 3, July 2013, Pages 1027–1054

23. Kaplow,L. (2006). On the undesirability of commodity taxation even when income taxation is not optimal. Journal of Public Economics, 90(6–7), 1235–50.
24. Kocherlakota, Narayana, (2010). The New Dynamic Public Finance, 1 ed., Princeton University Press.
25. Laroque, G. R. (2005). Indirect taxation is superfluous under separability and taste homogeneity:A simple proof. Economics Letters, 87(1), 141–144.
26. Ligon, Ethan. (1998). Risk Sharing and Information in Village Economics. Rev.Econ. Studies 65 (4): 847–64.
27. Mankiw NG, Weinzierl M, Yagan D.(2009),Optimal Taxation in Theory and Practice. Journal of Economic Perspectives. 2009;23 (4) :147-174.
28. Milgrom P, Segal I. (2002). Envelope theorems for arbitrary choice sets. Econometrica 70:583–601
29. Milgrom, P. Segal, I.(2002). Envelope Theorems for Arbitrary Choice Sets. EconometricaVolume70, Issue2 March 2002 Pages 583-601
30. Mirrlees, J. A. (1971). An exploration in the theory of optimal income taxation. Review of Economic Studies,38, 175–208.
31. Myerson, R., B.,(1981). Optimal Auction Design. Mathematics of Operations
32. Myerson,R.,(1991). Game Theory: Analysis of Conflict, Harvard University Press, Cambridge
33. Papoulis, A.(1984). Brownian Movement and Markoff Processes. Ch. 15 in Probability, Random Variables, and Stochastic Processes, 2nd ed. New York: McGraw-Hill, pp. 515-553, 1984.
34. Pavan, Alessandro, Ilya Segal, and Juuso Toikka.(2014). Dynamic mechanism design: a Myersonian approach. Econometrica 82, no. 2 (2014): 601–53. http://www.jstor.org/stable/24029270.
35. Rogerson, W. (1985). Repeated moral hazard. Econometrica 53:69–76.
36. Rogerson, W.P. (1985). Repeated Moral Hazard. Econometrica, 53, 69-76.
37. Saez,E.(2001).Using elasticities to derive optimal income tax rates, The Review of Economic Studies, 68(1), pp.205- 229.
38. Sheshinski, Eytan. (1972). The Optimal Linear Income-Tax. Rev. Econ. Studies 39 (3): 297–302
39. Shimer, R., & Werning, I. (2008). Liquidity and Insurance for the Unemployed. The American Economic Review, 98(5), 1922–1942. http://www.jstor.org/stable/29730157
40. Spear, S., and S. Srivastava. (1987). On repeated moral hazard with discounting. Review of Economic Studies 54: 599–617.
41. Stantcheva S. (2017). Optimal taxation and human capital policies over the life cycle. J. Political Econ. 125:1931–90
42. Stantcheva,S.(2020). Dynamic Taxation. Annual Review of Economics Vol. 12:801-831 (Volume publication date August 2020)
43. Stiglitz, J.E.(2018). Pareto efficient taxation and expenditures: pre- and re-distribution. NBER WORKING PAPER SERIES
44. Tuomala, M.(1990). Optimal Income Tax and Redistribution. New York: Oxford University Press.
45. Werning,I.(2007).Pareto efficient income taxation. NBER Public Economics meeting
46. Werning.I.(2002). Optimal Unemployment Insurance with Unobservable Savings. MIT, NBER and UTDT
47. Williams, N. (2011). Persistent private information. Econometrica, 79(4), 1233–1275. http://www.jstor.org/stable/41237860