ANNALS OF ECONOMICS AND FINANCE **16-1**, 199–229 (2015)

Optimal Income Taxations with Information Asymmetry: The Lagrange Multiplier Approach

Xiaojun Zhao

School of Economics, Peking University, Beijing, 100871, China
E-mail: Zhaoxiaojun@pku.edu.cn

This paper considers a general class dynamic optimal problem with incentive compatible constraints. The first-order conditions are derived using the Lagrange multiplier method. Applied the methods developed here, the optimal taxation problems studied by Mirrlees (1971) and Golosov et al. (2003) for more general form utility function are reexamined in this paper. The explicit solutions for optimal income taxations are derived in this paper. Finally, we present numerical solutions for optimal income taxations.

Key Words: Incentive compatible constraint; Asymmetric information; Lagrange functional.
JEL Classification Numbers: C61, D90, H00.

1. INTRODUCTION

Many interesting economic problems involve multiple decision makers in complex environments coupled with information asymmetry. For example, Mirrlees (1971) assumes asymmetric information in a planner's economy and derives the logistical relationship between optimal labor taxation and labor ability (private information). Rogerson (1985a, 1985b) use a dual approach to find the first-order optimal conditions in a static principle-agent problem. Golosov et al. (2003) and Golosov and Tsyvinski (2006) use the method developed by Rogerson (1985a, 1985b) to study optimal indirect taxation in a dynamic environment in which the agent's skill is private information and follows arbitrary stochastic processes. Their work extends Mirrlees' (1971) famous framework to a dynamic framework and shows that positive capital taxation is the Pareto optimal choice.

The dual approach developed by Rogerson (1985a, 1985b) should construct a minimized problem that delivers the same utility to all types as the candidate optimum and satisfies the resource feasibility in all periods.

1529-7373/2015

However, the method cannot be extended directly to a dynamic framework. To apply the dual approach, Golosov et al. (2003) and Golosov and Tsyvinski (2006) assume that the utility is separable in leisure and consumption to examine the optimal capital income taxation. Farhi and Werning (2010) consider the optimal insurance arrangements in a dynamic Mirrlees economy under the life cycle context. If the utility function is not separable, then a disturbance in consumption will not only affect the utility from consumption but also the labor supply, which in turn will affect production. Thus, the dual approach will not work. The separability of consumption and leisure in the utility function plays an important role in solving the problem.

On the other hand, Golosov et al. (2003), in contrast, consider only the optimal capital income tax. Recently, Golosov et al. (2010) consider the optimal labor taxation in the framework of Golosov et al. (2003). Applying the Mirrlees envelope theorem (Mirrlees, 1971), they derive the explicit solution for the optimal labor income taxation under the following specified utility function

$$U(c, l) = \frac{1}{\psi} \exp[-\psi(c - \frac{1}{\gamma} l^{\gamma})],$$

where ψ and γ are positive constants. However, Golosov et al. (2010) cannot depict the optimal taxation rules as the functions of agents' private information; they alternatively provide the relationship between the taxation rules and agents' income. However, Golosov et al. (2010) cannot deal with the problem with the consideration of both capital and labor income taxations. This is because the dual approach will not work because a disturbance in consumption will also affect the labor supply when both labor and capital income taxes are considered.

In fact, Mirrlees (1971), Golosov et al. (2003), and Golosov et al. (2010) can be also viewed as the extensional framework of optimal dynamic mechanism design, which is consistent with the classical principal-agent framework. Pavan et al. (2009) summarize the first order conditions in more general optimal dynamic mechanism design problems. He also uses the Mirrlees envelope theorem to derive the first-order conditions. Garrett and Pavan (2009) apply the same approach presented in Pavan et al. (2009) to investigate the optimal incentive scheme for a manager who faces costly effort decisions and stochastically varying ability over time. They assume that both the principal and agent are risk-neutral and derive the closed form solution. This paper focuses mainly on the evolution of the agent's ability and demonstrates how the shock affects the compensations of the manager.

Both the optimal capital and labor income taxations in a uniform framework, and the optimal income taxation under more general utility function are important problems that remain to be solved. These are the main aims of this paper. We use the Lagrange multiplier method to deal with a class of more general dynamic optimization problems with asymmetric information, and use the method developed here to reexamine the optimal labor income tax and capital income tax in the framework of Golosov et al. (2003).

The remainder of the paper is organized as follows. In Section 2, we introduce mathematic results of functional space. In Section 3, we use the Lagrange multiplier method to deal with general dynamic optimization problems with private information, in which we apply the truth-telling constraint. In Section 4, we re-examine optimal income taxation in both the static and dynamic frameworks. We obtain an optimal income taxation similar to that of Mirrlees (1971), Golosov et al. (2003), and Golosov and Tsyvinski (2006). However, we deal with more general forms of the utility function and with both capital income tax and labor income tax. Section 5 specifies the utility and production functions and presents the numerical solutions for optimal income taxation. We offer our conclusions in Section 5.

2. SOME MATHEMATICAL RESULTS

In this section, we review useful mathematical results that are important in the argument that follows.

2.1. Frechet Differentiation

Suppose that $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are two function spaces, and that N is the set of all linear maps from X to Y. A continuous map $\Lambda \in N$ is said to be differentiable at point $x \in X$ if there exists $A \in N$ such that

$$\Lambda(x') - \Lambda(x) = A(x' - x) + o(\|x' - x\|_X), \text{ for any } x' \in X. \tag{1}$$

A is the **Frechet differentiation** of Λ at x, which we denote as $\delta_{x^*}\Lambda$. Note that the Frechet differentiation is an extension of ordinary differentiation in the $\mathbb{R}^n$ space. When $Y = \mathbb{R}$, we call Λ a **functional**. We state some important properties of Λ in the following propositions.

PROPOSITION 1. *Suppose that $\Lambda : X \to R$ is differentiable on X and there exists $x^* \in U \subset X$ such that $\Lambda(x) \leq \Lambda(x^*)$ for all $x \in U$. Then,*

$$\delta_{x^*}\Lambda = 0. \tag{2}$$

Proof. The proof is easy and we omit it for simplicity. ∎

Proposition 1 is an extension of the traditional optimization problem in $\mathbb{R}^n$.

Example 2.1. If $X = C[0,1]$ is the set of all continuous functions on $[0,1]$ and $\Lambda(x) = f(x(1/2))$, then

$$\delta_x \Lambda(\eta) = f'(x(1/2))\eta(1/2) \quad \text{for all } \eta \in X.$$

Because $\delta_x \Lambda$ is also a functional, from Proposition 1, $\delta_x \Lambda = 0$ means that

$$f'(x(1/2))\eta(1/2) = 0 \quad \text{for all } \eta \in X,$$

which indicates that $f'(x(1/2)) = 0$.

Example 2.2. Suppose that X is the set of all integrable functions in the probability space $(\Omega, \mathcal{F}, \mu)$, and $\Lambda(x) = \int f(x) d\mu$. Then,

$$\delta_x \Lambda(\eta) = \int f'(x)\eta d\mu \quad \text{for all } \eta \in X.$$

Furthermore, $\delta_x \Lambda = 0$ means that $f'(x) = 0$.

Definition 2.1. Λ is a functional on X that is concave if for every $x, y \in X$ and any $\lambda \in [0,1]$ we have

$$\Lambda(\lambda x + (1-\lambda)y) \leq \lambda \Lambda(x) + (1-\lambda)\Lambda(y). \tag{3}$$

Furthermore, Λ is strictly concave on X if for any $x \neq y$ and $\lambda \in (0,1)$

$$\Lambda(\lambda x + (1-\lambda)y) < \lambda \Lambda(x) + (1-\lambda)\Lambda(y). \tag{4}$$

The following proposition connects the Frechet differentiation and this concavity.

Proposition 2. *If Λ is differentiable and concave on X, then for every x and $y \in X$*

$$\Lambda(x) - \Lambda(y) \geq \delta_x \Lambda(x-y). \tag{5}$$

Proof. The proof is similar to that for $X = \mathbb{R}^n$, and we thus omit it. ▌

2.2. The Lagrange Functional

In this section, we consider a discrete-time dynamic optimization problem. First, we define the basic space $\mathbb{N}_T \times \Omega$, where $\mathbb{N}_T = \{1, \ldots, T\}$ is the subset of natural numbers and Ω is the sample space. The $(\Omega, \mathcal{F}, \mu)$ is a probability space with the increasing filtration $\mathbb{F} = \{\mathcal{F}_t : t \in \mathbb{N}_T\}$, $\mathcal{F}$ is the union of all $\mathcal{F}_t$ in $\mathbb{F}$, and μ is the measurement on Ω. We define a sequence of function spaces

$$M_t = \{x : \Omega \to \mathbb{R}, x \text{ is bounded and } x \in \mathcal{F}_t\},\ t \in \mathbb{N}_T, \tag{6}$$

where $\mathbb{R}$ is a set of all real numbers. We then define the norm on M_t as $\|x\| = \max_{\omega \in \Omega} |x(\omega)|$. It follows from Stokey et al. (1989) and Stokey (2009) that $(M_t, \|.\|)$ is a Banach space.

The dynamic optimization problem is

$$\mathbf{P1}: \qquad \max_{x_t, u_t \in M_t} E[\sum_{t=1}^{T} \beta^{t-1} f(x_t, u_t)], \tag{7}$$

subject to

$$x_{t+1} = g(x_t, u_t, \varepsilon_{t+1}), \quad t = 1, \ldots, T-1, \tag{8}$$

with the given initial condition x_1.

Here, $x_t, u_t \in M_t$ are state and control variables, respectively. $\beta \in (0, 1)$ is the discount factor and $\varepsilon_t \in M_t$ is the stochastic disturbance. f and g are continuous and differentiable functions that are concave with respect to (x, u).

To solve the optimization problem, we define the Lagrange functional as

$$\mathcal{L}(x, u, z) = E\left[\sum_{t=1}^{T} \beta^{t-1} f(x_t, u_t)\right] + \sum_{t=1}^{T-1} \langle z_{t+1},\ g(x_t, u_t, \varepsilon_{t+1}) - x_{t+1}\rangle, \tag{9}$$

where $z_t \in M_t, t = 2, \ldots, T$ is the associated Lagrange multipliers and $\langle ., . \rangle$ is the inner product, which is defined as $\langle x, y\rangle = E(xy) = \int xyd\mu$.

This leads to the following proposition.

PROPOSITION 3. *If $(x_t^*, u_t^*), t = 1, \ldots, T$ is a solution for the optimization problem **(P1)**, then there exists $z_t^* \in M_t, t = 2, \ldots, T$ such that (x_t^*, u_t^*, z_t^*) is a local saddle point of the functional $\mathcal{L}(x, u, z)$, namely, there exists $\varepsilon > 0$ such that*

$$\mathcal{L}(x^*, u^*, z) \geq \mathcal{L}(x^*, u^*, z^*) \geq \mathcal{L}(x, u, z^*), \tag{10}$$

for any $x \in B(x^*, \varepsilon)$, $u \in B(u^*, \varepsilon)$, and $z \in B(x^*, \varepsilon)$.

The proof is presented in the Appendix.

We can use Frechet differentiations to find the optimal conditions for the optimization problem (**P1**). The first-order conditions are $\delta_{(x^*,u^*)}\mathcal{L} = 0$. The proof is rather easy, and we omit it.

3. DOPS WITH INCENTIVE COMPATIBLE CONSTRAINTS

In this section, we consider a generalized dynamic optimal problem with incentive compatible constraints. There are two decision makers, A and B. In examples of this problem, Golosov et al. (2003) consider an economy with a social planner and agents, and Rogenson (1985a, 1985b) considers an economy with a principle and an agent. For convenience, we call A the social planner and B the agent. The social planner allocates economic resources, and each agent has his or her own private information that is hidden to the social planner. However, the social planner wants each agent to report his or her private information. We call this strategy the *truth-telling* strategy, following the studies of Mirrlees (1971, 1976) and Diamond and Mirrlees (1986).

3.1. The Basic Settings

Suppose that there is an infinite number of long-lived agents situated in the real interval $[0,1]$. Each agent is indexed by j, $j \in [0,1]$ and has private information θ_t in period t. We suppose that θ_t is independent and identically distributed (i.i.d) across agents, and that $\{\theta_t\}, t = 1, \dots, T$ is defined on a probability space $(\Omega, \mathcal{F}, \mu)$ that is adapted to the increasing filtration $\mathbb{F} = \{\mathcal{F}_t\}, t = 1, \dots, T$, where $\mathcal{F}$ is the union of all of the elements in $\mathbb{F}$.

In each period, the agent with private information θ is rewarded $f(x, u, \theta)$ through the allocation x, u, where x is the state variable and can be deemed as the resource and u is the control variable.

As the social planner aims to maximize the sum of rewards for the agents in the economy, from Uhlig (1996) we obtain

$$\max_{x_t^j, u_t^j} \sum_{j\in[0,1]} \sum_{t=1}^{T} \beta^{t-1} f(x_t^j, u_t^j, \theta_t^j) = \max_{x_t, u_t} E_\theta[\sum_{t=1}^{T} \beta^{t-1} f(x_t, u_t, \theta_t)], \qquad (11)$$

where E_θ is the expectation on the random variables θ_t and $\beta \in (0,1)$ is the discount factor.

There are two types of constraints. The first comprises the resource constraints

$$x_{t+1} = g(x_t, \Lambda(u_t, \theta_t)); \quad t = 1, \dots, T-1, \qquad (12)$$

where $\Lambda(u_t, \theta_t) = (\Lambda_1(u_t, \theta_t), \ldots, \Lambda_m(u_t, \theta_t))$ is a functional of the random variables u_t and θ_t and is defined as the aggregate effect of private information and resource allocations on resource accumulation.

The other type of constraint for the social planner is the incentive compatible constraint (IC). Suppose that the social planner wants the agent to report his or her private information to enable the allocation of social resources. A report strategy is defined as maps $\sigma : \Theta^T \to \Theta^T$, and the set of all report strategies is denoted by Σ. Under report strategy σ, we denote the discounted gains for the agent with private information $\theta \in \Theta^T$ as

$$W(\sigma, \theta) = \sum_{t=1}^{T} \beta^{t-1} f(x_t(\sigma(\theta)), u_t(\sigma(\theta)), \theta_t). \tag{13}$$

σ^* is a *truth-telling* strategy if $\sigma^*(\theta) = \theta$ for all θ. Thus, an allocation is incentive compatible (IC) if

$$W(\sigma^*, \theta) \geq W(\sigma, \theta) \quad \text{for any} \quad \theta \in \Theta^T, \sigma \in \Sigma. \tag{14}$$

The dynamic optimization problem with IC constraints can then be summarized as

$$\textbf{P2:} \quad \max_{x_t, u_t} E_\theta \left[\sum_{t=1}^{T} \beta^{t-1} f(x_t, u_t, \theta_t) \right],$$

subject to the resource constraints (12) and the IC constraints (14) with the given initial condition x_1.

In the next subsection, we use the Lagrange method to solve the optimization problem (**P2**). For simplification, we assume that $x_t \in \mathbb{R}$ and $u_t \in \mathbb{R}^n$.

3.2. Optimal Conditions

In this section, we use the Lagrange method to find the optimal conditions for the dynamic optimization problem (P2). First, the IC constraint (14) can be rewritten as

$$\sum_{t=1}^{T} \beta^{t-1} f(x_t(\theta'), u_t(\theta'), \theta_t) \leq \sum_{t=1}^{T} \beta^{t-1} f(x_t(\theta), u_t(\theta), \theta_t), \tag{15}$$

for any $\theta \in \Theta^T$ and $\theta' \in \Theta^T$.

To solve the optimization problem, we define the Lagrange functional

$$\begin{aligned} \mathcal{L}(x, u) &= E_\theta[\sum_{t=1}^{T} \beta^{t-1} f(x_t, u_t, \theta_t)] + \sum_{t=1}^{T} \beta^{t-1} \lambda_t (g(x_t, \Lambda(u_t, \theta_t)) - x_{t+1}) \\ &+ E_{\theta', \theta} \lambda(\theta', \theta) \sum_{t=1}^{T} \beta^{t-1} (f(x_t(\theta), u_t(\theta), \theta_t) - f(x_t(\theta'), u_t(\theta'), \theta_t)), \end{aligned} \tag{16}$$

where $\lambda(\theta',\theta) : \Theta^T \times \Theta^T \longrightarrow \mathbb{R}$ are the Lagrange multipliers[1] for the IC constraint (15) and $\lambda_t,\ t = 1, 2, \ldots, T$ are the Lagrange multipliers for the resource constraints (12). It is easy to find that $\lambda(\theta',\theta)$ is measurable on $\mathcal{F} \times \mathcal{F}$, and that λ_t is a non-stochastic number.

This leads to the following proposition.

Proposition 4. *The optimal conditions for the optimization problem* ***P2*** *are*

$$\begin{aligned}\lambda_t &= \beta\lambda_{t+1} g_x(x_{t+1}(\theta), \Lambda(u_{t+1}(\theta), \theta_{t+1})) \\ &+ \beta E[(1 + E_{\theta'}\lambda(\theta',\theta)) f_x(x_{t+1}(\theta), u_{t+1}(\theta), \theta_{t+1})] \\ &- \beta E\left[\lambda(\theta,\theta') f_x(x_{t+1}(\theta), u_{t+1}(\theta), \theta'_{t+1})\right]\end{aligned} \tag{17}$$

and

$$\begin{aligned}&\lambda_t \sum_{j=1}^{m} g_{\Lambda_j}(x_t, \Lambda(u_t,\theta_t)) \delta_{u^i}\Lambda_j(u_t,\theta_t) \\ &= -(1 + E_t[E_{\theta'}\lambda(\theta',\theta)]) f_{u^i}(x_t(\theta), u_t(\theta), \theta_t) \\ &+ E_t\left[E_{\theta'}\left[\lambda(\theta,\theta') f_{u^i}(x_t(\theta), u_t(\theta), \theta'_t)\right]\right],\end{aligned} \tag{18}$$

where $\delta_{u^i}\Lambda_j$ is the Rize reprentative of the Frechet differentiation of Λ_j with respect to u^i at t.

Proof. The Lagrange functional (16) can be rewritten as

$$\begin{aligned}\mathcal{L}(x,u) &= \sum_{t=1}^{T} \beta^{t-1}\left[(1 + E_{\theta'}\left[\lambda(\theta',\theta)\right]) f(x_t(\theta), u_t(\theta), \theta_t)\right] \\ &- \sum_{t=1}^{T} \beta^{t-1} E_{\theta'}\left[\lambda(\theta,\theta') f(x_t(\theta), u_t(\theta), \theta'_t)\right] \\ &+ \sum_{t=1}^{T} \beta^{t-1}\lambda_t (g(x_t, \Lambda(u_t,\theta_t)) - x_{t+1}).\end{aligned}$$

We then obtain the optimialities by taking the Frechet differentiations of $\mathcal{L}$ with respect to x_t and $u_t^i,\ i = 1, \ldots, n$.

$$\begin{aligned}&\langle \lambda_t - \beta\lambda_{t+1} g_x(x_{t+1}, \Lambda(u_{t+1}, \theta_{t+1})), \eta\rangle \\ &= \langle \beta(1 + E_{\theta'}\left[\lambda(\theta',\theta)\right]) f_x(x_{t+1}(\theta), u_{t+1}(\theta), \theta_{t+1}), \eta\rangle \\ &- \langle \beta E_{\theta'}\left[\lambda(\theta,\theta') f_x(x_{t+1}(\theta), u_{t+1}(\theta), \theta'_{t+1})\right], \eta\rangle\end{aligned} \tag{19}$$

[1]Here, $\lambda(\theta',\theta)$ means that an agent has private information θ but reports θ'.

$$\begin{aligned}&\langle (1 + E_{\theta'}\left[\lambda(\theta', \theta)\right]) f_{u^i}(x_t(\theta), u_t(\theta), \theta_t) - E_{\theta'}\left[\lambda(\theta, \theta') f_{u^i}(x_t(\theta), u_t(\theta), \theta'_t)\right] \\ + \ & \lambda_t \sum_{j=1}^{m} g_{\Lambda_j}(x_t, \Lambda(u_t, \theta_t)) \delta_{u^i} \Lambda_j(u_t, \theta_t), \xi \rangle = 0 \qquad (20)\end{aligned}$$

for all $\xi \in \mathcal{F}_t$ and $\eta \in \mathbb{R}$.

From equations (19) and (20), it is easy to see that equations (17) and (18) hold. ▌

Proposition 4 can be used to deal with more general dynamic optimization problems with incentive compatible constraints. From equations (17) and (18), we can derive the "inverse Euler equation" similar to Golosov et al. (2003) and many other existing literatures. To understand these conditions more intuitional, we define

$$\widetilde{\lambda}(\theta, \theta_t) = 1 + E_{\theta'}[\lambda(\theta', \theta) - \lambda(\theta, \theta') \frac{f_x(x_t(\theta), u_t(\theta), \theta'_t)}{f_x(x_t(\theta), u_t(\theta), \theta_t)}],$$

and

$$\mu_i(\theta, \theta_t) = 1 + E_{\theta'}[\lambda(\theta', \theta) - \lambda(\theta, \theta') \frac{f_{u^i}(x_t(\theta), u_t(\theta), \theta'_t)}{f_{u^i}(x_t(\theta), u_t(\theta), \theta_t)}].$$

Then, equations (17) and (18) can be rewritten as

$$\lambda_t = \beta\left[\lambda_{t+1} g_x(x_{t+1}(\theta), \Lambda(u_{t+1}(\theta), \theta_{t+1}))\right] + \beta E[\widetilde{\lambda}(\theta, \theta_{t+1}) f_x(x_{t+1}(\theta), u_{t+1}(\theta), \theta_{t+1})],$$

and

$$\lambda_t \sum_{j=1}^{m} g_{\Lambda_j}(x_t(\theta), \Lambda(u_t(\theta), \theta_t)) \delta_{u^i} \Lambda_j(u_t(\theta), \theta_t) = -E_t\left[\mu_i(\theta, \theta_t)\right] f_{u^i}(x_t(\theta), u_t(\theta), \theta_t).$$

Furthermore, we define

$$\varphi_i(t) = \frac{f_{u^i}(x_t(\theta), u_t(\theta), \theta_t)}{\sum_{j=1}^{m} g_{\Lambda_j}(x_t(\theta), \Lambda(u_t(\theta), \theta_t)) \delta_{u^i} \Lambda_j(u_t(\theta), \theta_t)},$$

and

$$\begin{aligned}\psi_i(t) \ &= \ g_x(x_t(\theta), \Lambda(u_t(\theta), \theta_t)) \\ &- \ \sum_{j=1}^{m} g_{\Lambda_j}(x_t(\theta), \Lambda(u_t(\theta), \theta_t)) \delta_{u^i} \Lambda_j(u_t(\theta), \theta_t) \frac{f_x(x_t(\theta), u_t(\theta), \theta_t) E_t[\widetilde{\lambda}(\theta, \theta_t)]}{f_{u^i}(x_t(\theta), u_t(\theta), \theta_t) E_t\left[\mu_i(\theta, \theta_t)\right]}.\end{aligned}$$

Therefore, we have

$$\lambda_t = \beta \left[\lambda_{t+1} E \psi_i(t+1) \right], \tag{21}$$

and

$$\lambda_t = -E_t \left[\mu_i(\theta, \theta_t) \right] \varphi_i(t). \tag{22}$$

From equations (21) and (22), we can derive a generalized inverse Euler equation similar to that in Golosov et al. (2003) and Golosov and Tsyvinski (2006). However, the method developed here is more general, and can be deal woth more general dynamic problems with asymmetric information. In the next section, we will apply this method to study the optimal income taxations in an general economic system with incentive compatible constraints.

4. THE OPTIMAL INCOME TAXATIONS

In public finance and macroeconomics research, attempts to incorporate information asymmetry into economic models to study the optimal fiscal policy begin with Mirrlees (1971). He derives the optimal labor income tax and asserts the logistic relationship between the optimal labor income tax rate and the agent's ability (private information). His seminal findings have encouraged many scholars to step into this area. Golosov et al. (2003) extend Mirrlees' (1971) static framework to a dynamic framework to analyze the optimal capital income tax. Golosov et al. (2003) do not consider the optimal labor income tax, and we cannot compare the optimal labor income tax in the static and dynamic frameworks. In this section, we apply the Lagrange method to re-examine Mirrlees' (1971) optimal labor income tax in a static environment and the Golosov et al. (2003) optimal income taxes (capital income tax and labor income tax) in a dynamic environment.

4.1. The Static Mirrlees Model

The framework of the static Mirrlees model is simple. The agent's labor ability θ is a random variable on the measurable space $(\Omega, \mathcal{F}, \mu)$, and the range of θ is Θ. Thus, an agent's effective labor supply l is $l \times \theta$, which we denote as y. The agent's utility function is defined by his or her consumption c and the labor supply l, which mathematically is $U(c, y/\theta)$.

4.1.1. Social planner's economy

Suppose that the production function is $H(Y)$, where $Y = \int y(\theta) d\mu$ is the aggregate effective labor supply in the economy.

The optimal allocation problem for the social planner is

$$\max_{c(\theta),y(\theta)} \int U(c(\theta), y(\theta)/\theta) d\mu \tag{23}$$

subject to the resource constraint

$$\int c(\theta) d\mu = H(\int y(\theta) d\mu), \tag{24}$$

and IC constraints

$$U(c(\theta), y(\theta)/\theta) \geq U(c(\theta'), y(\theta')/\theta), \text{ for any } \theta, \theta' \in \Theta. \tag{25}$$

From equations (17) and (18), the optimal conditions for this optimization problem can be easily obtained as

$$\lambda = (1 + E_{\theta'}\lambda(\theta', \theta))U_c(c(\theta), y(\theta)/\theta) - E_{\theta'}\left[\lambda(\theta, \theta')U_c(c(\theta), y(\theta)/\theta')\right], \tag{26}$$

and

$$\begin{aligned} -\lambda H'(Y) &= (1 + E_{\theta'}\lambda(\theta', \theta))U_l(c(\theta), y(\theta)/\theta)/\theta \\ &\quad - E_{\theta'}\left[\lambda(\theta, \theta')U_l(c(\theta), y(\theta)/\theta')/\theta'\right], \end{aligned} \tag{27}$$

where λ is the Lagrange multiplier for the resource constraint (24) and $\lambda(\theta', \theta)$ is the Lagrange multipliers for the IC constraints (25).

To obtain some interesting results, we place further restrictions on $U(c, l)$.

Case 1. $U(c(\theta), y(\theta)/\theta') = u(c(\theta), y(\theta)/\theta)v(\theta, \theta')$ *with* $v(\theta, \theta) = 1$.

In this case, we let $\lambda(\theta) = 1 + E_{\theta'}[\lambda(\theta', \theta)(1 - v(\theta, \theta'))]$, which means that equations (26) and (27) become

$$\lambda = \lambda(\theta)u_c(c(\theta), y(\theta)/\theta) \text{ and } -\lambda H'(Y) = \lambda(\theta)u_l(c(\theta), y(\theta)/\theta)/\theta.$$

This gives

$$-u_l(c(\theta), y(\theta)/\theta)/\theta = u_c(c(\theta), y(\theta)/\theta)H'(Y). \tag{28}$$

Case 2. $U(c(\theta), y(\theta)/\theta') = u(c(\theta)) - v(y(\theta)/\theta)w(\theta, \theta')$ *with* $w(\theta, \theta) = 1$.

In this case, equations (26) and (27) become

$$\lambda = (1 + E_{\theta'}(\lambda(\theta', \theta) - E_{\theta'}\lambda(\theta, \theta'))u'(c(\theta)),$$

and

$$\lambda H'(Y) = (1 + E_{\theta'}\lambda(\theta', \theta) - E_{\theta'}\left[\lambda(\theta, \theta')w(\theta, \theta')\right])v'(y(\theta)/\theta)/\theta.$$

This gives

$$v'(y(\theta)/\theta)/\theta = \frac{1 + E_{\theta'}[\lambda(\theta', \theta) - \lambda(\theta, \theta')]}{1 + E_{\theta'}[\lambda(\theta', \theta) - \lambda(\theta, \theta')w(\theta, \theta')]} u'(c(\theta))H'(Y). \tag{28}$$

Equation (29) is the same as the optimal conditions in Mirrlees (1971) with the the separable utility function.

Case 3. $U(c(\theta), y(\theta)/\theta') = u(c(\theta))v(y(\theta)/\theta')$.

In this case, we let

$$\lambda(\theta, y) = 1 + E_{\theta'}\left[\lambda(\theta', \theta) - \lambda(\theta, \theta')\frac{v(y/\theta')}{v(y/\theta)}\right]$$

and

$$\mu(\theta, y) = 1 + E_{\theta'}\left[\lambda(\theta', \theta) - \lambda(\theta, \theta')\frac{v'(y/\theta')\theta}{v'(y/\theta)\theta'}\right].$$

The optimal conditions (26) and (27) can be deduced to

$$\frac{v'(y(\theta)/\theta)/\theta}{v(y(\theta)/\theta)} = \frac{\lambda(\theta, y)}{\mu(\theta, y)}\frac{u'(c(\theta))}{u(c(\theta))}H'(Y). \tag{29}$$

Equations (28), (29), and (30) determine the optimal conditions under the three specified utility function, we will use them to derive the optimal labor income taxation accordingly.

4.1.2. Decentralized economy

To derive the implementation mechanism, we consider a decentralized economy. Suppose that there is a representative firm in the economy and that agents work in it to earn income. The government levies a labor income tax in its budget.

Agents.

An agent with ability θ aims to maximize his/her utility under the budget constraint, that is,

$$\max_{c,y} U(c(\theta), y(\theta)/\theta)$$

subject to the budget constraint

$$c(\theta) = (1 - \tau^w)\omega y(\theta) + \chi(\theta), \tag{30}$$

where ω is the wage rate, τ^w is the labor incoem tax rate, and $\chi(\theta)$ is government's transfer.

It is easy to derive the optimal condition

$$-U_l(c(\theta), y(\theta)/\theta)/\theta = (1-\tau^w)\omega U_c(c(\theta), y(\theta)/\theta). \tag{31}$$

Firms.
The firm's profit maximization $\pi = \max_Y H(Y) - \omega Y$ indicates that

$$\omega = H'(Y). \tag{32}$$

Government.
Under the balance budget constraint, government's transfer equals the tax revenue, namely,

$$\int \chi(\theta) d\mu = \int \tau^w \omega y(\theta) d\mu + \pi. \tag{33}$$

Macroeconomic equilibrium.
Combining equations (31), (32), (33), and (34), we obtain the optimal conditions in the macroeconomic equlibrium

$$-U_l(c(\theta), y(\theta)/\theta)/\theta = (1-\tau^w)H'(Y)U_c(c(\theta), y(\theta)/\theta), \tag{34}$$

and

$$\int c(\theta) d\mu = H(\int y(\theta) d\mu). \tag{35}$$

Equations (35) and (36) can be used to derive the optimal labor income taxation.

4.1.3. Optimal taxations

In this subsection, we will derive the optimal taxations by setting the tax policy in the decentralizaed economy implement the social optimum. We will consider three cases respectively.

In case 1. Equation (35) is compared to equation (28) to give an optimal labor income tax rate of zero.

In case 2. Equation (35) is compared to equation (29) to give an optimal labor income rate of

$$\tau^w(\theta) = \frac{E_{\theta'}\left[\lambda(\theta, \theta')(1 - w(\theta, \theta'))\right]}{1 + E_{\theta'}[\lambda(\theta', \theta)(1 - w(\theta, \theta'))]}, \tag{36}$$

which is dependent on agent's labor ability θ.

In case 3. Under the specified utlity form in this case, the optimal condition (35) can be rewritten as

$$\frac{v'(y(\theta)/\theta)/\theta}{v(y(\theta)/\theta)} = (1-\tau^w)\frac{u'(c(\theta))}{u(c(\theta))}H'(Y).$$

Comparing the above equation with equation (30), the optimal labor income tax rate can be easily obtained as

$$\tau^w = \frac{E_{\theta'}[\lambda(\theta,\theta')(\frac{v(y/\theta')}{v(y/\theta)} - \frac{v'(y/\theta')\theta}{v'(y/\theta)\theta'})]}{1+E_{\theta'}[\lambda(\theta',\theta)-\lambda(\theta,\theta')\frac{v'(y/\theta')\theta}{v'(y/\theta)\theta'}]}. \tag{37}$$

Remark 4.1. Under the specified utility function in case 1, the optimal labor income tax rate is zero, which is the same as in many studies, such as those of, Chamley (1986) and Judd (1985) etc. Many familiar utility functions satisfy assumption in case 1. For example, suppose that

$$U(c,l) = \frac{(c^\omega l^{-\varepsilon})^{1-\sigma}}{1-\sigma},$$

where ω, ε, and $\sigma > 0$ are constants. Then,

$$U(c,y(\theta)/\theta') = \frac{(c^\omega(y(\theta)/\theta')^{-\varepsilon})^{1-\sigma}}{1-\sigma} = \frac{(c^\omega(y(\theta)/\theta)^{-\varepsilon})^{1-\sigma}}{1-\sigma}(\theta/\theta')^{-\varepsilon(1-\sigma)}.$$

It is easy to find that the optimal labor income tax is zero with this utility function.

Remark 4.2. Under the separable utility in case 2, equation (38) gives the optimal labor income tax rate, which is similar to that of Mirrlees (1971). For example, suppose that the utility function

$$U(c,l) = u(c) - l^\alpha,$$

for $0 < \alpha < 1$. Then,

$$U(c,y(\theta)/\theta') = u(c) - (y(\theta)/\theta)^\alpha(\theta/\theta')^\alpha,$$

and $w(\theta,\theta') = (\theta/\theta')^\alpha$.

Thus, for the agent with the greatest ability, where $\theta \geq \theta'$, we have $\tau^w(\theta) \leq 0$, and for the agent with the least ability we have $\tau^w(\theta) \geq 0$. We can also use numerical simulation to find the logistic relationship between the labor income tax rate and the agent's ability, but for simplicity we omit this discussion and instead consider the dynamic Mirrlees model.

Remark 4.3. Case 3 presents the optimal labor income tax for more general utility funtions, for example, similar to Golosov et al. (2010),

$$U(c, l) = -\frac{1}{\psi} \exp[-\psi(c - \frac{1}{\gamma} l^{\gamma})],$$

where γ and ψ are positive constant.

4.2. The Dynamic Mirrlees Model

In this section, we re-examine the model introduced in Golosov et al. (2003) and give its explicit solution. Golosov et al. (2003) consider an environment in which an agent's skill is private information. Following Golosov et al. (2003), we define an agent's utility by his or her consumption c_t and the amount of time spent working in period t by l_t, which mathematically is $U(c_t, l_t)$. We assume that θ_t is the agent's skill vector in period t and that both θ_t and l_t are private information of the agent. We present the framework briefly in the following. More details can be found in the Golosov et al. (2003).

4.2.1. Decentralized economy

The foregoing sections tell the story for a planner's economy. In this section, we implement the optimal allocation from the planner's economy in a decentralized economy, following the framework of Chamley (1986).

Agents.

Suppose that the government levies a capital income tax and a labor income tax, denoted as τ_t^k and τ_t^w, respectively. Further, suppose that an agent at time t only knows his or her own ability θ_t, but does not know his or her ability θ_s for all $s > t$. The agent thus faces future uncertainty, and must determine the consumption and labor supply by solving the following problem

$$\max_{c_t, y_t} E[\sum_{t=1}^{\infty} \beta^{t-1} U(c_t, y_t/\theta_t)], \tag{38}$$

subject to

$$A_{t+1} = (1 - \tau_t^k) r_t A_t + (1 - \tau_t^w) w_t y_t - c_t + \chi_t, \tag{39}$$

with the given initial asset A_0.

Here, A_t is the agent's asset at time t, r_t and w_t are the market interest rate and wage rate, respectively, and χ_t is the government's transfer.

The first-order conditions can be easily derived as

$$U_c(c_t, y_t/\theta_t) = \beta E_t\{(1 - \tau^k_{t+1})r_{t+1}U_c(c_{t+1}, y_{t+1}/\theta_{t+1})\}, \tag{40}$$

and

$$(1 - \tau^w_t)w_t U_c(c_t, y_t/\theta_t) = -U_l(c_t, y_t/\theta_t)/\theta_t, \tag{41}$$

where U_c and U_l are the partial differentiations of consumption c and labor $l = y/\theta$, respectively.

Firms.

Suppose that output is produced by firms with capital and efficient labor input. The production function is thus neoclassical. The profit maximization problem for firms is then

$$\max_{k_t, y_t} f(k_t, \int y_t d\mu) - r_t k_t - w_t \int y_t d\mu + (1 - \delta)k_t,$$

where k_t and $\int y_t d\mu$ are the aggregate capital and aggregate labor force, respectively, and δ is the rate of depreciation of capital.

The first-order conditions are

$$r_t = f_k(k_t, \int y_t d\mu) + 1 - \delta, \qquad w_t = f_y(k_t, \int y_t d\mu), \tag{42}$$

where $f_y(k_t, \int y_t d\mu)$ is the partial differentiation of the aggergate effective labor $\int y_t d\mu$ at time t.

Equilibrium.

In macro-equilibrium, $k_t = \int A_t d\mu$ and $\int \chi_t d\mu = 0$, and we thus have

$$k_{t+1} = (1 - \tau^k_t)r_t k_t + (1 - \tau^w_t)w_t \int y_t d\mu - \int c_t d\mu. \tag{43}$$

This is just the resource constraint of the economy.

Substituting equation (43) into equations (41) and (42), we arrive at

$$U_c(c_t, y_t/\theta_t) = \beta E_t\{(1-\tau^k_{t+1}) \left[f_k(k_{t+1}, \int y_{t+1} d\mu) + 1 - \delta\right] U_c(c_{t+1}, y_{t+1}/\theta_{t+1})\}, \tag{44}$$

and

$$(1-\tau_t^w) f_y(k_t, \int y_t d\mu) U_c(c_t, y_t/\theta_t) = -\frac{U_l(c_t, y_t/\theta_t)}{\theta_t}. \tag{45}$$

The dynamics for a competition economy are thus described by equations (44), (45), and (46).

4.2.2. Social planner's economy

Similar to Golosov et al. (2003), the social planner can observe the effective labor $y_t = l_t\theta_t$ for the agent and use the *truth-telling* mechanism to obtain private information. The social planner maximizes the social welfare subject to the social resource constraints and incentive compatible constraints. This gives the following dynamic optimization problem.

$$\max_{c_t, y_t} E_\theta[\sum_{t=1}^{\infty} \beta^{t-1} U(c_t, y_t/\theta_t)], \tag{46}$$

subject to

$$k_{t+1} = f(k_t, \int y_t d\mu) - \int c_t d\mu + (1-\delta)k_t, \tag{47}$$

and

$$W(\sigma^*, \theta) \geq W(\sigma, \theta), \tag{48}$$

with the given initial condition k_1, where $W(\sigma, \theta)$ is defined as

$$W(\sigma, \theta) = E_\theta[\sum_{t=1}^{\infty} \beta^{t-1} U(c_t(\sigma(\theta_t)), y_t(\sigma(\theta_t))/\theta_t)].$$

Similar to equations (17) and (18), the optimal conditions can be easily derived as

$$\lambda_t = \beta\lambda_{t+1} R_{t+1}, \tag{49}$$

$$\begin{aligned} \lambda_t &= E_t\left[(1 + E_{\theta'}\lambda(\theta', \theta))\right] U_c(c_t(\theta), y_t(\theta)/\theta_t) \\ &- E_t\left[E_{\theta'}\left[\lambda(\theta, \theta') U_c(c_t(\theta), y_t(\theta)/\theta'_t)\right]\right], \end{aligned} \tag{50}$$

and

$$\begin{aligned} -\lambda_t f_y(k_t, \int y_t d\mu) &= E_t\left[1 + E_{\theta'}\lambda(\theta', \theta)\right] U_l(c_t(\theta), y_t(\theta)/\theta_t)/\theta_t \\ &- E_t\left[E_{\theta'}\left[\lambda(\theta, \theta') U_l(c_t(\theta), y_t(\theta)/\theta'_t)/\theta'_t\right]\right], \end{aligned} \tag{51}$$

where $R_t = f_k(k_t, \int y_t d\mu) + 1 - \delta$.

Equations (50), (51), and (52) determine the competitive equilibrium. Similar to the static Mirrlees economy, we specify the following cases.

Case 1. $U(c(\theta), y(\theta)/\theta') = u(c(\theta), y(\theta)/\theta)v(\theta, \theta')$ *with* $v(\theta, \theta) = 1$.

We denote $\lambda(\theta) = E_{\theta'}[\lambda(\theta', \theta) - \lambda(\theta, \theta')]$ and $\widetilde{\lambda}(\theta_t, \theta'_t) = 1 + E_{\theta'}[\lambda(\theta', \theta) - \lambda(\theta, \theta')v(\theta_t, \theta'_t)]$. Equations (50), (51), and (52) then become

$$\lambda_t = \beta \lambda_{t+1} R_{t+1}, \tag{52}$$

$$\lambda_t = E_t\left[\lambda(\theta)\right] u_c(c_t(\theta), y_t(\theta)/\theta_t), \tag{53}$$

and

$$-\lambda_t f_y(k_t, \int y_t d\mu) = E_t\left[\lambda(\theta)\right] u_l(c_t(\theta), y_t(\theta)/\theta_t)/\theta_t. \tag{54}$$

Note that $\lambda_t \in \mathbb{R}$, and we thus obtain

$$u_c(c_t(\theta), y_t(\theta)/\theta_t) = \beta R_{t+1} \frac{1}{E_t[1/u_c(c_{t+1}(\theta), y_{t+1}(\theta)/\theta_{t+1})]}, \tag{55}$$

and

$$-u_l(c_t(\theta), y_t(\theta)/\theta_t)/\theta_t = f_y(k_t, \int y_t d\mu) u_c(c_t(\theta), y_t(\theta)/\theta_t). \tag{56}$$

Equation (56) is an inverse Euler equation that is the same as used by Golosov et al. (2003). Comparing equations (45) and (46) with equations (56) and (57) gives

$$\tau_t^w = 0 \tag{57}$$

and

$$E_t\left[(1 - \tau_{t+1}^k) R_{t+1} u_c(c_{t+1}(\theta), y_{t+1}(\theta)/\theta_{t+1})\right] = R_{t+1} \frac{1}{E_t[1/u_c(c_{t+1}(\theta), y_{t+1}(\theta)/\theta_{t+1})]}.$$

Thus,

$$\begin{aligned} \tau_{t+1}^k &= 1 - \frac{1}{u_c(c_{t+1}(\theta), y_{t+1}(\theta)/\theta_{t+1}) E_t[1/u_c(c_{t+1}(\theta), y_{t+1}(\theta)/\theta_{t+1})]} \\ &+ \frac{\epsilon}{u_c(c_{t+1}(\theta), y_{t+1}(\theta)/\theta_{t+1})}, \end{aligned} \tag{58}$$

where ϵ satisfies $E_t\left[\epsilon\right] = 0$.

If $\tau^k_{t+1} \in \mathcal{F}_t$, then τ^k_{t+1} has the unique solution

$$\tau^k_{t+1} = 1 - \frac{1}{E_t\left[u_c(c_{t+1}(\theta), y_{t+1}(\theta)/\theta_{t+1})\right] E_t[1/u_c(c_{t+1}(\theta), y_{t+1}(\theta)/\theta_{t+1})]}. \tag{59}$$

Remark 4.4. Equation (58) indicates that the optimal labor income tax rate is zero under case 1, which is similar to the static case previously stated. Equation (59) presents the optimal capital income tax rate, which states that many tax policies can be applied to implement the social optimum. This conclusion can be compared with that of Zhu (1992), who studies optimal taxation in an economy with production shocks.

Furthermore, if $\tau^k_{t+1} \in \mathcal{F}_t$, then we derive the same optimal capital income tax rate as Golosov et al. (2003). Similarly, because

$$1 - \frac{1}{E_t\left[u_c(c_{t+1}(\theta), y_{t+1}(\theta)/\theta_{t+1})\right] E_t[1/u_c(c_{t+1}(\theta), y_{t+1}(\theta)/\theta_{t+1})]} > 0,$$

we know that $\tau^k_{t+1} > 0$. However, the utility form used here is non-separable.

Case 2. $U(c(\theta), y(\theta)/\theta') = u(c(\theta)) - v(y(\theta)/\theta)w(\theta, \theta')$ *with* $w(\theta, \theta) = 1$.

This case is similar to that of Golosov et al. (2003), except that the utility is separable. We let $\widetilde{\lambda}(\theta, \theta') = 1 + E_{\theta'}[\lambda(\theta', \theta) - \lambda(\theta, \theta')w(\theta_t, \theta'_t)]$ and define $\lambda(\theta)$ as that of case 1, the optimal conditions are then

$$u'(c_t(\theta)) = \beta R_{t+1} \frac{1}{E_t[1/u'(c_{t+1}(\theta))]}, \tag{60}$$

and

$$v'(y(\theta_t)/\theta_t)/\theta_t = \frac{E_t\left[\lambda(\theta)\right]}{E_t[\widetilde{\lambda}(\theta, \theta_t)]} F_y(k_t, \int y_t d\mu) u'(c_t(\theta)). \tag{61}$$

Comparing equations (45) and (46) with equations (61) and (62) gives

$$\tau^k_{t+1} = 1 - \frac{1}{E_t\left[u'(c_{t+1}(\theta))\right] E_t[1/u'(c_{t+1}(\theta))]}, \tag{62}$$

and

$$\tau^w_t = \frac{E_t\{E_{\theta'}\left[\lambda(\theta, \theta')(1 - w(\theta_t, \theta'_t))\right]\}}{1 + E_t\{E_{\theta'}\left[\lambda(\theta', \theta)\right] - E_{\theta'}\left[\lambda(\theta, \theta')w(\theta_t, \theta'_t)\right]\}}. \tag{63}$$

Remark 4.5. With the same function as that of Golosov et al. (2003), equation (63) states the same optimal capital income tax. However, equation (64) presents the same optimal labor income tax rate as that of Mirrlees (1971).

For these two cases, we derive the same capital income tax rate but different optimal labor income tax rates. Under the non-separable utility function, the disturbance of consumption and labor supply not only affect the resource constraint, but also the IC constraints.

Case 3. $U(c(\theta), y(\theta)/\theta') = u(c(\theta))v(y(\theta)/\theta')$.

In this case, the first-order conditions (45) and (46) for the decentralized economy can be reduced as:

$$u'(c(\theta_t))v(y(\theta_t)/\theta_t) = \beta(1-\tau^k_{t+1})R_{t+1}E_t[u'(c(\theta_{t+1}))v(y(\theta_{t+1})/\theta_{t+1})], \tag{64}$$

and

$$(1-\tau^w_t)F_y(k_t, \int y_t d\mu)u'(c(\theta_t))v(y(\theta_t)/\theta_t) = -\frac{u(c(\theta_t))v'(y(\theta_t)/\theta_t)}{\theta_t}. \tag{65}$$

As for the social planner economy, let

$$\widetilde{\lambda}(\theta, y_t, \theta'_t) = 1 + E_{\theta'}[\lambda(\theta', \theta) - \lambda(\theta, \theta')\frac{v(y_t/\theta'_t)}{v(y_t/\theta_t)}],$$

and

$$\mu(\theta, y_t, \theta'_t) = 1 + E_{\theta'}[\lambda(\theta', \theta) - \lambda(\theta, \theta')\frac{v'(y_t/\theta'_t)\theta_t}{v'(y_t/\theta_t)\theta'_t}].$$

Then, we can rewrite equations (50), (51), and (52) as

$$\lambda_t = \beta\lambda_{t+1}R_{t+1},$$

$$\lambda_t = E_t[\widetilde{\lambda}(\theta, y_t, \theta'_t)]u'(c(\theta_t))v(y(\theta_t)/\theta_t),$$

and

$$-\lambda_t F_y(k_t, \int y_t d\mu) = E_t\left[\mu(\theta, y_t, \theta'_t)\right]u(c(\theta_t))v'(y(\theta_t)/\theta_t)/\theta_t.$$

Therefore, we further have

$$u'(c(\theta_t))v(y(\theta_t)/\theta_t) = \beta R_{t+1}\frac{1}{E_t[1/u'(c(\theta_{t+1}))v(y(\theta_{t+1})/\theta_{t+1})]}, \tag{66}$$

and

$$\frac{E_t[\widetilde{\lambda}(\theta, y_t, \theta_t')]}{E_t\left[\mu(\theta, y_t, \theta_t')\right]} F_y(k_t, \int y_t d\mu) u'(c(\theta_t)) v(y(\theta_t)/\theta_t) = -\frac{u(c(\theta_t)) v'(y(\theta_t)/\theta_t)}{\theta_t}. \tag{67}$$

Equation (67) restates the "Inverse Euler Equation" as that in Golosov et al. (2003) and many existing literatures. Comparing it with (65) and assuming $\tau_{t+1}^k \in \mathcal{F}_t$, then

$$\tau_{t+1}^k = 1 - \frac{1}{E_t\left[u_c(c_{t+1}(\theta), y_{t+1}(\theta)/\theta_{t+1})\right] E_t[1/u_c(c_{t+1}(\theta), y_{t+1}(\theta)/\theta_{t+1})]}. \tag{68}$$

Similarly, comparing equation (68) with equation (66), we have

$$\tau_t^w = \frac{E_t\{E_{\theta'}[\lambda(\theta, \theta')(\frac{v(y_t/\theta_t')}{v(y_t/\theta_t)} - \frac{v'(y_t/\theta_t')\theta_t}{v'(y_t/\theta_t)\theta_t'})]\}}{1 + E_t\{E_{\theta'}\left[\lambda(\theta', \theta)\right] - E_{\theta'}[\lambda(\theta, \theta')\frac{v'(y_t/\theta_t')\theta_t}{v'(y_t/\theta_t)\theta_t'}]\}}. \tag{69}$$

Remark 4.6. Equations (70) gives the similar optimal labor income taxation to Golosov et al. (2010), where he specifies the utility function as

$$U(c, l) = -\frac{1}{\psi} \exp[-\psi(c - \frac{1}{\gamma} l^\gamma)].$$

In fact, we can rewrite this utility function as

$$U(c, l) = -\frac{1}{\psi} \exp(-\psi c) \exp(\psi l^\gamma/\gamma),$$

which is just the same forms as Case 3 here.

Furthermore, we not only give the optimal labor income taxation but also the optimal capital income taxation. Using the dual approach developed by Rogerson (1985), we cannot deal with both the capital and labor income taxations. So in case 3, we have included the case in Golosov et al. (2010) and provide all income taxation by applying the Lagrange multiplier.

5. NUMERICAL RESULTS

In this section, we use the second-order approximation method presented by Schmitt-Grohe and Uribe (2004) to simulate the mentioned static and dynamic Mirrlees models. First, we give the optimal labor income taxation in both the static and dynamic Mirrlees models. We also present the optimal capital income taxation in the dynamic Mirrlees model. Second, we provide comparative static results and give the effects of some important parameters on the optimal taxation rules. Finally, we calculate the tax burdens for different population groups.

To simulate the model, we specify the utility function as

$$u(c,l) = \frac{c^{1-\sigma} - 1}{1-\sigma} - l^{\gamma}, \tag{70}$$

where $\sigma > 0$ and $\gamma > 0$ are positive constants.

In the static Mirrless model, we specify the production function as $H(y) = y^{\alpha}$, where $0 < \alpha < 1$ is a constant. For the dynamic Mirrlees model, we specify the production function as $f(k,l) = k^{\alpha} l^{1-\alpha}$.

We assume that the stochastic process $\{\theta_t\}$ is uniformly distributed on the interval $[2, 22]$. To make the simulation results more convincing, we identify the common parameters in the model as $T = 30$ being the periods for which the economy lasts, the discountor for utility, $\beta = 0.9$, and $\gamma = 2$. To allow comparison, we set the parameters $\sigma = 2$ and $\sigma = 3$, $\alpha = 0.3$, and $\alpha = 0.6$. We set the representative time as $t = 5$, $t = 10$, $t = 15$, and $t = 20$ to test the trends of the policies as times go.

Figure 1 presents the optimal labor income taxation in the static Mirrlees model.

FIG. 1. The Optimal Labor Income Tax in the Static Mirrlees Model

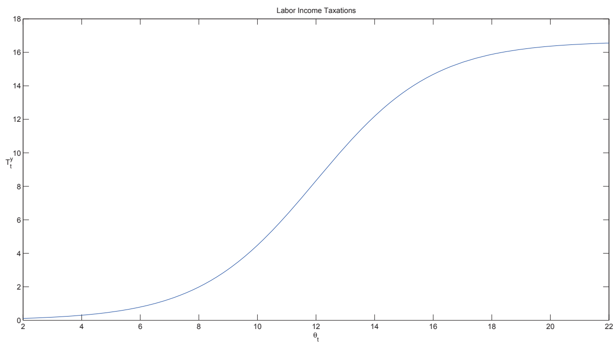


Figure 1 depicts the relationship between labor income taxation and labor ability in the static Mirrlees (1971) model; the corresponding parameters are set as $\sigma = 2$, $\gamma = 0.5$, and $\alpha = 0.7$. The logistic curve states that the optimal taxation rules are to give agents incentives in two ways. The marginal tax of a person with labor ability $\theta \geq 16$ will decrease as their labor ability increases, which will give the person the incentive to provide more effective labor. The marginal tax of the rest of people will increase as their labor ability increases. Agents with medium labor ability pay the highest marginal tax.

In the case of the dynamic Mirrlees model, we suppose that the government's transfer policy is set to equalize each person in society. The associated capital income tax $T_t^k = \tau_t^k r_t \overline{k}$, is indeed a function of τ_t^k. Figures 2 and 3 state the optimal capital and labor income tax rates via the agent's labor ability, respectively.

FIG. 2. The Optimal Capital Income Tax in the Dynamic Mirrlees Model

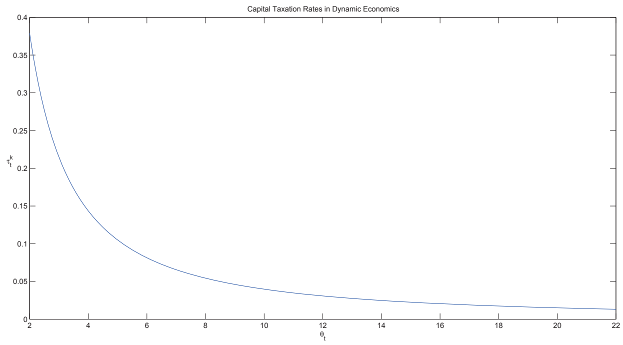
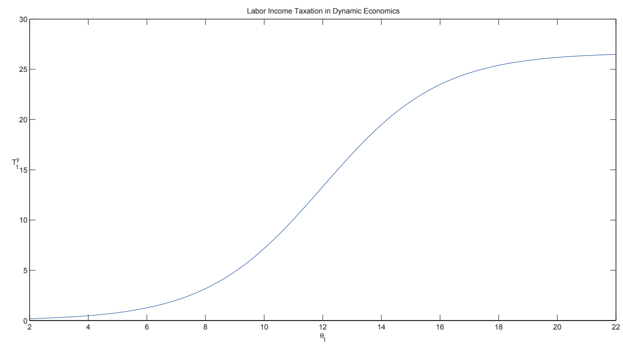


FIG. 3. The Optimal Labor Income Tax in the Dynamic Mirrlees Model



Similar to the previous discussion, both the optimal labor income tax and capital income tax are positive. Furthermore, Figure 2 shows that the optimal capital income tax rate is a decreasing function of the agent's labor ability. Similar to Figure 1 in the static Mirrlees model, Figure 3 also states a logistic relationship between the optimal labor income tax rate and the agent's labor ability.

The reason for the negative relationship between the government's capital income tax and the agent's labor ability is that the government should give agents with higher labor ability the incentive to accumulate more. The logistic relationship between the optimal labor income tax rate and the agent's labor ability can be explained in line with Mirrlees (1971) as that, the government should give agents with higher labor ability the incentive to work more, because the marginal labor income tax decreases with the agents' labor ability.

5.1. Comparative Static Results

In this subsection, we report the effects of certain parameters on the optimal taxations. For simplicity, we consider two parameters, relative risk aversion σ and capital/output ratio α, and inspect their effects on the optimal taxation rules. We set $\sigma = 2$, $\sigma = 3$, and $\alpha = 0.3$, $\alpha = 0.6$ for the experiment.

5.1.1. *Optimal capital income taxation via σ and α*

Figures 4 and 5 present the effects of relative risk aversion σ and capital/output ratio α on the optimal capital income taxation.

FIG. 4. The Effects of σ to the Capital Income Tax

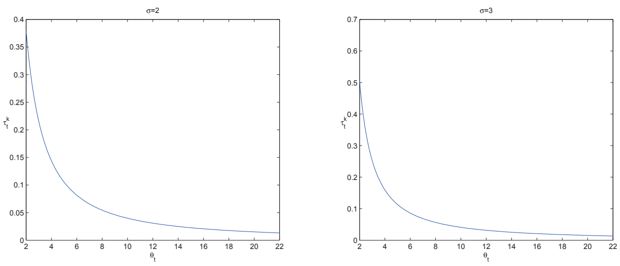
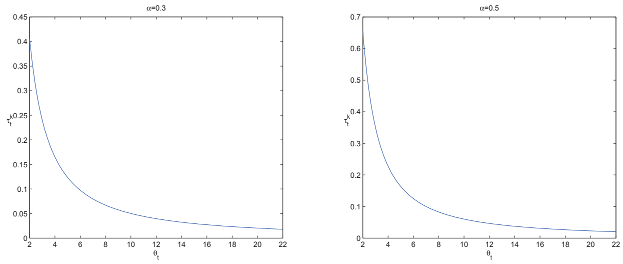


Figure 4 depicts that the curves of the labor ability and capital income tax become convex to the origin as the coefficient of the relative risk aversion σ increases. Because we set the time $t = 10$, the two subfigures are decreasing curves. Note that the curve with $\sigma = 3$ lies above that with $\sigma = 2$, which shows that as the relative risk aversion increases, the tax burden will increase. The effects of capital/output ratio α on the optimal capital income tax is similar to the effects of σ, which are depicted in Figure 5.

FIG. 5. The Effects of α to the Capital Income Tax



5.1.2. Optimal labor income taxation via σ and α

The following figures consider the effects of parameters σ and α on the optimal labor income taxation.

FIG. 6. The Effects of σ to the Labor Income Tax

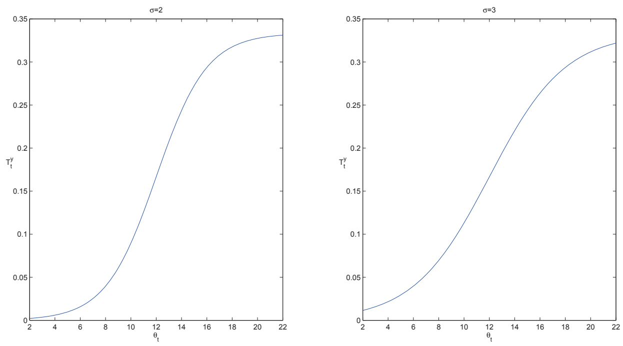


FIG. 7. The Effects of α to the Labor Income Tax

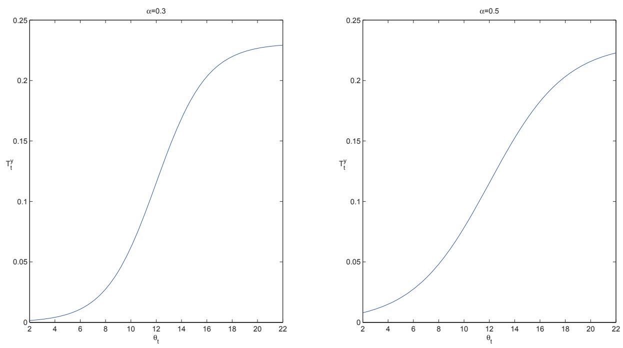


Figure 6 illustrates the effects of the coefficient of relative risk aversion σ. The logistic curve become flatter as σ increases, and the tax increases at the same time. With σ increasing, the agents care more about their consumption and thus the decreasing marginal labor income tax for the agent with medium labor ability raises that actor's confidence about consumption. Hence, the curve becomes more flat. With the same reasoning, we also report the effects of capital/output ratio α in Figure 7. As α increases, the labor income decreases relatively, which in turn decreases the marginal labor income tax for the agent with medium labor ability and the logistic curve again appears to be flatter.

5.2. The Distribution of the Tax Burden

In this subsection, we investigate the tax burden for different groups of the population. Letting $t = 10$ and π_t be the ratio of the tax burden for the agent with labor ability θ_t that lies in $[12 - \gamma, 12 + \gamma]$, the relationship between γ and π_t is summarized in the Table 1.

TABLE 1.

The relationship between γ and π_t.

γ	1	2	3	4	5	6	7
π_t	21.9%	40.2%	62.3%	86.4%	90.2%	93.1%	94.3%

From Table 1, we know that agents with labor ability $\theta_t \in [11, 13]$ contribute 21.9% of the tax in the economy. Agents with labor ability $\theta_t \in [8, 16]$ contribute over 85% of the tax in the economy. It is obvious that the tax burden is increasing with the increasing amount of people. However, after the agents with $\theta_t \in [8, 16]$, the magnitude of tax burden is increasing slow. For agents with labor ability from the interval $[11, 13]$ to $[8, 16]$, the tax burden increases from 21.9% to 86.4%. However, for agents with labor ability from $[8, 16]$ to $[5, 19]$, the tax burden only increases from 86% to 94%. This is consistent with the logistic relationship of the labor income tax rate and the agent's labor ability.

Letting $\gamma = 5$, Tables 2 and 3 present the relationship between the tax burden and other parameters.

TABLE 2.

The tax burden for the agents with $\theta_t \in [7, 17]$.

σ	1.5	2	2.5	3	4	5	6	7
π_t	94.3%	90.2%	89.3%	87.3%	86.9%	86.0%	85.4%	83.0%

From Table 2, we know that an increase in σ, the tax burden for a population group decreases. This conclusion is consistent with the above comparative static conclusion, which states that both the capital and labor income tax decrease as σ increases.

TABLE 3.

The tax burden for the agents with $\theta_t \in [7, 17]$.

α	0.3	0.35	0.4	0.45	0.5	0.55	0.6
π_t	90.2%	90.4%	91.3%	91.8%	93%	93.8%	96.0%

Table 3 presents the effects of capital/output ratio α on the tax burden for the agents with labor ability $\theta_t \in [7, 17]$. As the capital income share increases, the tax burden for this group also increases.

6. CONCLUSIONS

In this paper, we use the Lagrange multiplier method to find the first-order conditions for a class of general dynamic optimization problems with incentive compatible (IC) constraints. Using the method developed, we re-examine the optimal income tax of the Mirrlees (1971) economy and the Golosov et al. (2003) economy for a more general economy. First, we consider both capital income tax and the labor income tax rate. The optimal capital income tax is the same as that derived by Golosov et al. (2003), and the optimal labor income tax under the static and dynamic frameworks are similar. Second, we consider an economy with a non-separable utility function. We find that the optimal labor income tax rate under both the static and dynamic frameworks is zero. Furthermore, the optimal income tax rate may not be unique, as there is a large class of policies that can be used to implement the social optimum. This is similar to the conclusion derived in Zhu's (1992) model with production shocks.

The method developed here is more general, it can be used to deal with many interesting problems in economics that involve information asymmetry. For example, we could consider a more complex information structure, or extend the method to an economy with multi-level governments. In the latter case, there would be three kinds private information: that from agents via both local and fiscal governments, and that between governments. Similarly, there would be three kinds of IC constraints. This is an interesting topic for further study. We could also discuss an economy with evolving private information. However, applying the method to the situation in which the assumption of homogenous agents in this economy is extended to heterogeneous agents, which has been accomplished by Costa and Werning (2001), would prove rather difficult.

Further research can also focus on the implementation mechanism of the optimal taxation rules. Golosov et al. (2006) and Kocherlakota (2005) point out that the government issues the optimal tax policies; however, there still exist the cheating behavior. Even if they have provided the implementation mechanism, the resulting implementation procedure is too complicated to be applied directly. In fact, we can follow Kocherlakota (2005) to investigate the labor ability recognition scheme, which is dependent only on the current observable variables-the consumption and the income. According to the optimal taxation rules, we can finalize the concise implementation mechanism. Also, Pavan (2009)'s implementation mechanism for the principal-agent economy can be adopted for the further research.

APPENDIX: PROOF OF PROPOSITION 3

Proof. Because (x^*, u^*) is a solution to problem (**P1**), we have $x^*_{t+1} = g(x^*_t, u^*_t, \varepsilon_{t+1}, t)$, and then

$$\mathcal{L}(x^*, u^*, z) = \mathcal{L}(x^*, u^*, z^*) = E[\sum_{t=1}^{T} \beta^{t-1} f(x^*_t, u^*_t)]. \tag{A.1}$$

Thus, the first inequality of equation (10) is proven.

Denote $M = \Pi_{t=1}^T M_t \times \Pi_{t=1}^T M_t$ and $N = \Pi_{t=2}^T M_t$, and define

$$\Lambda_1 : \quad \Lambda_1(x, u) = E[\sum_{t=1}^{T} \beta^{t-1} f(x_t, u_t)], \tag{A.2}$$

$$\Lambda_2 : \quad \Lambda_2(x, u)_{t+1} = g(x_t, u_t, \varepsilon_{t+1}, t) - x_{t+1}, \quad t = 1, \ldots, T-1. \tag{A.3}$$

As f and g are continuous and differentiable functions, $\Lambda_1 : M \to \mathbb{R}$ and $\Lambda_2 : M \to N$ are differentiable functionals.

For any $z \in N$, define

$$\mathcal{L}(x, u, z) = \Lambda_1(x, u) + \langle z,\ \Lambda_2(x, u)\rangle. \tag{A.4}$$

Because f and g are concave, Λ_1 and Λ_2 are concave with respect to (x, u).

Thus, $\delta_{(x^*,u^*)}\Lambda_1$ is linear functional from M to $\mathbb{R}$. From the Rieze Representation Theorem, we know that there exists $\eta^* \in M$ such that for any $\xi \in M$,

$$\delta_{(x^*,u^*)}\Lambda_1(\xi) = \langle \eta^*, \xi\rangle. \tag{A.5}$$

In contrast, $\delta_{(x^*,u^*)}\Lambda_2$ is a linear map from M to N. Let A be the accompanying element, which means that it is a linear map from N to M that satisfies for $\xi \in M$ and $\zeta \in N$,

$$\langle \zeta, \delta_{(x^*,u^*)}\Lambda_2(\xi)\rangle = \langle A(\zeta), \xi\rangle. \tag{A.6}$$

We can thus choose $z^* \in N$, such that $A(-z^*) = \eta^*$, which gives

$$\mathcal{L}(x, u, z^*) = \Lambda_1(x, u) + \langle z^*, \Lambda_2(x, u)\rangle. \tag{A.7}$$

Because (x, u) is an element of the vector space M, we abbreviate it as $m = (x, u)$. The concavity of Λ_1 implies that

$$\Lambda_1(m^*) - \Lambda_1(m) \geq \delta_{m^*}\Lambda_1(m^* - m) = \langle \eta^*, m^* - m\rangle. \tag{A.8}$$

Substituting $A(-z^*) = \eta^*$ into equation (A.8) gives

$$\langle \eta^*, m^* - m\rangle = \langle A(-z^*), m^* - m\rangle = \langle -z^*, \delta_{m^*}\Lambda_2(m^* - m)\rangle, \tag{A.9}$$

for any $z^* \in N$.

As the functional $\langle z^*, \Lambda_2(m)\rangle$ is concave, from proposition 2 we obtain

$$\langle z^*, \Lambda_2(m^*)\rangle - \langle z^*, \Lambda_2(m)\rangle \geq \langle z^*, \delta_{m^*}\Lambda_2(m^* - m)\rangle. \tag{A.10}$$

Because $\Lambda_2(m^*) = 0$, we rewrite (A.10) as

$$\langle z^*, \Lambda_2(m)\rangle \leq \langle -z^*, \delta_{m^*}\Lambda_2(m^* - m)\rangle. \tag{A.11}$$

Thus, we arrive at

$$\Lambda_1(m^*) - \Lambda_1(m) \geq \langle z^*, \Lambda_2(m)\rangle, \tag{A.12}$$

which implies that

$$\mathcal{L}(x^*, u^*, z^*) \geq \mathcal{L}(x, u, z^*). \tag{A.13}$$

This completes the proof. ▌

REFERENCES

Atkinson, A. and J. E. Stiglitz, 1976. The Design of Tax Structure: Direct versus Indirect Taxation. *Journal of Public Economics* **6**, 55-75.

Atkinson, A. and J. E. Stiglitz, 1980. *Lectures on Public Economics* (New York, NY: McGraw-Hill).

Billingsley, P., 1995 *Probability and Measure* (New York: John Wiley and Sons).

Chamley, C. 1986. Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. *Econometrica* **54**, 607-622.

Da Costa, C. and I. Werning, 2001. On the Optimality of the Friedman Rule with Heterogeneous Agents and Non-Linear Income Taxation. University of Chicago manuscript.

Diamond, P. and J. A. Mirrlees, 1986. Payroll-Tax Financed Social Insurance with Variable Retirement. *Scandinavian Journal of Economics* **88**, 25-50.

Farhi, E. and I. Werning, 2010. Insurance and Taxation over the Life Cycle. Working paper.

Golosov, M. and O. Tsyvinski, 2006. Designing Optimal Disability Insurance. *Journal of Political Economy* **114**, 257-279.

Golosov, M., N. Kocherlakota, and O. Tsyvinski, 2003. Optimal Indirect and Capital Taxation. *Review of Economic Studies* **70**, 569-587.

Golosov, M., M. Troshkin, and O. Tsyvinski, 2010. Optimal Dynamic Taxes. Working paper.

Garrett, D. and A. Pavan, 2009. Dynamic Managerial Compensation: a Mechanism Design Approach. Working paper.

Judd, K., 1985. Redistributive Taxation in a Simple Perfect Foresight Model. *Journal of Public Economics* **28**, 59-83.

Kocherlakota, N., 1998. The Effects of Moral Hazard on Asset Prices When Financial Markets are Complete. *Journal of Monetary Economics* **41**, 39-56.

Kocherlakota, N., 2005. Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation. *Econometrica* **73**, 1587-1621.

Laffont, J.-J. and J. Tirole, 1994. *A Theory of Incentives in Procurement and Regulation* (Cambridge, MA: MIT Press).

Luenberger, D., 1969. *Optimization by Vector Space Methods* (New York: John Wiley and Sons).

Mirrlees, J., 1971. An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies* **38**, 175-208.

Mirrlees, J., 1976. Optimal Tax Theory: A Synthesis. *Journal of Public Economics* **6**, 327-58.

Phelan, C. and R. Townsend, 1991. Comparing Multi-Period Information Constrained Optima. *Review of Economic Studies* **58**, 853-881.

Prescott, E. C. and R. Townsend, 1984. Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard. *Econometrica* **52**, 21-45.

Pavan, A., I. Segal, and J. Tokka, 2009. Dynamic Mechanism Design: Incentive Compatibility, Profit–Maximization and Information Disclosure. Working paper.

Roberts, K., 1984. The Theoretical Limits of Redistribution. *Review of Economic Studies* **51**, 177-95.

Rogerson, W., 1985a. Repeated Moral Hazard. *Econometrica* **53**, 69-76.

Rogerson, W., 1985b. The First-Order Approach to Principal-Agent Problems. *Econometrica* **53**, 1357-67.

Schmitt-Grohe S. and M. Uribe, 2004. Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function. *Journal of Economic Dynamics and Control* **28**, 755-775.

Stokey, N. L., 2009. *The Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton University Press. Princetion and Oxford.

Stokey, N. L., R. E. Lucas, and E. C. Prescott, 1989. *Recursive Methods in Economic Dynamics*. Harvard University Press, Cambridge MA.

Thomas, J. and T. Worrall, 1990. Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem. *Journal of Economic Theory* **51**, 367-90.

Uhlig, H., 1996. A Law of Large Numbers for Large Economies. *Economic Theory* **8**, 41-50.

Werning, I., 2001. Optimal Dynamic Taxation. University of Chicago manuscript.

Zhu, Xiaodong, 1992. Optimal Fiscal Policy in a Stochastic Growth Model. *Journal of Economic Theory* **58(2)**, 250-289.