Complexity vs Progressivity*

Jaden Y. Chen† Maxim Troshkin‡

August 18, 2018

Abstract

We study optimal income taxation of individuals who lack precise knowledge of the tax schedule they face: Given an income level, an individual is ambiguous about their tax liability, and is averse to this ambiguity; more complex tax schedules lead to more ambiguity in a sense we make precise. We show that sufficiently imprecise knowledge of tax schedules by taxpayers implies the optimality of progressive taxation throughout the income distribution. This is in contrast with the standard optimality of U-shaped marginal taxes that imply regressivity in the lower half of the income distribution. We derive an optimal tax formula accounting for the imprecise knowledge of the tax, and show that the relationship to the standard optimal tax is additively separable, i.e., the ambiguity about tax liabilities manifests as an implicit additional tax. An implication is that the government that accounts for the ambiguity when designing taxation will have a motivation to reduce complexity.

1 Introduction

Standard approaches to optimal income taxation are built around the presumption that taxpayers have precise knowledge of the tax schedule they face, regardless of how non-linear or otherwise complex this schedule is. In other words, the taxpayers are assumed to never make mistakes when it comes to their tax liabilities. Beyond bracketing the optimal marginal tax rates between zero and 100 percent, truly general qualitative conclusions about the optimal tax properties are limited (see Mirrlees 1971). However, given that empirical income distributions are typically single peaked and feature thick right tails, an insight that does appear general is that the optimal marginal tax is U-shaped as a function of individual

*We are indebted to Todd Lensman for comments, suggestions, and overall indispensable research assistance. This is a preliminary draft.

†Cornell University, yc2325@cornell.edu

‡Cornell University, troshkin@cornell.edu.

income. That is, the optimal tax schedule is regressive for lower income taxpayers, with the lowest marginal rates around the median of the income distribution, and the schedule is progressive in the top part of the income distribution (see Diamond 1998, Saez 2001).[1]

At the same time, a growing body of evidence from empirical public finance, as well as from quasi-experimental and experimental behavioral literatures, suggests that taxpayers are typically unaware of their exact marginal tax liabilities. This lack of precise knowledge of the tax schedule they face extends to holding beliefs inconsistent with a probabilistic view of tax liabilities and effectively leaving money on the table.[2] This evidence also suggests that lower income taxpayers are more likely to make mistakes and all taxpayers are more likely to make mistakes when tax policy is more complicated.

In this paper we move away from the presumption of exact understanding of tax liabilities by taxpayers, and study optimal income taxation of individuals who lack exact or probabilistically sophisticated knowledge of the tax schedule they face. The environment we consider is the Mirrlees model, where agents are heterogeneous in their skills (productivity levels), which are assumed to be private information unobservable by the government. Unlike in the standard environment, we assume that agents hold ambiguous beliefs about tax schedules and are averse to this ambiguity. We model this aversion to ambiguity about tax liabilities following the maxmin expected utility representation (Gilboa and Schemeidler 1989). That is, the agents would like to make choices that deliver acceptable utility with all tax liabilities they believe are possible, which requires making choices that deliver the maximum utility with the worst tax liability. Each agent forms a set of beliefs, i.e., a collection of distributions about their tax liability at each income level. The belief sets of higher-skilled agents are smaller to reflect less ambiguity about their tax liabilities. All agents' belief sets depend on the tax function that the government chooses: The more complex a tax function is, the larger are the agents' belief sets to reflect ambiguity that increases with complexity of the tax.

We first derive an optimal tax formula accounting for this form of imprecise knowledge of the tax. The formula extends the Diamond-Saez formulas to an environment where individuals are ambiguous about their tax liability and are averse to this ambiguity. We use the formula to show that the relationship between the standard optimal income tax and the optimal income tax accounting for ambiguity is additively separable. This implies that the ambiguity manifests itself as an implicit additional tax. An immediate interpretation

[1]See also Golosov, Troshkin, and Tsyvinski (2011) for a close connection between a static Mirrlees model with two goods and a recursive formulation of dynamic optimal taxation, suggesting that dynamic labor wedges similarly feature both regressive and progressive components.

[2]See, for example, Aghion, Akcigit, Lequien, and Stantcheva 2017, Rees-Jones and Taubinsky 2017, and references therein.

of this observation is that the government that is aware of the ambiguity when designing income taxes will have a motivation to simplify the tax schedule. For instance, if Mirrleesian tax happens to be "simple", meaning that the agents are not ambiguous about it, then the optimal income tax is just the Mirrleesian optimal tax. Otherwise, the optimal income tax is "simpler" than the Mirrleesian tax, but not perfectly simple. In other words, the complexity of optimal income tax falls between simple and the complexity level of the Mirrleesian tax.

We then show that when there is enough ambiguity, this motivation for simplicity results in taxes that are progressive throughout the income distribution. The mechanism behind this result is intuitive. To account for the implicit tax from the ambiguity, the government finds it optimal to lower taxes at all income levels, but more so at lower incomes where marginal welfare is higher. This stands in contrast with standard insights: One implication of the standard U-shaped optimal marginal taxes is that they are regressive for a large part of the income distribution, typically up to the mean of the distribution. We illustrate that this contrast is independent of the exact notion of tax simplicity by showing that this is the case when all tax schedules result in the same exogenous amount of ambiguity.

The rest of this section reviews related literature. Section 2 describes the general environment. In section 3, we characterize the optimal complexity level and derive the optimal income tax formulas. Section 4 considers the case of no wealth effects and exogenous ambiguity. In section 5, we discuss the extension where agents' belief sets depends on agents' type.

Related literature. This paper contributes most directly to the optimal taxation literature cited above. Besides the traditional interest in the considerations of non-linearity and progressivity of fiscal policies, the literature has more recently emphasized the limitations potentially imposed by the extent to which common knowledge is assumed. One example that moves away from the standard assumptions of common knowledge is Kocherlakota and Phelan (2009). They consider an endowment shock economy in which the government, but not the agents, is uncertain about the data-generating process. They derive conditions under which the government cannot improve on the competitive equilibrium allocation. That is, they assume ambiguity on the part of the mechanism designer while we focus on the case where agents are ambiguous about the mechanisms. Kocherlakota and Song (2018) examine mechanisms for the provision of a public good in economies where agents are ambiguous about the distribution of private valuations. Similarly to our results, they find that ambiguity can lead to less complex efficient policies.

Our approach is related to the emerging literature on the design of ambiguous mechanisms, e.g., Di Tillio, Kos, and Messner (2017). They show that introducing ambiguity into the agents' beliefs can improve the performance of a mechanism. A key difference between the way we model the government's problem and their focus is that in their environment mechanism designer is assumed to be able to directly manipulate the agents' belief sets. In a sense, we focus on the opposite case where the government that cannot directly influence how taxpayers' belief sets are formed given a tax function, instead the government can only design a more or less complex tax policy.

Similarly related is the broader mechanism design literature under ambiguity, where agents' beliefs are typically assumed to have at least some independence from the mechanism (see, e.g., Bodoh-Creed 2012, Wolitzky 2016). Bodoh-Creed (2012) considers a bilateral trade problem where agents are ambiguous about other agents' types. He shows that ambiguity can decrease budget deficit thus improving the efficiency of the mechanism. Wolitzky (2016) further shows that under some conditions with maxmin agents efficient trade is implementable.

2 Setup

The economy is static and consists of a continuum of agents with heterogeneous *types* (productivities) $\theta \in \left[\underline{\theta}, \overline{\theta}\right] \subset \mathbb{R}_{++}$. Nature draws a type θ for each agent according to the distribution F, and agents exert *labor* $l \in [0, L]$, $L < \infty$, to earn *income* y in proportion to their types: $y = \theta l$. Income is publicly observable, but agents are privately informed about their types and their labor supply. The agents use their incomes to purchase a single consumption good, and their preferences over *consumption* c and labor l are additively separable:

$$U(c, l) = u(c) - v(l),$$

where u and v are twice continuously differentiable with $u', v' > 0$ and $u'', -v'' \leq 0$. It will also be convenient to assume that u is bounded below, with the normalization $u(0) = 0$. The agents face an *income tax* $T : \mathbb{R}_+ \to \mathbb{R}$, where $T(y)$ gives the tax liability of any agent with income y. Given a tax T, an agent with type θ who exerts labor l faces the budget constraint

$$c \leq \theta l - T(\theta l).$$

The tax T must satisfy $T(y) \leq y$ so that consumption is non-negative.

There are no constraints on the functional form of the tax policy other than the requirement that it depend only on income, so T may be arbitrarily nonlinear and complex. However, the agents have a limited ability to understand complex tax policies, and they may not be able to perfectly compute tax liabilities at each income level. Instead, for each income level y, the agents form a set $\Pi(y,T)$ of probability measures $\pi \in \Delta((-\infty, y])$ that represent their *beliefs* about the tax liability $T(y)$, where $\Delta((-\infty, y])$ denotes the set of Borel measures on $(-\infty, y]$. To relate these belief sets (and the agents' ambiguity about tax liabilities) to the policy T, we associate to each tax policy T a *complexity level* or *ambiguity level* $K(T) \geq 0$ that measures the agents' difficulties with understanding the policy. Our results below do not require a specific functional form for K, but it can be thought of as a measure of the nonlinearity of policies.[3] A policy is called *simple* when its complexity level is zero, $K(T) = 0$, and this means that agents understand the policy perfectly. We require only that K be non-negative, and for convenience we take K to be continuously Fréchet differentiable in T.

Given the complexity measure K, agents' belief sets are defined by

$$\Pi(y,T) \equiv \left\{\pi \in \Delta((-\infty, y]) \,\middle|\, D\left(\delta_{T(y)} \| \pi\right) \leq K(T)\right\}, \tag{1}$$

where $D(\cdot\|\cdot)$ denotes relative entropy and $\delta_{T(y)}$ is the Dirac measure that places unit mass on $T(y)$.[4] Under this definition $\Pi(y,T)$ is comprised of all distributions π such that the relative entropy of $\delta_{T(y)}$ with respect to π does not exceed the complexity level $K(T)$ of the tax policy. Thus the agents entertain ambiguity about the tax liability for every income level, and the degree of their ambiguity (i.e., the size of the belief set $\Pi(y,T)$) is increasing in the complexity $K(T)$ of the tax policy. Under a simple tax policy with $K(T) = 0$, the only element of $\Pi(y,T)$ is the measure $\delta_{T(y)}$ that places unit mass on $T(y)$, so the agents can perfectly compute tax liabilities.

The agents are averse to their ambiguity about tax liabilities, and we represent their preferences using the maxmin expected utility model axiomatized by Gilboa and Schmeidler (1989). Given a type θ and a tax policy T, an agent's perceived utility from income y is

[3]An example of $K(T)$ is

$$K(T) = \int_0^{\overline{y}} (T(y) - G)^2 dy,$$

where $\bar{y} \equiv \bar{\theta}L$ is the maximum amount of output that can be produced by any agent. In this case, a simple tax T corresponds to a flat tax of G on all agents.

[4]For any two measures $P, Q \in \Delta((-\infty, y])$ such that P is absolutely continuous with respect to Q, the relative entropy $D(P\|Q)$ is defined by

$$D(P\|Q) \equiv \int \log\left(\frac{dP}{dQ}\right) dP,$$

where $\frac{dP}{dQ}$ is the Radon-Nikodym derivative of P with respect to Q. When relative entropy $D(P\|Q)$ is used, it is understood that P is absolutely continuous with respect to Q, so $\frac{dP}{dQ}$ is well-defined.

defined by

$$V(\theta, y, T) \equiv \inf_{\pi \in \Pi(y,T)} \mathbb{E}_\pi \left[u\left(y - \mathscr{T}\right)\right] - v\left(\frac{y}{\theta}\right),$$

where $\mathscr{T} \in (-\infty, y]$ is a random variable that represents the agent's tax liability. In the Appendix, we show that the infimum in $V(\theta, y, T)$ is attained, and we provide an alternative characterization of the agent's perceived utility:

$$V(\theta, y, T) = e^{-K(T)} u\left(y - T(y)\right) - v\left(\frac{y}{\theta}\right).$$

Given a type θ and a tax policy T, the agent chooses an income level $y \geq 0$ to maximize $V(\theta, y, T)$.

The economy has a government that chooses the tax policy T to finance public expenditures $G \geq 0$ and provide a degree of redistribution. The government's redistribution motive is described by a differentiable welfare weight $\alpha : \left[\underline{\theta}, \bar{\theta}\right] \to \mathbb{R}_+$, where we assume $\alpha' \leq 0$ so that the government weakly prefers redistributing to lower types, and we normalize $\int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta)\, dF(\theta) = 1$. Given a tax policy T and the agents' chosen income levels y, the government's social welfare function is defined by

$$W(y, T) \equiv \int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) \left[u\left(y(\theta) - T\left(y(\theta)\right)\right) - v\left(\frac{y}{\theta}\right)\right] dF(\theta).$$

Thus the government evaluates welfare based on agents' true utilities from consumption and labor, weighted by α.

A standard revelation principle holds, so we can assume that the government chooses the tax policy $T(y)$ and income levels $y(\theta)$, and asks agents to report their private types θ. Given the tax T, the income function y, a true type θ, and a reported type θ', the agent's perceived utility is given by

$$V(\theta, y, T)(\theta') \equiv e^{-K(T)} u\left(y(\theta') - T\left(y(\theta')\right)\right) - v\left(\frac{y(\theta')}{\theta}\right).$$

The government must incentivize truthful reporting, so it chooses T and y to maximize $W(y, T)$, subject to a budget constraint

$$\int_{\underline{\theta}}^{\bar{\theta}} T\left(y(\theta)\right) dF(\theta) \geq G, \tag{2}$$

an incentive-compatibility constraint

$$V(\theta, y, T)(\theta) \geq V(\theta, y, T)(\theta') \quad \forall \theta, \theta' \in \left[\underline{\theta}, \bar{\theta}\right], \tag{3}$$

and non-negativity of income $y(\theta)$ and after-tax consumption $y(\theta) - T(y(\theta))$.

Compared with standard optimal taxation literature, there is an additional dimension of trade-off in the government's problem. When the government is designing the tax policy, in addition to considering standard forces, it also considers the complexity level of the tax policy. Intuitively, with a more complex tax function, the government can have higher redistributional power and it can expect to achieve a higher welfare level. On the other hand, agents might get confused by it, which may reduces the effectiveness of the tax policy.

Before characterizing the optimal tax policy T, it is instructive to consider an example in which agents have a fixed ambiguity level $\hat{K}$, where $K(T) = \hat{K} \geq 0$ for all T. Their belief sets become $\Pi(y, T) = \{\pi \in \Delta((-\infty, y])|d(\delta_{T(y)}, \pi) \leq \hat{K}\}$. With such belief set, we then solve for government's problem. The welfare-maximizing tax policy is denoted as $T(\hat{K})$ and the maximum welfare is denoted as $W^{ex}(\hat{K})$. $U^{ex}(\theta, \hat{K})$ denotes the corresponding utility level for agent θ under $T(\hat{K})$. We can see when $\hat{K} = 0$, corresponding to the case where agents are not ambiguous. The optimal taxation in this case $T(0)$ is just Mirrleesian tax.

Lemma 1. *$W^{ex}(\hat{K})$ is weakly decreasing in ambiguity level $\hat{K}$.*

The proof of Lemma 1 is shown in appendix. Intuitively, we can think of ambiguity as an additional "tax" on taxpayers: if ambiguity level is higher, taxpayers' perceived marginal tax rate is also higher. To see this, let $\tau(y) \equiv T'(y)$ denote the marginal tax rate, and note that the government's first order conditions for an optimal tax schedule are

$$u'(c)\left[1 - \left(e^{-\hat{K}}\tau(y) + 1 - e^{-\hat{K}}\right)\right] = \frac{1}{\theta}v'\left(\frac{y}{\theta}\right).$$

We define $\tau^p(y) \equiv e^{-K}\tau(y) + 1 - e^{-K}$, which can be interpreted as the perceived marginal tax rate at income y. Note that $\tau^p(y) > \tau(y)$ as long as $\hat{K} > 0$ and $\tau(y) < 1$. The Mirrleesian marginal tax is bounded above by 1, so for a fixed tax policy T, a higher ambiguity level is going to create additional distortions for each taxpayer (thus inducing lower actual utility), also discouraging people from supplying labor (thus inducing lower tax revenue collected). It's easy to see that the same tax policy T is going to achieve a higher welfare level in the less ambiguous case, thus $W^{ex}(\hat{K})$ decreases with ambiguity level $\hat{K}$.

Lemma 2. *Denote θ_c as the productivity level such that $\omega(\theta_c) = 1$, where $\omega(\theta) = \frac{u'(c(\theta))\alpha(\theta)}{\int_{\underline{\theta}}^{\overline{\theta}} u'(c(\theta))\alpha(\theta)dF(\theta)}$, we have $\frac{\partial U^{ex}(\theta, \hat{K})}{\partial \hat{K}} \geq 0$ for $\theta < \theta_c$ and $\frac{\partial U^{ex}(\theta, \hat{K})}{\partial \hat{K}} \leq 0$ for $\theta > \theta_c$.*

Here $\omega(\theta)$ represents the social value of individual θ's consumption. This result implies that increasing ambiguity level is welfare-enhancing for agent with productivity level below θ_c but it's welfare-detrimental for agents with productivity level above θ_c. In particular, when $\theta < \theta_c$ and $\omega(\theta) > 1$, the social value for the agent's consumption is higher than the social value for government's budget, and the government prefers leaving the money to taxpayers than collecting taxes from them. When $\theta < \theta_c$ and $\omega(\theta) < 1$, government prefers the opposite. When $\hat{K}$ is larger, for a fixed marginal tax rate, the perceived tax rate is higher, so the government tends to impose a lower marginal tax rate. As a result, marginally less revenue is collected from all agents. This implies that social welfare increases for agents $\theta < \theta_c$ but decreases for agents $\theta > \theta_c$.

3 Characterization and Implications for Simplicity

In this section, we characterize the problem of the government that is designing taxes while being aware of the ambiguity that a tax schedule can create among the agents. We derive an optimal tax formula accounting for this form of imprecise knowledge of the tax, and show that the relationship between the standard optimal tax and the tax accounting for ambiguity is additively separable, i.e., the ambiguity manifests itself as an implicit additional tax. An implication is that the government that is aware of the ambiguity when designing income taxes will have a motivation to simplify the tax schedule.

3.1 Optimal income tax formulas

Proposition 1. *Let $\tau^*(\theta)$ be the optimal marginal tax rate that taxpayer θ is facing. Denote $\varepsilon(\theta)$ as agent θ's Frisch elasticity of labor supply $\varepsilon(\theta) = \frac{v'(l(\theta))}{v''(l(\theta))}\frac{1}{l(\theta)}$, we have:*

$$\frac{\tau^*(\theta)}{1-\tau^*(\theta)} = \frac{\tilde{\tau}^M(\theta)}{1-\tilde{\tau}^M(\theta)} - X(\theta) + Y(\theta) \tag{4}$$

where

$$X(\theta) = \frac{e^{K(T^*)}-1}{e^{K(T^*)}}\omega(\theta)$$

$$Y(\theta) = \frac{(1+\frac{1}{\varepsilon})\int_\theta^{\overline{\theta}} \frac{\partial K}{\partial V(x)} dF(x) + \frac{\theta}{v'(l(\theta))}\frac{\partial K}{\partial l(\theta)}}{\theta f(\theta)}\Omega(V,l)$$

where $\Omega(V,l) = \frac{\partial W}{\partial K} < 0$ is the change of social welfare due to the increase in agents' ambiguity level.

The expression for optimal income tax formula is a variation of standard Mirrleesian formula (Diamond (1998), Saez (2001)). $\tilde{\tau}^M(\theta)$ in (4) takes the same form as Mirrleesian tax, where

$$\frac{\tilde{\tau}^M(\theta)}{1-\tilde{\tau}^M(\theta)} = [1+\frac{1}{\varepsilon(\theta)}]\frac{1-F(\theta)}{\theta f(\theta)}\frac{1}{1-F(\theta)}\int_{\theta}^{\overline{\theta}}[1-\omega(x)]dF(x)$$

We can see the optimal taxation formula (4) deviates from Mirrleesian tax because of $X(\theta)$ and $Y(\theta)$ terms. $X(\theta)$ is the effect of "exogenous ambiguity", which represents the modifications of standard model because agents are "exogenously" ambiguous at level $K(T^*)$. $Y(\theta)$ is the effect of "endogenous ambiguity", which gives additional modification of Mirrleesian formula because the ambiguity level is "endogenously" determined by the tax policies. When the government is choosing a tax policy, it is also affecting agents' ambiguity level, which brings changes to the optimal tax formula as well.

Consider first a special case: If $K(T) = 0$ for any T : $\forall\theta \in [\underline{\theta}, \overline{\theta}]$, we have $X(\theta) = \frac{e^0-1}{e^0}\omega(\theta) = 0$ so there is no exogenous ambiguity effect; also, $Y(\theta) = 0$, so there is no endogenous ambiguity effect as well. (6) becomes

$$\frac{\tau^*(\theta)}{1-\tau^*(\theta)} = \frac{\tilde{\tau}^M(\theta)}{1-\tilde{\tau}^M(\theta)} = [1+\frac{1}{\varepsilon(\theta)}]\frac{1-F(\theta)}{\theta f(\theta)}\frac{1}{1-F(\theta)}\int_{\theta}^{\overline{\theta}}[1-\omega(x)]dF(x) \tag{5}$$

which means, in this case, optimal income tax degenerates to Mirrleesian tax.

Exogenous ambiguity effect

Suppose $K(T) = K(T^*) > 0$ for all T, which means the ambiguity level is fixed regardless of tax policies. In this case, there is only "exogenous ambiguity" effect and (4) becomes

$$\frac{\tau^*(\theta)}{1-\tau^*(\theta)} = \frac{\tilde{\tau}^M(\theta)}{1-\tilde{\tau}^M(\theta)} - \underbrace{\frac{e^{K(T^*)}-1}{e^{K(T^*)}}\omega(\theta)}_{X(\theta)} \tag{6}$$

Since $\omega(\theta) > 0$ and $\frac{e^{K(T^*)}-1}{e^{K(T^*)}} > 0$, we have $X(\theta) > 0$. As a result, $\tau^*(\theta) < \tilde{\tau}^M(\theta)$, so the government should impose less marginal tax rate than what Mirrleesian model suggests. This is because extra ambiguity is equivalent to additional tax rate on agents, the government has incentives to cut marginal tax rate to encourage more labor supply.

Note that $\omega(\theta) = \frac{u'(c(\theta))\alpha(\theta)}{\int_{\underline{\theta}}^{\overline{\theta}} u'(c(\theta))\alpha(\theta)dF(\theta)}$ represents the social value of individual θ's consumption. It's easy to verify that $\omega'(\theta) \leq 0$, thus $X(\theta)$ is decreasing in θ: the government has stronger incentives to cut marginal tax rate for agents with lower-productivity level.

Endogenous ambiguity effect

If we also have endogenous ambiguity, optimal tax formula becomes:

$$\frac{\tau^*(\theta)}{1-\tau^*(\theta)} = \frac{\tilde{\tau}^M(\theta)}{1-\tilde{\tau}^M(\theta)} - X(\theta) - \underbrace{\frac{(1+\frac{1}{\varepsilon})\int_\theta^{\bar{\theta}} \frac{\partial K}{\partial V(x)} dF(x) + \frac{\theta}{v'(l(\theta))}\frac{\partial K}{\partial l(\theta)}}{\theta f(\theta)}\Omega(V,l)}_{Y(\theta)}$$

Denote $\tau^*_{ex}(\theta)$ as the optimal marginal tax rate only with exogenous ambiguity, denote $\tau^*_{en}(\theta)$ as the optimal tax rate when there is also endogenous ambiguity. Holding other factors fixed, since $\Omega < 0$, if $\frac{\partial K}{\partial l(\theta)} > 0$, we have $\tau^*_{en}(\theta) > \tau^*_{ex}(\theta)$: there is an incentive to increase marginal tax rate compared with exogenous case. Although the government discourages agent's labor supply ($l(\theta) \downarrow$) with a higher tax rate, the tax policy corresponding to this labor supply is simpler ($\frac{\partial K}{\partial l(\theta)} > 0$) and agents are less ambiguous with government's tax policy. As a result, the government will impose a higher tax rate. Similarly, if $\frac{\partial K}{\partial V(x)} > 0$ for any $x \geq \theta$, there is also an incentive to increase marginal tax rate compared with exogenous case. This is because a higher marginal tax rate for agent θ corresponds to a lower $V(x)$ for $x \geq \theta$. Since $\frac{\partial K}{\partial V(x)} > 0$, the government has incentive to decrease ambiguity level by using a higher marginal tax rate.

Corollary 1. *If a lower marginal tax rate for agent θ implies a simpler tax policy, then we have $\tau^*_{en}(\theta) < \tau^*_{ex}(\theta) < \tilde{\tau}^M(\theta)$.*

For a lower marginal tax rate $\tau(\theta)$, we would expect a higher labor supply $l(\theta)$ and a higher perceived utility level $V(x)$ for $x \geq \theta$. Since a lower marginal tax rate implies a simpler tax policy, in other words, we have $\frac{\partial K}{\partial l(\theta)} < 0$ and $\frac{\partial K}{\partial V(x)} < 0$ for any $x \geq \theta$. Follow similar arguments above, we have $\tau^*_{en}(\theta) < \tau^*_{ex}(\theta) < \tau^M(\theta)$.

3.2 Income tax complexity

Next, we derive the implications of the optimal tax formulas for the complexity of the optimal income tax. In particular, we ask if Mirrleesian tax is no longer optimal, is the optimal taxation simpler than the Mirrleesian tax? We also derive the bounds in the complexity of the optimal income tax accounting for the ambiguity. To provide such bounds, we must assume a condition on the complexity of optimal tax schedules when agents have exogenous ambiguity, i.e., when $K(T) = \hat{K} > 0$ for all tax policies T. Recall that $T(\hat{K})$ represents welfare-maximizing tax policy when the ambiguity level is $\hat{K}$, and $K(T(\hat{K}))$ gives the complexity level of $T(\hat{K})$. We impose the following assumption:

Assumption 1. *$K(T(\hat{K}))$ is a continuous and non-increasing function in $\hat{K}$, $\hat{K} \in \mathbb{R}_+$.*

This assumption says if the government is facing more ambiguous agents (a higher $\hat{K}$), the optimal income tax ($T(\hat{K})$) picked by the government is simpler ($K(T(\hat{K}))$ is smaller). When deciding how complex the tax policy to be, the government is indeed deciding how ambiguous agents become. If agents are more ambiguous ($\hat{K}$ is higher), the government expects to achieve a lower social welfare (since $W^{ex}(\hat{K})$ is weakly decreasing with $\hat{K}$) and the welfare maximizing-policy $T(\hat{K})$ has a lower redistribution power. Because of this effect, the government has incentives to make agents less ambiguous. However, there is another effect. When agents are more ambiguous, the tax policy $T(\hat{K})$ that achieves $W^{ex}(\hat{K})$ is simpler ($K(T(\hat{K}))$ smaller) and $T(\hat{K})$ has a higher effectiveness, which brings a higher actual social welfare. As a result, the government also has incentives to increase agents' ambiguity. With Assumption 1, we can see that there is a trade-off between redistribution and simplicity in government's problem.

We then have the following proposition:

Proposition 2. *(1) If Mirrleesian tax is simple, then the optimal complexity level $K(T^*) = 0$ and optimal tax function T^* is Mirrleesian tax, $T^*(y) = T^M(y)\ \forall y \in \mathbb{R}_+$.*

(2) If Mirrleesian tax is not simple, if Assumption 1 holds and $K(T(\hat{K}))$ is not a constant on $[0, K(T^M)]$, the optimal complexity level $K(T^*) \in (0, K_0] \subset (0, K(T^M))$ where K_0 is the fixed point of $K(T(\hat{K}))$.

Proof. In the appendix □

The first result is not surprising. Because social welfare is decreasing with ambiguity level, if Mirrleesian tax is simple, the government will can just pick Mirrleesian tax without incurring extra ambiguity. As a result, optimal income tax is just Mirrleesian tax.

The second result is built on Assumption 1, which says $K(T(\hat{K}))$ is weakly decreasing with $\hat{K}$. When $\hat{K}$ is greater than the fixed point K_0, we have $\hat{K} > K(T(\hat{K}))$. As a result, even if the government imposes $T(\hat{K})$, it will not cause an ambiguity level as high as $\hat{K}$. It is easy to see that optimal complexity should be smaller than $\hat{K}$. Since such argument applies to any $\hat{K} \geq K_0$, the optimal complexity level should be non-greater than K_0, $K(T^*) \in (0, K_0]$. Due to the fact that $K(T(\hat{K}))$ is non-increasing, we have the fixed point $K_0 \leq K(T^M)$, thus optimal tax should be simpler than Mirrleesian tax.

From Proposition 2,we can see that when Mirrleesian tax is not simple, the government will design a simpler tax than Mirrleesian tax, which will make people less ambiguous. However, it will not design a simple tax policy which moves all the ambiguity away. In the end, there should be some ambiguity in the optimal case, but agents are less ambiguous as what they would be if Mirrleesian tax is imposed.

3.3 Discussion

The results derived above are built on the specific structure of belief sets $\Pi(y, T)$. We next offer some reasons for why such belief structure is adopted and what it implies. We also discuss Lemma 1 and how it related to the results in Di Tillio et al (2017) that ambiguity could increase the performance of a mechanism.

Belief sets

When defining agents' belief set $\Pi(y, T)$, we want the belief set to possess the following three properties:

- **(Monotonicity with Complexity Level)** For any T_1, T_2 and any $y \in \mathbb{R}_+$, such that $T_1(y) = T_2(y)$, if T_2 is *more complex* than T_1, then taxpayers should be *more ambiguous* under T_2 in the sense that $\Pi(y, T_1) \subset \Pi(y, T_2)$;

- **(Non-ambiguity with Simple Tax)** For any T and any $y \in \mathbb{R}_+$, if T is a simple tax then $\Pi(y, T) = \{\delta_{T(y)}\}$.

- **(Correct Belief-holding)** For any T and any $y \in \mathbb{R}_+$, agents have *correct* beliefs in the sense that $\delta_{T(y)} \in \Pi(y, T)$.

The first two properties are quite intuitive. First, when tax function is more complicated, agents will become more ambiguous, thus having a larger belief set. Second, when faced with a simple tax, agents are not ambiguous, then the only possible belief is just true belief $\delta_{T(y)}$. The third property needs a bit more explanation. It says although agents may become ambiguous about tax liability, they don't delete the correct belief (which is $\delta_{T(y)}$) from their belief set, which says they form *ambiguous but correct* belief. Such specification has been adopted in some ambiguity literature. For example, in Bodoh-Creed (2012), the "true" belief is assumed to be in agents' belief set. If it's not in the belief set, he said "we would be able to generate arbitrary revenue rankings by creatively choosing a prior for the bidders ($\Delta = \pi$)". In Di Tillio et al (2017), they have similar correct belief-holding assumption which "rules out the possibility that the principal completely deceives the agent". It's easy to show that the belief structure adopted in our paper $\Pi(y, T) \equiv \left\{\pi \in \Delta((-\infty, y]) \,\middle|\, D\left(\delta_{T(y)} \| \pi\right) \leq K(T)\right\}$ satisfies those three conditions.

Relative entropy as a measure of distance

In the definition of belief set, we define belief distance D to be relative entropy. Relative entropy is measuring how one probability distribution (distribution q) diverges from another distribution (distribution p). Indeed, there are many ways that individuals might adopt when forming their belief set and relative entropy is only one example of them. Here are some reasons explaining why relative entropy is a good choice.

First, some previous literature also picks relative entropy as a measure of belief distance. For example, the idea of using relative entropy as a belief distance is also used by Hansen and Sargent's work (Hansen & Sargent 2001). In this paper, relative entropy is adopted as a measure to capture the discrepancy between two beliefs.

Second, relative entropy has a very good implication regarding agents' perceived utility level. It indicates that an agent's perceived utility could be written as a convex combination of actual utility level and the utility level when he is extremely ambiguous. To see this, let's consider a more general setting. Suppose S_y denotes the set of possible tax liability at income y and $\Pi(y,T) \coloneqq \left\{\pi \in \Delta(S_y) \,\middle|\, D\left(\delta_{T(y)} \,\|\right)\pi \leq K(T)\right\}$. The model setup that we see before is a special case when $S_y = (-\infty, y]$. Let's assume that there is a maximum in S_y, denote $T^{\infty}(y) \coloneqq \max(S_y)$. If the belief distance is defined to be relative entropy, we have the following fact:

Fact 1. *$\forall y \in \mathbb{R}_+$ and $\forall \theta \in [\underline{\theta}, \overline{\theta}]$, if $\Pi(y,T) \coloneqq \left\{\pi \in \Delta(S_y) \,\middle|\, D\left(\delta_{T(y)} \,\|\right)\pi \leq K(T)\right\}$ where D is relative entropy, we have*

$$V(\theta, y, T) = e^{-K(T)} u(y - T(y)) + \left(1 - e^{-K(T)}\right) u(y - T^{\infty}(y)) - v\left(\frac{y}{\theta}\right) \tag{7}$$

where $V(y,\theta) = \min_{\pi \in \Pi(y,T)} \mathbb{E}_{\pi} u(y - \mathscr{T}) - v(\frac{y}{\theta})$ is the perceived utility for agent θ when income level is y.

When agent θ is not ambiguous ($K(T) = 0$), his perceived utility level is just actual utility level $u(y - T(y)) - v(\frac{y}{\theta})$. When agent θ is extremely ambiguous ($K(T) = +\infty$), his perceived utility at income y becomes $u(y - T^{\infty}(y)) - v(\frac{y}{\theta})$. Belief-distance being relative entropy indicates that for any complexity level $K(T)$, agent's perceived utility could be written as a convex of actual utility level ($u(y - T(y)) - v(\frac{y}{\theta})$) and perceived utility with infinite ambiguity level ($u(y - T^{\infty}(y)) - v(\frac{y}{\theta})$) with coefficients being $e^{-K(T)}$ and $1 - e^{-K(T)}$. The more complex tax function is (the higher $K(T)$ is), the higher weight is attached to $u(y - T^{\infty}(y)) - v(\frac{y}{\theta})$ and less weight is attached to actual utility level $u(y - T(y)) - v(\frac{y}{\theta})$. This property is consistent with people's intuition that the more ambiguous agents become, the closer they are to the extremely ambiguous case (the further they are away from the non-ambiguity case). Besides, in expression (7), ambiguity level has a clear interpretation, which corresponds to the weight that agents attach to the extremely ambiguous case.

Welfare effects of ambiguity

In Lemma 1, social welfare $W^{ex}(\hat{K})$ is decreasing with ambiguity level $\hat{K}$. However, some literature on ambiguous mechanism design shows that ambiguity can improve the performance of mechanism. For example, Di Tillio, et al. (2017) showed that sellers can increase their revenue by using an ambiguous mechanism. Indeed, these two results are not contradictory to each other. Here is a concise explanation:

- By manipulating *the way* people get ambiguous, the government can improve social welfare by injecting ambiguity;

- By increasing *the extent* of people's ambiguity level, the government is likely to decrease the social welfare.

The first sentence corresponds to Di Tillio, et al.'s (2017) results that introduction of ambiguity can improve the performance of mechanism. This holds because ambiguity can relax incentive compatibility constraints, so some allocations that achieve higher welfare can be implemented. However, the way that the government increases social welfare is by bringing a new option into agents' mind. For example, the government can increase social welfare by changing $T^{\infty}(y)$, thus changing *the way* people get ambiguous. (Recall that agents' perceived tax liability $T^p(y)$ is traveling along line segment $[T(y), T^{\infty}(y)]$ as they are getting more or less ambiguous.)

The second sentence corresponds to the result in Lemma 1. To understand this result, you can imagine that all taxpayers are endowed with a T^{∞}, such T^{∞} denotes people's perceived tax when they become extremely ambiguous. The sign of $\frac{\partial W^{ex}(\hat{K})}{\partial \hat{K}}$ only tells that as taxpayers' perceived tax function T^p "travels" from actual T to endowed T^{∞} (people get more and more ambiguous), how the social welfare is changing. This doesn't preclude the possibility that increased ambiguity can improve social welfare, since it only captures how social welfare changes with *the extent* of ambiguity along a specific way (agents' utility travels along a line segment with endpoints being $u(y-T(y))-v(\frac{y}{\theta})$ and $u(y-T^{\infty}(y))-v(\frac{y}{\theta})$)

We can see although both are increasing ambiguity, Di Tillio, et al. (2017) investigate how designer can manipulate ambiguity so as to increase mechanism performance; while $\frac{\partial W^{ex}(K)}{\partial K} < 0$ explores the welfare implication when agents are automatically becoming more ambiguous. They are indeed talking about two different things.

4 Implications for Progressivity

To investigate the progressivity of optimal taxes with ambiguity, we consider in this section the case of no wealth effects. We show that with enough ambiguity, optimal taxes are progressive throughout the income distribution. Intuitivel, to account for the implicit tax from the ambiguity, the government finds it optimal to lower taxes at all incomes, but moreso at lower incomes where marginal welfare is higher. To illustrate that this is independent of the exact notion of tax complexity, we show that this is the case when all tax schedules result in the same exogenous amount of ambiguity.

Suppose taxpayers' utility function is $U(c,l) = c - v(l)$, with $v'(l) > 0$, $v''(l) > 0$. We assume that taxpayers have constant Frisch labor supply elasticity of ε, thus $\frac{v''(l)}{v'(l)}l = \frac{1}{\varepsilon}$. Assume that ambiguity level is fixed at some $K \geq 0$.

4.1 Optimal income tax formulas

Proposition 3. *Denote $\tau_1(\theta)$ as the optimal marginal tax rate that taxpayer θ is facing, we have:*

$$\frac{e^K - 1}{e^K}\alpha(\theta) + \frac{\tau_1(\theta)}{1-\tau_1(\theta)} = (1+\frac{1}{\varepsilon})\frac{1-F(\theta)}{\theta f(\theta)}\frac{1}{1-F(\theta)}\int_{\theta}^{\overline{\theta}}[1-\alpha(x)]dF(x) \tag{8}$$

Denote the Mirrleesian marginal tax rate that θ is facing as $\tau^M(\theta)$. Then we have the following relationship between these two marginal tax rates:

$$\frac{\tau_1(\theta)}{1-\tau_1(\theta)} + \frac{e^K - 1}{e^K}\alpha(\theta) = \frac{\tau^M(\theta)}{1-\tau^M(\theta)} \tag{9}$$

Proof. In the appendix. □

This formula is a special case of (4) when $K(T) = K$, $\forall T$. Because agents have quasi-linear utility preference and the government attaches the fixed weight to the same agent, $\frac{\tilde{\tau}^M(\theta)}{1-\tilde{\tau}^M(\theta)} = \frac{\tau^M(\theta)}{1-\tau^M(\theta)}$. From this proposition, we have the following findings.

Optimal tax rate is lower than Mirrleesian

With quasi-linear utility, taxpayer θ's perceived utility becomes $e^{-K}(y - T(y)) - v(\frac{y}{\theta})$. First order condition of taxpayer θ's problem gives:

$$\theta[1-\tau^p(y)] = v'(\frac{y}{\theta}) \tag{10}$$

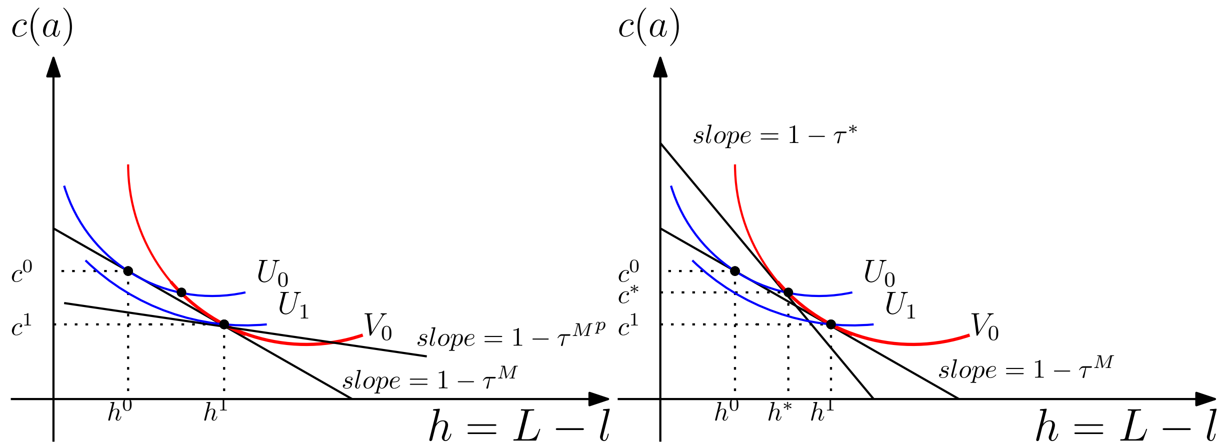

Figure 1: Optimal Taxation with and without ambiguity

Taxpayer θ's actual indifference curve is marked by U, $U = c(\theta) - v(L - h(\theta))$, where $h(\theta) = L - l(\theta)$ is leisure consumption. Taxpayer's perceived indifference curve is marked by V, $V = e^{-K}c(\theta) - v(1 - L(\theta))$.

where $\tau^p(y) = \tau(y) + (1 - e^{-K})(1 - \tau(y))$, which is the perceived marginal tax rate at income y. For $\tau(y) < 1$ and $K > 0$, perceived marginal tax rate $\tau^p(y)$ is always higher than actual marginal tax rate $\tau(y)$. As a result, ambiguity is equivalent to imposing an extra tax rate on taxpayers. In this case, agents are discouraged from supplying labor, thus labor supply is less than the optimal level. In order to encourage labor supply, the government should decrease marginal tax rate.

Figure 1 gives us an intuitive illustration. Left figure shows the non-optimality of Mirrleesian tax. When Mirrleesian tax rate τ^M is imposed, agent θ will act as if the tax rate is τ^{M^p}, where $\tau^{M^p} > \tau^M$. Instead of consuming (c^0, h^0) and getting a utility level of U^0, agent θ will consume (c^1, h^1) and get a lower utility level U^1. The right figure indicates that by adopting a lower tax rate to τ^*, agent reaches a higher utility level.

Marginal tax should be reduced more for low-income agents

We can write (9) as

$$\frac{\overline{\tau_1}(\theta)}{1 - \tau_1(\theta)} = \frac{\tau^M(\theta)}{1 - \tau^M(\theta)} \tag{11}$$

where

$$\overline{\tau_1}(\theta) = (1 - \alpha(\theta))\tau_1(\theta) + \alpha(\theta)\tau_1^p(\theta),$$

which is affine combination of actual marginal tax rate $\tau_1(\theta)$ and perceived marginal tax rate $\tau_1^p(\theta)$.

If θ is large (or if $\alpha(\theta)$ is small), $\overline{\tau_1}(\theta)$ is close to actual tax $\tau_1(\theta)$, so $\tau_1(\theta)$ is also close to

$\tau^M(\theta)$. If we have $\lim_{\theta\to\overline{\theta}} \alpha(\theta) = 0$, then $\tau_1(\theta)$ and $\tau^M(\theta)$ have the same limit. However, for small θ (in other words, large $\alpha(\theta)$), $\overline{\tau_1}(\theta)$ is dragged far away from $\tau_1(\theta)$, since $\tau_1^p(\theta) > \tau_1(\theta)$, $\overline{\tau_1}(\theta)$ is higher than $\tau_1(\theta)$. In order to make up for this effect, the government has incentives to cut $\tau_1(\theta)$.

There are two reasons for that. First, since $\alpha(\theta)$ is decreasing with productivity level, the government attaches a higher weight on low-productivity agents. As a result, it has incentives to cut more tax for low-income agents. Second, loosely speaking, low-income agents face a lower marginal tax rate than high-income agents (although there is a regressive part in Mirrleesian tax). Recall that $\tau^p(y) - \tau(y) = (1 - e^{-K})(1 - \tau(y))$, so the difference between perceived marginal tax rate and actual tax rate is generally higher for low-income agents. As a result, the government has stronger incentive to cut marginal tax rate for low-income agents.

Indeed, there is another reason which would magnify tax-cutting for low-income people. People with lower income appear empirically to be more likely to be mistaken about their tax liability (see, e.g., Aghion et al (2017)). That is, faced with the same tax policy, low-income agents may be more ambiguous than high-income people. Because of this extra-ambiguity, the government tends to further cut marginal tax rate for low-income people. We return to discuss this effect further in the extensions in the next Section, where we allow the belief set depends on productivity level with low-productivity agents having wider belief sets.

4.2 Progressivity

Since ambiguity about tax liability reduces optimal marginal tax rates and the reduction is larger for the low-income individuals, it is easy to see that the standard U-shapes of optimal schedules may be flattened. We show next that if such reduction is strong enough, the marginal tax rate will exhibit progressivity throughout the income distribution.

Definition 1. A tax policy T is *progressive* if $\forall y_1, y_2 \in \mathbb{R}_+$and $y_1 > y_2$ then the marginal tax rate at $y_1$is higher than marginal tax rate at y_2. ($\tau(y_1) > \tau(y_2)$)

The following proposition gives the condition for optimal income tax to be progressive.

Proposition 4. *The optimal income tax T_1 is progressive if* $\forall\theta \in [\underline{\theta}, \overline{\theta}]$

$$\frac{e^K - 1}{e^K}\zeta_{\alpha,\theta} > \frac{1-\alpha(\theta)}{\alpha(\theta)}(1+\frac{1}{\varepsilon}) + \frac{1+\theta\frac{f'(\theta)}{f(\theta)}}{\alpha(\theta)}\frac{\tau^M(\theta)}{1-\tau^M(\theta)} \tag{12}$$

where $\zeta_{\alpha,\theta} = -\frac{d\alpha(\theta)}{d\theta}\frac{\theta}{\alpha(\theta)}$.

Proof. In the appendix □

From the above proposition, we know the following factors are going to influence the progressivity of optimal income tax.

If $\zeta_{\alpha,\theta}$ is larger, which means the decreasing speed of $\alpha(\theta)$ is faster, optimal tax T_1 is more likely to exhibit progressivity. This is because if $\zeta_{\alpha,\theta}$ is large, the convergence speed of τ_1 to τ^M is faster. As a result, the shape ofτ_1 is steeper, thus more likely to exhibit progressivity.

If ambiguity level is higher, it's more likely to have progressive tax. Recall that $\tau^p(y) - \tau(y) = (1 - e^{-K})(1 - \tau(y))$, if K is smaller, the difference between $\tau^p(y)$ and $\tau(y)$ is also smaller, and the government has less incentives to cut marginal tax rate. τ_1 is closer to τ^M, which is not progressive. An extreme case is when $K = 0$, $\tau_1 = \tau^M$.

If Mirrleesian tax rate $\tau^M(\theta)$ is higher, it becomes harder to have progressive tax. To see this, we can also rewrite (9) as the following

$$\frac{\tau_1(\theta)}{1 - \tau_1(\theta)} = \frac{\overline{\tau^M}(\theta)}{1 - \tau^M(\theta)}$$

where $\overline{\tau^M}(\theta) = (1 + \alpha(\theta))\overline{\tau^M}(\theta) - \alpha(\theta)\overline{\tau^M}^p(\theta)$. Since $\tau^{M^p}(\theta) - \tau^M(\theta) = \frac{e^K - 1}{e^K}(1 - \tau^M(\theta))$, if $\tau^M(\theta)$ is larger, the difference between $\tau^{M^p}(\theta)$ and $\tau^M(\theta)$ is smaller. As a result, $\tau_1(\theta)$ is closer to $\tau^M(\theta)$ and is thus more likely to be non-progressive.

4.3 Numerical illustration

We suppose that agents' utility function is quasi-linear

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

and the social welfare function is

$$W = \int_{\underline{\theta}}^{\overline{\theta}} \alpha(\theta, \rho) U(\theta) dF(\theta),$$

where

$$\alpha(\theta, \rho) = \frac{\theta^{-\rho}}{\int \theta^{-\rho} dF(\theta)}$$

Here we assume $\rho > 0$, which means the government attaches higher weights to taxpayers with lower productivity. A higher ρ represents a higher tendency to redistribute from high ability to low ability.

Parameters

Following Heathcote et.al. (2017), we normalize θ to be the relative wage rate (relative to relative wage) and assume that log wage follows exponentially modified Gaussian distribution, $\log(\theta) \sim EMG(\mu_\theta, \sigma_\theta^2, \lambda_\theta)$, in other words, θ follows Pareto log-normal distribution with $(\mu_\theta, \sigma_\theta^2, \lambda_\theta)$.

- $\theta \in [0.12, 74]$, $\underline{\theta} = 0.12$ corresponding to \$5, which is less than federal minimum wage in 2007 (\$5.85); $\overline{\theta} = 74$ corresponds to household labor income at 99.99th percentile of SCF income distribution.
- $v_\theta = 0.466$, which is the variance of log hourly wages for men in 2005;
- $\lambda_\theta = 2.2$, which is picked in their paper in order to match the empirical density of log labor income from SCF.
- $\sigma_\theta^2 = 0.2594$, since $\sigma_\theta^2 = v_\theta - 1/\lambda_\theta^2$
- $\mu_\theta = -0.7358$ to let $\mathbb{E}(\theta) = 1$ (mean wage = average wage)
- $\rho = 0.8$, so social welfare weight for θ is $\alpha(\theta) = \frac{\theta^{-0.8}}{\int \theta^{-0.8} dF(\theta)}$.
- $\sigma = 2$, thus the labor supply elasticity $\varepsilon = 0.5$.

Simulation results

Here we simulate the change of marginal tax rate with different complexity levels (as showed by Figure 2).

In Figure 2, the black dashed line represents the Mirrleesian tax, which is "U-shaped" and the red solid line is the optimal tax rate under ambiguity. As we can see, optimal income tax is progressive when $K = 0.3$. As we are decreasing the value of K, the shape of optimal tax is approaching Mirrleesian tax. When $K = 0.01$, it's already very close to Mirrleesian tax.

When calculating the social welfare, we take government's expenditure $G = 0$. Figure 3 shows how do the social welfare and government's lump-sum transfer change with ambiguity level K_0.

The first graph of Figure 3(the red solid line) shows the relationship between social welfare and ambiguity level K. The second graph (the blue dashed line) shows the relationship between government's transfer and ambiguity level. It also exhibits a negative relation between these two. This pattern could give some explanations on why social welfare decreases

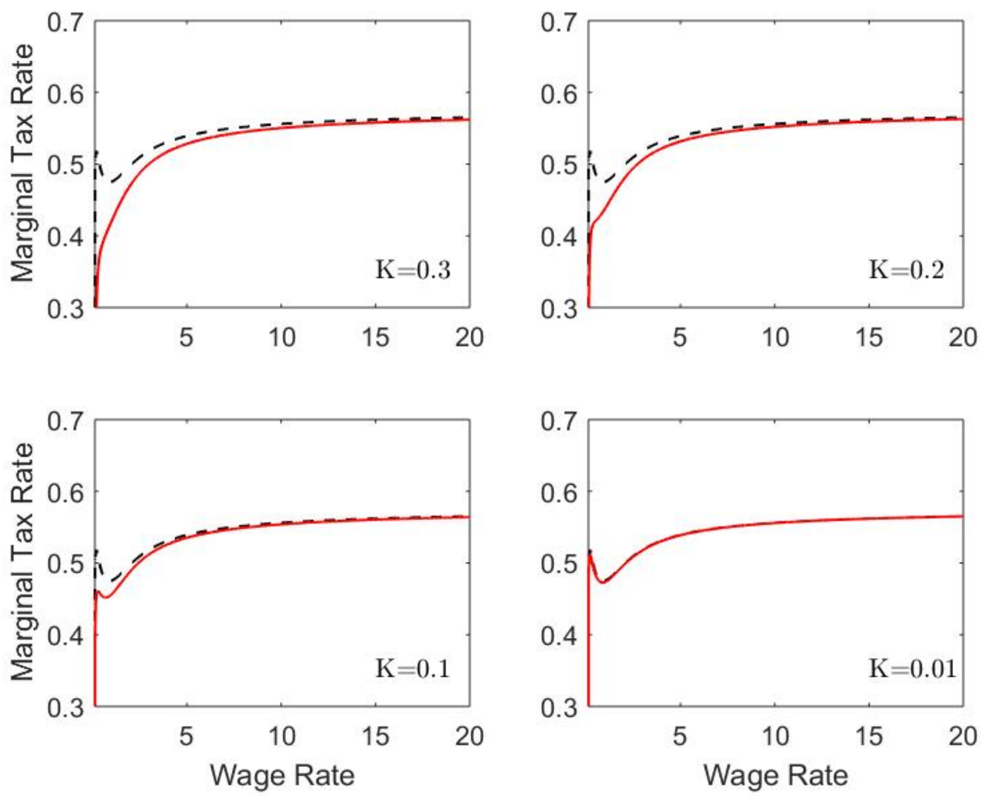

Figure 2: Optimal Marginal Tax Rate under Different Complexity Levels K

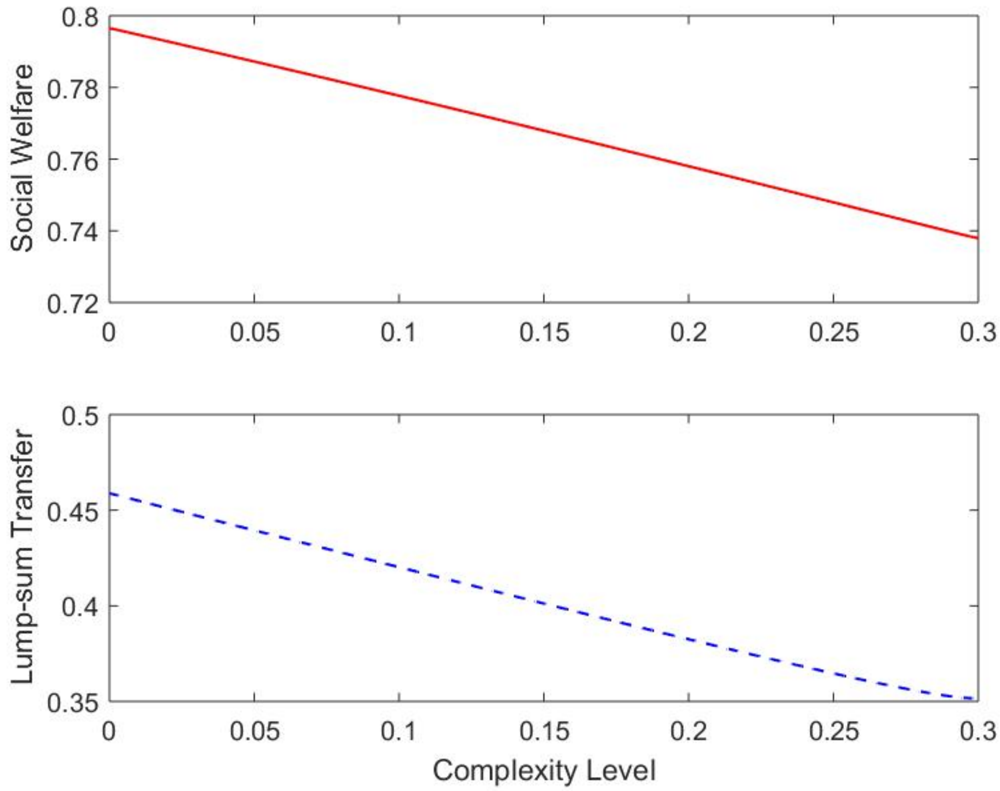

Figure 3: Social Welfare and Government's Transfer under Different Complexity Levels

with ambiguity level. If taxpayers are more ambiguous, the government has incentives to decrease the marginal tax rate for all individuals. In order to collect same amount of tax, the government will impose a higher lump-sum tax (thus a lower lump-sum transfer) on all people. Intuitively, lump-sum tax has weaker distributional power and when the government will rely on more lump-sum tax and less on distributional tax, the social welfare is expected to decrease.

4.4 Discussion: exogenous ambiguity

Above, we focus on exogenous ambiguity. We next discuss why exogenous case is relevant and when the results we get can be extended to the general model.

Recall that optimal income tax formula for a general case is $\frac{\tau^*(\theta)}{1-\tau^*(\theta)} = \frac{\tilde{\tau}^M(\theta)}{1-\tilde{\tau}^M(\theta)} - X(\theta) + Y(\theta)$ and we know that $Y(\theta)$ depends crucially on how we define our complexity function $K(T)$. As a result, the sign and magnitude of this term will differ according to different definitions. Notice that

$$\lim_{\frac{\partial K(T^*)}{\partial T(\theta)} \to 0} Y(\theta) = 0$$

if at the optimum, small perturbations of tax function have very small effect on complexity level, then the endogenous effect is going to be very small.

Furthermore, if the majority part of agents' ambiguity comes from exogenous source, or $X(\theta) \gg Y(\theta)$, we have $\frac{\tau^*(\theta)}{1-\tau^*(\theta)} \approx \frac{\tilde{\tau}^M(\theta)}{1-\tilde{\tau}^M(\theta)} - EX(\theta) = \frac{\tau_1(\theta)}{1-\tau_1(\theta)}$, thus optimal income tax under exogenous case T_1 is a good approximation to optimal income tax in general case T^*.

Also, from Proposition 2, we know that optimal complexity level $K(T^*) \leq K_0$ where K_0 is the fixed point of $K(T(\hat{K}))$. If the government has strong incentives to increase complexity level up to K_0, implied by Proposition 2, the optimal complexity level is just K_0, $K_0 = K(T^*)$. We have the following proposition:

Proposition 5. *If optimal complexity level $K(T^*) = K_0$ where K_0 satisfies $K_0 = K(T(K_0))$, then the optimal income tax $T^* = T(K_0)$, with marginal tax rate satisfying:*

$$\frac{e^{K_0}-1}{e^{K_0}}\alpha(\theta) + \frac{\tau^*(\theta)}{1-\tau^*(\theta)} = (1+\frac{1}{\varepsilon})\frac{1-F(\theta)}{\theta f(\theta)}\frac{1}{1-F(\theta)}\int_\theta^{\overline{\theta}}[1-\alpha(x)]dF(x)$$

The idea of this proposition is not difficult. Since K_0 is the optimal complexity level, we have $K_0 = K(T^*) = K(T(K_0))$. On one hand, because $T(K_0)$ is the optimal tax policy when ambiguity level is exogenously fixed at K_0 and $T^*$is the optimal tax policy with ambiguity level endogenously determined and equal to K_0, so we have $W(T(K_0)) \geq W(T^*)$, where

$W(T)$ represents the social welfare level when income tax is T.[5] On the other hand, since $K(T(K_0)) = K_0$, thus $T(K_0)$ is also in the choice set of government when ambiguity level is endogenously equal to K_0, due to the optimality of T^*, we also have $W(T(K_0)) \leq W(T^*)$. To summarize, $W(T(K_0)) = W(T^*)$, thus $T(K_0)$ is the optimal income tax.

Proposition 5 tells us that if optimal complexity level is the fixed point K_0, then the optimal income tax in general case is just the optimal income tax in exogenous ambiguity case. As a result, by looking at exogenous ambiguity case, we are also looking at some general cases where optimal complexity level is K_0.

5 Extensions

So far we have focused on belief sets $\Pi(y, T)$ that are independent of productivity level θ. It may be reasonable to expect that taxpayers' belief sets also depend on their productivity. We discuss the extension to type-dependent beliefs next.

Definition 2. Denote $\Pi(\theta, y, T)$ as the belief set of taxpayer θ, if his income level is y and the tax function is T.

$$\Pi(\theta, y, T) := \left\{\pi \in \Delta\left((-\infty, y]\right) \middle| D\left(\delta_{T(y)} \| \pi\right) \leq K(T) + \eta(\theta)\right\} \tag{13}$$

where D is relative entropy and $\eta \in \mathbb{R}_+^{[\underline{\theta}, \overline{\theta}]}$ is continuously differentiable satisfying $\eta'(\theta) < 0$, $\eta''(\theta) > 0$.

Compared with our original definition of $\Pi(y, T)$, we have a new term $\eta(\theta)$, which represents heterogeneity of belief set across different type individuals. $\eta'(\theta) < 0$ represents that individuals with higher productivity θ tend to be less ambiguous: if $\theta_1 > \theta_2$, we have $\eta(\theta_1) < \eta(\theta_2)$, thus $\Pi(\theta_1, y, T) \subset \Pi(\theta_2, y, T)$, high-type agent has a smaller belief set.

Suppose taxpayers have quasi-linear utility function, taxpayer θ's problem becomes

$$\max_{y \in \mathbb{R}^+} e^{-\tilde{K}(T,\theta)}[y - T(y)] - v(\frac{y}{\theta}) \tag{14}$$

where $\tilde{K}(T, \theta) = K(T) + \eta(\theta)$. Following similar process as in previous sections, we want to look at optimal income tax in exogenous ambiguity case. Here, "exogenous" means

[5] This is because in exogenous ambiguity case where ambiguity level is K_0, government can pick any tax policy. However, in endogenous ambiguity case with ambiguity level being K_0, the set of tax policies government can choose must satisfy $K(T) = K_0$. We can see there is one more constraint in endogenous case, so the welfare level in endogenous case is weakly less than in exogenous case.

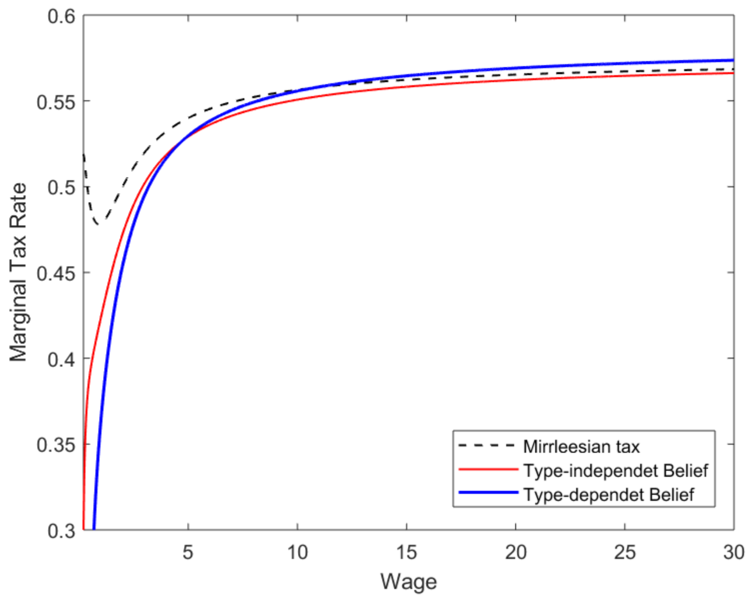

Figure 4: Optimal Income Tax with Type-dependent Belief Set

Type-independent belief is $\Pi(y,T) := \left\{\pi \in \Delta((-\infty,y]) \middle| D\left(\delta_{T(y)} \| \pi\right) \leq K\right\}$ and type-dependent belief is$\Pi(\theta,y,T) := \left\{\pi \in \Delta((-\infty,y]) \middle| D\left(\delta_{T(y)} \| \pi\right) \leq \tilde{K}(\theta)\right\}$, where $\tilde{K}(\theta) = K + \eta(\theta)$ with $K = 0.3$ and $\eta(\theta) = 0.1 \times \log(\frac{\bar{\theta}}{\theta})$. Choice of other key parameters follows section 4.3.

$K(T) \equiv K$ for some non-negative K, but agents' ambiguity level still depends on θ. Denote $\tilde{K}(\theta) = K + \eta(\theta)$, optimal income tax formula is as below

Proposition 6. *If taxpayers' belief set is type-dependent like* (13) *and utility function is quasi-linear. Denote* $\tau_2(\theta)$ *as the optimal marginal tax rate that taxpayer* θ *is facing, we have*

$$\frac{e^{\tilde{K}(\theta)} - 1}{e^{\tilde{K}(\theta)}}\alpha(\theta) + \frac{\tau_2(\theta)}{1 - \tau_2(\theta)} = (1 + \frac{-\theta\eta'(\theta)\varepsilon}{1 + \varepsilon})\frac{\tau^M(\theta)}{1 - \tau^M(\theta)} \tag{15}$$

where $\tilde{K}(\theta) = K + \eta(\theta)$.

Proof. In the Appendix □

From this proposition, we can have the following findings. First, when belief-set is type-dependent, government is more likely to cut marginal tax rate for low-income people. This is because $\tilde{K}(\theta)$ is larger for low-type agents, thus inducing a higher perceived tax rate. Because of this additional force, the government has stronger incentives to cut marginal tax rate for low-type agents compared with the case when belief-set is independent of type. Second, optimal income tax may not have the same limit as Mirrleesian tax. Unlike type-independent belief set case, where τ_1 and τ^M have the same limit as long as $\lim_{\theta \to \bar{\theta}} \alpha(\theta) = 0$. With type-dependent belief, even if $\lim_{\theta \to \bar{\theta}} \alpha(\theta) = 0$, optimal income tax will not converge to Mirrleesian tax as productivity level increases. Indeed, it should have a higher limit marginal tax rate than Mirrleesian tax.

As showed in Figure 4, compared with type-independent case (red solid line), type-independent case exhibits higher progressivity: the government imposes lower marginal tax rates for low-income people but higher tax rates tax rate for high-income earners. Finally, the figure also shows that the limit of the optimal income tax is higher than under the Mirrleesian tax.

6 Conclusions

This paper studied optimal income taxation of individuals who may possess imprecise knowledge of the tax schedule. We derived an optimal tax formula accounting for this form of imprecise knowledge of the tax, and argued that the relationship between the standard optimal tax and the optimal tax accounting for ambiguity is additively separable. This implies that the ambiguity manifests itself as an implicit additional tax. Another implication is that the government that is aware of the ambiguity when designing income taxes will have a motivation to reduce complexity of the tax system. With enough ambiguity this motivation leads to taxes that are progressive throughout the income distribution. We further illustrated that this is independent of the exact notion of tax complexity by showing that this is the case when all tax schedules result in the same exogenous amount of ambiguity.

References

[1] Aghion, P., Akcigit, U., Lequien, M., and Stantcheva, S. (2017). "Tax simplicity and heterogeneous learning", NBER Working Paper, No. 24049, National Bureau of Economic Research

[2] Bodoh-Creed, A. L. (2012). "Ambiguous beliefs and mechanism design," Games & Economic Behavior, 75(2), 518-537.

[3] Diamond, P. A. (1998). "Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates." American Economic Review, 88(1), 83-95.

[4] Golosov, M., Troshkin, M., and Tsyvinski, A., "Optimal Dynamic Taxes," NBER Working Paper No.17642

[5] Di Tillio, A., Kos, N., and Messner, M. (2016). "The Design of Ambiguous Mechanisms." The Review of Economic Studies, 84(1), 237-276.

[6] Hansen, L., and Sargent, T. J. (2001). "Robust control and model uncertainty," American Economic Review, 91(2), 60-66.

[7] Heathcote, J., and Tsujiyama, H. (2015). "Optimal income taxation: Mirrlees meets Ramsey".

[8] Kocherlakota, N., and Phelan, C. (2009). "On the robustness of laissez-faire". Journal of Economic Theory, 144(6), 2372-2387.

[9] Mankiw, N. G., Weinzierl, M., and Yagan, D. (2009). "Optimal taxation in theory and practice", Journal of Economic Perspectives, 23(4), 147-74.

[10] Rees-Jones, Alex, and Dmitry Taubinsky (2017) "Taxing Humans: Pitfalls of the Mechanism Design Approach and Potential Resolutions," NBER Working Paper, No. 23980, October 2017

[11] Saez, E. (2001), "Using Elasticities to Derive Optimal Income Tax Rates", The Review of Economic Studies, 68(1), 205-229.

[12] Wolitzky, A. (2016). "Mechanism design with maxmin agents: Theory and an application to bilateral trade." Theoretical Economics, 11(3), 971-1004.

Appendix

A Derivation of agents' perceived utility

Proposition. *If agents' belief sets are defined by* (1)*, an agent with type* θ *and income* y *has perceived utility* $V(y,\theta,T) = e^{-K(T)}u(y - T(y)) - v(\frac{y}{\theta})$.

Proof. We first provide an alternative characterization of an agent's belief set $\Pi(y,T)$. Given a tax policy T and an income level y, recall that an agent's belief set is defined by

$$\Pi(y,T) = \left\{\pi \in \Delta((-\infty, y]) \middle| D\left(\delta_{T(y)} \| \pi\right) \leq K(T)\right\}.$$

For any measure $\pi \in \Pi(y,T)$, $\delta_{T(y)}$ is absolutely continuous with respect to π. This implies $\pi(\{T(y)\}) > 0$, and it is easy to see that the Radon-Nikodym derivative $\frac{d\delta_{T(y)}}{d\pi}$ is

$$\frac{d\delta_{T(y)}}{d\pi}(x) = \begin{cases} \frac{1}{\pi(\{T(y)\})} & x = T(y), \\ 0 & \text{else.} \end{cases}$$

This implies that the relative entropy of $\delta_{T(y)}$ with respect to π is

$$\begin{aligned} D\left(\delta_{T(y)} \| \pi\right) &= \int \log\left(\frac{d\delta_{T(y)}}{d\pi}\right) d\delta_{T(y)} \\ &= \log\left(\frac{d\delta_{T(y)}}{d\pi}(T(y))\right) \\ &= -\log \pi(\{T(y)\}). \end{aligned}$$

We can then characterize $\Pi(y,T)$ as follows:

$$\begin{aligned} \Pi(y,T) &= \{\pi \in \Delta((-\infty, y]) | -\log \pi(\{T(y)\}) \leq K(T)\} \\ &= \left\{\pi \in \Delta((-\infty, y]) \middle| \pi(\{T(y)\}) \geq e^{-K(T)}\right\}. \end{aligned} \tag{16}$$

Given type θ, income y, and tax policy T, an agent's perceived utility is defined by

$$V(\theta, y, T) = \inf_{\Pi(y,T)} \mathbb{E}_\pi [u(y - \mathscr{T})] - v\left(\frac{y}{\theta}\right).$$

We will show that the infimum is attained and characterize the minimizing measure $\pi^* \in$

$\Pi(y,T)$. Define $\pi^* \in \Delta((-\infty, y])$ such that

$$\pi^*(\{T(y)\}) = e^{-K(T)},$$
$$\pi^*(\{y\}) = 1 - e^{-K(T)}.$$

The condition $K(T) \geq 0$ implies that π^* is a probability measure, and it is the measure that places maximal weight on the agent's income y while satisfying the inequality $\pi^*(\{T(y)\}) \geq e^{-K(T)}$. With the characterization of $\Pi(y,T)$ in (16), we immediately have $\pi^* \in \Pi(y,T)$. To see that π^* attains the infimum in the agent's perceived utility, fix any $\pi \in \Pi(y,T)$. Then

$$\begin{aligned}
\mathbb{E}_\pi [u(y - \mathscr{T})] &= \int u(y - \mathscr{T})\, d\pi \\
&= \pi(\{T(y)\})\, u(y - T(y)) + \int_{\mathcal{T} \neq T(y)} u(y - \mathscr{T})\, d\pi \\
&\geq e^{-K(T)} u(y - T(y)) + \left[1 - e^{-K(T)}\right] u(0) \\
&= \mathbb{E}_{\pi^*} [u(y - \mathscr{T})].
\end{aligned}$$

The third line holds because u is strictly increasing, so we can lower the expected utility by shifting as much mass onto the set $\{\mathscr{T} = y\}$ as possible while maintaining mass $e^{-K(T)}$ on $\{\mathscr{T} = T(y)\}$. This implies that π^* attains the infimum in the agent's perceived utility, so with the normalization $u(0) = 0$, we find

$$\begin{aligned}
V(\theta, y, T) &= \inf_{\Pi(y,T)} \mathbb{E}_\pi [u(y - \mathscr{T})] - v\left(\frac{y}{\theta}\right) \\
&= e^{-K(T)} u(y - T(y)).
\end{aligned}$$

□

Note: In the proofs below, the notations for relative entropy $D(\cdot \| \cdot)$ and the agents' value functions $V(\theta, y, T)$, $V(\theta, y, T)(\theta')$ are not used. In general, $d(\cdot, \cdot) \equiv D(\cdot \| \cdot)$, $V(\theta) \equiv V(\theta, y, T)$, and $V(\theta', \theta) \equiv V(\theta, y, T)(\theta')$.

B Proof of Proposition 3

Proposition. *Denote $\tau_1(\theta)$ as the optimal marginal tax rate that taxpayer θ is facing, we have:*

$$\frac{e^K - 1}{e^K}\alpha(\theta) + \frac{\tau_1(\theta)}{1-\tau_1(\theta)} = (1+\frac{1}{\varepsilon})\frac{1-F(\theta)}{\theta f(\theta)}\frac{1}{1-F(\theta)}\int_{\theta}^{\overline{\theta}}[1-\alpha(x)]dF(x)$$

Denote $\tau^M(\theta)$ as the Mirrleesian marginal tax rate that θ is facing, then we have the following relationship between these two marginal tax rates:

$$\frac{\tau_1(\theta)}{1-\tau_1(\theta)} + \frac{e^K - 1}{e^K}\alpha(\theta) = \frac{\tau^M(\theta)}{1-\tau^M(\theta)}$$

Proof. The government's problem is

$$\max_{\{c(\theta),y(\theta)\}} \int_{\underline{\theta}}^{\overline{\theta}} \alpha(\theta)U(\theta)dF(\theta)$$

$$s.t. \int_{\underline{\theta}}^{\overline{\theta}} c(\theta)dF(\theta) + G \leq \int_{\underline{\theta}}^{\overline{\theta}} y(\theta)dF(\theta) \tag{17}$$

$$e^{-K}[y(\theta) - T(y(\theta))] - v(\frac{y(\theta)}{\theta}) \geq e^{-K}[y(\theta') - T(y(\theta'))] - v(\frac{y(\theta')}{\theta}) \quad \forall\theta,\theta' \tag{18}$$

(17) is government's budget constraint ("BC") and (18) is incentive compatible constraint ("IC") for each individual.

Let $V(\theta',\theta) = e^{-K}c(\theta') - v(\frac{1}{\theta}y(\theta'))$, denoting the perceived utility of taxpayer θ if his report is θ'.

IC implies:

$$\frac{\partial V(\theta,\theta')}{\partial\theta'}|_{\theta'=\theta} = e^{-K}c'(\theta) - \frac{1}{\theta}v'(\frac{1}{\theta}y(\theta))y'(\theta) = 0 \tag{19}$$

$$\frac{\partial^2 V(\theta,\theta')}{\partial\theta'^2}|_{\theta'=\theta} < 0 \tag{20}$$

Let $V(\theta) = e^{-K}c(\theta) - v(\frac{1}{\theta}y(\theta))$, which is the perceived utility of taxpayer θ given he is truth-telling, we have:

$$V'(\theta) = e^{-K}c'(\theta) - \frac{1}{\theta}v'(\frac{1}{\theta}y(\theta))y'(\theta) + \frac{1}{\theta^2}v'(\frac{1}{\theta}y(\theta))y(\theta)$$

thus (19) is equivalent to

$$V'(\theta) = \frac{1}{\theta^2}v'(\frac{1}{\theta}y(\theta))y(\theta) = \frac{1}{\theta}v'(l(\theta))l(\theta)$$

From (19), we know $\frac{\partial^2 V(\theta,\theta')}{\partial\theta'^2} + \frac{\partial^2 V(\theta,\theta')}{\partial\theta\partial\theta'}|_{\theta'=\theta} = 0$, so (20) is equivalent to $\frac{\partial^2 V(\theta,\theta')}{\partial\theta\partial\theta'}|_{\theta'=\theta} > 0$

$$\frac{\partial^2 V(\theta,\theta')}{\partial\theta\partial\theta'}|_{\theta'=\theta} = y'(\theta)\frac{v'(l(\theta))}{\theta^2}[1 + \frac{v''(l(\theta))}{v'(l(\theta))}l(\theta)] > 0$$

so $y'(\theta) > 0$. Thus, we have

$$IC \Rightarrow \begin{cases} V'(\theta) = v'(l(\theta))\frac{1}{\theta}l(\theta) \\ y'(\theta) > 0 \end{cases}$$

we can show that these two conditions are also sufficient conditions for IC.

$$\begin{aligned} \frac{\partial V(\theta,\theta')}{\partial\theta'} &= e^{-K}c'(\theta') - \frac{1}{\theta}v'(\frac{1}{\theta}y(\theta'))y'(\theta') \\ &= \frac{1}{\theta'}v'(\frac{1}{\theta'}y(\theta'))y'(\theta') - \frac{1}{\theta}v'(\frac{1}{\theta}y(\theta'))y'(\theta') \quad \text{(implied by } V'(\theta) = v'(l(\theta))\frac{1}{\theta}l(\theta)) \\ &= [\frac{1}{\theta'}v'(\frac{y(\theta')}{\theta'}) - \frac{1}{\theta}v'(\frac{y(\theta')}{\theta})]y'(\theta') \end{aligned}$$

Since $y'(\theta) > 0, \forall\theta$

if $\theta' > \theta$, $\frac{1}{\theta'}v'(\frac{y(\theta')}{\theta'}) < \frac{1}{\theta}v'(\frac{y(\theta')}{\theta})$, $\frac{\partial V(\theta,\theta')}{\partial\theta'} < 0$, taxpayer has incentives to decrease his report from θ';

if $\theta' < \theta$, $\frac{1}{\theta'}v'(\frac{y(\theta')}{\theta'}) > \frac{1}{\theta}v'(\frac{y(\theta')}{\theta})$, $\frac{\partial V(\theta,\theta')}{\partial\theta'} < 0$, taxpayer has incentives to increase his report from θ';

To sum up, taxpayer θ reaches his highest perceived utility level by truthfully reporting, so these two conditions are also sufficient for IC. Thus,

$$IC \Rightarrow \begin{cases} V'(\theta) = v'(l(\theta))\frac{1}{\theta}l(\theta) \\ y'(\theta) > 0 \end{cases}$$

so we can rewrite government's problem as:

$$\max_{\{V(\theta),l(\theta)\}} \int_{\underline{\theta}}^{\overline{\theta}} \alpha(\theta)[e^K V(\theta) + (e^K - 1)v(l(\theta))]dF(\theta) \tag{21}$$

$$s.t. \int_{\underline{\theta}}^{\overline{\theta}} [\theta l(\theta) - e^K(V(\theta) + v(l(\theta)))]dF(\theta) \geq G \tag{22}$$

$$V'(\theta) = v'(l(\theta))\frac{1}{\theta}l(\theta) \tag{23}$$

$$y'(\theta) > 0 \tag{24}$$

For now, we just ignore the $y'(\theta) > 0$ condition and assume this to be true.

Let Hamiltonian equation $\mathcal{H}(\theta) = \alpha(\theta)[e^K V(\theta)+(e^K-1)v(l(\theta))]f(\theta)+\lambda[\theta l(\theta)-e^K(V(\theta)+v(l(\theta)))]f(\theta)+\mu(\theta)v'(l(\theta))\frac{1}{\theta}l(\theta)$, where λ is the multiplier for (22), and $\mu(\theta)$ is the multiplier for (23). Optimality condition implies:

$$\frac{\partial \mathcal{H}(\theta)}{\partial l(\theta)} = \alpha(\theta)(e^K - 1)v'(l(\theta))f(\theta) + \lambda(\theta - e^K v'(l(\theta)))f(\theta) + \mu(\theta)\frac{1}{\theta}v'(l(\theta))[1 + \frac{v''(l(\theta))}{v'(l(\theta))}l(\theta)] = 0 \tag{25}$$

$$\frac{\partial \mathcal{H}(\theta)}{\partial V(\theta)} = \alpha(\theta)e^K f(\theta) - \lambda e^K f(\theta) = -\mu'(\theta) \tag{26}$$

Since here we have quasi-linear utility function, $\lambda = \int_{\underline{\theta}}^{\overline{\theta}} \alpha(\theta)dF(\theta) = 1$. Assume that taxpayers have constant labor supply elasticity ε, $\frac{v''(l(\theta))}{v'(l(\theta))}l(\theta) = \frac{1}{\varepsilon}$, also from taxpayer's FOC, we know $e^{-K}(1 - \tau(y(\theta))) = \frac{1}{\theta}v'(l(\theta))$, thus we can simplify (25) as

$$\frac{e^K - 1}{e^K}\alpha(\theta) + \frac{\tau_1(\theta)}{1 - \tau_1(\theta)} = -\frac{\mu(\theta)}{\lambda}\frac{1}{\theta f(\theta)}(1 + \frac{1}{\varepsilon})e^{-K} \tag{27}$$

where $\tau_1(\theta) := T_1'(y_1(\theta))$, where $y_1(\theta)$ is the optimal income allocation for θ and T_1 is the optimal tax function. (26) implies:

$$-\frac{\mu(\theta)}{\lambda}e^{-K} = \int_{\theta}^{\overline{\theta}} (1 - \alpha(\theta))dF(\theta) \tag{28}$$

Combining (27) and (28), we get

$$\frac{e^K - 1}{e^K}\alpha(\theta) + \frac{\tau_1(\theta)}{1 - \tau_1(\theta)} = (1 + \frac{1}{\varepsilon})\frac{1}{\theta f(\theta)}\int_{\theta}^{\overline{\theta}} (1 - \alpha(\theta))dF(\theta) = \frac{\tau^M(\theta)}{1 - \tau^M(\theta)}$$

where τ^M denotes Mirrleesian tax rate. □

C Proof of Proposition 4

Proposition. *The optimal income tax T_1 is progressive if $\forall\theta \in [\underline{\theta}, \overline{\theta}]$*

$$\frac{e^K - 1}{e^K}\zeta_{\alpha,\theta} > \frac{1-\alpha(\theta)}{\alpha(\theta)}(1+\frac{1}{\varepsilon}) + \frac{1+\theta\frac{f'(\theta)}{f(\theta)}}{\alpha(\theta)}\frac{\tau^M(\theta)}{1-\tau^M(\theta)} \tag{29}$$

where $\zeta_{\alpha,\theta} = -\frac{d\alpha(\theta)}{d\theta}\frac{\theta}{\alpha(\theta)}$.

Proof. Denote $g_1(\theta) = \frac{\tau_1}{1-\tau_1}$, $g_0(\theta) = \frac{\tau^M}{1-\tau^M} = (1+\frac{1}{\varepsilon})\frac{1}{\theta f(\theta)}\int_\theta^{\overline{\theta}}(1-\alpha(\theta))dF(\theta)$

$$\begin{aligned} T_1(\theta) \text{ is progressive} &\Leftrightarrow g_1'(\theta) > 0 \\ &\Leftrightarrow g_0'(\theta) - \frac{e^K-1}{e^K}\alpha'(\theta) > 0 \\ &\Leftrightarrow -\frac{e^K-1}{e^K}\alpha'(\theta) > \frac{f(\theta)+\theta f'(\theta)}{\theta f(\theta)}\frac{\tau^M}{1-\tau^M} + (1+\frac{1}{\varepsilon})\frac{1-\alpha(\theta)}{\theta} \\ &\Leftrightarrow -\frac{e^K-1}{e^K}\frac{\alpha'(\theta)\theta}{\alpha(\theta)} > \frac{f(\theta)+\theta f'(\theta)}{\alpha(\theta)f(\theta)}\frac{\tau^M}{1-\tau^M} + (1+\frac{1}{\varepsilon})\frac{1-\alpha(\theta)}{\alpha(\theta)} \end{aligned}$$

Thus, we get the proposition result. □

D Proof of Proposition 1

Proposition 7. *Let $\tau^*(\theta)$ be the optimal marginal tax rate that taxpayer θ is facing. Denote $\omega(\theta) = \frac{u'(c(\theta))\alpha(\theta)}{\int_{\underline{\theta}}^{\overline{\theta}} u'(c(\theta))\alpha(\theta)dF(\theta)}$, $\varepsilon(\theta)$ is agent θ's Frisch elasticity of labor supply, we have:*

$$\frac{\tau^*(\theta)}{1-\tau^*(\theta)} = \frac{\tilde{\tau}^M(\theta)}{1-\tilde{\tau}^M(\theta)} - \underbrace{\frac{e^{K(T^*)}-1}{e^{K(T^*)}}\omega(\theta)}_{X(\theta)} + \underbrace{(\frac{[1+\frac{1}{\varepsilon(\theta)}]\int_\theta^{\overline{\theta}} g^V_{T^*}(x)dF(x)}{\theta f(\theta)} + g^v_{T^*}(\theta)\frac{\tau^*(\theta)}{1-\tau^*(\theta)})}_{Y(\theta)}$$

where $g^V_{T^*}(\theta) = \frac{\partial K}{\partial V(\theta)}(1-\omega(\theta))u(c(\theta))$ and $g^v_{T^*}(\theta) = \frac{\partial K}{\partial v(l(\theta))}(1-\omega(\theta))u(c(\theta))$, which captures the effects of policies on ambiguity level

$$\frac{\tilde{\tau}^M(\theta)}{1-\tilde{\tau}^M(\theta)} = [1+\frac{1}{\varepsilon(\theta)}]\frac{1-F(\theta)}{\theta f(\theta)}\frac{1}{1-F(\theta)}\int_\theta^{\overline{\theta}}[1-\omega(x)]dF(x)$$

, which has the same expression as Mirrleesian tax rate τ^M.

Proof. In the proof of Proposition 3, we change the choice variable from $\{c(\theta), y(\theta)\}_{\theta=\underline{\theta}}^{\overline{\theta}}$ to

$\{V(\theta), l(\theta)\}_{\theta=\underline{\theta}}^{\overline{\theta}}$. In the endogenous case, we also need to express the complexity level as a function of $\{V(\theta), l(\theta)\}_{\theta=\underline{\theta}}^{\overline{\theta}}$ instead of $\{T(\theta)\}_{\theta=\underline{\theta}}^{\overline{\theta}}$.

Since u is strictly increasing, it's invertible. $T(\theta) = y(\theta) - c(\theta) = \theta l(\theta) - u^{-1}[e^K V(\theta) + e^K v(l(\theta))]$, so we have

$$K = K(\{T(\theta)\}_{\theta=\underline{\theta}}^{\overline{\theta}}) = K(\{\theta l(\theta) - u^{-1}[e^K V(\theta) + e^K v(l(\theta))]\}_{\theta=\underline{\theta}}^{\overline{\theta}}) \tag{30}$$

Equation (30) gives us an implicit function of $K = K(V, l)$. Taking partial derivative w.r.t $V(\theta)$ and $l(\theta)$ on both sides, we have

$$\frac{\partial K}{\partial V(\theta)} = \frac{\partial K(T)}{\partial T(\theta)}(-e^K V(\theta)\frac{\partial K}{\partial V(\theta)} - e^K)\frac{1}{u'(c(\theta))} \tag{31}$$

$$\frac{\partial K}{\partial l(\theta)} = \frac{\partial K(T)}{\partial T(\theta)}(\theta - \frac{1}{u'(c(\theta))}e^K v'(l(\theta)) - \frac{1}{u'(c(\theta))}e^K v(l(\theta))\frac{\partial K}{\partial l(\theta)}) \tag{32}$$

From equation (31) and (32), we have

$$\frac{\partial K}{\partial V(\theta)} = \frac{-\frac{\partial K(T)}{\partial T(\theta)}e^K\frac{1}{u'(c(\theta))}}{1 + \frac{\partial K(T)}{\partial T(\theta)}e^K V(\theta)\frac{1}{u'(c(\theta))}} \tag{33}$$

$$\frac{\partial K}{\partial l(\theta)} = \frac{(\theta - e^K v'(l(\theta))\frac{1}{u'(c(\theta))})\frac{\partial K(T)}{\partial T(\theta)}}{1 + \frac{\partial K(T)}{\partial T(\theta)}e^K v(l(\theta))\frac{1}{u'(c(\theta))}} = \frac{\theta\frac{\partial K(T)}{\partial T(\theta)}\tau(\theta)}{1 + \frac{\partial K(T)}{\partial T(\theta)}e^K v(l(\theta))} \tag{34}$$

the second equality in equation (34) comes from taxpayer's F.O.C: $e^{-K}u'(c)(1 - \tau(y)) = \frac{1}{\theta}v'(\frac{y}{\theta})$.

Follow the similar proof as Proposition 3, the government's problem can be written as

$$\max_{\{V(\theta), l(\theta)\}} \int_{\underline{\theta}}^{\overline{\theta}} \alpha(\theta)[e^{K(V,l)}V(\theta) + (e^{K(V,l)} - 1)v(l(\theta))]dF(\theta)$$

$$s.t. \int_{\underline{\theta}}^{\overline{\theta}} (\theta l(\theta) - u^{-1}[e^{K(V,l)}(V(\theta) + v(l(\theta)))])dF(\theta) \geq G \tag{35}$$

$$V'(\theta) = v'(l(\theta))\frac{1}{\theta}l(\theta) \tag{36}$$

$$y'(\theta) > 0 \tag{37}$$

Let $\mathcal{H}(\theta) = \alpha(\theta)[e^{K(V,l)}V(\theta) + (e^{K(V,l)} - 1)v(l(\theta))]f(\theta) + \lambda(\theta l(\theta) - u^{-1}[e^{K(V,l)}(V(\theta) + v(l(\theta)))])f(\theta) + \mu(\theta)v'(l(\theta))\frac{1}{\theta}l(\theta)$, we have:

$$\frac{\partial \mathcal{H}(\theta)}{\partial l(\theta)} = \alpha(\theta)(e^K - 1)v'(l(\theta))f(\theta) + \lambda(\theta - e^K v'(l(\theta))\frac{1}{u'(c(\theta))})f(\theta)$$
$$+ (\alpha(\theta) - \frac{\lambda}{u'(c(\theta))})(v(l(\theta)) + V(\theta))e^K \frac{\partial K}{\partial l(\theta)} f(\theta) + \mu(\theta)\frac{1}{\theta}v'(l(\theta))[1 + \frac{v''(l(\theta))}{v'(l(\theta))}l(\theta)] = 0 \tag{38}$$

$$\frac{\partial \mathcal{H}(\theta)}{\partial V(\theta)} = (\alpha(\theta) - \frac{\lambda}{u'(c(\theta))})e^K f(\theta) + (\alpha(\theta) - \frac{\lambda}{u'(c(\theta))})(v(l(\theta)) + V(\theta))e^K \frac{\partial K}{\partial V(\theta)} f(\theta) = -\mu'(\theta) \tag{39}$$

where $\lambda = \int_{\underline{\theta}}^{\overline{\theta}} \alpha(\theta)u'(c(\theta))dF(\theta) = 1$, and $\frac{v''(l(\theta))}{v'(l(\theta))}l(\theta) = \frac{1}{\varepsilon}$. Denote $\omega(\theta) = \frac{u'(c(\theta))\alpha(\theta)}{\int_{\underline{\theta}}^{\overline{\theta}} u'(c(\theta))\alpha(\theta)dF(\theta)}$, after some simplifications (these are just similar as the process showed in proof to Proposition 3), we can get

$$[1-g^v_{T^*}(\theta)]\frac{\tau^*(\theta)}{1-\tau^*(\theta)} + \frac{e^{K(T^*)}-1}{e^{K(T^*)}}\omega(\theta) = [1+\frac{1}{\varepsilon(\theta)}]\frac{1-F(\theta)}{\theta f(\theta)}\frac{1}{1-F(\theta)}\int_{\theta}^{\overline{\theta}}[1-\omega(x)+g^V_{T^*}(x)]dF(x)$$

where

$$g^V_{T^*}(\theta) = \frac{-\frac{\partial K(T^*)}{\partial T^*(\theta)}e^{K(T^*)}(1-\omega(\theta))u(c(\theta))}{u'(c(\theta)) + \frac{\partial K(T^*)}{\partial T^*(\theta)}e^{K(T^*)}V(\theta)} = \frac{\partial K}{\partial V(\theta)}(1-\omega(\theta))u(c(\theta))$$
$$g^v_{T^*}(\theta) = \frac{\frac{\partial K(T^*)}{\partial T^*(\theta)}e^{K(T^*)}(1-\omega(\theta))u(c(\theta))}{u'(c(\theta)) + \frac{\partial K(T^*)}{\partial T^*(\theta)}e^{K(T^*)}v(l(\theta))} = \frac{\partial K}{\partial v(l(\theta))}(1-\omega(\theta))u(c(\theta))$$

rearrange the terms, we have

$$\frac{\tau^*(\theta)}{1-\tau^*(\theta)} = \frac{\tilde{\tau}^M(\theta)}{1-\tilde{\tau}^M(\theta)} - \underbrace{\frac{e^{K(T^*)}-1}{e^{K(T^*)}}\omega(\theta)}_{X(\theta)} + \underbrace{(\frac{[1+\frac{1}{\varepsilon(\theta)}]\int_{\theta}^{\overline{\theta}} g^V_{T^*}(x)dF(x)}{\theta f(\theta)} + g^v_{T^*}(\theta)\frac{\tau^*(\theta)}{1-\tau^*(\theta)})}_{Y(\theta)}$$

□

E Proof of Lemma 1

The government's problem with exogenous ambiguity K becomes:

$$\begin{aligned}
W^{ex}(K) = \max_{\{V(\theta),l(\theta)\}} & \int_{\underline{\theta}}^{\overline{\theta}} \alpha(\theta)[e^K V(\theta) + (e^K - 1)v(l(\theta))]dF(\theta) \\
s.t. & \int_{\underline{\theta}}^{\overline{\theta}} (\theta l(\theta) - u^{-1}[e^K(V(\theta) + v(l(\theta)))])dF(\theta) \geq G \\
& V'(\theta) = v'(l(\theta))\frac{1}{\theta}l(\theta) \\
& y'(\theta) > 0
\end{aligned}$$

Applying envelope theorem, we can get

$$\begin{aligned}
\frac{\partial W^{ex}(K)}{\partial K} &= \int_{\underline{\theta}}^{\overline{\theta}} (\alpha(\theta) - \frac{\lambda}{u'(c(\theta))})e^K(V(\theta) + v(l(\theta)))dF(\theta) \\
&= \lambda \int_{\underline{\theta}}^{\overline{\theta}} (\omega(\theta) - 1)\frac{u(c(\theta))}{u'(c(\theta))}dF(\theta) \qquad (40)
\end{aligned}$$

Recall $\omega(\theta) = \frac{u'(c(\theta))\alpha(\theta)}{\lambda}$. Notice that IC condition implies $y'(\theta) > 0$, since $c(\theta) = y(\theta) - T(y(\theta))$, we have $c'(\theta) = (1 - T'(y(\theta)))y'(\theta) > 0$ (since $T'(y(\theta)) < 1$). As a result, $u'(c(\theta))$ is non-increasing with θ. Since $\alpha(\theta)$ is non-increasing with θ, we have $\omega(\theta)$ is non-increasing in θ.

Denote θ_c as the ability level such that $\omega(\theta_c) = 1$. Because of continuity of $\alpha(\theta)$ and $\int_{\underline{\theta}}^{\overline{\theta}} \omega(\theta)dF(\theta) = 1$, we know such θ_c exists. When $\theta < \theta_c$, $\omega(\theta) \geq 1$; when $\theta > \theta_c$, $\omega(\theta) \leq 1$.

We can rewrite (40) as

$$\frac{\partial W^{ex}(K)}{\partial K} = [\int_{\underline{\theta}}^{\theta_c} (\omega(\theta) - 1)\frac{u(c(\theta))}{u'(c(\theta))}dF(\theta) - \int_{\theta_c}^{\overline{\theta}} (1 - \omega(\theta))\frac{u(c(\theta))}{u'(c(\theta))}dF(\theta)]\lambda \qquad (41)$$

Besides, we have $\frac{u(c(\theta))}{u'(c(\theta))}$ is non-decreasing in θ, so(41) becomes:

$$\begin{aligned}
\frac{\partial W^{ex}(K)}{\partial K} &\leq [\int_{\underline{\theta}}^{\theta_c} (\omega(\theta) - 1)\frac{u(c(\theta_c))}{u'(c(\theta_c))}dF(\theta) - \int_{\theta_c}^{\overline{\theta}} (1 - \omega(\theta))\frac{u(c(\theta_c))}{u'(c(\theta_c))}dF(\theta)]\lambda \\
&= [\int_{\underline{\theta}}^{\overline{\theta}} (\omega(\theta) - 1)dF(\theta)]\frac{u(c(\theta_c))}{u'(c(\theta_c))}\lambda = 0
\end{aligned}$$

thus we have $\frac{\partial W^{ex}(K)}{\partial K} \leq 0$.

F Proof of Proposition 6

Proposition. *If taxpayers' belief set is type-dependent like* (13) *and utility function is quasi-linear. Denote* $T_2'(\theta)$ *as the optimal marginal tax rate that taxpayer* θ *is facing, we have*

$$\frac{e^{\tilde{K}(\theta)}-1}{e^{\tilde{K}(\theta)}}\alpha(\theta)+\frac{\tau_2(\theta)}{1-\tau_2(\theta)}=(1+\frac{-\theta\eta'(\theta)\varepsilon}{1+\varepsilon})\frac{\tau^M(\theta)}{1-\tau^M(\theta)}$$

where $\tilde{K}(\theta)=K+\eta(\theta)$.

Proof. The proof to this proposition is similar to Proposition 3, the government's problem is

$$\max_{\{c(\theta),y(\theta)\}}\int_{\underline{\theta}}^{\bar{\theta}}\alpha(\theta)U(\theta)dF(\theta)$$

$$s.t.\int_{\underline{\theta}}^{\bar{\theta}}c(\theta)dF(\theta)+G\leq\int_{\underline{\theta}}^{\bar{\theta}}y(\theta)dF(\theta) \quad (42)$$

$$e^{-K-\eta(\theta)}[y(\theta)-T(y(\theta))]-v(\frac{y(\theta)}{\theta})\geq e^{-K-\eta(\theta)}[y(\theta')-T(y(\theta'))]-v(\frac{y(\theta')}{\theta}) \quad \forall\theta,\theta' \quad (43)$$

Let $V(\theta,\theta')=e^{-K-\eta(\theta)}c(\theta')-v(\frac{1}{\theta}y(\theta'))$, denoting the perceived utility of taxpayer θ if his report is θ'.

$$IC\Rightarrow\begin{cases}\frac{\partial V(\theta,\theta')}{\partial\theta'}|_{\theta'=\theta}=e^{-K-\eta(\theta)}c'(\theta)-\frac{1}{\theta}v'(\frac{1}{\theta}y(\theta))y'(\theta)=0\\ \frac{\partial^2 V(\theta,\theta')}{\partial\theta'^2}|_{\theta'=\theta}<0\end{cases}$$

Recall $V(\theta)=e^{-K-\eta(\theta)}c(\theta)-v(\frac{1}{\theta}y(\theta))$, thus

$$\begin{aligned}V'(\theta)&=e^{-K-\eta(\theta)}c'(\theta)-e^{-K-\eta(\theta)}c(\theta)\eta'(\theta)-\frac{1}{\theta}v'(\frac{1}{\theta}y(\theta))y'(\theta)+\frac{1}{\theta^2}v'(\frac{1}{\theta}y(\theta))y(\theta)\\&=-e^{-K-\eta(\theta)}c(\theta)\eta'(\theta)+\frac{1}{\theta}v'(l(\theta))l(\theta)>\frac{1}{\theta}v'(l(\theta))l(\theta)\end{aligned}$$

Notice that when the belief set is $\Pi(\theta,y,T)$, $V'(\theta)>\frac{1}{\theta}v'(l(\theta))l(\theta)$, which is the increment of perceived utility ($V'(\theta)$) when the belief set is $\Pi(y,T)$. This is means if allocations are fixed, increasing ability level by a same amount implies a higher perceived utility increase. This is because $\eta'(\theta)<0$, in addition to allocation changes, a higher utility level also makes taxpayers less ambiguous, thus brings more perceived utility increase.

Since $\frac{\partial^2 V(\theta,\theta')}{\partial\theta'^2} + \frac{\partial^2 V(\theta,\theta')}{\partial\theta\partial\theta'}|_{\theta'=\theta} = 0$, so $\frac{\partial^2 V(\theta,\theta')}{\partial\theta'^2}|_{\theta'=\theta} < 0$ is equivalent to $\frac{\partial^2 V(\theta,\theta')}{\partial\theta\partial\theta'}|_{\theta'=\theta} > 0$.

$$\frac{\partial^2 V(\theta,\theta')}{\partial\theta\partial\theta'}|_{\theta'=\theta} = \frac{1}{\theta}v'(l(\theta))(-\eta'(\theta) + \frac{1}{\theta}(1+\frac{1}{\varepsilon}))y'(\theta) > 0$$

which is equivalent to $y'(\theta) > 0$. (since $v'(l) > 0$, $\eta'(\theta) < 0$)

Also,

$$\begin{aligned}\frac{\partial V(\theta,\theta')}{\partial\theta'} &= e^{-K-\eta(\theta)}c'(\theta') - \frac{1}{\theta}v'(\frac{1}{\theta}y(\theta'))y'(\theta') \\ &= [e^{\eta(\theta')-\eta(\theta)}\frac{1}{\theta'}v'(\frac{1}{\theta'}y(\theta')) - \frac{1}{\theta}v'(\frac{1}{\theta}y(\theta'))y'(\theta')]y'(\theta')\end{aligned}$$

if $\theta' > \theta$, $\frac{1}{\theta'}v'(\frac{y(\theta')}{\theta'}) < \frac{1}{\theta}v'(\frac{y(\theta')}{\theta})$, $e^{\eta(\theta')-\eta(\theta)} < 1$, thus $e^{\eta(\theta')-\eta(\theta)}\frac{1}{\theta'}v'(\frac{1}{\theta'}y(\theta')) - \frac{1}{\theta}v'(\frac{1}{\theta}y(\theta'))y'(\theta') < 0$;

if $\theta' < \theta$, $\frac{1}{\theta'}v'(\frac{y(\theta')}{\theta'}) > \frac{1}{\theta}v'(\frac{y(\theta')}{\theta})$, $e^{\eta(\theta')-\eta(\theta)} > 1$, thus $e^{\eta(\theta')-\eta(\theta)}\frac{1}{\theta'}v'(\frac{1}{\theta'}y(\theta')) - \frac{1}{\theta}v'(\frac{1}{\theta}y(\theta'))y'(\theta') > 0$;

so we have

$$IC \Leftrightarrow \begin{cases} V'(\theta) = -e^{-K-\eta(\theta)}c(\theta)\eta'(\theta) + \frac{1}{\theta}v'(l(\theta))l(\theta) \\ y'(\theta) > 0 \end{cases}$$

the government's problem becomes

$$\max_{\{V(\theta),l(\theta)\}} \int_{\underline{\theta}}^{\overline{\theta}} \alpha(\theta)[e^{K+\eta(\theta)}V(\theta) + (e^{K+\eta(\theta)} - 1)v(l(\theta))]dF(\theta)$$

$$s.t. \int_{\underline{\theta}}^{\overline{\theta}} [\theta l(\theta) - e^{K+\eta(\theta)}(V(\theta) + v(l(\theta)))]dF(\theta) \geq G \tag{44}$$

$$V'(\theta) = -(V(\theta) + v(l(\theta))\eta'(\theta) + \frac{1}{\theta}v'(l(\theta))l(\theta) \tag{45}$$

$$y'(\theta) > 0 \tag{46}$$

Optimality condition implies:

$$\frac{\partial\mathcal{H}(\theta)}{\partial l(\theta)} = \alpha(\theta)(e^{K(\theta)} - 1)v'(l(\theta))f(\theta) + \lambda(\theta - e^{K(\theta)}v'(l(\theta)))f(\theta) + \mu(\theta)\frac{1}{\theta}v'(l(\theta))[-\theta\eta'(\theta) + (1+\frac{1}{\varepsilon})] = 0 \tag{47}$$

$$\frac{\partial\mathcal{H}(\theta)}{\partial V(\theta)} = \alpha(\theta)e^{K(\theta)}f(\theta) - \lambda e^{K(\theta)}f(\theta) - \mu(\theta)\eta'(\theta) = -\mu'(\theta) \tag{48}$$

where $\lambda$is the multiplier for (44) and $\mu(\theta)$ is the multiplier for (45) and we assume $\frac{v''(l(\theta))}{v'(l(\theta))}l(\theta) = \frac{1}{\varepsilon}$. Since we have $\lambda = \int_{\underline{\theta}}^{\overline{\theta}} \alpha(\theta)dF(\theta) = 1$, and from taxpayer θ's FOC we know $v'(l(\theta)) = \theta e^{-K(\theta)}(1 - T_2'(\theta))$. (47) becomes

$$\frac{e^{K(\theta)} - 1}{e^{K(\theta)}}\alpha(\theta) + \frac{\tau_2(\theta)}{1 - \tau_2(\theta)} = -\frac{\mu(\theta)}{\theta f(\theta)}e^{-K(\theta)}[-\theta\eta'(\theta) + (1 + \frac{1}{\varepsilon})] \tag{49}$$

Equation (48) becomes

$$\mu'(\theta) - \eta'(\theta)\mu(\theta) = (1 - \alpha(\theta))e^{K(\theta)}f(\theta) \tag{50}$$

which is a first-order linear differential equation, and its solution is

$$\begin{aligned}\mu(\theta) &= \int_{\underline{\theta}}^{\theta} e^{\eta(\theta)-\eta(\theta)}(1 - \alpha(\theta))e^{K(\theta)}dF(\theta) \\ &= \int_{\underline{\theta}}^{\theta} e^{K(\theta)}(1 - \alpha(\theta))dF(\theta)\end{aligned}$$

(50) becomes

$$\begin{aligned}\frac{e^{K(\theta)} - 1}{e^{K(\theta)}}\alpha(\theta) + \frac{\tau_2(\theta)}{1 - \tau_2(\theta)} &= [-\theta\eta'(\theta) + (1 + \frac{1}{\varepsilon})]\frac{1}{\theta f(\theta)}\int_{\theta}^{\overline{\theta}}[1 - \alpha(\theta)]dF(\theta) \\ &= \frac{-\theta\eta'(\theta) + (1 + \frac{1}{\varepsilon})}{1 + \frac{1}{\varepsilon}}(1 + \frac{1}{\varepsilon})\frac{1}{\theta f(\theta)}\int_{\theta}^{\overline{\theta}}[1 - \alpha(\theta)]dF(\theta) \\ &= (1 + \frac{-\theta\eta'(\theta)\varepsilon}{1 + \varepsilon})\frac{\tau^M(\theta)}{1 - \tau^M(\theta)}\end{aligned}$$

which finishes the proof. □