Optimal Income Taxation with a Stationarity Constraint in a Dynamic Stochastic Economy

Berliant, Marcus and Fujishima, Shota

Washington University in St. Louis, Tokyo University of Science

26 May 2016

Optimal Income Taxation with a Stationarity Constraint in a Dynamic Stochastic Economy*

Marcus Berliant† Shota Fujishima‡

May 26, 2016

Abstract

We consider the optimal nonlinear income taxation problem in a dynamic, stochastic environment when the government cannot change the tax rule as uncertainty resolves. Due to such a stationarity constraint, our taxation problem is reduced to a static one over an expanded type space that incorporates type evolution. We strengthen the argument in the static model that the zero top marginal tax rate result is of little practical importance because it only applies to the top of the expanded type space. If the maximal type increases over time, the person with top ability in any period but the last has a positive marginal tax rate.

JEL classification: H21

Keywords: Optimal income taxation; New dynamic public finance

1 Introduction

Since the *New Dynamic Public Finance* was inaugurated, progress has been made in clarifying what the optimal dynamic nonlinear income tax looks like. This

*We are grateful to John Conley, an associate editor, three referees, Eric Stephens, Yoshihiro Takamatsu, Jonathan Hamilton, Steven Slutsky, participants at the Southern Economic Association meeting in New Orleans and the Japanese Economic Association meeting at the University of Toyama, and the seminar audience at Boston College for their helpful comments and suggestions. Shota Fujishima gratefully acknowledges travel support from the Center for Research in Economics and Strategy (CRES) at the Olin Business School, Washington University in St. Louis. Any remaining errors are our own.

†Department of Economics, Washington University in St. Louis, Campus Box 1208, One Brookings Drive, St. Louis, MO 63130-4899 USA. Phone: +1 314 935 8486, Fax: +1 314 935 4156, Email: berliant@wustl.edu

‡School of Management, Tokyo University of Science, 1-11-2, Fujimi, Chiyoda, Tokyo 102-0071, Japan. Phone/Fax: +81 3 3288 2506, Email: sfujishima@rs.tus.ac.jp

agenda aims to extend the seminal work of Mirrlees (1971), who studies optimal income taxation in a static environment, to dynamic, stochastic environments.[1] Dynamic tax rules are in effect dynamic contracts because taxpayers have private information about their labor productivity, so the optimal dynamic income tax rule is generally complicated: it is non-stationary and depends on the entire history of income declared for any taxpayer. However, it is questionable whether governments can implement such complex tax rules because making tax rules time-dependent and tracking histories of income would entail large administrative and compliance costs. Indeed, neither of our governments (i.e., the US and Japanese governments) is tracking income histories for labor income taxation.

In view of this observation, we contribute to the New Dynamic Public Finance literature by considering optimal dynamic income taxation when the government faces a stationarity constraint that the tax rule cannot be changed over time. That is, the government can use only stationary tax rules. Moreover, stationarity of tax rule implies that the tax cannot depend on histories of income. Indeed, the government using a stationary tax rule can look at only current incomes, just as it can only look at current incomes in the initial period. Naturally, we also assume that the government makes a full commitment to its (stationary) tax rule. That is, once the tax rule is determined in the initial period, the government cannot switch to another stationary tax rule afterwards. We are assuming that such commitment is not only possible, but perhaps unavoidable, due to political deadlock over the issue of tax policy, as in the US right now.[2] Thus, we may interpret our planner's problem on a politician's short time-scale. Although our assumptions might be extreme, we believe that it is important and useful to have a sense about what the optimal dynamic income tax looks like when the set of tax rules is limited to ones that are feasible in practice. Moreover, as we shall discuss below, stationarity brings with it tractability.

We consider a finite horizon discrete time model in which the government would like to maximize the equal-weight utilitarian social welfare function. Our economy is heterogeneous, as we begin with a non-trivial exogenous type dis-

[1]See Kocherlakota (2010) for an overview of this literature.

[2]Indeed, the US government has not changed its income tax system in a major way since 1986. The Japanese government is more flexible, but it has not changed its income tax system in a major way since 2007. Therefore, once the tax systems are fixed, they persist for some time.

tribution where type here is people's earning ability.[3] People's type is subject to stochastic shocks in each period. We focus on idiosyncratic first-order Markov shocks that are i.i.d. among people. Regarding intertemporal resource allocation, we assume that the government can save or borrow from an outside party. However, we assume that agents cannot save or borrow. We shall discuss this assumption further below.

Although the analytical characterizations and even numerical analysis of the optimal dynamic tax system are difficult in general, we can analytically characterize the optimal tax system because our problem can be reduced to a static one owing to the stationarity of the tax rule.[4] Specifically, this is because the tax rule depends on only the current income under the stationarity constraint and individual saving or borrowing is not allowed, so we can regard an agent living for T periods, who has a time-separable preference, as distinct agents in each period and for each shock by appealing to the law of large numbers. Moreover, because the government can save or borrow, the law of large numbers also implies a single aggregate resource constraint. We can then directly apply the arguments for static models to our model.

A famous result in static optimal income taxation is that the top marginal tax rate is zero. That is, the top earner's marginal tax rate is zero. However, we cast doubt on its policy relevance. In our dynamic stochastic economy, the support of types will move over time, and a direct application of the static arguments implies that the marginal tax rate is zero at the top of the *expanded* type space, or the union of supports over time. As a result, we can consider a structure of type evolution such that the zero top marginal tax result does not apply for periods before the last period, and that the standard result applies at the last period. It has been argued in static models that the zero top marginal tax rate result is not important because the fraction of people who face a zero marginal tax rate is small, and our result strengthens this. Indeed, if the time horizon is large, it is not really relevant.

The policy relevance of the zero top marginal tax rate result has also been

[3]If we do not fix the initial type distribution, the model has identical agents facing uncertainty, which is like a macro model. However, as long as we consider the equal-weight utilitarian social welfare function, the distinction is not essential for the optimal tax rule as Farhi and Werning (2013) illustrate.

[4]Naturally, gaining tractability in this way widens the analytical insights about optimal dynamic income taxation we could derive. For example, if we assumed quasi-linear utility, we could conduct comparative static analysis as in Weymark (1987).

questioned from other perspectives. Diamond (1998) considers unbounded skill distributions and shows that the asymptotic marginal tax rate is generally nonzero (see also Diamond and Saez, 2011). Tuomala (1990, chapter 9) considers uncertainties in income and type, respectively, and examines how the insurance motive affects the optimal income tax. He numerically shows that, in either case, the marginal tax rate is increasing in income up to the 99th percentile of the income distribution. In their arguments, the zero top marginal tax result does not hold. In our model, on the other hand, it does hold but only at the top of the expanded type space.

The stationarity constraint has a non-negligible impact on equilibrium outcomes of new dynamic public finance models. Battaglini and Coate (2008), who consider history-dependent non-stationary tax rules, show that the marginal tax rate is zero if an agent is currently, or has at some point been, the top earner. Thus, an individual who is the top earner in all periods faces a zero marginal tax rate in *every* period. Evidently, their tax rule takes full advantage of the fact that it can be history-dependent and non-stationary. In our model, as we have stated, such an individual faces a zero marginal tax rate only in the terminal period. Moreover, we observe that the structure of the stochastic shock can also have a non-negligible impact. Because Battaglini and Coate (2008) consider a two-state Markov process, the support of types is fixed over time. Thus, their result implies that the fraction of people whose allocation is distorted vanishes as the time horizon increases. On the other hand, in our model, the support of types generally moves over time, and all people's allocations are almost surely distorted in any period.

Regarding the past literature that is relevant to our work, one of the most general treatments of optimal nonlinear income taxation in a dynamic, stochastic economy is Kocherlakota (2005). In his model, both idiosyncratic and aggregate shocks are present, and no restriction is imposed on the processes of shocks. Albanesi and Sleet (2006) consider optimal taxation in a dynamic stochastic economy with i.i.d. idiosyncratic shocks. They show that the constrained-efficient allocation can be implemented as a competitive equilibrium with an indirect mechanism that depends on only current wealth and current labor income. Battaglini and Coate (2008), as we have already mentioned, consider a dynamic stochastic economy where idiosyncratic shocks evolve as a two-state Markov process. Although the stochastic structure is simplified, they address the effects of people's risk attitude and the time-consistency of the optimal tax rule.[5]

There has been some work that shares our motivation and studies tax rules that are more realistic than fully optimal rules in dynamic economies. It has been found that simple tax rules can achieve sizable welfare gains. Weinzierl (2011) considers history-independent non-stationary tax rules, which he calls age-dependent tax rules, in dynamic economies.[6] He shows that, if an agent is the top earner in every period, she faces no distortion in every period. On the contrary, in our stationary case, the top of the expanded type space where no distortion arises is generally not attained in any period except the last, even if an agent is always the top earner. Golosov *et al*. (2013) study a realistic pension system in which the income tax is history-independent and retirement benefits are inspired by the actual US system. Farhi and Werning (2013) consider a life cycle model with idiosyncratic shocks that evolve as a Markov process, and study history-independent non-stationary tax rules. Compared to the work above, we consider an even simpler case where the tax is stationary.

The rest of this paper proceeds as follows. In Section 2, we state the basic structure of the model, present our problem, and characterize the second-best tax rule. Section 3 contains our conclusions and discusses subjects for future research. Proofs omitted from the main text are provided in an Appendix.

2 The Model

We consider a finite horizon model with a unit mass of agents. The economy lasts for $T+1$ periods. In period 0, each agent is endowed with type $w_0 \in W_0 \subseteq \mathbb{R}_{++}$, distributed with density function f^0. We view our model as a heterogeneous economy where the type distribution is f^0. However, there are idiosyncratic shocks to the agents' types in the subsequent periods. The draw of the shock is independent over agents. We assume that the type w_t follows a first-order Markov process with conditional density $f^t(w_{t+1} \,|\, w_t)$. Let $W_t = \{w_t : \exists w_{t-1} \in W_{t-1}, f^t(w_t \,|\, w_{t-1}) > 0\}$. W_t is the type space in period t. We assume that W_t is a non-degenerate closed interval in $\mathbb{R}_{++}$. For later use, we also define the expanded type space $W = \bigcup_{t=0}^{T} W_t$. This is

[5]In a two-period deterministic environment, Berliant and Ledyard (2014) study a history-dependent non-stationary tax rule while addressing time-consistency.

[6]Gaube (2010) also discusses age-dependent tax rules. Because his main interest lies in the time-consistency of tax rules, he focuses on a two-period model.

also assumed to be a non-degenerate closed interval.

The agents supply labor and consume the good produced under constant returns to scale in each period. In our model, type represents the earning ability of agents. That is, if the labor supply of agent w is ℓ, his gross income is given by $y = w\ell$. As is usual in optimal taxation models, the agents face a trade-off between consumption and leisure. The utility function is

$$U\left(\{c_t, \ell_t\}_{t=0}^T\right) = \sum_{t=0}^{T} \rho^t u(c_t, \ell_t) \tag{1}$$

where $\ell_t \in [0, 1]$ is labor in period t, c_t is consumption in period t, and $\rho > 0$ is the discount factor. We assume that $u(c, \ell)$ is separable in c and ℓ:

$$u(c, \ell) = \psi_c(c) - \psi_\ell(\ell), \tag{2}$$

where ψ_c and ψ_ℓ are strictly increasing and twice continuously differentiable. Moreover, ψ_c and $-\psi_\ell$ are concave where at least one of them is strictly so.[7]

The government would like to maximize social welfare. In this paper, we consider the following utilitarian social welfare function:

$$SW = \int_{W_0} E\left[U\left(\{c_t, y_t/w_t\}_{t=0}^T\right) \,\middle|\, w_0\right] f^0(w_0)\mathrm{d}w_0. \tag{3}$$

where the conditional expectation is taken with respect to

$$f^t(w_t \,|\, w_0) = \int_{W^{t-1}} f^t(w_t \,|\, w_{t-1}) f^{t-1}(w_{t-1} \,|\, w_{t-2}) \cdots f^1(w_1 \,|\, w_0)\mathrm{d}w_{t-1}\mathrm{d}w_{t-2} \cdots \mathrm{d}w_1$$

for each t. Under the assumptions imposed on the one-period utility function, it follows that redistribution is desirable under the utilitarian welfare function with, if necessary, a twice continuously differentiable, increasing, and strictly concave transformation of the one-period utility function (Hellwig, 2007, Corollary 3.2).[8]

[7] These are sufficient for u to be strictly quasi-concave.

[8] For general utility functions, we need the *single-crossing property* which requires that $\frac{\partial u(c,y/w)/\partial c}{|\partial u(c,y/w)/\partial y|}$ be increasing in w. For our separable utility function, we have

$$\frac{\mathrm{d}}{\mathrm{d}w} \frac{\partial u(c, y/w)/\partial c}{|\partial u(c, y/w)/\partial y|} = \frac{\psi_c'(c)\psi_\ell'(y/w) + \psi_c'(c)\psi_\ell''(y/w)y/w}{\psi_\ell'(y/w)^2} > 0, \tag{4}$$

The planner would like to carry out redistribution through lump-sum income taxes, but he cannot observe the agents' types. Thus, the government needs to design a mechanism that makes the agents reveal their true types (i.e., it needs to design an *incentive compatible* (IC) mechanism).

Specifically, we consider a direct mechanism in which agents report their types, and the labor income $y_t(\cdot)$ and consumption $c_t(\cdot)$ are specified for each report in each period. In the following, we assume that people cannot save or borrow, and hence focus on labor income taxation (i.e., capital income taxation is excluded). Although the no saving assumption is strong for dynamic models, it is not uncommon in the NDPF literature.[9] A model that is consistent with this assumption is obtained from primitives when we assume quasi-linear utility and the discount and interest rates coincide. In this case, if we tax savings at all, nobody saves because people are indifferent about when to consume, and there is no revenue from taxing savings. The tax then becomes a labor income tax. A quasi-linear utility function is consistent with the other assumptions we shall use.

Although the allocation rule could depend on histories of types in general, one important consequence of excluding saving is that any incentive compatible mechanism is history independent. To see this, suppose that allocation rule $(y_t(\cdot), c_t(\cdot))_{t=0}^T$ is incentive compatible. Then, in the last period,

$$\forall w^T \in W^T, u\left(c_T(w^T), \frac{y_T(w^T)}{w_T}\right) \geq u\left(c_T(\hat{w}^T), \frac{y_T(\hat{w}^T)}{w_T}\right) \text{ for all } \hat{w}^T \in W^T, \tag{5}$$

where W^t is the set of feasible histories up to period t. That is,

$$W^t = \{(w_0, w_1, ..., w_t) : w_0 \in W_0, f^s(w_s \mid w_{s-1}) > 0 \text{ for all } s = 1, 2, ..., t\}. \tag{6}$$

However, because the IC condition (5) depends on only w_T, $(c_T(\cdot), y_T(\cdot))$ is history-independent. Next, the IC constraint in period $T-1$ is

$$\forall w^{T-1} \in W^{T-1}, u\left(c_{T-1}(w^{T-1}), \frac{y_{T-1}(w^{T-1})}{w_{T-1}}\right) + \rho E\left[u\left(c_T(w^T), \frac{y_T(w^T)}{w_T}\right) \mid w_{T-1}\right]$$

because $\psi_c' > 0, \psi_\ell' > 0, \psi_c'' \leq 0, \psi_\ell'' \geq 0$, and $\psi_c'' = 0 \Rightarrow \psi_\ell'' > 0$. A twice continuously differentiable, increasing, and strictly concave transformation of u becomes necessary for the desirability of redistribution when ψ_c is not strictly concave.

[9] For example, Battaglini and Coate (2008) and Weinzierl (2011) make this assumption.

$$\geq u\left(c_{T-1}(\hat{w}^{T-1}), \frac{y_{T-1}(\hat{w}^{T-1})}{w_{T-1}}\right)+\rho E\left[u\left(c_T(\hat{w}^{T-1}, w_T), \frac{y_T(\hat{w}^{T-1}, w_T)}{w_T}\right) \mid w_{T-1}\right] \text{ for all } \hat{w}^{T-1} \in W^{T-1}. \tag{7}$$

However, because $(c_T(\cdot), y_T(\cdot))$ is history-independent, (7) reduces to

$$\forall w^{T-1} \in W^{T-1}, u\left(c_{T-1}(w^{T-1}), \frac{y_{T-1}(w^{T-1})}{w_{T-1}}\right) \geq u\left(c_{T-1}(\hat{w}^{T-1}), \frac{y_{T-1}(\hat{w}^{T-1})}{w_{T-1}}\right) \text{ for all } \hat{w}^{T-1} \in W^{T-1}. \tag{8}$$

Thus, $(c_{T-1}(\cdot), y_{T-1}(\cdot))$ is history-independent.[10] Continuing in this way, we induce that $(c_t(\cdot), y_t(\cdot))$ is history-independent for all t. Hence, in the following, we may focus on history-independent allocation rules. Moreover, because we will use optimal control theory to solve for the second-best allocation rule, defined below, we assume that allocation rules are piecewise continuously differentiable.

Given allocation rule $(y_t(\cdot), c_t(\cdot))_{t=0}^{T}$, the income tax rule $(\tau_t(\cdot))_{t=0}^{T}$, which is an indirect mechanism mapping declared labor incomes to taxes, is constructed. However, the planner faces a stationarity constraint for $(\tau_t(\cdot))_{t=0}^{T}$. That is, $\tau_t(\cdot) = \tau(\cdot)$ for all t and, as a consequence, $\tau(\cdot)$ is independent of history. Let $Y = \bigcup_{t=0}^{T} y_t(W_t)$ where $y_t(W_t)$ is the range of y_t. The income tax rule is then given by map $\tau : Y \to \mathbb{R}$.

We would like to consider stationary allocation rules for analytical convenience. To see under what environment a stationary τ implies a stationary allocation rule, note that the IC conditions in terms of τ must hold for any deviations in a single domain because the domain of τ is time-independent. Specifically, the IC constraint is that, for any t,

$$\forall w \in W_t,\ u\left(y_t(w) - \tau(y_t(w)), \frac{y_t(w)}{w}\right) \geq u\left(y' - \tau(y'), \frac{y'}{w}\right) \text{ for all } y' \in Y. \tag{9}$$

Thus, in terms of allocation rule, we have

$$\forall w \in W_t,\ u\left(c_t(w), \frac{y_t(w)}{w}\right) \geq u\left(c_s(w'), \frac{y_s(w')}{w}\right) \text{ for all } w' \in W_s \text{ and } s. \tag{10}$$

We then have the following result.

[10] Although (7) is an ex post condition, it follows that it is equivalent to ex ante IC constraints up to measure zero sets (Fernandes and Phelan, 2000).

Lemma 1. *Suppose that $\psi'_\ell(\ell)\ell$ is strictly increasing in ℓ. Let $(y_t(\cdot), c_t(\cdot))_{t=0}^T$ be history-independent and incentive compatible. Then, there exist maps $y : W \to \mathbb{R}$ and $c : W \to \mathbb{R}$ such that $y_t = y$ and $c_t = c$ almost everywhere on W_t for all t.*

Proof. See Appendix. □

We henceforth assume that $\psi'_\ell(\ell)\ell$ is strictly increasing in ℓ. For instance, this condition holds if $\psi_\ell(\ell) = \ell^\alpha$ for $\alpha \geq 1$. Therefore, we focus on allocation rules described by function $(y, c) : W \to \mathbb{R}^2$ such that the allocation rule in period t is the restriction of (y, c) to W_t.[11] By (10), the IC constraint is then

$$\forall w \in W,\ u\left(\frac{y(w)}{w}, c(w)\right) \geq u\left(\frac{y(w')}{w}, c(w')\right) \text{ for all } w' \in W. \tag{IC}$$

In addition to the IC constraints, the government faces a resource constraint: it needs to finance G in units of consumption good through the tax. This revenue could be used for a public good that is fixed in quantity (and thus in cost) or the public good could enter utility as an additively separable term. We assume that the government can borrow or save at rate ρ. Because the income tax collected from agent reporting w is $\tau(w) = y(w) - c(w)$, the government faces the following resource constraint (RC):

$$G \leq E\left[\sum_{t=0}^{T} \rho^t \left(y(w_t) - c(w_t)\right)\right]. \tag{RC}$$

Then, the planner's problem is given by

$$\begin{aligned} &\max_{c(\cdot), y(\cdot)} \int_{W_0} E\left[\sum_{t=0}^{T} \rho^t u\left(c(w_t), \frac{y(w_t)}{w_t}\right) \Big| w_0\right] f^0(w_0) \mathrm{d}w_0 \\ &\text{s.t.} \quad \text{(RC) and (IC).} \end{aligned} \tag{11}$$

We solve for the second-best allocation rule by using optimal control theory as in Hellwig (2007). For reference, the *first-best allocation rule* maximizes the utilitarian welfare function subject to the resource constraint only, assuming that the government knows the type of each agent at each time.

[11] We note that the government is aware of the stationarity constraint, so once it chooses its allocation rule, it knows the rule cannot be changed, and accounts for this when choosing the rule.

Before proceeding, let us summarize the regularity conditions we have imposed:

Assumption 1 (Regularity conditions)**.**

1. *The type follows a first-order Markov process with conditional density* $f^t(w_{t+1}\,|\,w_t)$ *such that* W_t *and* W *are non-degenerate closed intervals in* $\mathbb{R}_{++}$*;*
2. $u(c,\ell)$ *is separable in* c *and* ℓ *as* $u(c,\ell) = \psi_c(c) - \psi_\ell(\ell)$ *where* ψ_c *and* ψ_ℓ *are strictly increasing and twice continuously differentiable;* ψ_c *and* $-\psi_\ell$ *are concave where at least one of them is strictly so. Moreover,* $\psi'_\ell(\ell)\ell$ *is strictly increasing in* ℓ*.*

Here is the key idea of our work. When we solve the problem (11), we exploit the fact that our mechanism is time-invariant and does not depend on history, and we consider a time-separable utility function and the utilitarian social welfare function. Therefore, the problem can be reduced to a static problem in which the total mass of agents is expanded to $\sum_{t=0}^{T} \rho^t$. That is, each person in each period is considered to be a different person in the static model, though the mass of agents in period t is compressed to ρ^t owing to the planner's discounting. Utilitarianism with the time-separable utility gives us the equivalence.[12] Then, we take the standard approach for static optimal income taxation problems to solve the problem (11).[13] That is, we consider a relaxed problem in which the IC constraints are replaced with weaker conditions that address only downward deviations and invoke the fact that a solution to the relaxed problem is also a solution to the original problem.

Let $\underline{w} = \min W$ and $\overline{w} = \max W$ (thus, $W = [\underline{w}, \overline{w}]$). The main properties of the planner's allocation rule are then summarized in the following proposition.

Proposition 1. *Suppose that Assumption 1 holds and that people cannot save or borrow. Then,* $\tau'(y(\overline{w})) = 0$ *and if* $y(w)$ *is strictly increasing at* $w = \underline{w}$*,* $\tau'(y(\underline{w})) = 0$*. Moreover, if* $y(w) > 0$*, then* $\tau'(y(w)) \in (0,1)$ *for any* $w \in (\underline{w}, \overline{w})$*.*

Proof. See Appendix. □

The proposition states that the marginal tax rate is zero at the top of W and if income is strictly increasing at the bottom of W, the marginal tax rate is also zero

[12]The properties of the second-best tax rules we derive is robust, while the desirability of redistribution is unaffected, to a time-independent, twice continuously differentiable, increasing, and strictly concave transformation of instantaneous utilities.

[13]Because our problem is static, the taxation principle (Hammond, 1979) implies that characterizing the direct mechanism (y, c) is equivalent to designing a tax rule τ and letting each agent choose his income y_t and consumption $c_t = y_t - \tau(y_t)$.

there. On the other hand, if income is positive, the marginal tax rate is more than 0 but less then 1 in the interior of W. One remark is that this result also holds in more general environments, as long as the allocation rule is stationary. In particular, it is not necessary that the utility function is separable and the shock is Markovian. The shock simply has to be i.i.d. across agents.

Suppose that $y(w)$ is strictly increasing at $\underline{w}$. By Proposition 1, as long as everyone works so that $y(w) > 0$ for all $w \in W$, the marginal tax rate is zero only at the top and bottom of the *expanded* type space W. Thus, being the top earner in each period generally does not imply that his marginal tax rate is zero in every period. In particular, suppose that the type space monotonically expands over time (or $\max W_t$ increases and $\min W_t$ decreases over time). For example, suppose $T = 2$, $W_0 = [3, 4]$, and $w_{t+1} = w_t + s$ where s is uniformly distributed over $[-1, 1]$ for $t \geq 0$. In that case, we have $W_1 = [2, 5]$ and $W_2 = [1, 6]$. Then, *no one's marginal tax rate is zero in the first T periods nor the last period except when the type of an individual reaches* $\overline{w} = \max W$ *or* $\underline{w} = \min W$ *in the last period*. That is, the zero top marginal rate result does not apply for periods before time T, and that the standard result applies at time T. If T is large, this never really happens.

It is useful to compare our result with history-independent *non*-stationary tax studied by Weinzierl (2011). Under no individual saving or borrowing, it is shown that, if an agent is the top earner in every period, then he faces a zero marginal tax rate in *every* period. This generally does not hold for our stationary case because $\max W$ is not necessarily attained in every period even if an agent is the top earner in every period.[14]

Moreover, it would be worth pointing out that the results above are in sharp contrast with those of Battaglini and Coate (2008) in which the shock follows a Markov chain over two states (high and low). Under their tax rule, the allocation is distorted only when people's type is currently and has always been low. That is, the allocations of agents who are currently, or have at some point been, high types are first-best. Therefore, the fraction of people whose allocations are distorted is decreasing over time whereas, in our model, the allocation is distorted anywhere in the interior of W as long as income is positive.[15] Their results crucially depend

[14] Although the result is shown for the deterministic case, Weinzierl (2011) shows that making the wage path stochastic does not change the qualitative results as long as the tax is history-independent. See the technical appendix of that paper.

[15] Whether the allocation is distorted at the top depends on the utility function. For example, if

on the following facts: the support of types is discrete and fixed over time, and the tax rule can depend on history.[16] In our model, the support of types is continuous and generally *changes* over time, and the tax rule can depend on only the current income. As a result, all people's allocations are almost surely distorted in any period.

3 Conclusion

We consider the optimal dynamic income taxation problem faced by a government that cannot change the tax rule over time. Because of the stationarity constraint, we could reduce our problem to a static one and analytically characterize the second-best tax rule. We argued that the zero top marginal tax result is of little importance in practice because it only applies to the top of expanded type space that incorporates type evolution. For example, if we consider type spaces that monotonically expand over time, we ensure a positive marginal tax rate for the top type except for the last period.[17]

Regarding the stationarity of the tax rule, we have made an extreme assumption: the government cannot make its tax rule time-dependent and thus its tax rates cannot be history-dependent *at all*. It might be more realistic to consider the situation in which the government can make its tax rule time-dependent or look at past histories at some cost.

Moreover, because we considered idiosyncratic shocks that are i.i.d. among agents, we could obtain a single resource constraint by invoking the law of large numbers. Besides the stationarity of the tax rule, this was also crucial for our results. In fact, if the agents face the common aggregate shocks, their types are correlated with each other, and the analytical approach of this paper will fail to apply. These should be subjects of future research.

Finally, although we characterized an optimal tax rule, we did not address its existence. This can probably be proved using the results of Berliant and Page (2001) for static optimal income taxes.

the utility function is quasi-linear, the allocation is first-best at the top.

[16] Note that they also do not allow for saving or borrowing.

[17] In this paper, we consider a finite-horizon model. Technically speaking, we use optimal control theory, so by replacing terminal conditions with transversality conditions, we would be able to extend Proposition 1 to an infinite-horizon model.

Appendix

Proof of Lemma 1. Suppose that the interior of $W_t \cap W_s$ is not empty for some $t, s \geq 0$ with $t \neq s$.[18] Let w be an interior point of $W_t \cap W_s$. Then, by (10),

$$u\left(c_t(w), \frac{y_t(w)}{w}\right) = u\left(c_s(w), \frac{y_s(w)}{w}\right). \tag{12}$$

Moreover, we will show that

$$\left.\frac{\mathrm{d}u\left(c_t(w), \frac{y_t(w)}{w'}\right)}{\mathrm{d}w'}\right|_{w'=w} = \left.\frac{\mathrm{d}u\left(c_s(w), \frac{y_s(w)}{w'}\right)}{\mathrm{d}w'}\right|_{w'=w}. \tag{13}$$

To see this, note that

$$\left.\frac{\mathrm{d}u\left(c_t(w'), \frac{y_t(w')}{w'}\right)}{\mathrm{d}w'}\right|_{w'=w} = \left.\frac{\mathrm{d}u\left(c_t(w), \frac{y_t(w)}{w'}\right)}{\mathrm{d}w'}\right|_{w'=w} + \left.\frac{\mathrm{d}u\left(c_t(w'), \frac{y_t(w')}{w}\right)}{\mathrm{d}w'}\right|_{w'=w} = \left.\frac{\mathrm{d}u\left(c_t(w), \frac{y_t(w)}{w'}\right)}{\mathrm{d}w'}\right|_{w'=w}, \tag{14}$$

because $\left.\frac{\mathrm{d}u\left(c_t(w'), \frac{y_t(w')}{w}\right)}{\mathrm{d}w'}\right|_{w'=w} = 0$ by the IC constraint. Suppose (13) does not hold. Then,

$$\left.\frac{\mathrm{d}u\left(c_t(w'), \frac{y_t(w')}{w'}\right)}{\mathrm{d}w'}\right|_{w'=w} \neq \left.\frac{\mathrm{d}u\left(c_s(w), \frac{y_s(w)}{w'}\right)}{\mathrm{d}w'}\right|_{w'=w}. \tag{15}$$

Then, because (12) holds and w is in the interior of $W_t \cap W_s$, we can take $w' \in W_t \cap W_s$ such that

$$u\left(c_s(w), \frac{y_s(w)}{w'}\right) > u\left(c_t(w'), \frac{y_t(w')}{w'}\right). \tag{16}$$

However, because $w' \in W_t \cap W_s \subseteq W_t$ and $w \in W_t \cap W_s \subseteq W_s$, $(y_t(\cdot), c_t(\cdot))$ is not incentive compatible by (10).

Thus, we have

$$\psi'_\ell\left(\frac{y_t(w)}{w}\right)\frac{y_t(w)}{w} = \psi'_\ell\left(\frac{y_s(w)}{w}\right)\frac{y_s(w)}{w}. \tag{17}$$

Hence, we have $y_t(w) = y_s(w)$ if $\psi'_\ell(\ell)\ell$ is strictly increasing in ℓ. Note that, if $y_t(w) = y_s(w)$, we must have $c_t(w) = c_s(w)$ by the stationarity of τ. □

[18] Because $W = \bigcup_{t=0}^{T} W_t$ is assumed to be a closed interval, $W_t \cap W_s = \emptyset$ for any $t, s \geq 0$ with $t \neq s$ is not allowed. If the interior of $W_t \cap W_s$ is empty for any $t, s \geq 0$ with $t \neq s$, then we immediately attain the result.

Proof of Proposition 1. We show that, due to the stationarity constraint, our problem can be reduced to a static problem and then invoke the results of Hellwig (2007) who analyzes a static optimal taxation problem under the utilitarian welfare function. As in Hellwig (2007), we consider a relaxed problem by replacing the IC constraint with a weaker condition that is called the *downward IC constraint*:

$$\forall w \in W,\ u\left(c(w), \frac{y(w)}{w}\right) \geq u\left(c(w'), \frac{y(w')}{w}\right) \text{ for all } w' \in \{\tilde{w} \in W : \tilde{w} \leq w\}\,. \qquad \text{(IC}'\text{)}$$

Thus, the downward IC constraint takes care of only downward deviations. By Lemma 6.2 of Hellwig (2007), $(y(\cdot), c(\cdot))$ with nondecreasing $c(\cdot)$ satisfies (IC′) if and only if $\frac{\mathrm{d}u(c(w),y(w)/w)}{\mathrm{d}w} \geq \frac{\mathrm{d}u(c(w),y(w)/w')}{\mathrm{d}w'}|_{w'=w}$ for all $w \in W$. Thus, when we solve the problem, we impose the constraints that $c(w)$ is nondecreasing and $\frac{\mathrm{d}u(c(w),y(w)/w)}{\mathrm{d}w} \geq \frac{\mathrm{d}u(c(w),y(w)/w')}{\mathrm{d}w'}|_{w'=w}$ on W instead of the downward IC constraint.

Next, we rewrite the welfare function as

$$\int_{W_0} E\left[\sum_{t=0}^{T} \rho^t u\left(c(w_t), \frac{y(w_t)}{w_t}\right) \Big| w_0\right] f^0(w_0)\mathrm{d}w_0 = \sum_{t=0}^{T} \rho^t \int_{W_t} u\left(c(w_t), \frac{y(w_t)}{w_t}\right) f^t(w_t)\mathrm{d}w_t$$

where $f^t(w_t) = \int_{W^{t-1}} f^t(w_t \,|\, w_{t-1}) f^{t-1}(w_{t-1} \,|\, w_{t-2}) \cdots f^0(w_0)\mathrm{d}w_{t-1}\mathrm{d}w_{t-2} \cdots \mathrm{d}w_0$. Let $\overline{f^t}$ be an extension of f^t to W (i.e., $\overline{f^t}(w) = f^t(w)$ on W_t and $\overline{f^t}(w) = 0$ on $W \setminus W_t$). Then, the above expression reduces to

$$\int_{W} u\left(c(w), \frac{y(w)}{w}\right) g(w)\mathrm{d}w \qquad (18)$$

where $g(w) \equiv \sum_{t=0}^{T} \rho^t \overline{f^t}(w)$. Likewise, the resource constraint is reduced to

$$G \leq \int_{W} \left(y(w) - c(w)\right) g(w)\mathrm{d}w. \qquad (19)$$

Therefore, our relaxed problem is given by

$$\begin{aligned} &\max_{x(\cdot)} \int_W u\left(c(w), \frac{y(w)}{w}\right) g(w)\mathrm{d}w \\ &\text{s.t.} \quad G \le \int_W \left(y(w) - c(w)\right) g(w)\mathrm{d}w, \\ &\qquad c(w) \text{ is nondecreasing and } \tfrac{\mathrm{d}u(c(w),y(w)/w)}{\mathrm{d}w} \ge \tfrac{\mathrm{d}u(c(w),y(w)/w')}{\mathrm{d}w'}|_{w'=w} \text{ on } W. \end{aligned} \tag{20}$$

On the other hand, Hellwig (2007) considers a standard static optimal taxation problem. Specifically, in our notation, his problem is written as

$$\begin{aligned} &\max_{x(\cdot)} \int_{W_0} u\left(c(w), \frac{y(w)}{w}\right) f^0(w)\mathrm{d}w \\ &\text{s.t.} \quad G \le \int_{W_0} \left(y(w) - c(w)\right) f^0(w)\mathrm{d}w, \\ &\qquad c(w) \text{ is nondecreasing and } \tfrac{\mathrm{d}u(c(w),y(w)/w)}{\mathrm{d}w} \ge \tfrac{\mathrm{d}u(c(w),y(w)/w')}{\mathrm{d}w'}|_{w'=w} \text{ on } W_0. \end{aligned} \tag{21}$$

Hence, we can see that our problem can be viewed as a static problem in which the total mass of agents is $\sum_{t=0}^{T} \rho^t$, the support of type distribution is W, and the welfare weight for type w is $g(w)$, and therefore, the arguments of Hellwig (2007) directly apply. In particular, our claim follows from Theorems 4.1 and 6.1 of Hellwig (2007). □

References

Albanesi, S. and C. Sleet. (2006), "Dynamic optimal taxation with private information," *Review of Economic Studies*, **73**, 1-30.

Battaglini, M. and S. Coate. (2008), "Pareto efficient income taxation with stochastic abilities," *Journal of Public Economics*, **92**, 844-868.

Berliant, M. and J. Ledyard. (2014), "Optimal dynamic nonlinear income taxes with no commitment," *Journal of Public Economic Theory*, **16**, 196-221.

Berliant, M. and F. H. Page. (2001), "Income taxes and the provision of public goods: existence of an optimum," *Econometrica*, **69**, 771-784.

Diamand, P. (1998), "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rate," *American Economic Review*, **88**, 83-95.

Diamond, P. and E. Saez. (2011), "The case for a progressive tax: From basic research to policy recommendation," *Journal of Economic Perspectives*, **25**, 165-190.

Farhi, E. and I. Werning. (2013), "Insurance and taxation over the life cycle," *Review of Economic Studies*, **80**, 596-635.

Fernandes and Phelan (2000), "A recursive formulation for repeated agency with history dependence," *Journal of Economic Theory*, **91**, 223-247.

Gaube, T. (2010), "Taxation of annual income as a commitment device," Unpublished manuscript.

Golosov, M., A. Shourideh., M. Troshkin., and A. Tsyvinski. (2013), "Optimal pension systems with simple instruments," *American Economic Review: Papers & Proceedings*, **103**, 502-507.

Hammond, P. (1979), "Straightforward individual incentive compatibility in large economies," *Review of Economic Studies*, **46**, 263-282.

Hellwig, M. F. (2007), "A contribution to the theory of optimal utilitarian income taxation," *Journal of Public Economics*, **91**, 1449-1477.

Kocherlakota, N. (2010), *The New Dynamic Public Finance*, Princeton University Press.

Kocherlakota, N. (2005), "Zero expected wealth taxes: A Mirrlees approach to dynamic optimal taxation," *Econometrica*, **73**, 1587-1621.

Mirrlees, J. A. (1971), "An exploration in the theory of optimal income taxation," *Review of Economic Studies*, **38**, 175-208.

Tuomala, M. (1990), *Optimal Income Tax and Redistribution*, Clarendon Press.

Weinzierl, M. (2011), "The surprising power of age-dependent taxes," *Review of Economic Studies*, **78**, 1490-1518.

Weymark, J. A. (1987), "Comparative static properties of optimal nonlinear income taxes," *Econometrica*, **55**, 1165-1185.