MPRA
Munich Personal RePEc Archive

Optimal Taxation over the Life Cycle

Gorry, Aspen and Oberfield, Ezra

UC Santa Cruz, University of Chicago

26 May 2010

Online at https://mpra.ub.uni-muenchen.de/25297/
MPRA Paper No. 25297, posted 23 Sep 2010 15:23 UTC

Optimal Taxation over the Life Cycle*

Aspen Gorry

University of California, Santa Cruz

aspen@ucsc.edu

Ezra Oberfield

University of Chicago

ezrao@uchicago.edu

May 26, 2010

Abstract

We derive the optimal labor income tax schedule for a life cycle model with deterministic productivity variation and complete asset markets. An individual chooses whether and how much to work at each date. The government must finance a given expenditure and does not have access to lump sum taxation. We develop a solution method that combines incentive constraints into a single implementability constraint. The average tax rate determines when an individual will work while the marginal tax rate determines how much she will work. The optimal tax schedule has an increasing average tax rate at low levels of income to encourage labor market participation, even in the absence of redistributive concerns. In contrast to the Mirrleesian optimal taxation literature, the marginal tax rate at the top is strictly positive.

*Preliminary, comments welcome. We thank participants of the 2008 North American Summer Meetings of the Econometric Society, the 2009 SED Summer Meetings, the 2009 Midwest Macro meetings, and the Capital Theory Working Group at the University of Chicago for helpful comments. First version: September 25, 2007. All mistakes are our own.

1 Introduction

Individuals have substantial heterogeneity in earnings over their lifetimes. Estimates of life cycle wage patterns show that wages approximately double during a worker's life and individuals regularly shift between tax brackets[1]. Much of the theoretical work on optimal taxation treats individual productivity as static. These frameworks do not address important intertemporal trade-offs in individuals' labor market decisions. In reality, individual consumption, savings, and work decisions reflect both current and future income. Labor supply decisions at a given point in time are influenced by tax rates in several brackets. As a consequence, life cycle variation in earnings links segments of the tax schedule through both incentive constraints and individuals' lifetime budget constraints. Given these considerations it is important to understand how life cycle consumption and labor supply dynamics influence optimal tax policy.

This paper extends the existing tax literature by solving for the optimal nonlinear labor income tax schedule in a simple life cycle labor supply model. We model an individual who faces an exogenous productivity profile and a fixed cost of going to work. Fixed costs generate realistic patterns of labor supply in which an individual works only a fraction of her life. With hump shaped productivity, the model captures the basic pattern of lifetime labor supply: a period of non-participation followed by working and finally retirement.

In our model, labor taxes distort both the extensive and intensive margins of work. The average tax rate determines whether an individual works on a given date, while the marginal tax rate determines how much labor is supplied. This contrasts with the much of the optimal taxation

[1] Fullerton and Rogers (1991) argue that lifetime and annual income classifications are significantly different in their implications for tax incidence. In particular, using PSID data they compare annual income from 1984 to lifetime income for individuals by decile. They find that only 24.8% of individuals are in the same annual as lifetime decile in that given year and that only 56.1% are within one step of their lifetime decile.

literature where the marginal tax rate is the central object of interest[2].

Solving the optimal tax problem in the life cycle model generates two key results. First, the average tax rate is increasing at the bottom of the income distribution. It is optimal to have a low and rising average tax rate at the extensive margin in order to induce workers to enter the labor force earlier and retire later.

Second, the marginal tax rate at the highest realized income is strictly positive. This result stands in contrast to the optimal tax literature that follows Mirrlees (1971). While there is no redistributive motive, there is a close relationship between the Mirrleesian framework and this one. The Mirrleesian literature finds that the optimal marginal tax rate for the highest skilled individual is zero when the skill distribution is bounded (see Diamond (1998)). To understand the logic, suppose that the marginal tax at the highest income was not zero. Then if the tax schedule was extended beyond the top at a marginal rate of zero, the highest earner would choose to work and consume more. She would be better off while no tax revenue would be lost for the government. Therefore, the original tax schedule could not have been optimal. In a life cycle model this logic breaks down. Extending the tax function at a zero marginal rate still induces the worker to work more when she is most productive, but the increase in wealth induces the worker to work less at other points in her life, diminishing tax revenue.

These results are robust as long as key assumptions are satisfied. For the average tax rate to be increasing at the bottom, it is crucial that there is an active extensive margin. This is true as long as there are fixed costs of supplying labor. To generate a positive marginal tax rate at the top it is critical that individuals face a lifetime budget constraint. With any consumption smoothing,

[2] An important exception is Saez (2002), who studied the effect of the extensive margin on the bottom of the tax schedule in a static economy to contrast the earned income tax credit with a negative income tax. Diamond (1980) and Mulligan (2001) look at optimal tax policy focusing exclusively on the extensive margin.

the rate at the top distorts individuals' labor supply decisions over the life cycle. In particular, the positive marginal tax at the top is robust to adding other sources of wage heterogeneity as long as the most productive workers have income variation over their lifetime.

We calibrate the model to match stylized features of the US economy and compute the welfare gain from the optimal tax function relative to a flat tax. The optimal tax policy increases welfare by between 0.7% and 1.4%.

In addition to these results, this paper makes a methodological contribution. We show how to combine all incentive constraints into a single implementability constraint, greatly simplifying the characterization of the optimal tax schedule. Our approach shares features with the first order approach and, as is common when the first order approach is used, there is no guarantee that the constraint set of the planning problem is convex. We show that if the tax schedule described by the necessary conditions of the planning problem has the additional feature that the total tax paid is increasing in income, then the tax schedule is indeed an optimum.

Following Saez (2002), we study both the intensive and extensive margins of labor supply adjustment. Saez (2002) develops a simple expression relating the optimal tax rate to empirical local labor supply elasticities. Our study differs in that we explicitly motivate the agent's decisions at the extensive margin by modeling changes in productivity over the life cycle. Labor supply decisions depend on the entire tax function and conversely a local change in the tax schedule influences labor supply decisions over a range of wages. As a consequence, local elasticities cannot capture the full labor supply response to a local change in the tax schedule. We develop a simple expression for the optimal tax rate that incorporates wealth effects along with the relevant elasticities.

The problem posed here shares elements with two classic approaches to the problem of optimal taxation. Following Ramsey (1927) and Lucas and Stokey (1983), we focus on efficient collection of

tax revenue from an individual. We examine how a government should finance a given expenditure using only labor income taxes and constrain the tax policy so individuals can only be taxed when they earn income, ruling out lump sum taxation. Following Mirrlees (1971), we allow for a nonlinear tax schedule so that marginal and average tax rates can vary with income.

The recent dynamic Mirrlees literature (see for example Golosov et al. (2003), Kocherlakota (2005), and Golosov et al. (2006)) examines a related problem. By conditioning on past as well as current income, the government can solve the complete intertemporal hidden information problem. While these papers show that these policies can be extremely powerful, no current tax system comes close to them[3]. We do not take a stand on whether or not the tax system should have memory, but instead explore the shape of the optimal tax schedule when the government is constrained to condition taxes only on current income.

In related work, Weinzierl (2007) focuses on the partial reform of conditioning the tax schedule on both income and age. In showing that age dependent taxes are a powerful reform he focuses on heterogenous individuals. In contrast our model focuses on a single agent who has variable productivity over the life cycle and faces a constant fixed cost of working at each date [4]. This allows us to abstract away from redistributive concerns that are present in the traditional framework. Methodologically, he uses the traditional Mirrleesian approach of solving the model with a sequence of incentive compatibility constraints. Our approach is novel in that we reduce the problem to a single implementability constraint.

This paper compliments a literature that looks at capital taxation in a life cycle model. Recent

[3] Grochulski and Kocherlakota (2007) argue that the dynamic approach can be implemented with a flat tax on current income and history contingent payment after retirement that resembles social security.

[4] Age dependent taxes are trivial in our framework as the life cycle productivity profile that we use is indexed by a worker's age. Allowing for age dependent taxes, the government can in effect impose taxation lump-sum, eliminating all distortions.

papers in this literature including Erosa and Gervais (2002) and Mathieu-Bolh (2008) show that in life cycle models capital taxes can be positive. Specifically, Erosa and Gervais (2002) show that in an OLG framework capital taxes can be positive in equilibrium because the individual's optimal consumption-liesure plan is not constant. Mathieu-Bolh (2008) extends this framework to consider borrowing constraints and changes in productivity over the life cycle. With these extensions, it is shown that taxation of capital is optimal and that optimal taxation involves some redistribution to the poor. We focus on the effects of the life cycle on optimal labor taxation in a model without capital.

Section 2 presents the model and defines an equilibrium. Section 3 summarizes properties of the agent's decision problem that are used in Section 4 to solve the optimal tax problem. Section 5 presents some numerical examples and Section 6 discusses how the results are robust to many extensions of the model. Section 7 concludes.

2 Model

We consider the partial equilibrium problem of a single agent who faces an invariant labor income tax schedule. The government must choose a tax schedule to raise an exogenous amount over her life. For simplicity we examine a model with no discounting and a zero interest rate[5]. The agent lives for one unit of time, decides how much to work at each date, and faces a lifetime budget constraint.

[5] Analogous results can be obtained in a model with discounting and interest rates by appropriately scaling objects of interest, as discussed in section 6.1. Section 6.2 sets the model in general equilibrium.

Preferences of the agent are given by:

$$\int_0^1 [u(c(a)) - v(h(a))]\, da$$

where $c(a)$ is consumption at age a and $h(a)$ is hours worked at age a. We assume that u is twice continuously differentiable and strictly concave. Additionally, let v be given by:

$$v(h) = \tilde{v}(h) + \chi \mathbb{I}_{h>0}$$

We assume that $\tilde{v}$ is twice continuously differentiable and strictly convex. χ denotes the fixed utility cost of working at any given date. The fixed cost could represent the cost of commuting, waking up early, or making coffee[6]. This nonconvexity in the production technology is crucial for generating an extensive margin of labor adjustment. In equilibrium a worker finds it optimal to avoid this fixed cost by not working on dates when productivity is low. Consequently, the agent works only a fraction of her life. With a single peaked productivity profile, this corresponds to a labor market entry and retirement decision.

Output is produced using only labor. Hours worked by an individual of age a, $h(a)$, translate into output as follows:

$$y(a) = w(a)h(a)$$

$w(a)$ represents a worker's hourly wage at age a and varies exogenously and deterministically over an individual's lifetime.

[6]The fixed utility cost could also be written as a fixed time cost as in Rogerson and Wallenius (2007), which might be a more appropriate representation of a commuting cost. The fixed utility cost is slightly more tractable, and produces little difference in the qualitative or quantitative predictions of the model.

The government must finance a given expenditure, G, by picking a single nonlinear tax schedule that the individual will face over her life. $\tau(y)$ is the *total* tax paid at a date where the worker has income y. By assumption, the tax schedule can depend only on income at each date and is subject to the constraint $\tau(0) \leq 0$. The government budget constraint is given by:

$$G \leq \int_0^1 \tau(y(a))da \tag{1}$$

Given the tax function chosen by the government, the individual chooses an allocation of consumption and hours over her lifetime to solve:

$$\max_{\{c(a),h(a)\}_{a\in[0,1]}} \int_0^1 \left[u(c(a)) - v(h(a))\right] da \tag{2}$$

subject to her lifetime budget constraint:

$$\int_0^1 \left[y(a) - \tau(y(a)) - c(a)\right] da \geq 0 \tag{3}$$

In writing the budget constraint we implicitly assume that there are perfect capital markets and no binding borrowing constraints; the individual is able to perfectly smooth consumption over her lifetime.

An equilibrium is an allocation $\{c(a), h(a)\}_{a\in[0,1]}$ and a governement policy $\tau(\cdot)$ such that:

1. Given the government policy, the allocation solves the individual's optimization problem described by equations (2) and (3).

2. The government satisfies its budget constraint.

3. The allocation is feasible.

The government's optimal tax problem is to select the equilibrium that maximizes the agent's welfare.

3 Properties of the Agent's Decision Problem

Because the interest rate is equal to the discount rate the individual will choose to perfectly smooth consumption over her life. Additionally, the agent's labor allocation at a given age depends only on the wage she faces at that age. If she faces the same wage at two different ages during her life, she will choose the same labor allocation at those dates. Let $F(\cdot)$ be the CDF of wages over the agent's lifetime with support on $[\underline{w}, \bar{w}]$. We assume F is twice continuously differentiable. The individual's problem can be written as:

$$\max_{c,\{h(w)\}_{w=\underline{w}}^{\bar{w}}} u\left(c\right) - \int_{\underline{w}}^{\bar{w}} v\left(h\left(w\right)\right) dF(w) \tag{4}$$

$$\text{subject to: } \int_{\underline{w}}^{\bar{w}} \left[y\left(w\right) - \tau(y(w))\right] dF(w) - c \geq 0 \tag{5}$$

The government budget constraint is:

$$G \leq \int_{\underline{w}}^{\bar{w}} \tau\left(y\left(w\right)\right) dF(w) \tag{6}$$

The propositions below describe the properties of the allocation for any tax function[7]. To solve the problem, note that for any tax schedule the solution to the individual's labor supply problem

[7] The proofs of all propositions in this section are contained in the appendix.

has a reservation property: she will only choose to work if her wage is above a threshold. This is formalized in the following proposition:

Proposition 1 *In any equilibrium, there exists a wage level w^* such that $w > w^*$ implies $h(w) > 0$ and $w < w^*$ implies $h(w) = 0$.*

We restrict attention to the set of tax functions for which the marginal tax rate is not more than 100%. Proposition 2 shows that this restriction is without loss of generality:

Proposition 2 *Any allocation that results from an arbitrary tax function can also be the result of a tax function for which after-tax income is weakly increasing in pretax income.*

The formal proof of Proposition 2 borrows heavily from Mirrlees (1971). The intuition behind the result is simple: if the marginal tax rate is ever above 100%, no one will choose to supply labor in the region where the marginal tax paid is larger than the marginal income earned. Given that no one chooses to supply labor in the given region, a tax function with marginal tax rates weakly less than 100% can be constructed that yields an identical allocation.

Next we show that output is weakly increasing in the wage:

Proposition 3 *In any equilibrium, $y(w) = wh(w)$ is weakly increasing in w.*

Proposition 3 follows from Proposition 2 and individual preferences. Since the marginal tax rate is less than 100%, a more productive individual can earn the same amount as a less productive individual with less effort. Therefore, output must be weakly increasing in the wage.

Lastly we derive an implementability constraint in the style of Lucas and Stokey (1983). Let $\frac{1}{\eta(w)} \equiv \frac{wF'(w)}{1-F(w)}$ be the elasticity of the counter-cumulative wage distribution.

Proposition 4 *In any equilibrium, the agent chooses an allocation that satisfies*

$$u'(c)c - \int_{w^*}^{\bar{w}} \left[v(h(w)) + \eta(w)v'(h(w))h(w)\right] dF(w) \geq 0$$

4 Solving the Optimal Tax Problem

We now turn to the government's problem, which can be written as

$$\max_{\tau(\cdot),c,w^*,\{h(w)\}_{w=w^*}^{\bar{w}}} u(c) - \int_{w^*}^{\bar{w}} v(h(w))dF(w)$$

subject to:

$$c, w^*, \{h(w)\}_{w=w^*}^{\bar{w}} \text{ are optimal for the agent given } \tau(\cdot)$$

$$G \leq \int_{w^*}^{\bar{w}} \tau(y(w))dF(w)$$

$$c + G \leq \int_{w^*}^{\bar{w}} wh(w)dF(w)$$

We replace the individual's optimality conditions and the government budget constraint with the implementability condition to write down a relaxed version of the government's problem. The government maximizes the individual's lifetime utility subject to the implementability condition and feasibility:

$$\max_{c,w^*,\{h(w)\}_{w=w^*}^{\bar{w}}} u\left(c\right) - \int_{w^*}^{\bar{w}} v(h(w))dF(w)$$

subject to:

$$u'(c)c - \int_{w^*}^{\bar{w}} \big(v(h(w)) + \eta(w)v'(h(w))h(w)\big)\, dF(w) \geq 0$$

$$c + G \leq \int_{w^*}^{\bar{w}} wh(w)dF(w)$$

This is a relaxed version of the government's problem because it is possible that there will not be a tax schedule that can induce the agent to choose the resulting allocation. The set of allocations that satisfy the implementability constraint is larger than the set of equilibrium allocations, so the implementability constraint is necessary but not sufficient condition for equilibrium. In particular, if the marginal tax rate is declining, after-tax income would be a convex function of income, so the agent's optimization problem may not be convex.

We first solve the relaxed government's problem and then examine when this solution will be optimal. Using θ and μ as the multipliers for the implementability and feasibility constraints respectively, the first order conditions are:

$$\begin{aligned}
c &: \quad u'(c) + \theta[u''(c)c + u'(c)] = \mu \\
h(w) &: \quad v'(h) + \theta[v'(h) + \eta(w)v''(h)h + \eta(w)v'(h)] = \mu w \\
w^* &: \quad v(h(w^*)) + \theta[v(h(w^*)) + \eta(w^*)v'(h(w^*))h(w^*)] = \mu w^* h(w^*)
\end{aligned}$$

Arguments of the functions are omitted for simplicity where they are clear.

The differentiability of the optimal tax function is given in the following proposition:

Proposition 5 *The optimal tax function, $\tau(y)$, is differentiable on $y \in (y(w^*), y(\bar{w}))$.*

To check that the allocation from the relaxed problem is in fact the solution to the government's

problem, we must verify that the resulting allocation is optimal for the agent given the tax policy.

Proposition 6 *The solution to the relaxed problem can be implemented if $y(w)$ is strictly increasing on $[w^*, \bar{w}]$. If $y(w)$ is not weakly increasing, it cannot be implemented.*

To verify that $y(w)$ is strictly increasing we can numerically evaluate $y(w)$ after solving the problem. Alternatively, we can make fairly weak assumptions on the distribution of wages and the disutility of income. Conditions sufficient for $y(w)$ to be strictly increasing include a restriction on the distribution of wages and standard assumptions on the elasticity of labor supply[8]. For the remainder of the paper we assume that these conditions are satisfied.

The first results about the optimal tax function are that the average tax rate at the bottom of the income distribution determines the individual's reservation wage while the marginal tax rate determines how much the individual works at each wage. These results follow from the worker's

[8]To find conditions under which y is increasing we examine the first order condition of the relaxed problem with respect to $h(w)$:

$$v'(h)\left(1+\frac{\theta}{1+\theta}\eta(w)\left[1+\frac{v''(h)}{v'(h)}h\right]\right)=\frac{\mu}{1+\theta}w$$

where $\eta(w)$ is defined as $\eta(w)=\frac{1-F(w)}{wF'(w)}$. Let $\gamma(h)\equiv\frac{v''(h)h}{v'(h)}$. For further calculations all arguments will be supressed. Taking logs, differentiating completely with respect to w, and solving for $h'(w)$ gives:

$$h'=\frac{1}{w}\frac{\frac{1+\theta}{\theta}\frac{1}{(1+\gamma)\eta}+1-\frac{w\eta'}{\eta}}{\frac{\gamma}{h}\left(\frac{1+\theta}{\theta}\frac{1}{(1+\gamma)\eta}+1+\frac{h}{\gamma}\frac{\gamma'}{1+\gamma}\right)}$$

Since $y(w)=wh(w)$, we can write:

$$y'=h+wh'=h+\frac{\frac{1+\theta}{\theta}\frac{1}{(1+\gamma)\eta}+1-\frac{w\eta'}{\eta}}{\frac{\gamma}{h}\left(\frac{1+\theta}{\theta}\frac{1}{(1+\gamma)\eta}+1+\frac{h}{\gamma}\frac{\gamma'}{1+\gamma}\right)}$$

The conditions $\frac{w\eta'}{\eta}\leq 1$ and $\frac{h}{\gamma}\frac{\gamma'}{1+\gamma}\geq -1$ are therefore more restrictive than needed for $\hat{y}$ to be an increasing function. To see if the first condition is reasonable, consider the even more restrictive condition $\eta'(w)\leq 0, \forall w\in(w^*,w)$. First, note that $\eta(w)$ is positive for most of its range while $\eta(\bar{w})=0$, so η must be decreasing on average. Moreover, $\frac{1-F(w)}{w}$ is strictly decreasing, so for η' to be positive, F' would have to be growing fast enough to outweigh the decline in $\frac{1-F(w)}{w}$. So, this condition shouldn't be too restrictive. It is satisfied for parabolic wage-age profile which we use in the numerical computations. For the second condition a more restrictive assumption is that $\gamma'\geq 0$. This is satisfied for models with a constant Frisch elasticity which is a common assumption in the literature.

first order conditions:

$$\frac{v'(h)}{u'(c)} = \left(1-\tau'(y)\right)w \tag{7}$$

$$\frac{v(h)}{u'(c)} = \left(1-\frac{\tau(y(w^*))}{y(w^*)}\right)y(w^*) \tag{8}$$

The first equation is the standard first order condition for an individual's intratemporal labor supply choice. The individual equates the disutility from the marginal hour worked to the marginal after tax return for time spent working. Here the individual's choice of hours is a function of the marginal tax rate $\tau'(y)$.

The second equation determines the reservation wage. At the extensive margin, the agent equates the disutility from working the optimal hour for a given day to the total after-tax income from working the additional day. Here, the decision to work is a function of the average tax rate on income earned at the extensive margin.

The marginal and average tax rates can be written as:

$$\begin{aligned}\tau'(y) &= 1-\frac{1}{wu'(c)}v'(h)\\ \frac{\tau(y(w^*))}{y(w^*)} &= 1-\frac{1}{wu'(c)}\frac{v(h)}{h}\end{aligned}$$

These equations will be helpful in proving the following propositions:

Proposition 7 *The average tax rate is increasing at the extensive margin.*

Proof. The first order conditions of the government's problem with respect to $h(w)$ and w^* can

be re-written as:

$$
\begin{aligned}
h(w) \quad &: \quad v' = \frac{\mu}{1+\theta} w - \frac{\theta}{1+\theta} \eta [v' + v'' h] \\
w^* \quad &: \quad \frac{v}{h} = \frac{\mu}{1+\theta} w - \frac{\theta}{1+\theta} \eta v'
\end{aligned}
$$

Next we will show that at the extensive margin the marginal tax is greater than the average tax:

$$
\begin{aligned}
\left(\tau'(y) - \frac{\tau(y)}{y} \right) \Big|_{y^*} &= \frac{1}{wu'(c)} \left[\frac{v(h)}{h} - v'(h) \right] \\
&= \frac{1}{wu'(c)} \frac{\theta}{1+\theta} \eta v'' h > 0
\end{aligned}
$$

Differentiating the average tax rate and evaluating at y^* gives:

$$
\frac{d}{dy} \left(\frac{\tau(y)}{y} \right) \Big|_{y^*} = \frac{1}{y} \left(\tau'(y) - \frac{\tau(y)}{y} \right) \Big|_{y^*} > 0
$$

■

The fixed utility cost of working χ drives this result. It implies that a worker at the extensive margin works a strictly positive number of hours. Since the worker is indifferent to working additional dates at the extensive margin while the government gets sizable tax revenue from the entire income earned by working, it is optimal for the government to lower the average tax rate at the extensive margin to induce the individual to work a greater fraction of her life.

Proposition 8 *The marginal tax is positive at all realized levels of income if utility is strictly concave in consumption. The marginal tax attains a minimum at $w = \bar{w}$.*

Proof. As shown in the appendix the marginal tax rate can be written as:

$$\tau'(y) = \frac{\sigma(c) + \eta(w)[1 + \gamma(h)]}{\frac{1+\theta}{\theta} + \eta(w)[1 + \gamma(h)]} > 0 \tag{9}$$

where $\sigma(c) \equiv -\frac{cu''}{u'}$ and $\gamma(h) \equiv \frac{v''h}{v'}$.

As shown in Proposition 2, $\tau' \leq 1$, so τ' is increasing in the quantity $\eta(w)[1 + \gamma(h)]$. Since $\eta(w)|_{w=\bar{w}} = 0$, τ' attains a minimum at $w = \bar{w}$. The marginal tax at the top is:

$$\tau'(y)\big|_{w=\bar{w}} = \frac{\theta}{1+\theta}\sigma(c) > 0 \tag{10}$$

■

Instead of the well-known zero marginal tax at the top of the income distribution result, the marginal tax attains its minimum at the top, but this minimum is positive. In optimal tax models without life cycle productivity variation, as in Mirrlees (1976), the marginal tax rate at any income has the same sign as the multiplier on the incentive compatibility constraint of a worker at that income. This is positive for workers in the interior and zero for the worker at the top. Here, instead of a sequence of incentive compatibility constraints as in the Mirrlees problem, we have a single implementability constraint (which corresponds to a lifetime budget constraint) with a positive multiplier. This multiplier determines the marginal tax at the top and the minimum marginal tax across all realized levels of income.

Equation (9) describes the determinants of the marginal tax rate. As shown in the proposition, $\tau'(y)$ is increasing in $\eta(w)[1+\gamma(h)]$. $\gamma(h)$ is the inverse of the elasticity of labor supply and $\eta(w) = (1-F(w))/wF'(w)$ is the inverse of the elasticity of the counter-cumulative wage distribution. These

objects are familiar from past studies of optimal taxation. The inverse elasticity of labor supply is related to Ramsey models of taxation in which the optimal distortion on a good is inversely related to the elasticity of deadweight loss. In the Mirrleesian framework the counter-cumulative wage distribution relates a change in the marginal tax rate at one income level to changes in the total tax and incentive constraints at all higher income levels. Given the features of both the Ramsey and Mirrleesian framework in our model, both are important determinants of the tax schedule.

A new object is the inverse intertemporal elasticity of substitution, $\sigma(c)$. As a benchmark, it is useful to consider a quasilinear utility function, $c - \int_0^1 v(h(a))da$. When the utility of consumption is linear $\sigma(c) = 0$. In this case, the marginal tax at the top is also zero. This specification brings the model back to the standard Mirrleesian framework. Positive marginal taxes at the top arise as we depart from the quasilinear specification because additional earnings generate a wealth effect, reducing the marginal utility of consumption. This induces the agent to work fewer hours at all wages. Indeed, equation (10) shows that $\sigma(c)$ has a direct effect on the tax at the top. In the Mirrleesian model, there are no wealth effects across wages: more income for the top worker has no effect on the decisions of workers at lower wages. A crucial difference between this model and the Mirrleesian model are these wealth effects.

5 Numerical Examples

In this section of the paper the model is calibrated to match salient features of the United States economy. With these calibrated parameters we solve for the optimal tax functions and make welfare comparisons to a model that uses a flat tax of 30%. We see these examples as instructive to show the qualitative features of optimal tax functions when life cycle features are taken into account.

5.1 Calibration

We pick simple functional forms that exhibit key features of the model. The functions for utility and technological parameters are as follows:

$$u(c) = \log(c)$$

$$v(h) = \alpha \frac{h^{1+\gamma}}{1+\gamma} + \chi \mathbb{I}_{h>0}$$

$$w(a) = \bar{w} - (\bar{w} - \underline{w})(2a-1)^2$$

The choices of utility functions are common and consistent with balanced growth preferences. χ is a fixed cost to starting work at each date. The size of the fixed cost is chosen to induce an equilibrium fraction of life spent working by individuals in the model. The wage profile by age is chosen to be a quadratic function that is symmetric over the first and second half of the individual's life. While not matching all features of observed wage age profiles, it captures the initial rise and later decline of wages.

We calibrate the model for three different values of γ: .5, 1, and 3. This covers a broad range of potential labor supply elasticities. In every case, given the assumption of log utility, $\bar{w}$ can be normalized to one without loss of generality. The time endowment is also normalized to one in each period. The remaining parameters α, χ, and $\underline{w}$ are chosen to match features of the US economy for each γ.

Following Rogerson and Wallenius (2007), we calibrate the model to a constant average tax rate in the US of 30%. Individuals work about $\frac{2}{3}$ of their lives, which this implies $1-F(w^*) = \frac{2}{3}$. Second, we calibrate the wage profile so that highest realized hourly wage is twice the lowest realized wage.

Lastly, because we have assumed that disutility of labor $\tilde{v}$ has constant elasticity, we can choose units of h so that the agent works 1 unit of time at the highest wage. The calibrated values of α, χ, and $\underline{w}$ for each value of γ are reported in Table 1.

γ	0.5	1	3
α	2.36	2.08	1.89
χ	0.10	0.26	0.56
$\underline{w}$	-0.11	-0.11	-0.11

Table 1: Calibrated values of α, χ, and $\underline{w}$ for $\gamma = .5$, 1, and 3.

5.2 Optimal Tax Functions

For each calibration of the model the optimal tax functions are computed holding government revenues the same as they were for the calibrated case with a flat tax of 30%. Since all of the figures are similar, we show the graphs for the case of $\gamma = 1$ and discuss how they differ under alternate parameterizations.

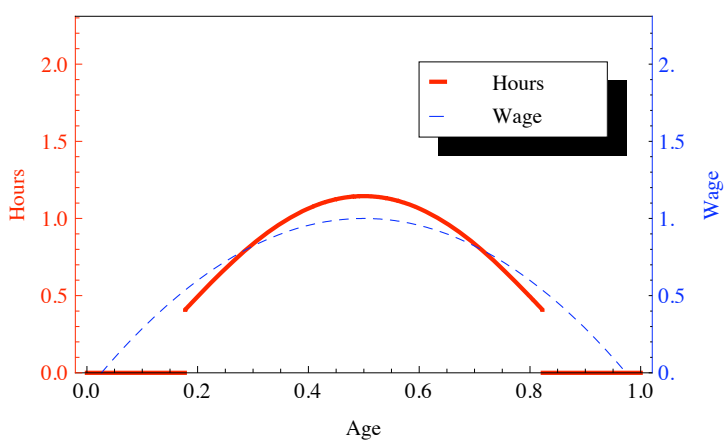

Figure 1: Hours worked and productivity by age under optimal tax policy.

First, Figure 1 depicts the hours worked and wage of workers in the economy by age. The wage

graph plots the function $w(a)$ while the hours worked are the individual's choices under the optimal tax policy. The curvature of the hours profile is largely determined by the labor supply elasticity $\frac{1}{\gamma}$. For $\gamma = .5$, the hours profile displays more curvature as there is a larger difference between hours worked from $w = w^*$ to $w = \bar{w}$. With $\gamma = 3$, the hours profile is quite flat.

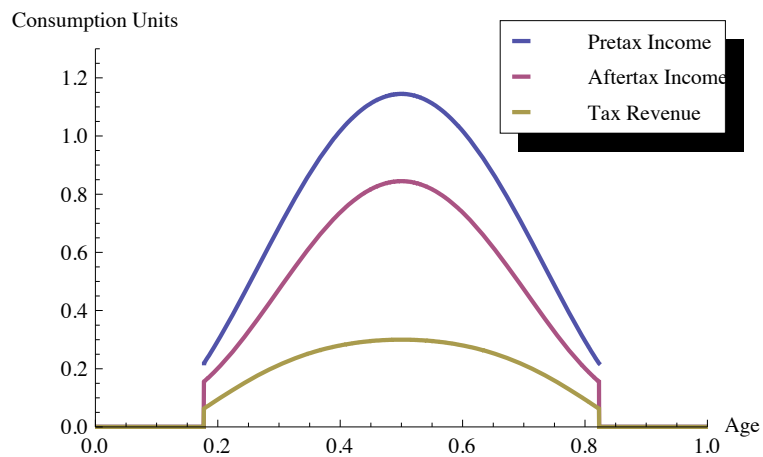

Figure 2: Income and taxes paid by age.

Next, Figure 2 shows pre-tax income, after-tax income, and tax revenue by age. Hours and income reach their peak when productivity is highest in the middle of an individual's life. From this picture we see that total income varies dramatically over the life cycle and most of this variance is reflected in takehome pay while the tax revenue is fairly flat.

The more surprising results come from examination of the tax functions. Figure 3 shows the optimal average and marginal tax rates by income. The figure confirms Proposition 7 as it can be seen that the average tax rate is increasing at the bottom of the income distribution. This hump shaped average tax curve is a feature of the life cycle framework where the active extensive margin of labor supply implies that it is optimal to have a lower average tax. This lower average tax encourages individuals to work a larger portion of their life.

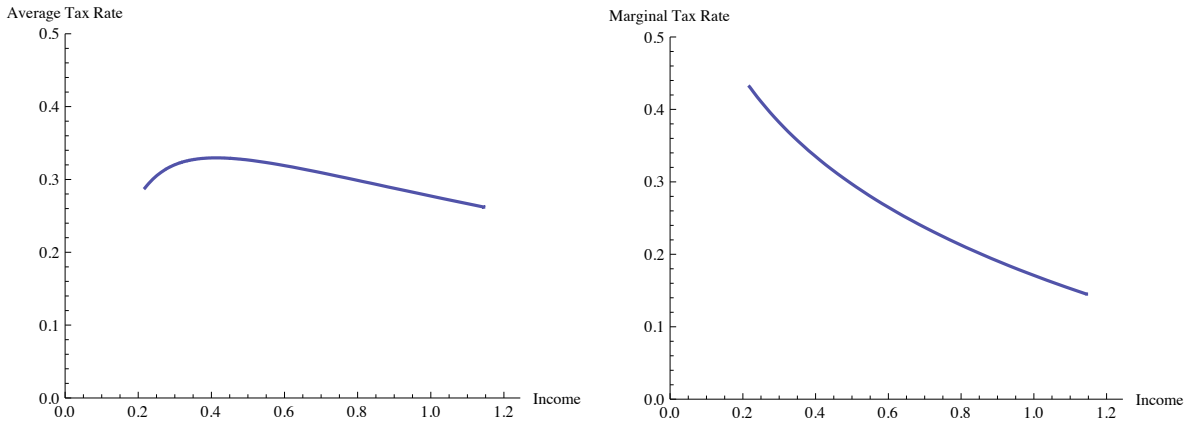

Figure 3: Average and marginal tax rates by income.

The second panel of Figure 3 shows that the marginal tax rate is decreasing in income and attains a minimum at the highest observed level of income in accordance with Proposition 8. The marginal rate declines so that individuals will be encouraged to work more hours during the most productive dates in their lifetime. Moreover, this minimum marginal rate is strictly greater than zero. In this example, we see that the marginal rate varies from about 43% at the bottom to about 15% at the top of the income distribution. Though the marginal rate is highest for the bottom of the income distribution, remember that the average tax rate is lower than the marginal rate.

5.3 Welfare Comparisons

Finally, we use the calibrated models to conduct welfare comparisons. In each case, we compare the optimal tax policy with a flat tax that collects an equal amount of revenue. We compute the percentage increase in consumption needed to make an individual facing the flat tax as well off as an individual facing the optimal tax policy.

Table 2 reports the welfare gain from using the optimal tax policy under all three parameterizations. The welfare gains range from 1.38% when $\gamma = .5$ to 0.74% when $\gamma = 3$. Next, the table

γ	0.5	1	3
Welfare Gain (%)	1.38	1.14	0.74
Consumption Change (%)	7.92	6.68	4.42
Fraction of Life Working (%)	0.62	0.65	0.68
$h(\bar{w})$	1.12	1.14	1.07
$h(w^*)$	0.20	0.41	0.67

Table 2: Welfare comparison and solution values for three calibrations.

reports the percent changes in consumption. Across all values of γ, the agent's consumption rises relative to a flat tax. For the cases of $\gamma = .5$ and $\gamma = 1$, the individual works a smaller fraction of her life under the optimal tax policy than under a flat tax. However, the welfare gains from this are offset by an increase in hours during the most productive periods in life. The optimal tax scheme induces individuals to economize on the fixed costs of work that they pay during their lifetime. For $\gamma = 3$, the individual works a slightly higher fraction of her life and consumes more than under the flat tax, leading to a smaller total welfare gain. The next three lines of Table 2 report the equilibrium values of $1 - F(w^*)$, $h(\bar{w})$, and $h(w^*)$ in the optimal tax model. Recall that the flat tax model is calibrated in each case so that $1 - F(w^*) = \frac{2}{3}$ and $h(\bar{w}) = 1$.

While the welfare gains associated with moving to the optimal tax are relatively small, the comparison is with a flat tax rate. One might claim that a flat tax is close to optimal. For a more meaningful comparison, the model should be extended to account for heterogeneity across workers in addition to human capital accumulation over the life cycle and calibrated to the current tax system.

6 Extensions

Thus far we have emphasized the optimal tax results that arise in a life cycle model with a partial equilibrium framework. In this section, we explore a number of extensions to show that the results of rising average tax rates at the extensive margin and positive marginal taxes at the top are robust. The key features for the results to hold are the active extensive margin and consumption smoothing. As long as these characteristics are present some form of the results will hold.

6.1 Discounting

We simplified the exposition of the original model to have no discounting. All of the results follow with discounting. With a zero interest rate, F could be written as $F(w) = \int_0^1 \mathbb{I}_{w(a)\leq w} da$. If the discount and interest rates were positive and equal, we could write the problem in the exact same way, replacing F in equations (4), (5), and (6) with:

$$\tilde{F}(w) = \frac{\int_0^1 e^{ra} \mathbb{I}_{w(a)\leq w} da}{1 - e^r}$$

The needed assumption would be that $\tilde{F}$ is twice continuously differentiable. A positive discount rate has a small quantitative effect on the solution to the optimal tax problem; the qualitative features of the tax schedule are unchanged.

Moreover, if the interest rate is different from the discount rate (e.g., in a growing economy) the same analysis can be used in a slightly modified form as long as we use balanced growth preferences ($u(c) = \log c$). With these preferences, consumption, wages, income, and the tax function can be appropriately weighted by $e^{-(r-\rho)a}$ to account for the growth in variables over the life cycle.

6.2 General Equilibrium

Under suitable assumptions, we could extend the framework to a general equilibrium setting. From an analytical standpoint, the easiest way to approach this problem is to consider a dynastic family. Within the dynastic family at each date, t, a unit mass of agents is born. Each agent lives for exactly one unit of time so that at date t there is a unit density of agents of age a. All members of the dynasty face the exact same productivity profile over their lifetimes. The dynasty decides how to allocate consumption and how much each individual works at each date and has preferences given by:

$$\int_0^\infty e^{-\rho t} \left(\int_0^1 u(c_t(a)) - v(h_t(a)) da \right) dt$$

where the inner integral integrates over all of the agent alive at time t. The important feature of this setup is that the distribution of wages in the cross section at each date is identical to the distribution of wages over each individual's lifetime. Given this, the problem facing the government is the same as the individual problem constructed earlier. Hence, all of the results carry over to this general equilibrium framework.

Alternatively we could consider a standard overlapping generations model in which agents have access to perfect asset markets so they can transfer their wealth over time. There is a stationary steady state equilibrium of this model for any given tax policy that is identical to the dynasty model discussed.[9]

[9] A complication arises when changes in tax policy are considered. Since agents hold assets, any proposed tax policy must take into account the transition dynamics of the distribution of assets in the population. (With the dynastic family we were able to avoid this because there were no state variables). This means that the dynamics of changes to tax policy in this model are nonstationary and difficult to compute. However, if, in addition to choosing the optimal tax schedule, we allowed the government a one time initial redistribution of asset holdings, the solution to this problem would be identical to that of the dynastic family.

6.3 Human Capital Accumulation

In taking the individual's lifetime wage profile as exogenous, we abstracted away from any type of human capital accumulation. In reality, the wage that individuals face at any given date depends on past work experience. In such a setting the government has an additional incentive to lower the average tax rate at the bottom. This would induce the agent to enter the labor force early, giving her experience and raising her productivity later in life, which increases potential tax revenue. In contrast, if instead individuals gained more human capital in activities outside of work, such as schooling, then having a low average tax rate would be less important.

6.4 Heterogeneous Individuals

Another natural extension would be to add multiple types of workers who face different productivity schedules. Solving these models is considerably more difficult. With multiple agents, the optimal tax schedule could contain kinks at the extensive margin of the more productive agents. When kinks arise workers with different productivities choose to earn the same income. This bunching presents difficulties for both analytical and numerical characterizations of the optimal allocation. Su and Judd (2006) discuss issues that are encountered with such problems and demonstrate numerical methods that can be used to solve them.

It is clear that the positive tax at the top result is robust to adding additional agent types. Because the most productive worker types in the economy still face a lifetime budget constraints, the tax at the top must still be positive for these workers in their most productive period in life. Moreover, the positive marginal tax at the top means that more revenue will be raised from the most productive workers which allows for lower tax rates and hence less distortion for other workers near the bottom of the tax schedule.

7 Conclusion

By taking into account life cycle labor supply decisions, this paper makes two primary contributions to the literature on optimal taxation. First, we find that the marginal tax for the most productive worker is positive. This differs from the standard result in the Mirrlees tax literature. If the most productive individual in the economy is smoothing consumption over her lifetime, then setting her marginal tax rate to zero will have adverse effects for her hours worked in other periods of her life. Therefore, taking this life cycle effect into account gives justification for positive marginal tax rates at the top of the income distribution.

Second, the life cycle framework emphasizes the extensive margin of labor adjustment. The model implies that average tax rates should be increasing at the extensive margin. While this result may be difficult to implement in a world with heterogeneous individuals, it provides guidelines for labor market policies. It favors policies that encourage people to join the workforce at dates where they may otherwise choose to stay home. This can be done through either tax policy or other labor market programs that reduce costs of entering the labor force. These conclusions should be viewed in a broad context that takes into account all policies that might incentivize work including not only taxes, but also welfare, social security, and unemployment insurance.

Appendix

For a number of proofs it will be useful to define $z(y) \equiv y - \tau(y)$ as after tax income.

A Proof of Proposition 1

Claim 1 *There exists a wage level w^* such that $w > w^*$ implies $h(w) > 0$ and $w < w^*$ implies $h(w) = 0$*

Proof. A necessary condition for an optimum is that there exists a $\lambda \geq 0$ such that $c, \{h\}$ maximize the Lagrangian:

$$\begin{aligned}
\mathcal{L}\left(c, \{h(w)\}_{w\in[\underline{w},\bar{w}]}, \lambda\right) &= u(c) - \int_{\underline{w}}^{\bar{w}} v(h(w))\, dF(w) + \lambda\left(\int_{\underline{w}}^{\bar{w}} z(w)\, dF(w) - c\right) \\
&= u(c) - \lambda c + \int_{\underline{w}}^{\bar{w}} \left(-v(h(w)) + \lambda z(w)\right) dF(w)
\end{aligned}$$

Towards a contradiction, assume that there exists $w_1 > w_2$ with $h(w_1) = 0 < \tilde{h} = h(w_2)$. Let $\hat{h} \equiv \frac{w_2\tilde{h}}{w_1}$ be the number of hours it would take at wage w_1 to earn $y(w_1)$. $w_1 > w_2$ implies $\hat{h} < \tilde{h}$ which in turn implies $v\left(\hat{h}\right) < v\left(\tilde{h}\right)$. It was assumed that $h(w_1) = 0$ and $h(w_2) = \tilde{h}$ were optimal choices, so it must be that

$$\begin{aligned}
-v(0) + \lambda z(0) &\geq -v\left(\hat{h}\right) + \lambda z\left(w_1\hat{h}\right) \\
-v\left(\tilde{h}\right) + \lambda z\left(w_2\tilde{h}\right) &\geq -v(0) + \lambda z(0)
\end{aligned}$$

Adding these together gives $v\left(\hat{h}\right) \geq v\left(\tilde{h}\right)$, a contradiction. ■

B Proof of Proposition 2

Claim 2 *Any allocation that results from an arbitrary tax function can also be the result of a tax function for which aftertax income is non-decreasing in pretax income.*

Proof. Define $z(y) \equiv y - \tau(y)$ as after tax income and let $\zeta(y) \equiv \sup_{y' \leq y} z(y')$. Lastly let $u \equiv \max_{c,\{h\}} U(c, \{h\})$ subject to $c \leq \int z(w)dF(w)$ be the maximum attainable utility.

Choose any c' and $\{h'\}$ such that $c' \leq \int \zeta(wh'(w))dF(w)$. $\forall \varepsilon > 0$, there exists a function $\hat{h}(w) \leq h'(w)$ such that $c' - \varepsilon \leq \int z\left(w\hat{h}(w)\right) dF(w)$. Since u is maximal, $U(c' - \varepsilon, \{\hat{h}\}) \leq u$. U is decreasing in h, so $U(c' - \varepsilon, \{h'\}) \leq u$, and taking the limit $\varepsilon \to 0$ gives $U(c', \{h'\}) \leq u$. Lastly, $c \leq \int \zeta(wh(w))\, dF(w)$ is a larger set than $c \leq \int z(wh(w))\, dF(w)$ but yields no more utility, so we can write

$u = \max_{c,\{h\}} U(c, \{h\})$ subject to $c \leq \int \zeta(wh(w))\, dF(w)$

Therefore, without loss of generality we can consider tax functions for which z is non-decreasing.

■

C Proof of Proposition 3

Claim 3 *In any equilibrium, $y(w) = wh(w)$ is non-decreasing in w.*

Proof. A necessary condition for an optimum is that there exists a $\lambda \geq 0$ such that $c, \{h\}$ maximize the Lagrangian:

$$\begin{aligned}
\mathcal{L}\left(c, \{h(w)\}_{w\in[w^*,\bar{w}]}, \lambda\right) &= u(c) - \int_{w^*}^{\bar{w}} v(h(w))\, dF(w) + \lambda\left(\int_{w^*}^{\bar{w}} z(w)\, dF(w) - c\right) \\
&= u(c) - \lambda c + \int_{w^*}^{\bar{w}} \left(-v(h(w)) + \lambda z(w)\right) dF(w)
\end{aligned}$$

Assume that there exists $w_1 > w_2$ with optimal choice of hours $h_1 = h(w_1)$ and $h_2 = h(w_2)$ respectively. Towards a contradiction, assume that $w_1 h_1 < w_2 h_2$. Since $w_1 > w_2$, it must be that $h_1 < h_2$, so that $v(h_1) < v(h_2)$. Note that:

$$w_2 h_1 < w_1 h_1 < w_2 h_2 < w_1 h_2$$

Let $\tilde{h}_1$ and $\tilde{h}_2$ be defined to satisfy $w_1 \tilde{h}_2 = w_2 h_2$ and $w_2 \tilde{h}_1 = w_1 h_1$. These definitions along with the above inequalities imply $h_1 < \tilde{h}_1 < h_2$ and $h_1 < \tilde{h}_2 < h_2$. Also, taking the ratio, we have $\frac{\tilde{h}_2}{h_2} = \frac{w_2}{w_1} = \frac{h_1}{\tilde{h}_1}$. Taking the log of both sides of the equation gives:

$$\log \tilde{h}_1 + \log \tilde{h}_2 = \log h_1 + \log h_2$$

The concavity of log along with the fact that h_1, h_2 are spread wider than $\tilde{h}_1, \tilde{h}_2$ imply that $\tilde{h}_1 + \tilde{h}_2 < h_1 + h_2$.

Second, since h_1 and h_2 are maximizing choices, it must be that:

$$\begin{aligned} -v(h_1) + \lambda z(w_1 h_1) &\geq -v\left(\tilde{h}_1\right) + \lambda z\left(w_1 \tilde{h}_1\right) \\ -v(h_2) + \lambda z(w_2 h_2) &\geq -v\left(\tilde{h}_2\right) + \lambda z\left(w_2 \tilde{h}_2\right) \end{aligned}$$

Adding these together gives:

$$v\left(\tilde{h}_1\right) + v\left(\tilde{h}_2\right) \leq v(h_1) + v(h_2)$$

Since v is strictly convex and since h_1, h_2 are spread wider than $\tilde{h}_1, \tilde{h}_2$, it must be that $\tilde{h}_1 + \tilde{h}_2 < h_1 + h_2$, a contradiction ■

D Proof of Proposition 4

First, for mathematical convenience we define:

$$x(w;\lambda) \equiv \sup_h \{-v(h) + \lambda z(wh)\}$$

λ is the shadow value of income. For an agent at w with shadow value of income λ, $x(w;\lambda)$ is the gain from working in units of utility. The following lemma guarantees certain regularity properties of x that are necessary to solve the optimal tax problem:

Lemma 1 *For any λ, in equilibrium $x(w;\lambda)$ is a weakly increasing, continuous, and differentiable function of w on $(w^*, \bar{w})$ with $x_w(w;\lambda) = \frac{1}{w}v'(h(w))h(w)$.*

The proof of differentiability borrows heavily from Mirrlees (2005).

The proof will proceed in three parts. First, we show that x is weakly increasing:

Proof. Let $w_2 > w_1$

$$\begin{aligned} x(w_1;\lambda) &= -v(h_1) + \lambda z(w_1 h_1) \\ &\leq -v(h_2) + \lambda z(w_1 h_2) \\ &\leq -v(h_2) + \lambda z(w_2 h_2) \\ &= x(w_2;\lambda) \end{aligned}$$

where the first inequality comes from the optimality of h_1 and the second inequality comes from the fact that $z(\cdot)$ is nondecreasing. ■

Second, we show that x is continuous:

Proof. Toward a contradiction, assume that $x(w;\lambda)$ is discontinuous at w_0. Let $\left\{w_i^+\right\}$ be an decreasing sequence and $\left\{w_i^-\right\}$ be a increasing sequence such that $\lim_{i\to\infty} w_i^+ = \lim_{i\to\infty} w_i^- = w_0$. Since x is weakly monotone in its first argument, $\left\{x\left(w_i^+;\lambda\right)\right\}$ and $\left\{x\left(w_i^-;\lambda\right)\right\}$ converge. By assumption, $\lim_{i\to\infty} x\left(w_i^+;\lambda\right) - \lim_{i\to\infty} x\left(w_i^-;\lambda\right) = \delta > 0$.

Define

$$\begin{aligned} h_i^+ &\equiv \arg\max_h -v(h) + \lambda z\left(w_i^+ h\right) \\ h_i^- &\equiv \arg\max_h -v(h) + \lambda z\left(w_i^- h\right) \end{aligned}$$

$\left\{h_i^+\right\}$ has a convergent subsequence, $\left\{h_{j(i)}^+\right\}$. Also define

$$\tilde{h}_{j(i)} \equiv \frac{w_{j(i)}^+ h_{j(i)}^+}{w_{(i)}^-}$$

Note that $\lim_{i\to\infty} \tilde{h}_i = \lim_{i\to\infty} h_{j(i)}^+$. Again, by assumption we have:

$$\left[-v\left(h_{j(i)}^+\right) + \lambda z\left(w_{j(i)}^+ h_{j(i)}^+\right)\right] - \left[-v\left(h_{j(i)}^+\right) + \lambda z\left(w_{j(i)}^+ h_{j(i)}^+\right)\right] \geq \delta$$

and by optimality we have

$$\left[-v\left(h_i^-\right) + \lambda z\left(w_i^- h_i^-\right)\right] - \left[-v\left(\tilde{h}_i\right) + \lambda z\left(w_i^- \tilde{h}_i\right)\right] \geq 0$$

Adding these together give

$$\begin{aligned}\left[-v\left(h_{j(i)}^{+}\right)+\lambda z\left(w_{j(i)}^{+}h_{j(i)}^{+}\right)\right]-\left[-v\left(\tilde{h}_i\right)+\lambda z\left(w_i^{-}\tilde{h}_i\right)\right] &\geq \delta \\ v\left(\tilde{h}_i\right)-v\left(h_{j(i)}^{+}\right) &\geq \delta\end{aligned}$$

Taking the limit of each sides gives $0 \geq \delta$, a contradiction ■

Finally, we show differentiability:

Proof. Define $\tilde{h}(y,w)=\frac{y}{w}$ to be hours needed to earn y with wage w.

For $\varepsilon>0$ define $\alpha_\varepsilon(w)\equiv\frac{1}{\varepsilon}\left[v(h(w))-v\left(\tilde{h}(y(w),w+\varepsilon)\right)\right]$

$$\lim_{\varepsilon\to 0}\alpha_\varepsilon(w)=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(v(h(w))-v\left(\tilde{h}(y(w),w+\varepsilon)\right)\right)=\lim_{\varepsilon\to 0}v'\left(\frac{wh(w)}{w+\varepsilon}\right)\frac{wh(w)}{(w+\varepsilon)^2}=\frac{1}{w}v'(h(w))\,h(w)$$

$$\begin{aligned}\varepsilon\int_{w_1}^{w_2}\alpha_\varepsilon(w)\,dw &= \int_{w_1}^{w_2}\left[v(h(w))-v\left(\tilde{h}(y(w),w+\varepsilon)\right)\right]dw \\ &= \int_{w_1}^{w_2}\left(\left[-v\left(\tilde{h}(y(w),w+\varepsilon)\right)+\lambda z(w)\right]-\left[-v(h(w))-\lambda z(w)\right]\right)dw \\ &\leq \int_{w_1}^{w_2}\left(\left[-v(h(w+\varepsilon))+\lambda z(w+\varepsilon)\right]-\left[-v(h(w))-\lambda z(w)\right]\right)de \\ &= \int_{w_1}^{w_2}\left[x(w+\varepsilon;\lambda)-x(w;\lambda)\right]dw \\ &= \int_0^{\varepsilon}\left[x(w_2+s;\lambda)-x(w_1+s;\lambda)\right]ds\end{aligned}$$

$$\lim_{\varepsilon\to 0^-}\frac{1}{\varepsilon}\int_0^{\varepsilon}\left[x(w_2+s;\lambda)-x(w_1+s;\lambda)\right]ds\leq\int_{w_1}^{w_2}\alpha_\varepsilon(w)\,dw\leq\lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}\int_0^{\varepsilon}\left[x(w_2+s;\lambda)-x(w_1+s;\lambda)\right]ds$$

x is continuous in w for any λ, so both limits exist and are equal to $x(w_2;\lambda)-x(w_1;\lambda)$

Take $\lim_{w_1 \to w^*}$

$$\begin{aligned} x(w^*;\lambda) - x(w_1;\lambda) &= \int_{w_1}^{w^*} \frac{1}{w} v'(h(w))\, h(w) \\ &\Rightarrow \quad x(w_1;\lambda) = x(w^*;\lambda) - \int_{w_1}^{w^*} \frac{1}{w} v'(h(w))\, h(w) \\ &\Rightarrow \quad \frac{\partial x(w;\lambda)}{\partial w} = \frac{1}{w} v'(h(w))\, h(w) \end{aligned}$$

■

Claim 4 *In any equilibrium, the agent chooses an allocation that satisfies*

$$u'(c)c - \int_{w^*}^{\bar{w}} \left[v(h(w)) + \eta(w) v'(h(w)) h(w) \right] dF(w) \geq 0$$

Now to derive the implementability constraint we use necessary conditions from the individual's optimization problem:

$$\begin{aligned} x\left(w^*; u'(c)\right) &= 0 \\ x_w(w; u'(c)) &= \frac{1}{w} v'(h(w)) h(w) \\ c &= \int_{w^*}^{\bar{w}} z(w) dF(w) \end{aligned}$$

We now combine these necessary conditions with the government budget constraint. The government budget constraint and feasibility give:

$$\int_{w^*}^{\bar{w}} z(w)\, dF(w) \leq c$$

Multiplying both sides by $u'(c)$ and adding and subtracting $v(h(w))$ gives:

$$\int_{w^*}^{\bar{w}} \left[v\left(h\left(w\right)\right) - v\left(h\left(w\right)\right) + u'\left(c\right) z\left(w\right)\right] dF\left(w\right) \leq u'\left(c\right) c$$

Using the definition of x we have:

$$\int_{w^*}^{\bar{w}} v(h(w))dF(w) + \int_{w^*}^{\bar{w}} x(w; u'(c))dF(w) \leq u'(c)c$$

Integration by parts on the second integral and imposing $x(w^*; u'(c)) = 0$ gives:

$$\int_{w^*}^{\bar{w}} v(h(w))dF(w) + \int_{w^*}^{\bar{w}} \left(1 - F\left(w\right)\right) x_w(w; u'(c))dw \leq u'\left(c\right) c$$

Using the equation for x_w and combining the integrals yields the implementability constraint:

$$\int_{w^*}^{\bar{w}} \left(v\left(h\left(w\right)\right) + \frac{\left[1 - F\left(w\right)\right]}{F'\left(w\right)} \frac{1}{w} v'\left(h\left(w\right)\right) h\left(w\right)\right) dF\left(w\right) \leq u'\left(c\right) c$$

E Proof of Proposition 5

Claim 5 *The optimal tax function, $\tau(y)$, is differentiable on $y \in (y(w^*), y(\bar{w}))$.*

Proof. The first order condition with respect to $h(w)$ implies that $h(w)$ is a differentiable function of w, which along with the differentiability of $x(w; \lambda)$ implies a differentiable tax function. ■

F Proof of Proposition 6

Let $\left(\hat{c}, \hat{h}(\cdot)\right)$ be the solution to the government's relaxed maximization problem with $\hat{y}(w) \equiv w\hat{h}(w)$. w^* is the reservation wage that comes from the relaxed problem.

Lemma 2 *Assume $\hat{y}(w)$ is strictly increasing on $[w^*, \bar{w}]$. Choose any budget feasible allocation $(\tilde{c}, \{\tilde{y}\})$. For all $w \in [\underline{w}, \bar{w}]$, if $\tilde{y}(w) \neq \hat{y}(w)$, the following inequality holds:*

$$-v\left(\frac{\hat{y}(w)}{w}\right) + u'(\hat{c})\, z(\hat{y}(w)) > -v\left(\frac{\tilde{y}(w)}{w}\right) + u'(\hat{c})\, z(\tilde{y}(w)) \tag{11}$$

Proof. Assuming that $\hat{y}(\cdot)$ is strictly increasing, the tax schedule is well defined on $[w^*, \bar{w}]$, by the equations

$$\begin{aligned} v\left(\frac{\hat{y}(w^*)}{w^*}\right) &= u'(c)\left[\hat{y}(w^*) - \tau(\hat{y}(w^*))\right] \\ v'\left(\frac{\hat{y}(w)}{w}\right) &= wu'(c)\left[1 - \tau'(\hat{y}(w))\right], \quad \forall w \in [w^*, \bar{w}] \end{aligned}$$

We can define $\omega(y)$ to be the inverse of $\hat{y}(w)$ to write the second equation as:

$$v'\left(\frac{y}{\omega(y)}\right) = \omega(y)\, u'(c)\, z'(y), \quad \forall y \in [\hat{y}(w^*), \hat{y}(\bar{w})] \tag{12}$$

Let w_0 be such that $\tilde{y}(w) \neq \hat{y}(w)$ and define $\tilde{y}_0 \equiv \tilde{y}(w_0)$ and $\hat{y}_0 \equiv \hat{y}(w_0)$. We now enumerate all possibilities.

1. $\hat{y}_0, \tilde{y}_0 \in [\hat{y}(w^*), \hat{y}(\bar{w})]$

$$
\begin{aligned}
& \left[-v\left(\frac{\hat{y}_0}{w_0}\right) + u'(\hat{c})\, z(\hat{y}_0)\right] - \left[-v\left(\frac{\tilde{y}_0}{w_0}\right) + u'(\hat{c})\, z(\tilde{y}_0)\right] \\
= & \int_{\tilde{y}_0}^{\hat{y}_0} \left[-v'\left(\frac{y}{w_0}\right)\frac{1}{w_0} + u'(\hat{c})\, z'(y)\right] dy \\
= & \int_{\tilde{y}_0}^{\hat{y}_0} -v'\left(\frac{y}{w_0}\right)\frac{1}{w_0} + v'\left(\frac{y}{\omega(y)}\right)\frac{1}{\omega(y)} dy \\
= & \int_{\tilde{y}_0}^{\hat{y}_0} \int_{\omega(y)}^{w_0} \left[v''\left(\frac{y}{w}\right)\frac{y}{w^3} + v'\left(\frac{y}{w}\right)\frac{1}{w^2}\right] dw dy \\
> & \ 0
\end{aligned}
$$

The first and third equalities follow from the fundamental theorem of calculus and the second equality uses (12). Finally the quantity is positive because (i) the integrand is positive and (ii) $w_0 - \omega(y)$ has the same sign as $\hat{y}_0 - \tilde{y}_0$ since $\omega(y)$ is strictly increasing.

2. $\hat{y}_0 \in [\hat{y}(w^*), \hat{y}(\bar{w})]$ and $\tilde{y}_0 = 0$

At the reservation wage, the tax schedule is contructed, so that $-v\left(\frac{\hat{y}(w^*)}{w^*}\right) + u'(\hat{c})\, z(\hat{y}(w^*)) = 0$. Further, for any $w_0 > w^*$ the value of earning $\hat{y}(w_0)$ is positive

$$
\begin{aligned}
-v\left(\frac{\hat{y}_0}{w_0}\right) + u'(\hat{c})\, z(\hat{y}_0) &= \int_{w^*}^{w_0} \left[-v'\left(\frac{\hat{y}(w)}{w}\right)\left(\frac{\hat{y}'(w)}{w} - \frac{\hat{y}(w)}{w^2}\right) + u'(\hat{c})\, z'(\hat{y}(w))\, \hat{y}'(w)\right] dw \\
&= \int_{w^*}^{w_0} \left[-v'\left(\frac{\hat{y}(w)}{w}\right)\left(\frac{\hat{y}'(w)}{w} - \frac{\hat{y}(w)}{w^2}\right) + \frac{1}{w} v'\left(\frac{\hat{y}(w)}{w}\right)\hat{y}'(w)\right] dw \\
&= \int_{w^*}^{w_0} v'\left(\frac{\hat{y}(w)}{w}\right)\frac{\hat{y}(w)}{w^2} dw \\
&> 0
\end{aligned}
$$

3. $\hat{y}_0 = 0$ and $\tilde{y}_0 \in [\hat{y}(w^*), \hat{y}(\bar{w})]$

Again we start with the constructed tax at the reservation wage

$$\begin{aligned} 0 &= -v\left(\frac{\hat{y}\left(w^{*}\right)}{w^{*}}\right)+u'\left(\hat{c}\right)z\left(y^{*}\left(w^{*}\right)\right) \\ &\geq -v\left(\frac{\tilde{y}_{0}}{w^{*}}\right)+u'\left(\hat{c}\right)z\left(\tilde{y}_{0}\right) \\ &> -v\left(\frac{\tilde{y}_{0}}{w_{0}}\right)+u'\left(\hat{c}\right)z\left(\tilde{y}_{0}\right) \end{aligned}$$

The first inequality follows from the optimality of $\hat{y}\left(w^{*}\right)$ (as shown above) while the second inequality comes from the monotonicity of v.

4. $\tilde{y}\in\left(0,\hat{y}\left(w^{*}\right)\right)\cup\left(\hat{y}\left(\bar{w}\right),\infty\right)$

   We can construct the tax schedule so that $\forall y\notin\left[\hat{y}\left(w^{*}\right),\hat{y}\left(\bar{w}\right)\right]$, $\tau\left(y\right)=y$ which would imply that

$$v\left(\frac{\tilde{y}_{0}}{w_{0}}\right)+u'\left(\hat{c}\right)z\left(\tilde{y}_{0}\right)<0\leq -v\left(\frac{\hat{y}_{0}}{w_{0}}\right)+u'\left(\hat{c}\right)z\left(\hat{y}_{0}\right)$$

■

Claim 6 *The solution to the relaxed problem can be implemented if $\hat{y}\left(w\right)$ is strictly increasing on $\left[w^{*},\bar{w}\right]$ If $\hat{y}\left(\cdot\right)$ is not weakly increasing, it cannot be implemented.*

Proof. Choose any budget feasible allocation $(\tilde{c}, \{\tilde{y}\})$. We now show that agent prefers the allocation given by the relaxed problem:

$$\begin{aligned}
& u(\tilde{c}) - \int_{\underline{w}}^{\bar{w}} v\left(\frac{\tilde{y}(w)}{w}\right) dF(w) \\
\leq\ & u(\hat{c}) + u'(\hat{c})(\tilde{c} - \hat{c}) - \int_{\underline{w}}^{\bar{w}} v\left(\frac{\tilde{y}(w)}{w}\right) dF(w) \\
\leq\ & u(\hat{c}) + u'(\hat{c}) \int_{\underline{w}}^{\bar{w}} \left[z(\tilde{y}(w)) - z(\hat{y}(w))\right] dF(w) \\
& - \int_{\underline{w}}^{\bar{w}} v\left(\frac{\hat{y}(w)}{w}\right) dF(w) - \int_{\underline{w}}^{\bar{w}} \left[v\left(\frac{\tilde{y}(w)}{w}\right) - v\left(\frac{\hat{y}(w)}{w}\right)\right] dF(w) \\
\leq\ & u(\hat{c}) - \int_{\underline{w}}^{\bar{w}} v\left(\frac{\hat{y}(w)}{w}\right) dF(w)
\end{aligned}$$

The first inequality follows from the concavity of u and is strict if $\hat{c} \neq \tilde{c}$. The next equality uses $\tilde{c} \leq \int_{\underline{w}}^{\bar{w}} z(\tilde{y}(w))\, dF(w)$ and $\hat{c} = \int_{\underline{w}}^{\bar{w}} z(\hat{y}(w))\, dF(w)$. The last inequality follows from Lemma 1, and is strict if $\tilde{y}(\cdot)$ differs from $\hat{y}(\cdot)$ on a subset of $[\underline{w}, \bar{w}]$ that has positive measure.

To prove the second part of the proposition, note that if $\hat{y}(\cdot)$ is not weakly increasing, it must differ from the agent's optimal choice which, as shown in Proposition 3, is always weakly increasing.

■

G Deriving the Marginal Tax Rate

The marginal tax rate is given by:

$$\tau'(y) = 1 - \frac{1}{u'(c)} \frac{v'(h)}{w} = 1 - \frac{1}{u'(c)} \frac{\mu}{1+\theta} + \frac{1}{w} \frac{1}{u'(c)} \frac{\theta}{1+\theta} \eta(w) \left[v' + v'' h\right]$$

Where the second equality comes from the first order condition with respect to h. Using the first order condition with respect to c we get:

$$\tau'(y) = \frac{\theta}{1+\theta}\left(-\frac{cu''}{u'}\right) + \frac{v'(h)}{wu'(c)}\frac{\theta}{1+\theta}\eta(w)\left[1+\frac{v''h}{v'}\right]$$

Letting $\sigma(c) \equiv -\frac{cu''}{u'}$ and $\gamma(h) \equiv \frac{v''h}{v'}$ we get:

$$\tau'(y) = \frac{\theta}{1+\theta}\left(\sigma(c) + \frac{v'(h)}{wu'(c)}\eta(w)[1+\gamma(h)]\right) \geq 0$$

Noting that $\frac{v'(h)}{wu'(c)} = 1 - \tau'(y)$, we can rearrange terms to get:

$$\tau'(y) = \frac{\sigma(c) + \eta(w)[1+\gamma(h)]}{\frac{1+\theta}{\theta} + \eta(w)[1+\gamma(h)]}$$

References

Diamond, P., "Income taxation with fixed hours of work," *Journal of Public Economics*, February 1980, *13* (1), 101–110.

Diamond, Peter A, "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, March 1998, *88* (1), 83–95.

Erosa, Andres and Martin Gervais, "Optimal Taxation in Life-Cycle Economies," *Journal of Economic Theory*, 2002, *105* (2), 338–369.

Fullerton, Don and Diane Lim Rogers, "Lifetime vs. Annual Perspectives on Tax Incidence," *National Tax Journal*, September 1991, *44* (3), 277–287.

Golosov, Mikhail, Aleh Tsyvinski, and Ivan Werning, "New Dynamic Public Finance: A User's Guide," in "NBER Macroannual" 2006.

_ , Narayana Kocherlakota, and Aleh Tsyvinski, "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 2003, *70* (3), 569–587.

Grochulski, Borys and Narayana Kocherlakota, "Nonseparable Preferences and Optimal Social Security Systems," NBER Working Papers 13362 2007.

Kocherlakota, Narayana R., "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 2005, *73* (5), 1587–1621.

Lucas, Robert Jr. and Nancy L. Stokey, "Optimal fiscal and monetary policy in an economy without capital," *Journal of Monetary Economics*, 1983, *12* (1), 55–93.

Mathieu-Bolh, Nathalie, "Optimal Taxation Over the Life-Cycle for the Poor and the Rich," February 2008. Transcript, University of Vermont.

Mirrlees, J. A., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38* (2), 175–208.

_ , "Optimal tax theory : A synthesis," *Journal of Public Economics*, November 1976, *6* (4), 327–358.

_ , "The theory of optimal taxation," in K. J. Arrow and M. D. Intriligator, eds., *Handbook of Mathematical Economics*, 2 ed., Vol. 3, Elsevier, 2005.

Mulligan, Casey B., "Aggregate Implications of Indivisible Labor," *Advances in Macroeconomics*, 2001, *1* (1).

Ramsey, F. P., "A Contribution to the Thory of Taxation," *Economic Journal*, 1927, *37*, 47–61.

Rogerson, Richard and Johanna Wallenius, "Micro and Macro Elasticities in a Life Cycle Model with Taxes," June 2007. Manuscript, Arizona State University.

Saez, Emmanuel, "Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses," *The Quarterly Journal of Economics*, August 2002, *117* (3), 1039–1073.

Su, Che-Lin and Kenneth L. Judd, "Optimal Income Taxation with Multidimensional Taxpayer Types," Computing in Economics and Finance 2006 471, Society for Computational Economics July 2006.

Weinzierl, Matthew, "The Surprising Power of Age-Dependent Taxes," December 2007. Job Market Paper, Harvard University.