Review of Economic Dynamics 15 (2012) 551–572

Contents lists available at SciVerse ScienceDirect

Review of Economic Dynamics

www.elsevier.com/locate/red

Optimal taxation over the life cycle

Aspen Gorry[a], Ezra Oberfield[b,*,1]

[a] Economics Department, University of California-Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, United States
[b] Federal Reserve Bank of Chicago, 230 S. LaSalle Street, Chicago, IL 60604, United States

ARTICLE INFO

Article history:
Received 10 April 2011
Revised 11 May 2012
Available online 22 May 2012

JEL classification:
H21
E60

Keywords:
Optimal taxation
Life cycle
Extensive margin
Ramsey problem
Non-linear taxation
Primal approach

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 uses the primal approach to solve for the optimal non-linear tax function. The average tax rate determines when an individual will work while the marginal tax rate determines how much she will work. Even in the absence of redistributive concerns, the optimal tax schedule has an increasing average tax rate at low levels of income to encourage labor market participation. The marginal tax rate at the top is strictly positive. Finally, the model is used to assess the effects of changing the current tax schedule to the optimal one. Under the preferred parameters, this delivers a welfare gain equivalent to 0.67 percent of lifetime consumption.

1. Introduction

Individuals exhibit substantial variation in earnings over their lifetimes. Estimates of life cycle wage patterns show that wages approximately double during workers' lives and individuals regularly shift between tax brackets.[2] 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 the individual's lifetime budget constraint. 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 non-linear labor income tax schedule in a simple life cycle labor supply model. As in Rogerson and Wallenius (2009) and French (2005), 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, this model captures the basic pattern of lifetime labor supply: a period of non-participation followed by working and finally retirement.

Following Ramsey (1927) and Lucas and Stokey (1983), we model a government that must raise revenue to finance a given level of expenditure but does not have access to lump sum taxes. In contrast to the Ramsey literature, the government

* Corresponding author.
E-mail addresses: aspen.gorry@gmail.com (A. Gorry), ezraoberfield@gmail.com (E. Oberfield).
[1] The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Chicago or the Federal Reserve System.
[2] 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 percent of individuals are in the same annual as lifetime decile in that given year and that only 56.1 percent are within one step of their lifetime decile.

1094-2025/$ – see front matter
http://dx.doi.org/10.1016/j.red.2012.05.002

can impose a non-linear labor income tax schedule.[3] In this 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.

To solve the problem, we generalize the standard primal approach to allow for a non-linear tax function. The incentive constraints are combined into a single implementability constraint. When the government chooses a feasible allocation subject to this implementability constraint, the allocation can in fact be implemented provided that income is strictly increasing in wages. With proportional taxes, this last step is trivial; here it is more difficult because with a non-linear tax schedule, the individual's constraint set is not generally convex.

Solving the optimal tax problem in the life cycle model generates two analytical results. First, the average tax rate is increasing in income at the extensive margin. 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. This is accomplished with a low average tax rate but a high marginal tax rate at the extensive margin.

Second, the marginal tax rate at the highest realized income is strictly positive. This result stands in contrast to much of the optimal tax literature that follows Mirrlees (1971). In a Mirrleesian environment 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 of zero marginal tax at the top, suppose that the marginal tax at the highest income was not zero. Then if the tax schedule is 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. We show that this logic breaks down in a life cycle framework. Extending the tax function at a zero marginal rate still induces the worker to work and earn 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. Because of this income effect, the optimal tax function in our model features a positive marginal tax rate at the top of the income distribution.[4]

These results are robust. 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. The positive marginal tax rate at the top is driven by wealth effects. This result is robust to adding other sources of wage heterogeneity as long as the most productive worker has income variation over her lifetime.

The model is parameterized to match the current tax and transfer system in the United States. The income tax system is approximated with the Gouveia and Strauss (1994) tax function and social security is modeled as a 12.4 percent proportional tax on labor income that is transferred back to individuals as lump sum payments after the age of 65. Holding the level of government revenues and transfer payments fixed, we consider reforms to the current tax system. First, moving from the current tax system to the optimal tax system increases welfare by between 0.42 and 1.6 percent in lifetime consumption equivalents depending on the parameter chosen for the Frisch elasticity of labor supply. For our preferred choice of the Frisch elasticity of 0.5 the welfare gain is 0.67 percent. Like the optimal tax schedule, the current tax code has low and rising average tax rates at the extensive margin. However, in contrast to the current tax code, the optimal schedule features high marginal tax rates at the extensive margin that are declining in income. The current tax policy is also compared to two simple tax reforms: constrained lump sum taxes, where the individual faces lump sum taxes only when working, and a proportional income tax. Constrained lump sum taxation places all of the distortion on the extensive margin. Such a policy generates large welfare losses of between 5.2 and 20.5 percent of lifetime consumption equivalents. The gains in moving to a proportional tax system are between 0.12 and 1.1 percent and generate about half of the gains of moving to the optimal tax system in the preferred parameterization.

When the elasticity of intertemporal substitution is smaller so that wealth effects have a more prominent role, the optimal tax schedule has a flatter profile of marginal tax rates and the consumption equivalent of moving to the optimal tax schedule is smaller. In contrast, with linear utility of consumption the optimal tax schedule features a steep profile of marginal tax rates that includes a zero marginal tax rate at the top.

This paper contributes to a growing literature that considers optimal taxation in life cycle environments. Our environment is most similar to Erosa and Gervais (2002), who study a life cycle economy with age dependent proportional taxes, Gervais (2009), who solves for the optimal tax among a particular family of functions, and Conesa and Krueger (2006), who consider the optimal progressivity of the tax code in a life cycle model where heterogenous agents face uninsurable productivity risk over their lifetime. We extend this literature by including an endogenous extensive margin of labor supply, solving for the optimal non-linear tax schedule, and clarifying the differing roles played by the average and marginal tax rates.

Following Saez (2002), we study both the intensive and extensive margins of labor supply adjustment.[5] Saez (2002) develops a simple expression relating the optimal tax rate to empirical local labor supply elasticities. We explicitly motivate

[3] Lump sum taxes are ruled out by imposing the restriction that individuals who earn no income cannot be taxed. Age dependent taxes are also ruled out by assumption as the tax schedule is only a function of individual labor income. With either lump sum taxation or non-linear age dependent taxation the government could achieve the first best allocation.

[4] Several other studies have qualified the classic Mirrleesian result of zero tax at the top. Tuomala (1984) shows numerically that marginal rates can remain far from zero until very close to the top income, while Diamond (1998) shows that when the distribution of skills is unbounded, the marginal tax rate may not converge to zero as income rises. Gaube (2007) shows that a top marginal rate can be positive in a multi-period model in which unobserved skill is constant over time but the tax schedule can vary by period. Other models in which a positive top rate is optimal include Varian (1980), in which income includes an exogenous stochastic component, and Oswald (1983) in which individuals care both about their own income as well as relative incomes.

[5] Diamond (1980) and Mulligan (2001) look at optimal tax policy focusing exclusively on the extensive margin.

the agent's labor supply decisions at both margins by modeling changes in productivity over the life cycle. Because of wealth effects, 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.

Related to our findings is a literature that considers taxation of couples. A number of papers including Kleven et al. (2009), Alesina et al. (2011) and Guner et al. (2011b) have studied optimal tax policy in two earner households where the second earner chooses whether to work. In particular, Guner et al. (2011b) find that moving from the current progressive tax system that taxes second earners at a higher rate to a proportional tax system is associated with sizable welfare gains. This paper demonstrates that a non-linear tax function can reduce distortions at the extensive margin by combining a low average tax rate with a high marginal rate.

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 the policies developed in these papers can be extremely powerful, no current tax system comes close to them.[6] 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 and does not have access to lump sum taxes. In related work, Weinzierl (2011) focuses on the partial reform of conditioning the tax schedule on both income and age and finds that age dependent taxes are a powerful reform. In our framework, allowing for age dependent taxes makes the government's problem trivial as the exogenous productivity profile with no heterogeneity allows age dependence to achieve the first best.[7]

Section 2 presents the model and defines the 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 discusses how the results are robust to many extensions of the model, and Section 6 presents some numerical examples. 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.[8] The agent lives for one unit of time, decides how much to work at each date, and faces a lifetime budget constraint.

Preferences of the agent are given by:

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

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.[9] The fixed cost could represent the cost of commuting, waking up early, or making coffee.[10] This nonconvexity in the production technology generates 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, there are endogenous dates of labor market entry and retirement.

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] Grochulski and Kocherlakota (2010) 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.

[7] At each age for which the individual chooses to work, the government could set a lump sum tax that is low enough that it does not discourage participation. Marginal tax rates would be zero, so the tax avoids distorting the intensive margin as well. We do not model the informational frictions or heterogeneity that would make it difficult to actually implement such a scheme.

[8] Analogous results can be obtained in a model with discounting and interest rates by appropriately scaling objects of interest, as discussed in Section 5.1. Section 5.2 extends the model to allow for transfer payments and Section 5.3 sets the model in general equilibrium.

[9] All of our results go through even if $\chi = 0$, as will be clear from the exposition. What is necessary is that there is an active extensive margin. If $\chi = 0$, an active extensive margin requires a utility function without Inada conditions on hours worked. When $\chi > 0$, hours worked at the extensive margin are positive.

[10] The fixed utility cost could also be written as a fixed time cost as in Rogerson and Wallenius (2009), 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 non-linear 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. We impose two ad hoc restrictions on the tax function. First, it can depend only on income at each date. This implies that the tax system is independent of age. Second, the tax schedule is subject to the constraint $\tau(0) \leqslant 0$. This restriction eliminates lump sum taxes. The government budget constraint is given by:

$$G \leqslant \int_0^1 \tau\big(y(a)\big)\, da \tag{1}$$

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

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

subject to her lifetime budget constraint:

$$\int_0^1 \big[y(a) - \tau\big(y(a)\big) - c(a)\big]\, da \geqslant 0 \tag{3}$$

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 government policy $\tau(\cdot)$ such that:

1. Given the government policy, the allocation solves the individual's optimization problem described by Eqs. (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 $[\underline{w}, \bar{w}]$. This can be constructed from the individual's age profile of wages using:

$$F(w) = \int_0^1 \mathbb{I}_{w(a)\leqslant w}\, da$$

We assume F is twice continuously differentiable. The individual's problem can be written as:

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

$$\text{subject to:} \quad \int_{\underline{w}}^{\bar{w}} \big[y(w) - \tau\big(y(w)\big)\big]\, dF(w) - c \geqslant 0 \tag{5}$$

The government budget constraint becomes:

$$G \leqslant \int_{\underline{w}}^{\bar{w}} \tau\big(y(w)\big)\, dF(w) \tag{6}$$

The propositions below describe properties of the allocation for any tax function.[11] For any tax schedule the solution to the individual's labor supply problem 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%, the individual will never 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.

Finally, we proceed to solve the model using the primal approach by deriving 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 the implementability constraint*:

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

Note that in the standard approach to a Ramsey problem an implementability constraint is usually a necessary and sufficient condition for implementability. However, in this environment, the condition in Proposition 4 is only a necessary condition; with a non-linear tax function the constraint set may not be convex.[12] This issue will be further addressed when we solve the optimal tax problem.

4. Solving the optimal tax problem

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

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

subject to:

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

$$G \leqslant \int_{w^*}^{\bar{w}} \tau\big(y(w)\big)\, dF(w)$$

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

[11] The proofs of all propositions are contained in Appendix A.

[12] In particular, if the marginal tax rate declines quickly enough with income, the marginal after-tax return to labor may increase faster than the marginal disutility of labor. Such a situation cannot happen with a proportional labor income tax.

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:

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

subject to:

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

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

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 may be 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 at a given wage would be a convex function of hours worked, so the constraint set in 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:

$$c: \quad u'(c) + \theta\big[u''(c)c + u'(c)\big] = \mu$$

$$h(w): \quad v'(h) + \theta\big[v'(h) + \eta(w)v''(h)h + \eta(w)v'(h)\big] = \mu w$$

$$w^*: \quad v\big(h\big(w^*\big)\big) + \theta\big[v\big(h\big(w^*\big)\big) + \eta\big(w^*\big)v'\big(h\big(w^*\big)\big)h\big(w^*\big)\big] = \mu w^* h\big(w^*\big)$$

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.*

This is the key result that allows our use of the primal approach as a solution method. The proof relies heavily on a single crossing property being satisfied. 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. Appendix A.7 gives one set of assumptions sufficient to guarantee that $y(w)$ is increasing. 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 first order conditions:

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

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

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. 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. 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:

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

These equations will be helpful in proving the following propositions:

Proposition 7. *The average tax rate is increasing in income 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:

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

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

$$\lim_{w \downarrow w^*} \left(\tau'\big(y(w)\big) - \frac{\tau(y(w))}{y(w)} \right) = \lim_{w \downarrow w^*} \frac{1}{wu'(c)} \left[\frac{v(h)}{h} - v'(h) \right]$$
$$= \lim_{w \downarrow w^*} \frac{1}{wu'(c)} \frac{\theta}{1+\theta} \eta v'' h > 0$$

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

$$\lim_{w \downarrow w^*} \frac{d}{dy} \left(\frac{\tau(y(w))}{y(w)} \right) = \lim_{w \downarrow w^*} \frac{1}{y(w)} \left(\tau'\big(y(w)\big) - \frac{\tau(y(w))}{y(w)} \right) > 0 \qquad \square$$

It is optimal for the government to have a lower average tax rate than marginal tax rate at the extensive margin to induce the individual to work a greater fraction of her life. This result holds even without a fixed cost of working as long as there is an active extensive margin of labor supply. In standard life cycle labor supply models such as Erosa and Gervais (2002) in which the proportional labor income tax rate can be conditioned on age, it is optimal that the tax rate at the extensive margin is low, and is higher at higher incomes. Such models feature a high (and declining) labor supply elasticity at the extensive margin, which motivates a low (and rising) proportional tax rate. Using the same model, Gervais (2009) studies a two parameter family of tax functions, and finds a similar result: welfare is higher when the tax schedule is progressive. A feature of the functions studied in Gervais (2009) is that the average tax rate is rising whenever the marginal rate is rising.

Our model sharpens these results by generalizing the tax function to allow the average and marginal rates to behave differently. The tax function can be characterized by a level of average taxes at the extensive margin and a profile of marginal tax rates at each observed level of income. The average tax rate at the extensive margin is low to induce the worker to enter the workforce. The high marginal tax rate at the bottom allows the planner to collect more taxes at higher income ages by only distorting hours worked when the worker is least productive. We show that is always optimal for average tax rate to be rising at low levels of income, whereas Section 6 provides examples in which it is optimal for marginal tax rates to be declining in income.

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 Appendix A 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, $\tau' \leqslant 1$. Therefore, $\tau'(y)$ is increasing in the quantity $\eta(w)[1+\gamma(h)]$. Since $\eta(w)|_{w=\bar{w}} = 0$, $\tau'(y)$ attains a minimum at $w = \bar{w}$. The marginal tax at the top is:

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

Eq. (9) describes the determinants of the marginal tax rate. As shown in Proposition 8, $\tau'(y)$ 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 (see for instance Erosa and Gervais, 2002). That is if the elasticity of labor supply is high at a particular h then the marginal tax rate should be low. As in a Mirrleesian framework, the counter-cumulative wage distribution plays a role in determining marginal tax rates. When setting the marginal tax rate for a given income the government trades off distorting the intensive margin of labor supply at that income against the additional revenue gained (without distortion) from parts of the productivity distribution that earn higher incomes. That is, raising the marginal rate at a given level of income generates more revenue from everyone with higher income without distorting their labor supply decision.[13]

A new object is the inverse elasticity of intertemporal substitution, $\sigma(c)$. As a benchmark, it is useful to consider a quasilinear utility function, $u(c) = c$. When the utility is linear in consumption, $\sigma(c) = 0$ and the marginal tax at the top is zero. Positive marginal tax rates at the top arise as we depart from the quasilinear specification because additional earnings generate a wealth effect. Higher earnings at any point during the worker's life reduce the marginal utility of consumption. This induces the agent to work fewer hours at all wages. Indeed, Eq. (10) shows that $\sigma(c)$ has a direct effect on the tax at the top.

Another interesting feature of the tax function arises when the individual has a constant Frisch elasticity of labor supply ($\gamma(h) = \bar{\gamma}$). In this case, changes in the marginal tax rate are completely determined by changes in $\eta(w)$, the inverse of the elasticity of the counter-cumulative wage distribution. Put differently, whether the marginal tax rate is increasing or decreasing at a particular point depends only on the wage distribution.

Alternatively, if we relax the restriction that $\bar{w}$ is finite and let wages follow a Pareto distribution, then $\eta(w)$ is constant. In this case, changes in the marginal tax rate are completely determined by changes in the Frisch elasticity of labor supply.

5. Extensions

Thus far we have emphasized the optimal tax results that arise in a simple life cycle model. In this section, we explore a number of extensions. We show how to adjust the solution method to accommodate several alternative assumptions. The qualitative results of rising average tax rates at the extensive margin and positive marginal taxes at the top are robust to these extensions. The key features required for these results to hold are an active extensive margin and consumption smoothing.

5.1. Discounting

We simplified the exposition of the original model by abstracting from discounting. All of the results follow with discounting. Recall that with a zero interest rate, F can be written as $F(w) = \int_0^1 \mathbb{I}_{w(a)\leqslant w}\,da$. If the discount and interest rates were positive and equal, we could write the problem in the exact same way, replacing F in Eqs. (4), (5), and (6) with:

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

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, that is $u(c) = \log c$. 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.

5.2. Government transfers

The model can easily accommodate exogenous lump sum government transfers such as social security. Suppose that the government, for reasons outside the model, decides on a policy to transfer $T(a)$ units of the consumption good to the agent

[13] It may also be helpful to think about the marginal tax rates in terms of the slope of the wage profile by age. If we make the assumption that the age wage profile, $w(a)$, is continuously differentiable, hump shaped, and $w(0) = w(1) = \underline{w}$, then for any wage $\hat{w} \in [\underline{w}, \bar{w})$ there exists $a_1 < a_2$ such that $w(a_1) = w(a_2) = \hat{w}$ and $w'(a_1) > 0 > w'(a_2)$. Note that $1 - F(\hat{w}) = a_2 - a_1$. Hence, by the implicit function theorem:

$$\eta(\hat{w}) = \frac{1-F(\hat{w})}{\hat{w}F'(\hat{w})} = \frac{a_2 - a_1}{\hat{w}\left[\frac{1}{w'(a_1)} - \frac{1}{w'(a_2)}\right]}$$

Therefore $\eta(\hat{w})$ is increasing in the magnitude of the slope of the age wage profile; steeper wage profiles are associated with higher marginal tax rates. The intuition is that there is less mass in the distribution at each wage so the cost of distorting labor supply at that location is lower.

at age a. In this case we can solve for the optimal tax schedule that raises revenue sufficient to cover both the transfers and the government expenditure G.

Letting $T \equiv \int_0^1 T(a)\,da$ be the cumulative transfer, the individual's budget constraint becomes $c \leqslant T + \int_{\underline{w}}^{\bar{w}} [y(w) - \tau(y(w))]\,dF(w)$ and the government's budget constraint is $G + T \leqslant \int_{\underline{w}}^{\bar{w}} \tau(y(w))\,dF(w)$. If we set $\tilde{G} \equiv G + T$, $\tilde{c} \equiv c - T$, and $\tilde{u}(\tilde{c}) \equiv u(\tilde{c} + T)$, we recover (4), (5), and (6) so that the model can be solved in the same way.

5.3. General equilibrium

Under suitable assumptions, we 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:

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

where the inner integral integrates over all of the agents 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.[14]

5.4. Other possible extensions

In taking the individual's lifetime wage profile as exogenous, we abstract away from 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 an individual gains more human capital in activities outside of work, such as schooling, then having a low average tax rate would be less important.

Another natural extension is 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 in the economy still faces a lifetime budget constraint, the tax at the top must still be positive for this worker in her most productive period in life.

6. Numerical examples

In this section, the model is parameterized to match salient features of the tax and social security transfer system in the United States. With these parameters we solve for the optimal tax function. We compute the welfare gain of moving from the current tax system to the optimal tax system and to other simple tax reforms.

6.1. Parameterization

This section parameterizes the version of the model with discounting as described in Section 5.1 and a government transfer system as described in Section 5.2. The annual discount rate is set to 0.03. To simulate the model, functional forms for the utility function and an empirical wage profile by age are needed. In the baseline case, utility functions consistent with balanced growth preferences are used:

[14] 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. 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.

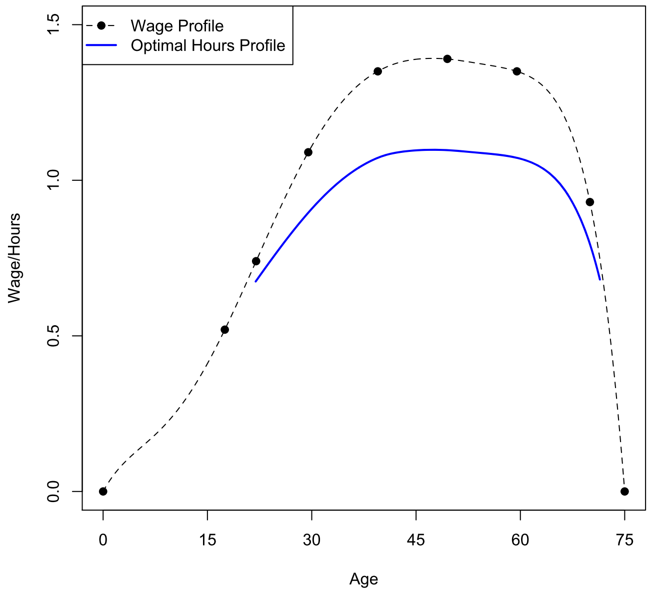


Fig. 1. This figure plots the wage profile by age generated from Hansen (1993) and the hours profile by age for a worker facing a Gouveia and Strauss (1994) tax function when the Frisch elasticity is 0.5.

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

χ is the fixed cost to starting work at each date.

To construct the wage profile by age, we normalize the lifespan for the worker to be 75 years and use the productivity profile from Hansen (1993). Hansen (1993) provides the relative productivities for seven different age groups. We use his estimates for males in 1987, letting each estimate correspond to the productivity at the midpoint of the age range. After adding the further restrictions that productivity is zero at ages 0 and 75, the profile is constructed by a sixth order interpolation that goes through all nine points.[15] The constructed wage profile along with the points used are shown in Fig. 1.

In the baseline case the model is parameterized to match the current levels of taxes and transfers meant to approximate the tax and social security system in the United States. The tax system is approximated using the Gouveia and Strauss (1994) tax function. This tax function is given by:

$$T(y) = a_0\left(y - \left(y^{-a_1} + a_2\right)^{-1/a_1}\right)$$

where a_0, a_1, and a_2 are parameters. For a useful summary of the properties of this tax function see Conesa and Krueger (2006). Briefly, a_0 determines the limiting level of the marginal and average tax rates as income approaches infinity, $a_1 \in [-1, \infty)$ determines the progressivity of the tax function, and a_2 is a scale parameter that depends on the level of income in the economy.

The government is assumed to run a balanced budget social security system. The system is approximated by a proportional 12.4 percent tax on labor income whose revenues are transferred back to the individual in a lump sum fashion starting at age 65. In the baseline parameterization, balanced growth preferences imply that the flat payroll tax leaves labor supply decisions unchanged. However, the benefit payments will influence employment decisions through a wealth effect: agents will enjoy more leisure and retire earlier. Once the model is parameterized to match the tax and transfer system, we solve for the optimal tax function that generates revenue sufficient to cover government expenditures and transfers.[16]

[15] The age groups for which productivity estimates are provided in Hansen (1993) are 16–19, 20–24, 25–34, 35–44, 45–54, 55–64, 65+. The relative productivity weights for males in 1987 that correspond to these ages provided in the paper are 0.52, 0.74, 1.09, 1.35, 1.39, 1.35, and 0.93. We assume these correspond with ages 17.5, 22, 29.5, 39.5, 49.5, 59.5, and 69.5. Gervais (2009) uses the same estimates and then approximates the productivity profile using a second order polynomial. Here we use a higher order interpolation to get a closer approximation to the profile since the numerical methods to solve for tax functions do not require a simple functional form.

[16] While the 12.4 percent payroll tax with a transfer captures the main features of the social security system, it abstracts away from some of the more complicated features of the program. See French and Jones (2011) for a discussion of key features of the social security system. As in Conesa and Krueger (2006) we abstract away from benefits being tied to the workers income by considering reforms that hold the level of transfers fixed. While considering how the progressivity of social security transfers impacts welfare is an important question, in our framework it is a cleaner exercise to consider the optimal

Table 1
Values of α, χ, and a_2 for Frisch elasticities of 0.25, 0.5, 1, and 2.

$1/\gamma$	0.25	0.5	1	2
α	2.13	2.19	2.31	2.50
χ	0.83	0.63	0.38	0.17
a_2	0.30	0.30	0.31	0.32

To parameterize the economy we use the Gouveia and Strauss (1994) estimates for the United States in 1989 to obtain $a_0 = 0.258$ and $a_1 = 0.768$.[17] This leaves the utility parameters γ, α, and χ and the tax parameter a_2 to be chosen to match relevant features of the economy. We parameterize the model for four different Frisch elasticities of labor supply: 0.25, 0.5, 1, and 2. These correspond to values of γ of 4, 2, 1, and 0.5 respectively.

For each γ, the α, χ and a_2 are chosen to target observed features of the economy for a worker facing the estimated Gouveia and Strauss (1994) tax function. First, a_2 is set by targeting the average tax rate for the median wage observed in the economy. We target the median realized average tax rate over the lifetime to match the median average tax rate from Gouveia and Strauss (1994) of 8.625 percent (not including payroll taxes).[18] Following Rogerson and Wallenius (2009), since individuals work about two-thirds of their life, we target $1 - F(w^*) = 0.67$. Second, because we have assumed that disutility of labor $\tilde{v}$ has constant elasticity, we normalize units of h so that $h = 1$ during the period where the worker has her maximum wage. The values of α, χ, and a_2 for each value of γ are reported in Table 1.

To get a sense of the magnitude of χ, we compute the increase in consumption that would be equivalent to setting χ to zero holding constant the profile of hours worked. The increase in consumption ranges from 7.2 percent when the Frisch elasticity is 2 to 47 percent when it is 0.25. In the preferred case when the labor supply elasticity is 0.5 the increase in consumption is 33 percent. The high fixed costs for low Frisch elasticities is a known feature of this calibration approach from Rogerson and Wallenius (2009). Finally, the hours profile generated by the parameterization when the Frisch elasticity is 0.5 is depicted along with the wage profile in Fig. 1.

6.2. Optimal tax functions

This section compares the parameterized Gouveia and Strauss (1994) tax policy with the optimal tax function, a constrained lump sum tax, and a proportional tax. All of the potential policies are parameterized to raise the tax revenue sufficient to cover the current level of government expenditures and transfers. The constrained lump sum tax policy implements a lump sum tax with the constraint that taxes must be zero if the agent has no income. This implies that it is a lump sum tax conditional on working. In this case there is no distortion on the intensive margin as the marginal tax rate is always zero. This is an interesting comparison as the entire tax distortion is concentrated at the extensive margin. The proportional tax provides another benchmark of a tax policy that is easily implemented.

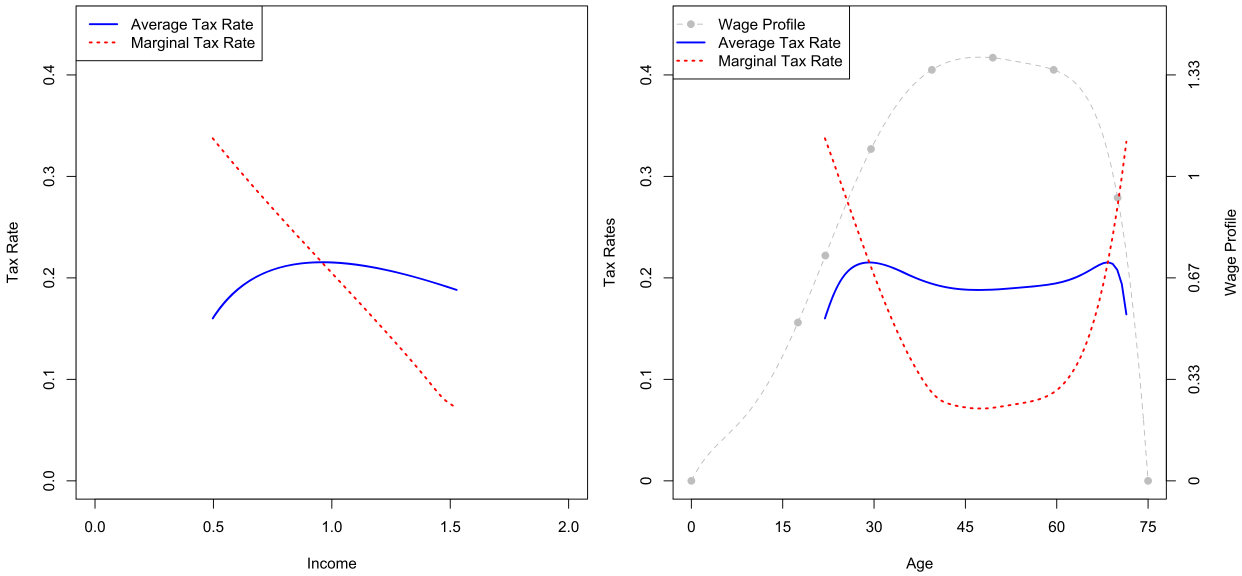
Fig. 2 plots the marginal and average tax rates from the optimal tax function by age and income. Panel (a) shows the optimal marginal and average tax functions by income. All of the figures in this paper use a Frisch elasticity of 0.5 (the figures look similar for other parameterizations). Panel (b) plots the optimal marginal and average tax rates by age faced over the optimal working life of the agent. For reference, the wage profile is replicated with a separate scale on the right axis. The figure confirms Proposition 7; the average tax rate is increasing at the bottom of the income distribution. The marginal tax rate is decreasing in income and, in accordance with Proposition 8, it attains a minimum at the highest observed level of income. 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 34 percent at the lowest observed income to about 7 percent at the highest. Though the marginal rate is highest for the bottom of the income distribution, the average tax rate at the bottom is lower at about 16 percent.

Fig. 3 compares the marginal and average tax rates by income for the four tax policies considered here. Panel (a) presents the marginal tax rates. The Gouveia and Strauss (1994) tax function displays increasing marginal tax rates by income that is consistent with the progressivity of the actual tax code. This structure contrasts with the optimal tax code that has declining marginal rates with income. By construction, the constrained lump sum tax schedule has a zero marginal tax rate and the proportional tax code displays a constant marginal tax rate for all observed incomes. It is worth noting that each tax function is plotted over slightly different ranges of income representing the range of incomes observed in the

tax policy holding the level of government spending and transfers constant. A social security system cannot improve welfare in our setup as the model without taxes and transfers is first best.

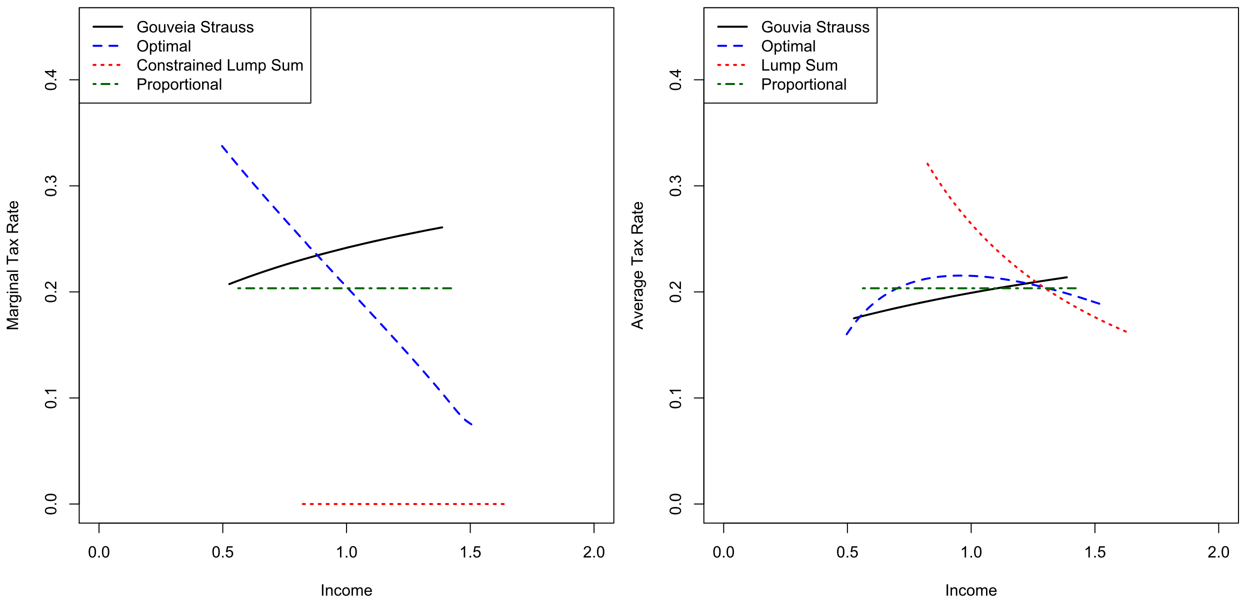
[17] It should be noted that Gouveia and Strauss (1994) estimated a tax as a function of total income, while we use it as a function of labor income.

[18] The lifetime wage profile only provides one source of income heterogeneity, so our model does not generate the full range of incomes from the estimated tax function. Essentially, this parameterization is targeting average taxes in our model to correspond with the average labor taxes paid by the median individual in the income distribution. By decile, the predicted average tax rate from the estimated Gouveia and Strauss (1994) tax rate ranges from about 3–17 percent. 8.625 percent is the mean of the average tax rates for the 5th and 6th income deciles.



(a) (b)

Fig. 2. This figure plots the optimal tax functions by age and income when the Frisch elasticity is 0.5. Panel (a) depicts the optimal marginal and average tax rates by income. Panel (b) depicts the optimal average and marginal tax rates by age with the wage profile plotted for reference with the scale on the right axis.



(a) (b)

Fig. 3. This figure plots the marginal and average tax rates from four tax policies when the Frisch elasticity is 0.5. Panel (a) depicts the marginal tax rates by income for the Gouveia and Strauss (1994) tax function, the optimal tax function, constrained lump sum taxes, and a proportional tax. Panel (b) depicts the average tax rate by income for the same four tax systems. Each tax function is plotted over the range of observed incomes for an individual facing that tax policy.

economy with that tax system in place. Panel (b) shows average tax rates by income. In the current tax code, average tax rates are increasing in income. The optimal tax policy also features a rising average tax rate at the extensive margin. The constrained lump sum tax function displays decreasing average tax rates. Under this policy the entire distortion is placed on the extensive margin with an average tax rate over 30 percent for the lowest income. The average proportional tax is again constant at the proportional rate.

Table 2
Welfare comparison for the optimal tax function, constrained lump sum taxes, and a proportional tax policy relative to the current tax system for each parameterization of the model. The table reports the welfare gain, change in lifetime consumption, and fraction of life spent work in percentage changes from the outcomes under the current tax system. The table also reports hours spent working when the individual is most productive and at the extensive margin of labor supply. For comparison, the current tax system is parameterized so that the fraction of the life spent working is 67 percent, $h(\bar{w}) = 1$, and $h(w^*)$ equals 0.85, 0.74, 0.56, and 0.33 when the Frisch elasticity of labor supply is 0.25, 0.5, 1, and 2 respectively.

Tax function	Optimal				Lump sum				Proportional			
$1/\gamma$	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
Welfare gain (%)	0.42	0.67	1.1	1.6	−5.2	−9.6	−15.0	−20.5	0.12	0.33	0.66	1.1
Consumption change (%)	3.1	4.3	6.7	9.6	5.0	10.0	16.7	24.0	0.31	1.2	2.6	4.3
Fraction working (%)	67.5	66.1	64.4	62.3	63.3	61.0	56.8	49.1	66.1	65.8	65.3	64.5
$h(\bar{w})$	1.05	1.09	1.16	1.25	1.10	1.18	1.31	1.52	1.02	1.03	1.06	1.09
$h(w^*)$	0.79	0.67	0.51	0.30	0.96	0.94	0.92	1.01	0.87	0.76	0.58	0.34

6.3. *Welfare comparisons*

Next, we use the model to conduct welfare comparisons. For each of the three alternative tax policies, we compute the percentage increase in consumption needed to make an individual facing the current tax code as well off as an individual facing the alternative tax policy.

Table 2 reports the change in welfare associated with moving from the current tax system to each of the three other tax policies. This exercise is done for each parameterization of the model to understand how the labor supply elasticity influences the results. In comparison to the baseline tax function, welfare gains of moving to the optimal policy range from 0.42 percent when the Frisch elasticity is 0.25 to 1.6 percent when it is 2. Moving from the current policy to a constrained lump sum tax system places all of the distortions on the extensive margin and is associated with large welfare costs. Moving to a proportional tax system generates gains in welfare between 0.12 and 1.1 percent. Proportional tax policies do a fairly good job balancing distortions at the extensive and intensive margins. Next, the table reports the percent change in consumption under the alternative policy relative to the current policy. Across all Frisch elasticities, the individual's consumption increases under each tax reform. There are large increases in consumption under constrained lump sum as workers work much more when they are most productive. However, the large distortions on the extensive margin are costly for the worker.

Finally, the table reports the fraction of the life spent working and hours at the top and bottom of the income distribution. With a higher Frisch elasticity, the optimal tax function places less distortion on the intensive margin, pushing instead on the extensive margin. Compared to the current tax system, the optimal tax schedule calls for workers to work a shorter fraction of their life when the labor supply elasticity is 0.5, 1, or 2 and a larger fraction when the elasticity is 0.25. Under the optimal tax policy workers work longer when they are most productive for all parameterizations. The optimal tax system features lower hours of work at the extensive margin than the current system, as the marginal tax rate is lower at that point. The current tax policy encourages workers to work a longer fraction of their lives and shorter hours when they are most productive compared to all of the other tax policies under consideration.

6.4. *The importance of wealth effects*

This section assesses the role of the elasticity of intertemporal substitution on the results. For smaller elasticities the marginal utility of income declines faster with wealth. Assume now that an individual has preferences with a constant elasticity of intertemporal substitution $1/\sigma$:

$$u(c) = \frac{c^{1-\sigma} - 1}{1 - \sigma}$$

The elasticity of intertemporal substitution is an important determinant of the optimal marginal tax rates as seen in Eq. (9). When wealth effects are more important, the optimal tax schedule features a flatter profile of marginal tax rates and that the current tax policy is closer to optimal. Fig. 4 plots the optimal marginal tax rates as a function of observed incomes for elasticities of intertemporal substitution of 0.5, 1, and 2. For lower elasticities of intertemporal substitution, the optimal marginal tax rates are flatter as a function of labor income. Earning more income at certain dates induces the individual to consume more leisure at other points in the wage distribution. This moderates the benefit to the government of having especially low marginal tax rates when the individual is most productive.

Table 3 presents the welfare comparisons from changing from the current tax system to the optimal tax system, constrained lump sum taxes, and a proportional tax for different values of the elasticity of intertemporal substitution. The table shows that higher values of the elasticity of intertemporal substitution are associated with larger welfare gains in moving from the current tax code to either the optimal code or a proportional tax system and lower welfare losses in moving to a constrained lump sum system. Overall, the magnitudes of the welfare results are similar to those in the baseline parameterization.

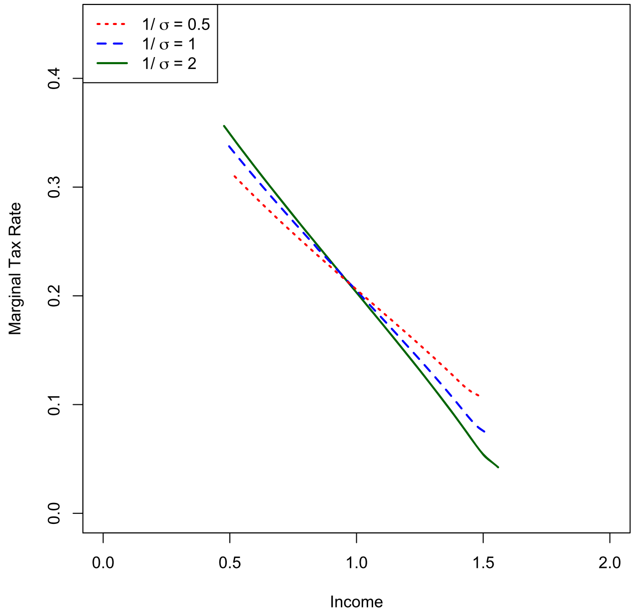


Fig. 4. This figure plots the optimal marginal rates for $1/\sigma$ equal to 0.5, 1, and 2. Each tax function is plotted over the range of observed incomes for an individual facing that tax policy.

Table 3
Welfare comparison for the optimal tax function, constrained lump sum taxes, and a proportional tax policy with the current tax and social security system for different values of the elasticity of intertemporal substitution in the case where the Frisch elasticity is 0.5. The table reports the welfare gain, change in lifetime consumption, and fraction of life spent work in percentage changes from the outcomes under the current tax system. The table also reports hours spent working when the individual is most productive and at the extensive margin of labor supply.

Tax function	Optimal			Lump sum			Proportional		
$1/\sigma$	0.5	1	2	0.5	1	2	0.5	1	2
Welfare gain (%)	0.41	0.67	0.96	−11.0	−9.55	−8.39	0.25	0.33	0.43
Consumption change (%)	2.3	4.3	6.5	11.2	10.0	8.7	0.79	1.2	1.6
Fraction working (%)	65.7	66.1	66.6	61.3	61.0	10.6	65.7	65.8	65.9
$h(\bar{w})$	1.07	1.10	1.12	1.19	1.18	1.18	1.03	1.03	1.03
$h(w^*)$	0.69	0.67	0.66	0.94	0.94	0.94	0.76	0.76	0.76

6.5. The level of tax rates

This section considers different levels of the tax rates paid by the individual over her lifetime. In the baseline parameterization, the median realized average tax rate over the lifetime was targeted at 8.625 percent to match the median average tax rate from tax returns in the United States as reported by Gouveia and Strauss (1994). Combining this with the 12.4 percent payroll tax for the social security system, the total average tax rate paid in the baseline case was 21.025 percent. These numbers are closely aligned with average labor tax rates as computed by Guner et al. (2011a). The choice of the average level of taxes determines the level of a_2 in the parameterization.

However, there exists significant heterogeneity in average tax rates paid by individuals who earn different levels of income in a given year. To get a sense of the sensitivity of our results to the average level of taxes (or the level of a_2), we target three different levels of the median realized average tax rate over the lifetime that correspond to the average tax rates of the 3rd, 8th, and 10th decile of income. Gouveia and Strauss (1994) report that the average tax rates for these deciles as 6.14, 11.27, and 17.27 percent which combined with the social security payroll taxes add up to total average tax rates of 18.54, 23.67, and 29.67 percent respectively. With these different tax rates, the parameterization generates a range of values of a_2 between 0.19 and 1.11.

Table 4 reports welfare comparisons for each of the tax reforms under the four different levels of average taxes with the baseline parameterization where the Frisch elasticity of labor supply is 0.5 and the elasticity of intertemporal substitution is 1. Higher levels of average taxes magnify the welfare gains in moving from the current tax system to the optimal one. For an individual whose median income matches the 3rd decile of income, the welfare gain of moving to the optimal tax schedule is 0.47 percent (relative to 0.67 percent in the baseline). For individuals in the 8th and 10th deciles the welfare gains in moving to the optimal tax policy are 0.88 and 1.33 percent respectively.

Table 4
Welfare comparison for the optimal tax function, constrained lump sum taxes, and a proportional tax policy with the current tax and social security system for different values of the total average tax rate in the economy including payroll taxes in the case where the Frisch elasticity is 0.5. The table reports the values of the parameters of the model, welfare gain, change in lifetime consumption, and fraction of life spent work in percentage changes from the outcomes under the current tax system. The table also reports hours spent working when the individual is most productive and at the extensive margin of labor supply. For average tax rates of 18.5, 21.0, 23.7, and 29.7 percent the values of a_2 in the parameterizations are 0.19, 0.30, 0.46, and 1.11. The values of α and χ are almost unchanged.

Tax function	Optimal				Lump sum				Proportional			
Average tax rate (%)	18.5	21.0	23.7	29.7	18.5	21.0	23.7	29.7	18.5	21.0	23.7	29.7
Welfare gain (%)	0.47	0.67	0.88	1.33	−8.8	−9.6	−10.4	−12.2	0.22	0.33	0.45	0.58
Consumption change (%)	3.6	4.3	5.0	6.2	9.2	10.0	10.8	12.2	0.92	1.21	1.44	1.52
Fraction working (%)	66.4	66.1	65.9	65.7	62.0	61.0	59.9	57.5	66.1	65.8	65.5	65.3
$h(\bar{w})$	1.08	1.10	1.11	1.13	1.16	1.18	1.21	1.26	1.02	1.03	1.04	1.04
$h(w^*)$	0.67	0.67	0.67	0.65	0.91	0.94	0.97	1.04	0.75	0.76	0.76	0.76

7. Conclusion

By taking into account life cycle labor supply decisions, this paper makes several contributions to the literature on optimal taxation. This paper extends the use of the primal approach to solve for an optimal non-linear tax function. We derive an implementability condition and show that the optimal allocation can be implemented provided that pretax income is increasing in the wage.

Next, we find that the marginal tax at the highest income is positive. This contrasts with the standard zero tax at the top result in Mirrleesian environments. Life cycle variation in earnings gives justification for positive marginal tax rates at the top of the income distribution. In numerical simulations, a lower elasticity of intertemporal substitution makes wealth effects stronger. This flattens the optimal profile of marginal tax rates and lowers the welfare gain of moving from the current tax code to the optimal tax schedule.

Finally, 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 by improving the total return to participation rather than improving the marginal return to working more hours at the extensive margin. 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.

Acknowledgments

We thank the editor, Gianluca Violante, and two anonymous referees along with Marco Bassetto, Emmanuel Farhi, Eric French, Ken Judd, Narayana Kocherlakota, Jean Lee, Robert Lucas, Robert Shimer, Nancy Stokey, Andrew Zuppann, 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. All mistakes are our own.

Appendix A

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

A.1. 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 $\lambda \geqslant 0$ such that $c, \{h\}$ maximize the Lagrangian:

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

$$= u(c) - \lambda c + \int_{\underline{w}}^{\bar{w}} \big(-v\big(h(w)\big) + \lambda z(w)\big)\, dF(w)$$

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 $v(\hat{h}) < v(\tilde{h})$. It was assumed that $h(w_1) = 0$ and $h(w_2) = \tilde{h}$ were optimal choices, so it must be that

$$-v(0) + \lambda z(0) \geqslant -v(\hat{h}) + \lambda z(w_1 \hat{h})$$
$$-v(\tilde{h}) + \lambda z(w_2 \tilde{h}) \geqslant -v(0) + \lambda z(0)$$

Adding these together gives $v(\hat{h}) \geqslant v(\tilde{h})$, a contradiction. □

A.2. 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 after-tax income is non-decreasing in pretax income.*

Proof. Let $\zeta(y) \equiv \sup_{y' \leqslant y} z(y')$ and let $U^* \equiv \max_{c,\{h\}} U(c, \{h\})$ subject to $c \leqslant \int z(w)\,dF(w)$ be the maximum attainable utility.

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

$$U^* = \max_{c,\{h\}} U\big(c, \{h\}\big) \quad \text{subject to} \quad c \leqslant \int \zeta\big(wh(w)\big)\,dF(w)$$

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

A.3. 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 $\lambda \geqslant 0$ such that $c, \{h\}$ maximize the Lagrangian:

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

Assume that there exist w_1 and w_2 such that $w_1 > w_2 > w^*$, 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 $h_1 < h_2$. 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 $w_1 \tilde{h}_1 = w_2 h_2$ and $w_2 \tilde{h}_2 = w_1 h_1$. These definitions along with the above inequalities imply $h_1 < \tilde{h}_1 < h_2$ and $h_1 < \tilde{h}_2 < h_2$, as well as $\frac{\tilde{h}_1}{h_2} = \frac{w_2}{w_1} = \frac{h_1}{\tilde{h}_2}$. Taking the log of both sides 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 $\tilde{h}_1, \tilde{h}_2$ implies that $\tilde{h}_1 + \tilde{h}_2 < h_1 + h_2$.[19]

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

$$-v(h_1) + \lambda z(w_1 h_1) \geqslant -v(\tilde{h}_1) + \lambda z(w_1 \tilde{h}_1)$$
$$-v(h_2) + \lambda z(w_2 h_2) \geqslant -v(\tilde{h}_2) + \lambda z(w_2 \tilde{h}_2)$$

Adding these together gives:

$$v(\tilde{h}_1) + v(\tilde{h}_2) \geqslant v(h_1) + v(h_2)$$

Since $\tilde{v}$ 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. □

[19] This is an application of Karamata's inequality: $\{\log h_1, \log h_2\}$ majorizes $\{\log \tilde{h}_1, \log \tilde{h}_2\}$ while $\exp(\cdot)$ is a convex function.

A.4. Proof of Proposition 4

First, for mathematical convenience we define:

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

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 $w_1 > w_2$

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

where the first inequality comes from the fact that $z(\cdot)$ is non-decreasing and the second inequality comes from the optimality of h_1. $\square$

Second, we show that x is continuous:

Proof. Toward a contradiction, assume that $x(w;\lambda)$ is discontinuous at w_0. Let $\{w_i^+\}$ be a decreasing sequence and $\{w_i^-\}$ be an 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, $\{x(w_i^+;\lambda)\}$ and $\{x(w_i^-;\lambda)\}$ converge. By assumption, $\lim_{i\to\infty} x(w_i^+;\lambda) - \lim_{i\to\infty} x(w_i^-;\lambda) = \delta > 0$.

For each i there is an h_i^+ such that

$$-v\bigl(h_i^+\bigr) + \lambda z\bigl(w_i^+ h_i^+\bigr) \geqslant x\bigl(w_i^+;\lambda\bigr) - \delta/2 \tag{11}$$

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

$$h_i^- \equiv \frac{w_{j(i)}^+ h_{j(i)}^+}{w_i^-}$$

Since $x(\cdot;\lambda)$ is weakly increasing, by assumption we have:

$$x\bigl(w_{j(i)}^+;\lambda\bigr) - x\bigl(w_i^-;\lambda\bigr) > \delta$$

and by optimality we have

$$x\bigl(w_i^-;\lambda\bigr) \geqslant \bigl[-v\bigl(h_i^-\bigr) + \lambda z\bigl(w_i^- h_i^-\bigr)\bigr]$$

Adding these together with (11) evaluated at $j(i)$ gives

$$\begin{aligned} &\bigl[-v\bigl(h_{j(i)}^+\bigr) + \lambda z\bigl(w_{j(i)}^+ h_{j(i)}^+\bigr)\bigr] - \bigl[-v\bigl(h_i^-\bigr) + \lambda z\bigl(w_i^- h_i^-\bigr)\bigr] \geqslant \delta/2 \\ &v\bigl(h_i^-\bigr) - v\bigl(h_{j(i)}^+\bigr) \geqslant \delta/2 \end{aligned}$$

Because $\lim_{i\to\infty} h_i^- = \lim_{i\to\infty} h_{j(i)}^+$, taking the limit of each side gives $0 \geqslant \delta/2$, a contradiction. $\square$

Finally, we show differentiability:

Proof. Let $h(w)$ be the optimal hours at w, and let $y(w) = wh(w)$. Define $\tilde{h}(y,w) = \frac{y}{w}$ to be hours needed to earn y with wage w. For $\varepsilon \neq 0$ define

$$\alpha_\varepsilon(w) \equiv \frac{1}{\varepsilon}\big[v\big(h(w)\big) - v\big(\tilde{h}\big(y(w), w+\varepsilon\big)\big)\big]$$

Taking the limit as $\varepsilon \to 0$ gives

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

We can also write

$$\begin{aligned}
\varepsilon\int_{w_1}^{w_2}\alpha_\varepsilon(w)\,dw &= \int_{w_1}^{w_2}\big[v\big(h(w)\big) - v\big(\tilde{h}\big(y(w), w+\varepsilon\big)\big)\big]\,dw \\
&= \int_{w_1}^{w_2}\big(\big[-v\big(\tilde{h}\big(y(w), w+\varepsilon\big)\big) + \lambda z\big(y(w)\big)\big] - \big[-v\big(h(w)\big) - \lambda z\big(y(w)\big)\big]\big)\,dw \\
&\leqslant \int_{w_1}^{w_2}\big(\big[-v\big(h(w+\varepsilon)\big) + \lambda z\big(y(w+\varepsilon)\big)\big] - \big[-v\big(h(w)\big) - \lambda z\big(y(w)\big)\big]\big)\,dw \\
&= \int_{w_1}^{w_2}\big[x(w+\varepsilon;\lambda) - x(w;\lambda)\big]\,dw \\
&= \int_{0}^{\varepsilon}\big[x(w_2+s;\lambda) - x(w_1+s;\lambda)\big]\,ds
\end{aligned}$$

where the inequality uses the optimality of $h(w+\varepsilon)$ when productivity is $w+\varepsilon$;

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

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

Taking the limit as $w_2 \to w^*$

$$x\big(w^*;\lambda\big) - x(w_1;\lambda) = \int_{w_1}^{w^*}\frac{1}{w}v'\big(h(w)\big)h(w)\,dw \quad\Rightarrow\quad x(w_1;\lambda) = x\big(w^*;\lambda\big) - \int_{w_1}^{w^*}\frac{1}{w}v'\big(h(w)\big)h(w)\,dw$$

$$\Rightarrow\quad \frac{\partial x(w;\lambda)}{\partial w} = \frac{1}{w}v'\big(h(w)\big)h(w) \qquad \square$$

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

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

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

$$x\big(w^*; u'(c)\big) = 0$$

$$x_w\big(w; u'(c)\big) = \frac{1}{w}v'\big(h(w)\big)h(w)$$

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

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

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

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

Using the definition of x we have:

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

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

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

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

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

A.5. 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. □

A.6. Proof of Proposition 6

Let $(\hat{c}, \hat{h}(\cdot))$ 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*:

$$-v\left(\frac{\hat{y}(w)}{w}\right) + u'(\hat{c})z\big(\hat{y}(w)\big) > -v\left(\frac{\tilde{y}(w)}{w}\right) + u'(\hat{c})z\big(\tilde{y}(w)\big) \tag{12}$$

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

$$v\left(\frac{\hat{y}(w^*)}{w^*}\right) = u'(\hat{c})\left[\hat{y}\big(w^*\big) - \tau\big(\hat{y}\big(w^*\big)\big)\right]$$

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

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

$$v'\left(\frac{y}{\omega(y)}\right) = \omega(y)u'(\hat{c})z'(y), \quad \forall y \in \left[\hat{y}\big(w^*\big), \hat{y}(\bar{w})\right] \tag{13}$$

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

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

The first and third equalities follow from the fundamental theorem of calculus and the second equality uses (13). 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 constructed so that $-v(\frac{\hat{y}(w^*)}{w^*})+u'(\hat{c})z(\hat{y}(w^*))=0$. Further, for any $w_0 > w^*$ the value of earning $\hat{y}(w_0)$ is positive

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

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

$$
\begin{aligned}
0&=-v\left(\frac{\hat{y}(w^*)}{w^*}\right)+u'(\hat{c})z\big(y^*\big(w^*\big)\big)\\
&\geqslant -v\left(\frac{\tilde{y}_0}{w^*}\right)+u'(\hat{c})z(\tilde{y}_0)\\
&> -v\left(\frac{\tilde{y}_0}{w_0}\right)+u'(\hat{c})z(\tilde{y}_0)
\end{aligned}
$$

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

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

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

$$
-v\left(\frac{\tilde{y}_0}{w_0}\right)+u'(\hat{c})z(\tilde{y}_0)<0\leqslant -v\left(\frac{\hat{y}_0}{w_0}\right)+u'(\hat{c})z(\hat{y}_0) \qquad \square
$$

Claim 6. *The solution to the relaxed problem can be implemented if $\hat{y}(w)$ is strictly increasing on $[w^*, \bar{w}]$. If $\hat{y}(\cdot)$ 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:

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

The first inequality follows from the concavity of u and is strict if $\hat{c} \neq \tilde{c}$. The next equality uses $\tilde{c} \leqslant \int_{\underline{w}}^{\bar{w}} z(\tilde{y}(w))\, dF(w)$ and $\hat{c} = \int_{\underline{w}}^{\bar{w}} z(\hat{y}(w))\, dF(w)$. The last inequality follows from Lemma 2, 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. □

A.7. Sufficient conditions for $y(w)$ to be increasing

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 suppressed. 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 $\frac{w\eta'}{\eta} \leqslant 1$ and $\frac{h}{\gamma}\frac{\gamma'}{1+\gamma} \geqslant -1$ 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 $\eta'(w) \leqslant 0$, $\forall w \in (w^*, w)$. 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 the wage–age profile which we use in the numerical computations. For the second condition a more restrictive assumption is that $\gamma' \geqslant 0$. This is satisfied for models with a constant Frisch elasticity which is a common assumption in the literature.

A.8. Deriving the marginal tax rate

The marginal tax rate is given by:

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

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:

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

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

Alesina, A., Ichino, A., Karabarbounis, L., 2011. Gender-based taxation and the division of family chores. American Economic Journal: Economic Policy 3 (2), 1–40.

Conesa, J.C., Krueger, D., 2006. On the optimal progressivity of the income tax code. Journal of Monetary Economics 53 (7), 1425–1450.

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

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

Erosa, A., Gervais, M., 2002. Optimal taxation in life-cycle economies. Journal of Economic Theory 105 (2), 338–369.

French, E., 2005. The effects of health, wealth, and wages on labour supply and retirement behaviour. Review of Economic Studies 72 (2), 395–427.

French, E., Jones, J.B., 2011. The effects of health insurance and selfinsurance on retirement behavior. Econometrica 79 (3), 693–732.

Fullerton, D., Rogers, D.L., 1991. Lifetime vs. annual perspectives on tax incidence. National Tax Journal 44 (3), 277–287.

Gaube, T., 2007. Optimum taxation of each year's income. Journal of Public Economic Theory 9 (1), 127–150.

Gervais, M., 2009. On the optimality of age-dependent taxes and the progressive U.S. tax system. Discussion Paper Series in Economics and Econometrics 0905, Economics Division, School of Social Sciences, University of Southampton.

Golosov, M., Kocherlakota, N., Tsyvinski, A., 2003. Optimal indirect and capital taxation. Review of Economic Studies 70 (3), 569–587.

Golosov, M., Tsyvinski, A., Werning, I., 2006. New dynamic public finance: A user's guide. In: NBER Macroannual.

Gouveia, M., Strauss, R.P., 1994. Effective federal individual income tax functions: An exploratory empirical analysis. National Tax Journal 47 (2), 317–339.

Grochulski, B., Kocherlakota, N., 2010. Nonseparable preferences and optimal social security systems. Journal of Economic Theory 145 (6), 2055–2077.

Guner, N., Kaygusuz, R., Ventura, G., 2011a. Income taxation of U.S. households: Basic facts. IZA Discussion Papers 5549, Institute for the Study of Labor (IZA).

Guner, N., Kaygusuz, R., Ventura, G., 2011b. Taxing women: A macroeconomic analysis. IZA Discussion Papers 5962, Institute for the Study of Labor (IZA).

Hansen, G.D., 1993. The cyclical and secular behaviour of the labour input: Comparing efficiency units and hours worked. Journal of Applied Econometrics 8 (1), 71–80.

Kleven, H.J., Kreiner, C.T., Saez, E., 2009. The optimal income taxation of couples. Econometrica 77 (2), 537–560.

Kocherlakota, N.R., 2005. Zero expected wealth taxes: A Mirrlees approach to dynamic optimal taxation. Econometrica 73 (5), 1587–1621.

Lucas, R.J., Stokey, N.L., 1983. Optimal fiscal and monetary policy in an economy without capital. Journal of Monetary Economics 12 (1), 55–93.

Mirrlees, J.A., 1971. An exploration in the theory of optimum income taxation. Review of Economic Studies 38 (2), 175–208.

Mirrlees, J.A., 2005. The theory of optimal taxation. In: Arrow, K.J., Intriligator, M.D. (Eds.), Handbook of Mathematical Economics, vol. 3. 2nd edition. Elsevier.

Mulligan, C.B., 2001. Aggregate implications of indivisible labor. Advances in Macroeconomics 1 (1).

Oswald, A.J., 1983. Altruism, jealousy and the theory of optimal non-linear taxation. Journal of Public Economics 20 (1), 77–87.

Ramsey, F.P., 1927. A contribution to the theory of taxation. Economic Journal 37, 47–61.

Rogerson, R., Wallenius, J., 2009. Micro and macro elasticities in a life cycle model with taxes. Journal of Economic Theory 144 (6), 2277–2292.

Saez, E., 2002. Optimal income transfer programs: Intensive versus extensive labor supply responses. Quarterly Journal of Economics 117 (3), 1039–1073.

Su, C.-L., Judd, K.L., 2006. Optimal income taxation with multidimensional taxpayer types. In: Computing in Economics and Finance 2006, 471, Society for Computational Economics.

Tuomala, M., 1984. On the optimal income taxation: Some further numerical results. Journal of Public Economics 23 (3), 351–366.

Varian, H.R., 1980. Redistributive taxation as social insurance. Journal of Public Economics 14 (1), 49–68.

Weinzierl, M., 2011. The surprising power of age-dependent taxes. Review of Economic Studies 78 (3).