Tax Smoothing with Redistribution*

Iván Werning

Massachusetts Institute of Technology,
National Bureau of Economic Research,
and Federal Reserve Bank of Minneapolis

ABSTRACT

This paper studies optimal taxation of labor and capital income in a dynamic economy subject to government expenditure and aggregate productivity shocks. The model relaxes two assumptions commonly adopted in Ramsey models: that there is a single representative agent and that all taxation is proportional. In contrast, to capture a redistributive motive for distortive taxation the model features differences in relative skills across workers and considers two main scenarios for tax instruments available to the government: (i) labor taxation is linear with arbitrary intercept and slope; and (ii) taxation is non-linear and unrestricted as in Mirrlesian models. The main result provides conditions for perfect tax smoothing: the marginal tax on labor income should remain constant over time and remain invariant in the face of government expenditure and aggregate technology shocks. In addition, the tax on capital income should be zero. I show that the model differs along several important dimensions with Ramsey settings such as: public debt management, the nature of the time inconsistency problem and the possibility of replicating complete markets without state contingent bonds. Finally, an extension of the basic model highlights movements in the distribution of relative skills as a potential source for variations in optimal marginal tax rates.

* Email: iwerning@mit.edu. This paper is a revised version of the first part of my Ph.D. dissertation at the University of Chicago (Chapters 2 and 3 in Werning, 2002). I am grateful to my advisors Fernando Alvarez, Gary Becker, Pierre-Andre Chippori and Robert Lucas for their support. For useful discussions and comments I also thank the editor, three anonymous referees and Manuel Amador, Marios Angeletos, Paco Bueras, Emmanuel Farhi, Mike Golosov and Pierre Yared. I am grateful for the hospitality during my visit at the Federal Reserve Bank of Minneapolis. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.
1 Introduction

How should a government set and adjust taxes on labor and capital over time in the face of shocks to government expenditure and aggregate productivity? Ramsey optimal tax theory provides two important insights into this question: taxes on labor income should be smoothed (Barro, 1979; Lucas and Stokey, 1983; Zhu, 1992) while taxes on capital should be set to zero (Chamley, 1986; Judd, 1985).

This paper addresses an important shortcoming in interpreting these cornerstone results. The Ramsey approach casts the optimal tax problem within a representative-agent paradigm; then, to avoid the first-best allocation, lump-sum taxes—or any combination of tax instruments that may replicate lump-sum taxes—are ruled out. The second-best problem chooses the right mix of distortive taxes to maximize the representative agent’s utility subject to the government’s intertemporal budget condition.

Societies may have their own good reasons for avoiding complete reliance on lump-sum taxation and resigning themselves to the use of distortionary taxes. Unfortunately, none of these reasons are explicitly captured by a representative-agent Ramsey framework. Although the first-best allocation is ruled out, an arbitrary second-best problem is set in its place. What confidence can we have that tax recommendations obtained this way accurately evaluate the trade-offs faced by society? If, for reasons unspecified by the model, lump-sum taxes are presumably undesirable for society yet still desirable within the model, how can we be sure that the tax prescriptions derived are not, for the same unspecified reasons, also socially undesirable?

Distributional concerns are a natural reason to resort to distortionary taxation (Mirrlees, 1971). If workers are heterogeneous with respect to their labor productivity, and if this trait is not observable, or if for some reason taxes cannot be conditioned upon them, then society cannot attain almost any of the first-best allocations. In contrast, by taxing observable differences such as income, redistribution is possible, albeit at a loss in efficiency. Such a trade-off between redistribution and efficiency provides a microfoundation for the role of distortionary taxes.

Thus, the representative-agent Ramsey framework, which simply assumes taxation is proportional and rules out lump-sum taxation, does not model the distributional concerns that motivate distortionary taxation in the first place. On the other hand, these models have the virtue of tractability that allow working with rich dynamics stochastic general equilibrium models, of the sort used in other areas of macroeconomics. With this in mind, this paper reexamines optimal taxation in dynamic economies close to those used representative-agent Ramsey models such as Chari, Christiano, and Kehoe (1994) and others, but modeling
distributional concern explicitly and allowing for a richer tax structure.

This model in this paper features workers that have heterogeneous skills in transforming their work effort into effective labor services used in production. The aggregate economy features a neoclassical production function that combines capital with total effective labor services. The economy is subject to exogenous fluctuations in government expenditures and technology. Two main scenarios for the set of available tax instruments are considered. In the first, labor income taxes are assumed linear, but an arbitrary lump-sum tax intercept in the schedule is allowed. In the second, no restrictions are placed on tax instruments, but the government is limited by an inherent asymmetry of information, as in Mirrlees’s (1971) nonlinear taxation model.

In the first case, the labor income tax schedule can be summarized at any moment by two variables: the intercept $T_t$ or lump-sum tax component, and the slope $\tau_t$ or marginal tax rate. Thus, this simple linear tax case suffices for incorporating the essential missing tax instrument in Ramsey models: the lump-sum tax component. Indeed, if all workers had the same skill, then the lump-sum tax component can be used to attain the first-best allocation. However, with heterogeneity a positive marginal tax is generally preferable, since more productive (“richer”) workers then bear a larger tax burden and alleviate the burden of less productive (“poorer”) workers. Since restrictions on the lump-sum tax component are hard to motivate, heterogeneity seems primordial for a well-motivated non-trivial tax problem.\(^1\)

The first main result is that **perfect tax smoothing is optimal** for a class of preferences with isolastic and separable utility. At the optimum, marginal income tax rates are constant over time and invariant to government expenditure and technology shocks. The government uses debt and the lump-sum tax component to smooth out these shocks. In addition, **the optimal tax rate on capital is zero**—a version of the Chamley-Judd result for our heterogeneous agent stochastic economy.

The intuition for the tax smoothing results is that, with heterogeneous workers and a lump-sum tax component, it is **distributional concerns** that determine the desired level of distortive taxation. At any point in time, the current tax rate is a measure of redistribution across workers, while the distribution of relative skills determines the desired level of redistribution. In the model, this level is constant over time and invariant to government expenditure and technology shocks because these shocks do not directly affect the distribution of relative skills across workers.

\(^1\) Indeed, income taxation in most countries feature a negative intercept because of deductibles or welfare programs. This will be the optimal outcome of our model if there is enough inequality in skills and a strong enough desire for redistribution.
For the unrestricted Mirrleesian tax scenario, I find that marginal tax rate should vary across workers, but that the tax rate for each worker should remain perfectly constant over time and across aggregate shocks to technology and government expenditure. This form of tax smoothing suggests a role for taxation based on lifetime earnings or income-averaging. I show that in a deterministic setting such a tax scheme implements the optimal allocation by automatically equating marginal tax rates over time while retaining the desired nonlinearity across workers. Vickery (1947), for different reasons, was an early proponent of such income-averaging taxation arrangements.

The normative model in this paper attributes a crucial role in the determination of optimal tax rates to the distribution of relative skills. This relates to positive political economy models, where the distribution of income has always played a prominent role—such as the median-voter model in Meltzer and Richard (1981). To bring distributional concerns to the forefront, I extend the linear and Mirrleesian tax model and consider aggregate shocks that affect the distribution of relative skills. I find that optimal tax rates on labor income do respond to these shocks, but remain invariant to government expenditure and technology shocks. Intuitively, tax rates rise when the dispersion of worker skills widens. This extension highlights a novel determinant for the evolution of optimal tax rates, one that cannot be addressed in a representative-agent Ramsey framework.

The model also has some novel implications for public debt management. Ramsey models break Ricardian equivalence, which otherwise renders government debt indeterminate. In particular, in Barro (1979) and Lucas and Stokey (1983) public debt plays a crucial role in allowing the government to smooth tax rates over time. In contrast, in this paper, despite distortionary taxation, Ricardian equivalence reemerges: as long as a lump-sum tax component is available the government can smooth marginal tax rates with various mixes of debt and lump-sum tax financing. Towards the end of the paper, I briefly speculate on extensions of the model that determine debt policy uniquely.

In standard Ramsey settings initial capital levies are desirable because they are not distortionary and perfectly mimic the missing lump-sum taxes. This leads to a time-inconsistency problem: capital should be untaxed in the long run, but it is always desirable to tax it in the short run. In contrast, in the model of this paper, lump-sum taxation is already available. As a result, the value to the government of initial capital levies, and the nature of the resulting time inconsistency problem, are quite different from standard Ramsey settings. In the model, the value of capital levies hinges on the distribution of capital wealth across workers. Similarly, a time consistency problem may still arise, but it is due to the evolution this wealth distribution.

Finally, the model also contrasts with standard Ramsey settings regarding the possibil-
ity of replicating complete market allocations without state-contingent bonds by exploiting state-contingent non-distortionary capital taxation (Zhu, 1992; Chari, Christiano, and Kehoe, 1994). In representative-agent models the role of complete markets is to provide insurance between the private and public sectors. Capital taxation, then provides a source for state-contingent revenue that can perfectly substitute for contractual transfers arranged in asset markets. However, with heterogeneous workers complete markets also serve the role of insurance between workers, within the private sector. This role cannot be replicated by capital taxation policies which simply provide state-contingent sources of revenue to the government.

The rest of this paper is organized as follows. Section 2 introduces the basic economic environment. Section 3 introduces the linear taxation scenario, while Section 4 derives the main tax smoothing and capital taxation results for this case. Section 5 then does the same for the Mirrleesian tax scenario. Section 6 concludes.

2 The Dynamic Economy

The main purpose of the model is to extend Ramsey frameworks in the direction of incorporating heterogeneity, in the spirit of Mirrlees’s (1971) private information model, in a simple and tractable way so that taxation can be motivated by a desire for redistribution.

To this end, the model economy is populated by a continuum of infinitely lived workers with heterogeneous skills. To focus on uncertainty at the aggregate level, I consider fixed skill differences across workers.² The population is divided into a finite number of types indexed by \( i \in I \), with skill level \( \theta^i \) and relative size \( \pi^i \). Thus, the distribution of skills is summarized by \( \{ \theta^i, \pi^i \}_{i \in I} \). For now, the distribution is taken to be fixed over time. However, in Section 4 aggregate shocks to the skill distribution are also considered. Abler workers produce more efficiency units of labor for any given level of work effort. Importantly, individual skill \( \theta^i \) and work effort \( n^i \) are private information to the worker. Only the product of the two, the efficiency units of labor \( L^i = \theta^i n^i \), is publicly observable. As a result, the government cannot levy discriminatory lump-sum taxes that condition on the worker’s type \( i \in I \), and first-best allocations are not generally attainable.³

Individuals have identical preferences represented by the additively separable utility func-

² This distinguishes this paper from an incipient and growing line of work that attributes to taxation the role of insurance of ongoing shocks to workers’ productivity (e.g. Golosov, Kocherlakota, and Tsivinsky, 2003; Albanesi and Sleet, 2004) by assuming that market cannot provide such insurance arrangements.

³ Alternatively, instead of modeling private information, one can simply assume that taxes cannot be conditioned on worker skills, but only on labor or capital income.
\[ \sum_{t=0}^{\infty} \beta^t \mathbb{E} [u(c_t) - v(n_t)] \] (1)

with an isoelastic specification for the utility functions: \( u(c) = c^{1-\sigma}/(1 - \sigma) \) and \( v(n) = \alpha n^\gamma/\gamma \), with \( \sigma, \alpha > 0 \) and \( \gamma > 1 \). This specification has the advantage of delivering clean analytical results that are likely to be useful benchmarks for other cases.

Aggregate uncertainty is introduced by a publicly observed state \( s_t \in S \) in period \( t \), where \( S \) is some finite set. Both government expenditure and the production function are functions of the state. Let \( s^t \equiv (s_0, s_1, s_2, ..., s_t) \in S^t \) denote the history of states, which is the available information in any period \( t \). No restriction is imposed on the probability distribution governing the evolution of states. Denote the unconditional probability of any history by \( \Pr(s^t) \).

An allocation specifies consumption, labor and capital in every period and history: \( \{c^t(s^t), \ L^t(s^t), \ K_{t+1}(s^t)\} \); aggregates are denoted by \( c(s^t) \equiv \sum_{i \in I} c^t(s^t)\pi^i \) and \( L(s^t) \equiv \sum_{i \in I} L^t(s^t)\pi^i \) (note the notational convention of omitting time subscripts whenever doing so does not lead to confusion). Production combines labor with capital using a constant returns to scale technology. The resource constraints are

\[ c(s^t) + K(s^t) + g_t(s_t) \leq F(L(s^t), K(s^{t-1}), s^t, t) + (1 - \delta)K(s^{t-1}) \] (2)

for all periods \( t = 0, 1, \ldots \) and histories \( s^t \in S^t \).\(^4\) Note that both government expenditures and the production function are allowed to depend on the history \( s^t \) (to capture the impact of uncertainty) as well as on the time period \( t \) (to capture growth or other deterministic changes). To simplify the exposition I have assumed that government expenditures are exogenous and not valued; towards the end of Section 4, however, I show that these assumptions can be relaxed without affecting the results.

### 3 Linear and Proportional Taxation

This section considers the case where in each period the tax schedule is a linear function of labor income: \( \tau(s^t)w_t(s^t)L^t(s^t) + T(s^t) \). The natural case is where the lump sum component \( T(s^t) \) is not restricted. For completeness, however, I also consider the proportional tax case where the lump-sum tax component \( T(s^t) \) is constrained to be zero. In addition to taxing labor income, the government can levy a proportional tax, denoted \( \kappa(s^t) \), on the net return

---

\(^4\) I present the model with a neoclassical production technology, but this could be extended to allow for other features, such as adjustment costs for investment, etc.
Asset markets are assumed to be complete, as in Lucas and Stokey (1983), Chari, Christiano, and Kehoe (1994), and many others. One literal interpretation of this envisions government debt as a rich set of Arrow-Debreu state-contingent bonds. An alternative, less literal, interpretation is provided by the fact that even with non-contingent debt there are other ways of replicating complete market outcomes. For example, Angeletos (2002) and Buera and Nicolini (2004) show how a portfolio composed of riskless bonds of different maturities might be used to this end.

Worker Problem. With complete markets each individual $i \in I$ can be seen as facing a single intertemporal budget constraint:

$$\sum_{t,s} p(s^t)(c^i(s^t) + K^i(s^t) - w(s^t)(1 - \tau(s^t)))L^i(s^t) - R(s^t)K^i(s^{t-1})) \leq B^i_0 - T, \quad (3)$$

where $T \equiv \sum_{t,s} p(s^t)T(s^t)$ is the present value of the lump-sum components of taxes and $B^i_0$ are the initial holding for worker $i \in I$ of short term government bonds. Here $p(s^t)$ represents the Arrow-Debreu price of the consumption good in period $t$ after history $s^t$, normalized so that $p(s_0) = 1$; the real wage is $w(s^t)$; and $R(s^t) \equiv 1 + (1 - \kappa(s^t))(r(s^t) - \delta)$ is the after-tax gross rate of return on capital, where $r(s^t)$ is the rental rate of capital.

Firms. Each period firms maximization profits, $F(L, K, s^t, t) - r(s^t)K - w(s^t)L$, leading to the usual first-order conditions:

$$r(s^t) = F_K(L(s^t), K(s^{t-1}), s^t, t), \quad (4)$$

$$w(s^t) = F_L(L(s^t), K(s^{t-1}), s^t, t). \quad (5)$$

Given constant returns to scale, profits are zero in equilibrium.

Government Budget Constraint. With complete markets the government can also be
seen as facing a single intertemporal budget constraint

$$\sum_{i \in I} B_i^0 + \sum_{t,s^t} p(s^t) g(s^t) \leq T + \sum_{t,s^t} p(s^t) (\tau(s^t) w(s^t) L(s^t) + \kappa(s^t) (r(s^t) - \delta) K(s^t-1))$$

(6)

A version of Walras law implies the government budget constraint holds with equality whenever the resource constraints (2) and the workers’ budget constraints (3) hold with equality.

**Competitive Equilibria.** A competitive equilibrium is a sequence of taxes \(\{T(s^t), \tau(s^t), \kappa(s^t)\}\), prices \(\{p(s^t), r(s^t), w(s^t)\}\), and quantities \(\{c^i(s^t), L^i(s^t), K(s^t)\}\), such that: (i) workers maximize utility: for all individuals \(i \in I\) consumption and labor choices \(\{c^i(s^t), L^i(s^t)\}\) maximize utility in (1) subject to the budget constraint (3), taking prices and taxes as given; (ii) firms maximize profits: the first-order conditions (4)–(5) hold; (iii) the government’s budget constraint (6) holds; (iv) markets clear: the resource constraints (2) hold for all periods \(t\) and histories \(s^t\).

**Characterizing Equilibrium Allocations.** Our first goal is to provide a useful characterization of the set of allocations that are sustainable by a competitive equilibrium for some taxes and prices. This later allows for a primal approach that formulates the taxation problem directly in terms of allocations, which generalizes the method popularized by Lucas and Stokey (1983).

The necessary and sufficient first-order conditions for the workers’ maximization are

$$w(s^t)(1 - \tau(s^t)) = \frac{1}{\theta^i} v'\left(\frac{L^i(s^t)}{\theta^i}\right),$$

(7)

$$\frac{p(s^t)}{p(s_0)} = \beta \frac{u'(c^i(s^t))}{u'(c^i(s_0))} \Pr(s^t),$$

(8)

$$p(s^t) = \sum_{s_{t+1} \in S} p(s^t, s_{t+1}) R(s^t, s_{t+1}),$$

(9)

together with the budget constraint (3) holding with equality. Equation (7) is the intratemporal optimality conditions, equating the marginal rate of substitution of consumption and labor with the real after-tax wage. Equation (8) is the standard intertemporal optimality condition. Equation (9) ensures no arbitrage for investment in capital.

The first-order conditions (7)–(8), combined with the assumption that \(u(c)\) and \(v(n)\) are power functions, imply that individual consumption and labor are proportional to their
aggregates

\[ c^i(s^t) = \omega^i_c c(s^t), \quad \text{(10)} \]

\[ L^i(s^t) = \omega^i_L L(s^t), \quad \text{(11)} \]

with some fixed shares \( \omega^i_c \) and \( \omega^i_L \) with \( \sum_{i \in I} \omega^i_j \pi^i = 1 \), for \( j = c, L \).

Substituting conditions (7)–(11) into the budget constraint (3), and using the fact that

\[ u'(c)c = (1 - \sigma)u(c) \quad \text{and} \quad v'(n)n = \gamma v(n) \]

yields the implementability constraint

\[ \sum_{t, s^t} \beta^t \left( (1 - \sigma)(\omega^i_c)^{1-\sigma}u(c(s^t)) - \gamma \left( \frac{\omega^i_L}{\theta^t} \right)^{\gamma} v(L(s^t)) \right) Pr(s^t) = (\omega^i_c)^{-\sigma} c^0_i (B^0_i + R_0 K^0_i - T). \quad \text{(12)} \]

for each worker type \( i \in I \).

With a representative agent the implementability condition and the resource constraints fully characterize the set of supportable allocations. In contrast, this setup requires an additional restriction to capture the fact that all workers face the same marginal tax on labor. Using the proportionality conditions from (10)–(11) in equation (7) implies

\[ \frac{1}{\theta^t \alpha} \frac{v'(\omega^i_c)}{u'(\omega^i_c)} = \frac{u'(c(s^t))}{v'(L(s^t))} w(s^t)(1 - \tau(s^t)), \]

where I use that with power functions: \( u'(xy) = u'(x)u'(y) \) and \( v'(xy) = v'(x)v'(y)/\alpha \). The left-hand side of this equation depends on \( i \), but not on \( t \) nor \( s^t \); the right-hand side depends on \( t \) and \( s^t \), but not on \( i \). It follows that both sides must equal some constant independent of \( t, s^t \) and \( i \). In particular, this implies a relation between the shares for consumption and labor:

\[ \omega^i_L = \frac{(\omega^i_c)^{\frac{\pi^i}{\gamma}} (\theta^i)^{\frac{\pi^i}{\gamma}}}{\Phi} \quad \text{for all } i \in I \quad \text{where} \quad \Phi \equiv \sum_{i \in I} (\omega^i_c)^{\frac{-\pi^i}{\gamma}} (\theta^i)^{\frac{-\pi^i}{\gamma}} \pi^i. \quad \text{(13)} \]

This is the additional consistency condition needed to characterize equilibrium allocations.

This shows that conditions (10)–(13) together with the resource constraint (2) are necessary for an equilibrium. It turns out that the converse is also true: these equations fully characterize allocations that can be supported as an equilibrium for some tax and price
sequences, obtained from the first-order conditions and the definition of $\Phi$:

$$\tau(s^t) = 1 - \Phi^{1-\gamma} \frac{v'(L(s^t))}{u'(c(s^t))} \frac{1}{F_L(L(s^t), K(s^{t-1}), s^t, t)},$$  \hspace{2cm} (14)$$

$$\frac{p(s^t)}{p(s_0)} = \beta^t \frac{u'(c(s^t))}{u'(c(s_0))} \Pr(s^t),$$  \hspace{2cm} (15)

Factor prices $r(s^t)$ and $w(s^t)$ are given by the marginal product conditions (4)–(5). The tax rates on capital income $\kappa(s^t)$ can be set in any way that satisfies condition (9) with prices (15).\(^8\)

**Proposition 1** An allocation \(\{c^i(s^t), L^i(s^t), K(s^{t-1})\}\) can be supported by a competitive equilibrium if and only if there are distributional shares \((\omega^i_L, \omega^i_c)\) and aggregates \(\{c(s^t), L(s^t)\}\) so that the resource constraint (2) and conditions (10)–(13) hold.

**Proof.** Condition (2) directly implies condition (iv) for an equilibrium. Factor prices given by (4)–(5) ensure that the firm maximizes, so that condition (ii) for an equilibrium is met. Conditions (10)–(13) ensure that consumers are maximizing given taxes and prices given by (14)–(15), so that condition (i) for a competitive equilibrium is met. Finally, condition (iii) is automatically met given that the resource constraints and budget constraints of all individuals hold with equality. \(\blacksquare\)

### 4 Tax Smoothing and Zero Capital Taxation

Based on **Proposition 1** the optimal tax problem can be written as maximizing a weighted sum of utilities

$$\sum_{i \in I} \lambda^i \pi^i \sum_{t, s^t} \beta^t \left( (\omega^i_c)^{1-\sigma} u(c(s^t)) - \left( \frac{\omega^i_L}{\theta^i} \right)^\gamma v(L(s^t)) \right) \Pr(s^t)$$  \hspace{2cm} (16)

subject to the resource constraint (2), the implementability condition (12) and the consistency condition (13).\(^9\) The maximization is performed over the distributional shares \((\omega^i_L, \omega^i_c)\), the aggregates \(\{c(s^t), L(s^t)\}\), if it is assumed to be available as a free variable, the lump-sum tax component $T$.

---

8 Note that with a single worker type $\omega_c = \omega_L = 1$ and $\Phi = 1$. Equation (14) is then the standard expression used in the standard Ramsey literature to back out tax rates.

9 Note that for each individual $i \in I$ the left-hand side of the implementability condition is comparable to their contribution in the objective function except that more relative weight is placed on the disutility of work than on consumption, since $\gamma > 1$ and $1 - \sigma < 1$. 

9
Here $\pi \lambda^i$ represents the Pareto weight placed on worker type $i \in I$. Since the analysis requires no special assumptions on the weights $\{\lambda^i\}$, it does not presume any particular direction for the desired redistribution. An interesting special case that does presume redistribution is desirable from the skilled to unskilled workers is the Utilitarian one, where $\lambda^i = 1$. This case can also be reinterpreted as insurance instead of redistribution: the objective function (16) is equivalent to expected utility behind the “veil of ignorance”—before the skill types $i \in I$ are realized with probability distribution $\{\pi^i\}$.

### 4.1 A Perfect Tax Smoothing Result

The arguments that follow only involve the optimality conditions with respect to aggregate variables, not those related to distributional shares. The first-order conditions with respect to $c(s^t)$ for $t \geq 1$, and $L(s^t)$ for any $t \geq 0$, yield

$$u'(c(s^t)) \sum_{i \in I} (\omega^i)^{1-\sigma} (\lambda^i - (1 - \sigma)\mu^i) \pi^i = \eta(s^t),$$

$$v'(L(s^t)) \sum_{i \in I} \frac{\omega^i}{\theta^i} (\lambda^i - \gamma \mu^i) \pi^i = \eta(s^t) F_L(L(s^t), K(s^{t-1}), s^t, t),$$

where the multiplier on the resource constraint is $\beta^t \eta(s^t) \Pr(s^t)$ and that on the left-hand side of the implementability condition for worker type $i \in I$ is $\mu^i \pi^i$.

The first result concerns the optimal taxation of capital. The first-order condition with respect to capital $K(s^t)$ is

$$\eta(s^t) = \beta \sum_{s_{t+1}} \eta(s^{t+1}) R^*(s^{t+1}) \Pr(s_{t+1}|s^t),$$

where $R^*(s^t) \equiv F_K(L(s^t), K(s^{t-1}), s_t, t) + 1 - \delta$ is the social marginal rate of return on capital. Condition (17) implies that $\eta(s^t)$ is proportional to $u'(c(s^t))$ for $t \geq 1$, so this gives

$$u'(c(s^t)) = \beta \sum_{s_{t+1}} u'(c_{t+1}(s^{t+1})) R^*(s^{t+1}) \Pr(s_{t+1}|s^t)$$

for $t \geq 1$. Thus, capital is undistorted: comparing this condition with condition (9), using prices from (15), reveals that the tax on capital can be set to zero $\kappa(s^t) = 0$ for all $t \geq 2$.

Turning to labor income taxation, dividing equation (18) by (17) and rearranging, one

---

10 The first-order condition for initial period consumption $c(s_0)$ is derived later. It is not crucial for any of our main results. As usual, it is slightly different due to the presence initial wealth.
obtains
\[
\frac{u'(L(s^t))}{u'(c(s^t))} \frac{1}{F_L(L(s^t), K(s^{t-1}), s^t, t)} = \frac{\sum_{i \in I} \omega_i^L(\omega_c^i)^{1-\sigma}(\lambda^i - (1 - \sigma)\mu^i)\pi_i}{\sum_{i \in I} (\frac{\omega_c^i}{\sigma})^\gamma(\lambda^i - \gamma\mu^i)\pi_i}
\]
for all \( t \geq 1 \). Combining this with equation (14) gives
\[
\tau(s^t) = \bar{\tau} \equiv 1 - \frac{\sum_{i \in I} u'(\omega_c^i)\omega_c^i(\lambda^i - (1 - \sigma)\mu^i)\pi_i}{\sum_{i \in I} u'(\omega_c^i)\omega_c^i(\lambda^i - \gamma\mu^i)\pi_i},
\]
which uses equation (13) to replace \( \left(\frac{\omega_c^i}{\sigma}\right)^\gamma \Phi^{-1} \) for \( u'(\omega_c^i)\omega_c^i \). Hence, labor income tax rates are constant across time and states. The next proposition summarizes both results.

**Proposition 2**  Perfect tax smoothing is optimal \( \tau(s^t) = \bar{\tau} \) given by equation (20), for \( t \geq 1 \). The optimal tax on capital is zero \( \kappa(s^t) = 0 \) for all \( t \geq 2 \). Both results hold with or without a lump-sum tax component.

One interpretation for the zero tax on capital result in Proposition 2 is based on the well-known uniform taxation principles due to Diamond and Mirrlees (1971). The assumption that the utility function is isoelastic implies an homotheticity in preferences over consumption paths. This, in turn, implies that consumption at different dates should be taxed uniformly, which is only possible if capital income is untaxed. Basically, the model does not upset the main logic of the Chamley-Judd result\(^{11} \) and allows one to generalize the results derived for representative-agent models with proportional taxation by Zhu (1992) and Chari, Christiano, and Kehoe (1994).

The intuition for the tax smoothing result is best conveyed by considering in turn the cases with and without the lump-sum tax component. With lump-sum taxes distortionary taxation is simply a redistribution mechanism. A positive marginal tax rate is the instrument by which the “rich” pay more taxes than the “poor”, which is desirable whenever redistribution is valued. The optimal tax rate at any point in time balances concerns for redistribution and efficiency. Tax smoothing emerges as long as the determinants of inequality are invariant to government expenditure or aggregate technology shocks. The desired amount of redistribution is then constant over time, and a perfectly constant optimal tax rate results.

In representative-agent Ramsey models tax smoothing results are often informally explained by the following intuition: in order to minimize the total cost from distortions it is optimal to equate the marginal cost of distortions over time by equating tax rates over time (Barro, 1979). The tax smoothing result derived here refines this intuition. Consider

\(^{11} \) Indeed, in different ways, zero capital tax results have been derived allowing for heterogeneity and some forms of redistribution (e.g., see Chamley, 1986; Judd, 1985).
first the more natural scenario where a lump-sum tax component is available. In this case, optimality dictates that the marginal cost of increased distortions be equated to the marginal benefit from increased redistribution. Since the latter is invariant to government expenditure and aggregate technology shocks it follows that the marginal cost from distortions should be equated over time and perfect tax smoothing is optimal. Although the tax rate is invariant to the changes in government expenditure over time, the process \( \{g_t(s^i)\} \) generally affect the desired level of constant taxation \( \bar{\tau} \) through its indirect effect on the resource constraint.

Interestingly, the results regarding tax smoothing and capital taxation hold even when a lump-sum tax component is not available. However, there are some important differences. To begin with, the overall level of taxation cannot be determined by distributional considerations. To take an extreme example, even in the absence of inequality—that is, if skills were identical across workers—distortive taxation is still required. Moreover, for any given distribution of relative skills the distributional concerns are irrelevant. For instance, even if equality is not valued—say, if the weights \( \lambda^i \) are higher for more productive workers—then a positive marginal tax is a necessity in order to meet the government’s budget constraint. In this sense, the level of taxation is driven primarily by budgetary needs, not distributional concerns. This contrasts with the more natural case where a lump-sum tax component is available where distributional concerns are at center stage determining the overall level of taxation.

Turning to the timing of taxes, debt becomes critical when no lump-sum tax component is available. If, instead, a government were forced to balance the budget each period, tax smoothing would simply be infeasible. Debt allows the government to spread out the collection of taxes and meet a single present-value budget constraint. Intuitively, smoothing tax rates is then optimal because it minimizes total efficiency costs by equating the marginal efficiency costs from distortions over time.

The model here nests the standard representative-agent Ramsey framework: restricting the lump-sum tax to zero and taking a degenerate distribution of skills, \( \theta^i = 1 \) for all \( i \). Restricted within this Ramsey case, our results echo Zhu’s (1992), who derived a perfect tax smoothing result with the same preference structure assumed here. Chari, Christiano, and Kehoe (1994) provided a quantitative analysis of optimal policy, within the same setting, but using different preferences. Their simulations, yield minuscule variations in labor income taxes, which they attribute intuitively to a tax smoothing motive. Their numerical results are suggestive of the relevance of the exact analytical result, when one moves away from the class of additively separable isoelastic preferences assumed to derive them.
### 4.2 Distributional Shocks

To bring out the importance of distributional concerns in determining the marginal tax rate, I now extend the model to allow the distribution of relative skills to vary over time, or with the aggregate state of the economy. The skill of a worker of type \( i \in I \) is now given by \( \theta^i_t(s_t) \) in period \( t \).

The first-order conditions from the worker’s maximization problem are now

\[
\alpha \left( \theta^i_t(s_t) \right)^{-\gamma} (L^i_t(s^t))^{\gamma-1} = \nu^i w(s^t)p(s^t)(1 - \tau(s^t)),
\]

(21)

\[
(c^i_t(s^t))^{-\sigma} = \nu^i p(s^t),
\]

(22)

where \( \nu^i \) is the multiplier individual \( i \in I \) has on its own budget constraint (3). It follows that \( c^i_t(s^t) = \omega^i_t c(s^t) \) and \( L^i_t(s^t) = \omega^i_L(s_t) L(s^t) \) with

\[
\omega^i_{L,t}(s_t) \equiv \frac{(\omega^i_t)^{-\gamma} (\theta^i_t(s_t))^{\gamma-1}}{\Phi_t(s_t)}.
\]

(23)

for all \( i \in I \), where \( \Phi_t(s_t) \equiv \sum_{i \in I} (\omega^i_t)^{-\gamma} (\theta^i_t(s_t))^{\gamma-1} \pi^i \), so that \( \sum_{i \in I} \omega^i_{L,t}(s_t) = 1 \). The shares of labor are determined in each period as a function of the skill distribution \( \{\theta_t(s_t)\} \) and the fixed shares of consumption are determined. The fixed shares for consumption \( \omega^i_s \) encapsulate the same information than the multipliers \( \nu^i \).

Dividing the worker’s two first-order conditions (21) and (22) implies

\[
\frac{1}{\theta^i_t(s_t)} \alpha \frac{v'(\omega^i_{L,t}(s_t))}{v'(\omega^i_t)} = \frac{u'(c(s^t))}{u'(L(s^t))} w_t(s^t) \left( 1 - \tau(s^t) \right) = \Phi_t(s_t)^{1-\gamma},
\]

(24)

which generalize the consistency conditions in equations (13) and (14).

The rest of the analysis proceeds along the same lines as before. The planning problem is to maximize (16) subject to the resource constraint (2), the implementability condition (12) and the restriction on individuals shares in (23) (implying the consistency condition in equation (24)). The first-order conditions with respect to the aggregates \( c(s^t) \) and \( L(s^t) \) are essentially identical to equations (17)–(18), as are the expression for \( w_t(s^t)u'(c(s^t))/v'(L(s^t)) \)

---

\(^{12}\) The equilibrium individual allocations \( \{c^i_t(s_t), L^i_t(s_t)\}_{i \in I} \) solve in each period \( t \) and history \( s^t \) the static subproblem of maximizing \( \sum_{i \in I} (\nu^i)^{-1} (u(c^i_t(s^t)) - v(L^i_t(s^t))) \) subject to \( \sum_{i \in I} L^i_t(s_t) = L_t(s^t) \) and \( \sum_{i \in I} c^i_t(s_t) = c_t(s_t) \), taking as given the aggregates \( c_t(s_t) \) and \( L_t(s_t) \). Linear taxation ensures that the distribution of consumption and labor across workers is efficient; inefficiency emerges in the aggregate choice between consumption and labor.
in equation (19). Using this expression in equation (24) yields
\[
\tau(s^t) = \bar{\tau}(s_t) \equiv 1 - \Phi_t(s_t) 1 - \frac{\sum_{i \in I} u'(\omega_i^c) \omega_i^c (\lambda^i - (1 - \sigma) \mu^i)}{\sum_{i \in I} \left( \frac{\omega_i^L(s_t)}{\theta_i^t(s_t)} \right)^\gamma (\lambda^i - \gamma \mu^i)}
\]
\[
\Rightarrow \bar{\tau}(s_t) = 1 - \frac{\sum_{i \in I} u'(\omega_i^c) \omega_i^c (\lambda^i - (1 - \sigma) \mu^i)}{\sum_{i \in I} u'(\omega_i^c) \omega_L^t(s_t)(\lambda^i - \gamma \mu^i)},
\]
where the second equality uses equation (24) to replace \((\omega_i^L(s_t)/\theta_i^t(s_t))^{\gamma} \Phi_t(s_t)^{\gamma-1}\) with \(u'(\omega_i^c)\omega_L^t(s_t)\).

The optimal tax representation in equation (25) generalizes that in equation (14). The only difference is that now an individual’s share of labor income \(\omega_L^t(s_t)\) potentially varies over time and with the current state, due to underlying variations in the skill distribution. Indeed, from equation (23) one sees that the shares \(\{\omega_i^L(s_t)\}\), and hence the tax rate \(\bar{\tau}(s_t)\), is solely a function of the current distribution of skills \(\{\theta_i^t(s_t)\}\). Thus, tax smoothing continues to be optimal in that tax rates remain unresponsive to shocks that affect only government expenditure or aggregate technology. The only source of variations in tax rates are changes in the distribution of relative skills. For this extended model, the results in Proposition 2 on zero capital taxation are unchanged. I summarize these results in the next proposition.

**Proposition 3** The optimal tax rate on labor income is history independent: \(\tau(s^t) = \bar{\tau}(s_t)\). It depends on time \(t\) or the current aggregate state \(s_t\) solely through the current distribution of skills \(\{\theta_i^t(s_t)\}_{i \in I}\). For any given skill distribution the optimal tax rate does not vary with government expenditure or aggregate technology. The optimal tax on capital is zero \(\kappa(s^t) = 0\) for all \(t \geq 2\). These results hold with or without a lump-sum tax component.

Movements in the distribution of relative skills turn out to be the only source for tax rate fluctuations in this model. This underscores the point made earlier, that a crucial determinant for tax rates is distributional concerns. Indeed, as discussed above, when a lump-sum tax component is available distributional concerns are the main determinant of the overall level of tax rates. Proposition 3 can be seen as generalizing this notion by showing that fluctuations in the distribution of skills lead to optimally fluctuating tax rates over time.

Intuitively, with a redistributive motive optimal tax rates will be higher during times of higher dispersion of relative skills. The reason is that the redistribution from “rich” to “poor” workers that can be engineered by labor income taxation is more powerful then. Recall the intuition that, with a lump-sum tax component, the marginal cost from distortions should equal the marginal benefit from increased redistribution in each period. Then, as long as the skill distribution does not vary, the marginal benefit from redistribution is unchanging.
Thus, the marginal cost from distortions should be equated over time, which in turn implies that tax rates should be constant. However, when the distribution of skills does shift, the marginal benefit from redistribution shift with it, so the marginal cost from distortions should not be equated over time. As a result, the optimal tax rate responds to such shifts.

Interestingly, changes in the distribution of skills affect tax rates with or without the availability of a lump-sum tax component. As argued previously, if no lump-sum tax component is available, then distributional concerns simply cannot shape the overall level of tax rates. They can, however, affect their timing. For example, during times of high skill dispersion equation (23) shows that labor income becomes more disperse, so that a greater fraction of taxes are paid by those who have the most. If redistribution is valued, then it is may be optimal to concentrate distortionary taxation then.

**Initial Period Taxation.** I now turn to the determinants of optimal taxation in the very first periods. Since only the very first period require special analysis, Proposition 2 and Proposition 3 already contain the most relevant implications for the dynamics properties of optimal tax rate paths. However, exploring initial taxation is important for understanding potential time inconsistency problems, some of which are addressed further in the next subsection.

There are two distortive taxes to consider. First, there is the initial tax rate on labor income $\tau_0$. Second, there is the capital income tax $\kappa(s^1)$ which distorts investment in the initial period. Finally, there is also the initial time-zero capital levy $\kappa_0$ which is not distortive, discussed in the next subsection.

The first-order condition for initial aggregate consumption $c_0$ contains a few extra terms relative to equation (17):

$$
u'(c_0) \sum_{i \in I} u'(\omega_c^i)(\omega_c^i(\lambda^i - \mu^i) + (B_0^i + R_0K_0^i + T)\mu^i)\pi^i = \eta_0.$$ 

Of course, if no lump-sum tax is available the same condition holds but with $T = 0$. When a lump-sum tax is available, the first-order condition with respect to $T$ is $\sum_{i \in I} u'(\omega_c^i)\mu^i\pi^i = 0$. In both cases, it follows that the term involving $T$ vanishes:

$$
u'(c_0) \sum_{i \in I} u'(\omega_c^i)\mu^i\pi^i = \eta_0.$$

Now, unless the term involving initial wealth $B_0^i + R_0K_0^i$ also drops out, the initial tax rate $\tau_0$, determined by equation (14), will be affected and differ from the constant tax rate $\bar{\tau}$ found for all other periods. Likewise, the tax on capital income $\kappa(s^1)$ will not be zero. However, the term involving capital does drop out in two important cases.
First, even if the initial capital levy is restricted, if there is no initial inequality in asset wealth so that \( B_i^0 + R_0 K_i^0 = B_j^0 + R_0 K_j^0 \), then the term involving wealth also drops out using the first-order condition for \( T \). This is an interesting benchmark as it corresponds to the canonical optimal taxation situation where heterogeneity is due solely to skill differences \( \text{(Mirrlees, 1971)} \). In this case the conclusions in Proposition 2 extend and \( \tau(s^t) = \bar{\tau} \) for all \( t \geq 0 \), and \( \kappa(s^t) = 0 \) for all \( t \geq 1 \). Moreover, in this case the time-zero capital levy \( \kappa_0 \) can be set to zero and any ad hoc restriction on initial wealth taxation is nonbinding.\(^{13} \)

Second, whenever the time-zero capital levy \( \kappa_0 \) is unrestricted, its first-order condition yields \( \sum_{i \in I}(\omega'_i)^{-\sigma} \mu_i^i K_i^0 \pi^i = 0 \). As a result, if there is no differences in initial bond holdings \( B_i^0 = B_j^0 \), then the first-order condition for \( c_0 \) is identical to that of any other periods. Once again, the conclusions from Proposition 2 extend to all periods.

### 4.3 Some Implications and Differences with Ramsey

The previous two subsections have shown that the heterogenous agent model delivers tax smoothing and zero capital results. Although these results largely support the conclusions of representative-agent model, I stressed the important difference by highlighting how distributional concerns determine tax rates when a lump-sum tax is available. In this subsection I discuss additional implications of the model, focusing on three issues that differ with standard representative-agent Ramsey settings.

**Debt Management.** Since Barro \( \text{(1979)} \), second-best tax problems have been used to avoid the neutrality results implied by Ricardian equivalence. In Ramsey models the optimal timing of taxes implies an optimal management of debt. Barro was the first to argue that distortionary tax rates should be smoothed: by analogy with permanent income theory, tax rates should be set with an eye towards permanent government spending, as opposed to current spending. As a result, government debt should be used to buffer any resulting deficits and surpluses. Lucas and Stokey \( \text{(1983)} \) extended this argument by allowing state-contingent debt: then taxes should also be smoothed across states of the world, as well as across time. Both models share the essential feature that the solution to the tax problem determines a debt management policy. This is the case because, with proportional taxation, average and marginal taxes coincide. The same is true in the present model when the lump-sum tax component is arbitrarily ruled out. Government debt is then key to smoothing tax rates over time, just as in representative-agent Ramsey models. Optimal tax policy then uniquely determines an optimal debt management policy.

\(^{13} \) However, it seems difficult to justify restricting the taxation of initial wealth. Moreover, if explicit taxes on wealth are limited, consumption taxes could perfectly replicate their effects.
However, the link between revenue and marginal taxes is broken when a lump-sum component is available. Marginal tax rates do not determine current revenue: Ricardian equivalence is recovered, rendering the debt level indeterminate. Indeed, government debt is simply irrelevant in that nothing is lost if the government is required to balance its budget each period—the lump-sum component can do all the work.\footnote{However, note that even though the government may not need to issue bonds, the asset market may still be important to allow the heterogeneous workers to trade with each other.}

**Capital Levies and Time Inconsistency.** In Ramsey models a striking contrast emerges between long-run and short-run capital tax prescriptions: eventually capital should go untaxed, but initially it should be taxed heavily. Time-zero capital levies provide revenues without distortions, mimicking the desired missing lump-sum tax.\footnote{Hence, to avoid the first-best, most analyses proceed by imposing ad-hoc upper bounds on the amount of such levies. In contrast, here even when such levies are initially unrestricted a nontrivial distortionary taxation problem remains because of distributional concerns.} This tension, between long-run and short-run tax prescriptions, provides a source for time inconsistency of government policy.

In contrast, time-zero capital levies may be completely irrelevant in the present model. Indeed, the reason for their irrelevance is precisely what makes them so desirable in Ramsey models: capital levies that imitate lump-sum taxes bring nothing new to the table when lump-sum taxes themselves are already available. Thus, if positive time-zero capital levies turn out to be optimal, it must be for different reasons than in the standard Ramsey settings.

Capital levies cease to be neutral if initial asset holdings are unequal. For example, consider an extreme case where more productive workers are also wealthier, so that $\theta^i > \theta^j$ implies $K^i > K^j$. A proportional tax on initial wealth then acts as an ideal redistributive device, taking more from the rich, as income taxation does, but without introducing distortions. In such a case, as long as equality is valued, an initial wealth tax is desirable. Indeed, Pareto improvements may be possible if the tax on assets is coupled with a reduction in the distortionary tax rate on labor income.

In a nutshell, the standard Ramsey setup is about the need to “redistribute” from the private to the public sector, to finance the latter. Any initial wealth in the hands of the private sector is best expropriated. In contrast, in the present model the government also needs resources from the private sector, but the central tension is not getting these resources without distortions—which could always be done through the lump-sum tax. Rather, it is the distributional concern about who the government is extracting resources from. Instead of redistribution from private to public sector, it is redistribution within the private sector that crucial.

Thus, a desire for initial wealth taxation can also be generated in this framework. More-
over, this may create a source for time-inconsistent policy if more productive workers tend to accumulate more assets over time.\textsuperscript{16} However, the mechanism is entirely different and suggests new issues, that cannot be addressed by representative-agent Ramsey settings, such as the distribution of assets within the private sector.

Indeed, a complete characterization of the time-inconsistency issue requires address not only the wealth distribution and its evolution, but also about its composition. That is, heterogenous workers may take different positions in physical capital and Arrow-Debreu securities, or state-contingent bonds of various maturities.\textsuperscript{17} A complete study of the distribution of asset holdings, and their role in the time-inconsistency problem, is beyond the scope of this paper.

**Replicating Completing Markets with Taxes.** For representative-agent Ramsey settings, Zhu (1992) and Chari, Christiano, and Kehoe (1994) have show that capital taxation can help implement complete-market outcomes even without state-contingent bonds. When markets are complete the tax rate on capital can be set in advance, so that at date \( t \) it is conditioned only on \( s_t^{-1} \); the relevant tax rate is then known at the moment of investment. However, if one conditions further on \( s_t \), so that in addition to \( s_t^{-1} \) the tax now depends on \( s_t \), one can replicate any state-contingent profile of revenue, without introducing additional distortions to investment. As a result, state-contingent debt becomes inessential when the government can tax capital flexibly enough.

In the present model, however, such a scheme will generally not work. The reason is related to our previous discussion on the role that capital levies play here: redistribution. The lump-sum tax already provides a non-distortive source of state-contingent revenue. However, when markets are incomplete more is missing than a source of state-contingent revenue for the government. Complete markets also provide insurance across heterogeneous workers; replicating complete markets then requires also replicating these insurance arrangements. It is easy to see that, in general, a proportional tax on capital will not do the trick.

### 4.4 Two Generalizations

I now consider a two generalizations of the economic environment that do not affect our conclusions for optimal taxes. I begin by discussing shocks to preferences and then turn to endogenizing government expenditures.

\textsuperscript{16} A similar, more direct, time-inconsistency issue arises if taxation based on past income were possible. That is, an unanticipated tax on past income can redistribute across permanent skills without distorting present or future work effort choices.

\textsuperscript{17} Lucas and Stokey (1983) showed that in a representative-agent Ramsey model the maturity structure of bonds is key for determining the time-consistency of labor income taxes.
Taste Shocks. Suppose utility is given by
\[
\sum_{t,s} \beta^t \left( A(s^t, t) u(c^t_i(s^t)) - B(s^t, t) v(n^t_i(s^t)) \right) \Pr(s^t),
\]  
(26)
for some strictly positive \(A(\cdot)\) and \(B(\cdot)\) functions. This specification creates more flexibility along two dimensions. First, it allows for shocks to the marginal rate of substitution of work and consumption from movements in \(B(s^t, t)/A(s^t, t)\). Intuitively, these shocks shift the labor supply schedule. Formally, they are equivalent to the “wedges” often stressed in the business-cycle literature (e.g. Chari, Kehoe, and McGrattan, 2006), which account for a sizeable fraction of the variation in labor. Second, the more general formulation of utility also captures taste shocks to the marginal utility of consumption, in \(A(s^t, t)\). These shocks affect the stochastic discount factor, defined by \(A(s^t, t) u'(c(s^t))\), which is used for asset pricing. Thus, this additional flexibility could be used to ensure that the model is not add odds with asset-returns data.

The change in utility affects the implementability condition, and some other equations. Basically, in the preceding analysis wherever the algebra previously called for \(u'(c(s^t))\) it now requires \(A(s^t, t) u'(c(s^t))\), and similarly \(v'(L(s^t))\) should be substituted for \(B(s^t, t) v'(L(s^t))\). It turns out, that these changes do not affect the bottom line: the tax formulas (20) or (25) remain unchanged. Hence, the optimal tax rate should remain invariant to these aggregate taste shocks.

It is also possible to allow the curvature parameters \(\sigma\) and \(\gamma\) to vary over time and depend on the history of shocks and government expenditure. Once again, the optimal tax formulas (20) or (25) still apply, but they now imply that optimal taxes will react to shocks that affect these parameters. This is quite natural, since the curvature parameters affect, among other things, the desired amount of distribution and the elasticity of labor supply. As before, taxes remain invariant to technology or expenditure shocks that do not affect these parameters or the distribution of skills.

Non-wasteful and Endogenous Government Expenditures. Up to now, I have assumed that government expenditures are exogenous and not valued, for production or utility. I now show that these assumptions can easily be relaxed, without affecting the results.

First note that, although government expenditures do not enter the production function explicitly, the history of states \(s^t\) is allowed to affect production in a general way. It follows that this formulation implicitly captures any effect that the history of government expenditure \(g^t(s^t) \equiv (g_0(s_0), g_1(s^1), \ldots, g_t(s^t))\), which is simply a function of \(s^t\), may have on production possibilities at time \(t\). This captures, for example, the idea that the stock of public infrastructure, which is related to the investment in infrastructure, past and present,
can affect the efficiency of private production.

As for utility, all the results go through when government expenditures are valued according to: 
\[ A(g^t, s^t, t)u(c^t_i(s^t)) - B(g^t, s^t, t)v(n^t_i(s^t)) + w(g^t, s^t). \]
The separable utility function \( w(\cdot) \) plays no role whatsoever in worker decisions and, consequently, does not affect the results for optimal taxes. As for the multiplicative factors, the utility specification (26) already implicitly captures the effects of government expenditures since \( g^t(s^t) \) is a function of the history \( s^t \)

Finally, for simplicity I have also taken the process of government expenditures as exogenously given, but it can be modeled endogenously without affecting the results. In essence the planning problem considered above can be interpreted as the tax subproblem that results after government expenditure have been chosen optimally, yielding some solution \( g^*_t(s^t) \). That is, taking this solution as given I have characterized the subproblem that optimizes over tax policy.

5 Mirrleesian Taxation: Constrained Efficiency

This section considers the Mirrleesian scenario, no restrictions are placed on tax instruments and constrained-efficient allocations are characterized. The tax schemes required to implement these allocations are necessarily more complicated than the linear schemes considered thus far. The main goal is to characterize the shadow marginal tax rates for constrained efficient allocations, to examine the optimality of tax smoothing and zero capital taxation. Towards the end of this section, a simple tax scheme that implements the constrained efficient allocation for a version of the model with no uncertainty is discussed.

The results in this section no longer require assuming that the utility function from consumption be isoelastic. The disutility of work effort function, on the other hand, is still assumed isoelastic: \( v(n) = \alpha n^\gamma \). I also work directly with the extended environment that allows the distribution of skills to shift over time, or with the aggregate state, but abstract from the other extensions, such as aggregate taste shocks.

5.1 The Planning Problem

Invoking the revelation principle, one can set up a direct truth-telling mechanism: workers submit reports \( j \in I \) regarding their skill type \( i \in I \) and receive an allocation as a function of this report. To ensure that individuals report truthfully the following incentive compatibility

\[ * \] One can also generalize by modeling government expenditure as a vector, so that some elements may affect mostly production (e.g. infrastructure), while other affect mostly utility (e.g. public services such as security).
constraints must be imposed:

\[
\sum_{t,s^t} \beta^t \left( u(c^i(s^t)) - v \left( \frac{L^i(s^t)}{\theta^i_t(s_t)} \right) \right) \Pr(s^t) \geq \sum_{t,s^t} \beta^t \left( u(c^j(s^t)) - v \left( \frac{L^j(s^t)}{\theta^j_t(s_t)} \right) \right) \Pr(s^t), \tag{27}
\]

for all \( i, j \in I \).

The planning problem maximizes the weighted sum of utilities in (16) subject the resource constraints (2) and the incentive constraints (27). This problem allows the richest set of possible tax instruments, for example, non-linear taxation of income and capital—it is only constrained by the assumed asymmetry of information.\(^{19}\)

Let the multiplier on the incentive constraint for worker type \( i \) reporting to be type \( j \) be \( \psi^{i,j} \pi^i \). The Lagrangian that incorporates all these constraints is then

\[
\mathcal{L} \equiv \sum_{i,j,t,s^t} \beta^t \left( \lambda^i + \psi^{i,j} \right) \left( u(c^i(s^t)) - v \left( \frac{L^i(s^t)}{\theta^i_t(s_t)} \right) \right) - \psi^{i,j} \left( u(c^j(s^t)) - v \left( \frac{L^j(s^t)}{\theta^j_t(s_t)} \right) \right) \Pr(s^t) \pi^i.
\]

By exploiting the fact that \( v(L/\theta^i) = (\theta^i/\theta^j)^{\gamma} v(L/\theta^j) \), one can rewrite the Lagrangian more compactly as

\[
\mathcal{L} = \sum_{i,t,s^t} \beta^t \left( \varphi^i_c u(c^i(s^t)) - \varphi^i_{L,t}(s_t) v \left( \frac{L^i(s^t)}{\theta^i_t(s_t)} \right) \right) \Pr(s^t) \pi^i,
\]

where

\[
\varphi^i_c \equiv \lambda^i + \sum_j (\psi^{i,j} - \psi^{j,i}) \quad \text{and} \quad \varphi^i_{L,t}(s_t) \equiv \lambda^i + \sum_j \left( \psi^{j,i} - \left( \frac{\theta^i_t(s_t)}{\theta^j_t(s_t)} \right)^{\gamma} \psi^{i,j} \right).
\]

### 5.2 Tax Smoothing with Nonlinear Taxation

The first-order conditions for consumption, labor and capital are

\[
\varphi^i_c u'(c^i(s^t)) = \eta(s^t), \tag{28}
\]

\[
\varphi^i_{L,t}(s_t) \frac{1}{\theta^i_t(s_t)} v' \left( \frac{L^i(s^t)}{\theta^i_t(s_t)} \right) = \eta(s^t) F_L(L(s^t), K(s^{t-1}), s^t, t), \tag{29}
\]

\[
\eta(s^t) = \beta \sum_{s_{t+1}} \eta(s^{t+1}) R^*(s^{t+1}) \Pr(s_{t+1} \mid s^t), \tag{30}
\]

where \( \beta' \eta(s^t) \Pr(s^t) \) is the multiplier on the resource constraint at time \( t \) after history \( s^t \), and where \( R^*(s^t) \equiv F_K(L(s^t), K(s^{t-1}), s^t, t) + 1 - \delta \) is the marginal social return to capital.

\(^{19}\) An alternative interpretation is to take skills as observable, but to simply assume that taxation can only be based on labor and capital income, not directly on skills.
Combining the first-order conditions for consumption (28) with that for capital (30) yields
\[ u'(c'(s^t)) = \beta \sum_{s_{t+1}} u'(c'_{t+1}(s^{t+1})) R^*(s^{t+1}) \Pr(s_{t+1} \mid s^t), \tag{31} \]
the standard undistorted intertemporal Euler equation: capital accumulation is undistorted. \(^{20}\)

Define the implicit marginal tax rate on labor as the solution to
\[ \frac{1}{\theta_i^t(s_t)} \frac{v'(L^i(s^t) \theta_i^t(s_t))}{u'(c'(s^t))} = F_L(L(s^t), K(s^{t-1}), s^t, t)(1 - \tau^i(s^t)). \]
Combining the first-order conditions (28)–(29) gives
\[ \tau^i(s^t) = 1 - \frac{1}{\theta_i^t(s_t)} \frac{v'(L^i(s^t) \theta_i^t(s_t))}{u'(c'(s^t))} \frac{1}{F_L(L(s^t), K(s^{t-1}), s^t, t)} = 1 - \frac{\varphi_{ct}^i}{\varphi_{L,t}^i(s_t)} \equiv \bar{\tau}_i^t(s_t), \]
Note that \( \varphi_{L,t}^i(s_t) \) only depends on time \( t \) and on the state of the economy through the distribution of skills. The following proposition collects these results.

**Proposition 4** At the constrained efficient allocation: (a) capital is not distorted and the standard Euler equation (31) holds; (b) each worker type faces an implicit marginal tax rate on labor income \( \tau^i(s^t) = \bar{\tau}_i^t(s_t) \) for all \( t \) and \( s^t \), that depends only on the current skill distribution \( \{\theta_i^t(s_t)\} \). In particular, if the skill distribution is fixed then the optimal tax rate faced by each worker type is constant over time, \( \bar{\tau}_i^t(s_t) = \bar{\tau}^i \).

This result provides a simple benchmark for zero capital taxation and constant marginal tax rates on labor income. Tax smoothing is optimal here in the sense that marginal tax rates on each worker type are constant across time and states that do not affect the distribution of skills, as in Proposition 3. However, unlike the linear tax case, here the optimal tax schedule is nonlinear in that marginal tax rates generally vary across individuals.

It is useful to note that the optimal allocation maintains the ratio of marginal utilities \( u'(c'(s^t)) / u'(c'(s^t)) \), between any two worker types, constant. If one assumes a power utility function then this implies that \( c'(s^t) = \omega_{ct}^i(s^t) \) for some constant shares \( \omega_{ct}^i \). Similarly, the tax smoothing result implies that \( L(s^t) = \omega_{L,t}^i(s_t)L(s^t) \), where the share \( \omega_{L,t}^i(s_t) \) depends on time and the state only insofar as these affect the distribution of skills \( \{\theta_i^t(s_t)\} \). \(^{21}\) Although

\(^{20}\) Distorting the standard Euler condition is optimal if there are ensuing privately observed skill shocks at the individual level (Diamond and Mirrlees, 1977; Rogerson, 1985; Werning, 2002; Golosov, Kocherlakota, and Tsivinski, 2003; Furhi and Werning, 2005).

\(^{21}\) Using the workers’ first-order conditions one sees that \( \omega_{L,t}^i(s_t) \) is proportional to \( (\varphi_{ct}^i) \frac{\bar{\tau}^i(s_t)}{\bar{\tau}^i(s_t)} (\varphi_{ct}^i) \frac{\bar{\tau}^i(s_t)}{\bar{\tau}^i(s_t)} \) with the constant of proportionally set so that \( \sum_{i \in I} \omega_{L,t}^i(s_t) \pi^i \).
a similar feature held in the case of linear taxation there it was implied directly by the restrictions imposed by a competitive equilibrium; in contrast, here the planner could do otherwise, but simply finds such an arrangement optimal. Finally, it is useful to note that although the individual assignments can be characterized in terms of the shares \( \{\omega^i_c, \omega^i_L(s_t)\} \) of aggregates, these do not satisfy a consistency condition, such as (13) or (23), because the marginal tax here is not generally equalized across workers.

### 5.3 Income Tax Averaging

Suppose the skill distribution does not vary over time and that there is no aggregate uncertainty. The analysis then suggests tax schemes that somehow equates marginal tax rates over time, but allows these to vary across workers. One arrangement that works is taxation based on income averages, as opposed to taxes based on current income. Such rules were advocated by Vickery (1947) for different reasons.\(^{22}\)

This tax implementation works as follows. The government does not tax capital income. It sets a non-linear income tax payment as a function of the present value of lifetime earnings, \( \sum_{t=0}^{\infty} p_t w_t L^i_t \). In the first period workers must pay

\[
\Psi \left( \sum_{t=0}^{\infty} p_t w_t L^i_t \right),
\]

for some, potentially nonlinear, tax function \( \Psi \) (recall the normalization that \( p_0 = 1 \)). Workers of type \( i \in I \) maximize \( \sum_t \beta^t (u(c_t) - v(L^i_t/\theta^i)) \) subject to their budget constraint \( \sum_t p_t c_t \geq \sum_t p_t w_t L^i_t - \Psi \left( \sum_t p_t w_t L^i_t \right) \). Assuming differentiability of the tax function \( \Psi \), their first-order conditions yield

\[
\frac{1}{\theta^i} \frac{u'(L^i_t/\theta^i)}{u'(c^i_t)} = \left( 1 - \Psi' \left( \sum_{t=0}^{\infty} p_t w_t L^i_t \right) \right) w_t \quad t = 0, 1, \ldots
\]

This shows that the derivative \( \Psi'(\cdot) \) can play the role of the constant marginal tax rate in all periods \( \bar{\tau}^i \). In fact, one can always find a smooth tax function \( \Psi \) that works and set its derivatives, evaluated at the optimal equilibrium values of \( p_t w_t L^i_t \), to \( \bar{\tau}^i \). In this way, taxation based on income averages automatically ensures that workers’ marginal tax rates are constant over time, while allowing these to vary across workers; as required by

\(^{22}\)Vickery (1947) argued based of horizontal equity: if the tax schedule for current income is convex then individuals with highly fluctuating earnings would pay more taxes on average than individuals with steadier earnings. Here, in contrast, they serve to implement constant marginal income tax rates over time, while retaining the non-linearities across individuals.
Proposition 4. I summarize this implementation result in the next proposition.

**Proposition 5** In a deterministic economy and the skill distribution does not vary then the optimal allocation can be implemented as a competitive equilibrium with a nonlinear tax schedule \( \Psi(\cdot) \) on the present value of labor income and no tax on capital income.

One can also modify the tax scheme slightly so that, instead of paying taxes in the first period as a function of future labor earnings, the worker is taxed in each period as a function of the present value of past labor income only.

### 6 Conclusions

This paper provided a tractable framework to address issues of optimal taxation in dynamic economies. Unlike representative-agent Ramsey models, distortive taxation is microfounded by a concern for redistribution. This framework proved tractable for rich specifications of the dynamic economy, such as those used in representative-agent Ramsey analyses.

The results in this paper provide interesting benchmarks for perfect tax smoothing of labor income taxes and for zero taxation of capital income. Although the mechanisms and insights are quite different, it is of interest that a microfounded model of taxation does not disturb these two cornerstone results in Ramsey tax theory. The model does suggest a novel source for variations in optimal tax rates—that is, for deviating from perfect tax smoothing. In particular, movements in the relative skill distribution induce changes in the optimal amount of redistribution, and thus, in the optimal tax rate. In addition, compared to standard Ramsey settings, I show that the nature of the time inconsistency problem is different and that it is not possible to replicate complete market without state-contingent bonds by simply using state-contingent capital taxation.

Finally, unlike Ramsey settings without lump-sum taxation, the model presented here recovered a form of Ricardian equivalence that renders debt management indeterminate. I conclude by speculating that some extensions of the model that overcome perfect Ricardian neutrality are likely to provide a determinate theory of debt management, without affecting the central conclusions for tax rates. One such possibility is to model some individuals as having limited participation in asset markets.\(^{23}\) For example suppose, to take an extreme, that one of the worker types \( i \in I \) is hand-to-mouth, with no initial assets and no access to asset markets whatsoever. The desire to smooth consumption for such nonparticipants, and hence their income net of taxes, may then pin down the lump-sum tax component, and public debt with it.

\(^{23}\) Another interesting direction may be an overlapping generations framework.
References


Farhi, Emmanuel, 2005. “Capital Taxation and Ownership When Markets are Incomplete.” Mimeo. 6


