OPTIMAL FISCAL POLICY WITH REDISTRIBUTION*

IVÁN WERNING

I study the optimal taxation of labor and capital in a dynamic economy subject to government expenditure and technology shocks. Unlike representative-agent Ramsey models, workers are heterogenous and lump-sum taxation is not ruled out. I consider two tax scenarios: (a) linear taxation, with a lump-sum intercept and (b) nonlinear-Mirrleesian taxation. When taxes are linear, I derive a partial-equivalence result with Ramsey settings that provides a reinterpretation of such analyses. I find conditions for perfect tax smoothing of labor-income taxes and zero capital taxation. Implications that contrast with Ramsey are derived for public-debt management, for the nature of the time-inconsistency problem and for the viability of replicating complete markets without state-contingent bonds. Shifts in the distribution of skills provide a novel source for variations in tax rates. For the nonlinear tax scenario, I show that taxation based on income averages is optimal.

I. 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 offers two important insights into this question: taxes on labor income should be smoothed [Barro 1979; Lucas and Stokey 1983; Kingston 1991; 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 standard Ramsey approach adopts a representative-agent framework; then, to avoid a first-best outcome, lump-sum taxes— or any combination of tax instruments that may replicate them—are simply ruled out. Societies may have good reasons for avoiding complete reliance on lump-sum taxes, but none of these are captured by a representative-agent Ramsey framework. Although the first-best allocation is

* This paper is an extended and revised version of the first part of my Ph.D. dissertation at the University of Chicago [Werning 2002, Chapters 2 and 3], which later circulated as a working paper titled "Tax Smoothing with Redistribution." I am indebted to my thesis committee: Fernando Alvarez, Gary Becker, Pierre-André Chiappori, and Robert Lucas. I am grateful for helpful comments and suggestions from Manuel Amador, Marios Angeletos, Robert Barro, Paco Buera, Emmanuel Farhi, Narayana Kocherlakota, Mike Golosov, and Pierre Yared. Pablo Kurlat provided invaluable research assistance as well as influential suggestions. I thank Emily Gallagher for excellent editing assistance. The hospitality of the Federal Reserve Bank of Minneapolis during my visit there is greatly appreciated. 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.

The Quarterly Journal of Economics, August 2007

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 unspecified reasons, lump-sum taxes are presumed undesirable, yet still highly desirable within the model, how can we be sure that tax prescriptions derived are not, for the same unspecified reasons, also socially undesirable?

In contrast, distributional concerns provide a natural rationale for distortionary taxation [Mirrlees 1971; Sheshinski 1972]. For instance, when workers differ in their labor productivity, and this trait is not observable— or if, for some other reason, taxes simply cannot be conditioned upon them—then almost all first-best allocations are unattainable. A trade-off emerges between redistribution and efficiency, providing a foundation for distortionary taxes.

In its favor, one virtue of the more *ad hoc* Ramsey approach has been its tractability for the study of rich dynamic stochastic general equilibrium models, such as those common in the growth and business-cycle literatures. In contrast, most research incorporating heterogeneity works with relatively stylized environments. What is missing is a framework in which distortionary taxes arise naturally that is tractable within rich dynamic environments. With this in mind, this paper reexamines optimal taxation in dynamic economies close to those used by representative-agent Ramsey models—such as Chari, Christiano, and Kehoe [1994] and others—while modeling distributional concerns explicitly and allowing for a richer tax structure.

The model economy is inhabited by workers that differ in the productivity of their work effort. Technology is neoclassical, with capital and labor services combining to produce a single good that can be consumed or invested. The economy is subject to fluctuations in government expenditures and technology. I consider two scenarios for the set of available tax instruments: (i) taxation is linear, allowing for an arbitrary lump-sum tax intercept in the schedule; and (ii) no arbitrary constraints are imposed except for the asymmetry of information of workers' skills, as in Mirrlees's nonlinear-taxation model.

In the first scenario, the labor-income tax schedule can be summarized at any moment by two numbers: the intercept or lump-sum tax, T_t, and the slope or marginal tax rate, τ_t. This simple tax structure is enough to incorporate the essential missing instrument in Ramsey models: the lump-sum tax. For the special case in which all workers share the same skill, the lump-sum tax can be used to attain the first-best allocation with a zero tax rate. However, with skill inequality a positive tax rate ensures that more productive, richer workers bear a heavier tax burden and alleviate that of less productive, poorer workers.[1] Restrictions on lump-sum taxation are hard to justify, so heterogeneity seems essential to motivate tax distortions.[2]

Optimal policy can be fully characterized for two preference specifications. For separable and isoelastic utility, I show that the tax rate on capital income should be zero and that perfect tax smoothing is optimal: labor-income tax rates are constant over time and unresponsive to either government expenditure or technology shocks. The government uses a combination of debt and lump-sum taxation to smooth out its financing needs. With heterogeneous workers and a lump-sum tax, it is distributional concerns that determine the optimal tax rate. Since the desired level of redistribution is pinned down by the constant distribution of relative skills across workers, a constant tax rate is optimal. I also characterize policy for the class of utility functions consistent with balanced-growth (used by Chari, Christiano, and Kehoe [1994] and others) and provide closed-form expressions for the sensitivity of the tax rate to shocks. Although tax rates are not perfectly constant in this case, my analysis suggests that the results with separable-isoelastic utility provide a useful benchmark.

As a methodological by-product of the analysis, I uncover a partial-equivalence result between my model and the representative-agent Ramsey framework that can be used as a foundation or reinterpretation for the latter. The result states that both frameworks lead to the same first-order optimality conditions and tax rate rules, except in the very first period. This provides a useful connection with a large body of previous theoretical and quantitative work based on the representative-agent Ramsey framework.

Turning to the nonlinear Mirrleesian tax scenario, I find that, when the disutility of work is isoelastic, workers should face

1. Sheshinski [1972] and Hellwig [1986] study optimal linear taxation in a static setting, focusing on finding conditions for the optimal tax rate to be strictly positive.

2. Most countries are best described as having a negative lump-sum intercept in the schedule due to income-tax deductions or transfers from welfare programs. In my model, a negative lump-sum tax is optimal with enough inequality or concern for the poor.

different marginal tax rates but that these should remain perfectly constant over time and unresponsive to shocks. To the best of my knowledge, this is the first tax-smoothing result with nonlinear taxation in a dynamic economy with aggregate uncertainty. Previous work focuses on static settings or dynamics settings with idiosyncratic uncertainty instead of aggregate uncertainty. Implementing this nonlinear form of tax smoothing suggests a role for taxation based on lifetime earnings or income averaging. Indeed, I prove that such a tax scheme fully implements any constrained-efficient allocation. Vickrey [1947] was an early proponent of income-tax averaging rules, although his reasons were different, having to do primarily with considerations of horizontal equity.

In both the linear and the nonlinear case, my model has several implications that contrast with Ramsey analyses. First, the model attributes a crucial role in the determination of tax rates to the skill distribution. To bring this to the forefront, I extend the model and consider shocks to the distribution of relative skills. Tax rates do respond to these shocks—typically rising when inequality rises—while remaining invariant to government expenditure and technology shocks. This extension highlights a source for tax fluctuations that cannot be addressed by a representative-agent model.

Second, the implications for public-debt management differ dramatically. Ramsey models break Ricardian equivalence by ruling out lump-sum taxes, and public debt becomes crucial for the government to smooth tax rates over time. In contrast, here Ricardian equivalence reemerges: the government can smooth tax rates using any mix of debt and lump-sum tax financing. I briefly speculate on a variation, based on imperfect participation in asset markets, that makes debt-management policy determinate.

Third, the source of time-inconsistency problems is different from that in Ramsey models. These models stress the desirability of initial capital levies, as they mimic the missing lump-sum taxes. This leads to a time-inconsistency problem since capital should eventually not be taxed, but it is always desirable to tax it in the short run. In contrast, with heterogeneous agents and lump-sum taxation, the optimum may be time consistent in some special cases. More generally, a time-inconsistency problem may arise, but I show that it depends on the distribution of wealth across workers and on its evolution over time.

Finally, Ramsey analyses have stressed that complete-market allocations can be replicated without using state-contingent bonds by exploiting state-contingent capital taxation [Kingston 1991; Zhu 1992; Chari, Christiano, and Kehoe 1994]. I show how this logic relies heavily on a representative-agent framework and generally fails with heterogeneous workers.[3]

Throughout the paper, I focus on innate differences across workers and distributional concerns as the motives for distortive taxation. Although I allow for idiosyncratic skill uncertainty in Section V, I assume that asset markets are complete, which provides workers with insurance opportunities against such shocks. This contrasts with another line of work that attributes an important role in the insurance of idiosyncratic skill shocks to taxation by assuming that markets cannot provide such arrangements [Golosov, Kocherlakota, and Tsyvinski 2003; Albanesi and Sleet 2006; Farhi and Werning 2005]. While adding imperfect insurance against privately observed idiosyncratic shocks is an interesting step for future work, there are at least three reasons to first focus on the distributional motive for taxation. First, although no consensus exists, heterogeneity appears to be a major contributor in the observed variations of lifetime earnings: most studies place its contribution above 50 percent, with some as high as 90 percent.[4] Moreover, since no attempt is made to discern idiosyncratic shocks that are publicly observable from those that are not, these numbers potentially overstate the contribution of idiosyncratic risk that is uninsurable. Second, *ex ante* heterogeneity provides a clearer role for government policy: it is less clear what the role of market insurance versus government taxation should be in providing insurance against idiosyncratic shocks. Third, to date, models with privately-observed idiosyncratic shocks are not tractable enough for the purposes of the present study. In particular, optimal labor-income tax rates have only been characterized for simple skill processes that also abstract from a fully dynamic economy subject to aggregate shocks.

Section II introduces the model environment. Section III defines and characterizes the linear tax problem. Section IV de-

3. Similar remarks apply to replication schemes based on inflations that devalue nominal claims held by the private sector.

4. Using a model, Keane and Wolpin [1997] estimate the contribution of heterogeneity to be 90 percent. In their statistical analysis, Storesletten, Telmer, and Yaron [2004] attribute about 50 percent to heterogeneity. Hugget, Ventura, and Yaron [2006] reach intermediate conclusions.

rives tax-smoothing and capital-taxation results for two common preference specifications; the partial-equivalence result with Ramsey is also discussed there. Section V extends the model to incorporate shocks to the distribution of skills. Three implications that contrast with the Ramsey case are discussed in Section VI. The nonlinear-Mirrleesian tax problem is analyzed in Section VII; taxation based on income averaging is also discussed there. Section VIII concludes.

II. The Dynamic Economy

The economy is populated by a continuum of infinitely-lived workers divided into a finite number of types $i \in I$ of relative size π^i. Preferences for workers of type $i \in I$ are given by the utility function

$$\sum_{t=0}^{\infty} \beta^t \mathbb{E}[U^i(c_t,L_t)], \tag{1}$$

where $c_t \geq 0$ is consumption and $L_t \geq 0$ is labor in efficiency units. The leading case is when the sole source of heterogeneity is differences in productivity: workers of type i have relative skill θ^i, normalized so that $\sum_{i \in I} \theta^i \pi^i = 1$, and everyone shares some underlying utility function $U(c,n)$ over consumption c and work time n so that the implied utility function over consumption and effective labor units is $U^i(c,L) = U(c,L/\theta^i)$.

Importantly, workers know their own type $i \in I$, but this information is not publicly observable. As a result, the government cannot levy the sort of discriminatory lump-sum taxes that condition on the worker's type $i \in I$ that are needed to achieve any first-best allocation. Equivalently, one can simply assume that taxes cannot be conditioned on a worker's type instead of using private information as a motivation for this assumption.

Uncertainty is captured by a publicly observed state $s_t \in S$ in period t, where S is some finite set; let $\Pr(s^t)$ denote the probability of any history $s^t = (s_0,s_1, \ldots ,s_t)$. An allocation specifies consumption, labor, and capital in every period after every history: $\{c^i(s^t),L^i(s^t),K(s^t)\}$; aggregates are denoted by $c(s^t) \equiv \sum_{i \in I} c^i(s^t)\pi^i$ and $L(s^t) \equiv \sum_{i \in I} L^i(s^t)\pi^i$. Production combines labor with capital using a constant-returns-to-scale technology; capital depreciates at rate δ. The resource constraints are

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

for all s^t and $t = 0,1, \ldots$. Both government expenditures and the production function are allowed to depend on the history s^t (to capture the impact of uncertainty) and the time period t (to capture growth or other deterministic changes).

III. Linear and Proportional Taxation

I start with the case where the tax schedule is linear in labor income $\tau(s^t)w_t(s^t)L^i(s^t) + T(s^t)$ in each period. The natural case is where the lump-sum tax $T(s^t)$ is not restricted, but, for completeness and to relate my results to the standard Ramsey case, I also consider the proportional tax case where $T(s^t)$ is constrained to zero. The government taxes capital income at rate $\kappa(s^t)$. Taxes on initial wealth are also allowed. Consumption taxes are superfluous and can be ignored without loss in generality.

III.A. Competitive Equilibria with Taxes

Markets are assumed to be competitive and complete, as in Lucas and Stokey [1983], Chari, Christiano, and Kehoe [1994], and many others. One interpretation of this envisions government debt as a rich set of Arrow-Debreu state-contingent bonds. A less literal interpretation is provided by the fact that, even with noncontingent debt available, assets may span the necessary payoffs to complete the market.[5]

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

$$\sum_{t,s^t} p(s^t)(c^i(s^t) + k^i(s^t) - w(s^t)(1 - \tau(s^t))L^i(s^t) - R(s^t)k^i(s^{t-1})) \leq (1 - \kappa_B(s_0))B^i(s_0) - T.$$

Here $p(s^t)$ represents the Arrow-Debreu price of consumption in period t after history s^t, normalized so that $p(s_0) = 1$; the real

5. For example, Angeletos [2002] and Buera and Nicolini [2004] show that a portfolio of riskless bonds of various maturities may be used to this end. Nevertheless, in this paper I adopt the complete-market assumption not for its realism, but for its simplicity and to focus on the extension of Ramsey models to heterogeneity and lump-sum taxation. Recent work featuring incomplete markets include Aiyagari et al. [2002], Werning [2005], and Farhi [2005].

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; $T \equiv \Sigma_{t,s^t}\, p(s^t)T(s^t)$ is the present value of the lump-sum components of taxes; finally, $B^i(s_0)$ represents some given initial holdings of short-term government bonds, which are taxed at rate $\kappa_B(s_0) \in [0,1]$.[6]

Ruling out arbitrage opportunities requires

$$p(s^t) = \sum_{s_{t+1}} p(s^{t+1})R(s^{t+1}), \tag{3}$$

simplifying the budget constraint to

$$\sum_{t,s^t} p(s^t)(c^i(s^t) - w(s^t)(1 - \tau(s^t))L^i(s^t)) \leq R(s_0)k_0^i + (1 - \kappa_B(s_0))B^i(s_0) - T. \tag{4}$$

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

$$r(s^t) = F_K(L(s^t),K(s^{t-1}),s^t,t), \tag{5}$$

$$w(s^t) = F_L(L(s^t),K(s^{t-1}),s^t,t). \tag{6}$$

Since the production function has constant returns to scale, profits are zero in equilibrium.

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

$$(1 - \kappa_B(s_0)) \sum_{i \in I} B^i(s_0) + \sum_{t,s^t} p(s^t)g(s^t) \leq T + \sum_{t,s^t} p(s^t)(\tau(s^t)w(s^t)L(s^t) + \kappa(s^t)(r(s^t) - \delta)K(s^{t-1})). \tag{7}$$

A version of Walras' law applies: the government budget con-

6. The single intertemporal budget constraint is equivalent to the sequence of budget constraints

$$c^i(s^t) + k^i(s^t) + \sum_{s_{t+1}} \frac{p(s^t,s_{t+1})}{p(s^t)} B(s^t,s_{t+1}) \leq (1 - \tau(s^t))w(s^t)L^i(s^t) - T(s^t) + R(s^t)k^i(s^{t-1}) + (1 - \kappa_B(s^t))B(s^t)$$

for all $t = 0,1,\ldots$ and histories s^t as well as the no-Ponzi condition $\lim_{t\to\infty} \Sigma_{s^t}\, p(s^t)B(s^t) = 0$.

straint (7) holds with equality when the resource constraints (2) and the workers' budget constraints (4) hold with equality.

DEFINITION 1. Given initial capital and bond holdings $\{k_0^i, B^i(s_0)\}$, a *competitive equilibrium* is a sequence of taxes $\kappa_B(s_0)$ $\{T(s^t),\tau(s^t),\kappa(s^t)\}$, prices $\{p(s^t),r(s^t),w(s^t)\}$, and nonnegative quantities $\{c^i(s^t),L^i(s^t),K(s^t)\}$, such that (i) workers maximize utility: consumption and labor choices $\{c^i(s^t),L^i(s^t)\}$ maximize (1) subject to the budget constraint (4) taking prices and taxes that satisfy (3) as given; (ii) firms maximize profits: the first-order conditions (5) and (6) hold; (iii) the government's budget constraint (7) holds; and (iv) markets clear: the resource constraints (2) hold for all periods t and histories s^t.

III.B. *A Simple Characterization*

I now characterize the set of aggregate allocations that are sustainable by an equilibrium for some sequence of prices and taxes. This leads to a primal approach, which formulates the planning problem directly in terms of the aggregate allocation, dropping taxes and prices. It generalizes the method popularized by Lucas and Stokey [1983] within the representative-agent Ramsey model to a setting with heterogeneity and lump-sum taxation.

With linear taxation, all workers face the same after-tax prices for consumption, $\{p(s^t)\}$, and labor, $\{-p(s^t)w(s^t)(1 - \tau(s^t))\}$; as a result, marginal rates of substitution are equated across workers. Thus, any equilibrium delivers an efficient assignment of individual consumption and labor $\{c^i(s^t),L^i(s^t)\}$ given the allocation for aggregates $\{c(s^t),L(s^t)\}$; in other words, all inefficiencies due to distortive taxation are confined to the determination of aggregates $\{c(s^t),L(s^t)\}$.

Formally, for any equilibrium there exist "market" weights $\varphi \equiv \{\varphi^i\}$, with $\varphi^i \geq 0$ and the normalization that $\sum_{i\in I} \varphi^i\pi^i = 1$, so that individual assignments solve the static subproblem

$$(8)\quad U^m(c,L;\varphi) \equiv \max_{\{c^i,L^i\}} \sum_{i\in I} \varphi^i U^i(c^i,L^i)\pi^i \quad \text{subject to} \quad \sum_{i\in I} c^i\pi^i = c \quad \text{and} \quad \sum_{i\in I} L^i\pi^i = L,$$

where the superscript m stands for "market." Letting

$$h^i(c,L;\,\varphi) = (h^{i,c}(c,L;\,\varphi), h^{i,L}(c,L;\,\varphi))$$

be the solution to this problem for worker type $i \in I$, an equilibrium must satisfy

$$(c^i(s^t), L^i(s^t)) = h^i(c(s^t), L(s^t);\,\varphi) \tag{9}$$

for some market weights φ.

Equilibrium after-tax prices can be computed *as if* the economy were populated by a fictitious representative-agent with the utility function $U^m(c,L;\,\varphi)$:

$$w(s^t)(1 - \tau(s^t)) = \frac{-U_L^m(c(s^t),L(s^t);\,\varphi)}{U_c^m(c(s^t),L(s^t);\,\varphi)}, \tag{10}$$

$$\frac{p(s^t)}{p(s_0)} = \beta^t \frac{U_c^m(c(s^t),L(s^t);\,\varphi)}{U_c^m(c(s_0),L(s_0);\,\varphi)} \Pr(s^t). \tag{11}$$

The envelope condition for the static subproblem (8) is $U_c^m(c,L;\,\varphi) = \varphi^i U_c^i(h^i(c,L;\,\varphi))$ and $U_L^m(c,L;\,\varphi) = \varphi^i U_L^i(h^i(c,L;\,\varphi))$, so that equations (10) and (11) hold with U^i in place of U^m, and workers' marginal rates of substitution are equated to after-tax prices.

In equilibrium, each worker's budget constraint (4) must hold with equality. Using equations (10) and (11) to substitute out prices and taxes gives the *implementability conditions*

$$\begin{aligned}\sum_{t,s^t} \beta^t (U_c^m(c(s^t),L(s^t);\,\varphi) h^{i,c}(c(s^t),L(s^t);\,\varphi) \\ + U_L^m(c(s^t),L(s^t);\,\varphi) h^{i,L}(c(s^t),L(s^t);\,\varphi)) \Pr(s^t) \\ = U_c^m(c(s_0),L(s_0);\,\varphi)(R_0 k_0^i + (1 - \kappa_B(s_0))B^i(s_0) - T)\end{aligned} \tag{12}$$

for all $i \in I$. These constraints (12) are expressed entirely in terms of the aggregate allocation $\{c(s^t),L(s^t)\}$ and the market weights φ.

Summing up, a competitive equilibrium implies that its aggregate allocation $\{c(s^t),L(s^t)\}$ must satisfy the resource constraints (2) and the implementability conditions (12) for some market weights φ. The converse is also true.

PROPOSITION 1. Given initial individual wealth $\{R_0 k_0^i + (1 - \kappa_B(s_0))B^i(s_0)\}$, an aggregate allocation $\{c(s^t),L(s^t),K(s^t)\}$ can be supported by a competitive equilibrium if and only if the resource constraints (2) hold and there exist market weights φ and a lump-sum tax T so that the implementability conditions (12) hold for all $i \in I$. Individual allocations can then be computed using equation (9), prices and taxes can be computed using equations (3), (5), (6), (10), and (11).

III.C. Planning Problem

Applying Proposition 1, the set of all competitive equilibria defines a set $\mathscr{U}$ of attainable lifetime utilities $\{u^i\}$ such that $u^i = \sum_{t,s^t} \beta^t U^i(h^i(c(s^t),L(s^t);\ \varphi))\ \Pr(s^t)$, the resource constraints (2) are satisfied, and the implementability conditions (12) hold for all $i \in I$. The optimal tax problem is to reach the northeastern frontier of this set: maximize u^j subject to $u^n \geq \bar{u}^n$ for all $n \neq j$ and $\{u^i\} \in \mathscr{U}$ for any feasible lower bounds $\{\bar{u}^n\} \in \mathscr{U}$. The necessary first-order conditions can be derived by considering the weighted sum of utilities

$$\sum_{t,s^t,i \in I} \beta^t \lambda^i U^i(h^i(c(s^t),L(s^t);\ \varphi))\ \Pr(s^t)\pi^i, \tag{13}$$

where the Pareto weights $\lambda^i \geq 0$ are rescaled versions of the multipliers on the $u^i \geq \bar{u}^i$ constraints, normalized so that $\sum_{i \in I} \lambda^i \pi^i = 1$. The analysis that follows only exploits first-order necessary conditions and does not presume convexity of the planning problem, or of $\mathscr{U}$, in any way.[7]

The planning problem is over aggregate variables $\{c(s^t), L(s^t),K(s^t)\}$, market weights φ, and the lump-sum tax T (whenever not restricted to zero). In the special case with no inequality and the restriction that lump-sum taxation be zero, $T = 0$, the problem is identical to the primal approach in the representative-agent Ramsey model [see Lucas and Stokey 1983; Chari, Christiano, and Kehoe 1994; Atkeson, Chari, and Kehoe 1999]. For the general case, the analysis shows that one can retain the tractability of an aggregate formulation even when worker heterogeneity and lump-sum taxation are present.[8]

7. In general, the planning problem and the set of utilities $\mathscr{U}$ may not be convex, so one cannot claim that every point on the frontier is characterized by maximizing some weighted sum of utilities, such as (13). (The converse statement is true.) Instead, expression (13) is simply a stepping stone to the full Lagrangian in (14) below that can be used to obtain the necessary first-order conditions for any frontier point even when $\mathscr{U}$ is non-convex. Thus, convexity of the planning problem, or of $\mathscr{U}$, is not needed in any way.

8. Chari and Kehoe [1999] adopt a different, but related, primal approach. They study long-run capital taxation in a deterministic setting, allowing for agent heterogeneity but not lump-sum taxation. Their formulation maximizes over individual allocations and imposes that marginal rates of substitution be equalized across agents as additional constraints on the planning problem. The aggregate formulation pursued here reduces the dimensionality of the problem by solving for individual allocations in terms of aggregates and market weights φ.

The choice over market weights φ is key in determining the level of tax rates. For instance, more equal weights imply a more equal consumption allocation, which requires a more equal after-tax income, which, in turn, requires higher tax rates. If the optimal market weights $\{\varphi^i\}$ happen to equal the Pareto weights $\{\lambda^i\}$, then optimal tax rates are zero; this corresponds to the unique point on the utility frontier $\mathscr{U}$ where the government is entirely financed by lump-sum taxation. Anywhere else on the frontier, the two sets of weights do not coincide, and distortionary taxes are employed.

The analysis does not presume any desired direction for redistribution, so that any point on the frontier of $\mathscr{U}$ is characterized; in other words, no assumptions are required on the Pareto weights $\{\lambda^i\}$. However, one special case that deserves mention is the Utilitarian specification with $\lambda^i = 1$, where redistribution from rich to poor is desirable. In this case, the planning problem can be reinterpreted as one of optimal insurance behind the "veil of ignorance": the objective in (13) interpreted as the expected utility before skill types $i \in I$ are realized with probabilities $\{\pi^i\}$.

III.D. *Optimal Tax Rates*

It is useful to set up the Lagrangian that incorporates the implementability conditions (12) with multipliers $\{\mu^i\pi^i\}$

$$(14)\quad \sum_{t,s^t} \beta^t W(c(s^t),L(s^t);\,\varphi,\mu,\lambda)\ \Pr(s^t) \\ - U_c^m(c(s_0),L(s_0);\,\varphi) \sum_{i\in I} \mu^i(R_0 k_0^i + (1-\kappa_B(s_0))B^i(s_0) - T)\pi^i,$$

where $\mu \equiv \{\mu^i\}$ and the pseudo-utility function $W(c,L;\,\varphi,\mu,\lambda)$ is defined by

$$W(c,L;\,\varphi,\mu,\lambda) \equiv \sum_{i\in I} \pi^i(\lambda^i U^i(h^i(c,L;\,\varphi)) \\ + \mu^i(U_c^m(c,L;\,\varphi)h^{i,c}(c,L;\,\varphi) + U_L^m(c,L;\,\varphi)h^{i,L}(c,L;\,\varphi))).$$

First-Order Conditions. Except for the initial period term, everything is conveniently summarized by the pseudo-utility function $W(c,L;\,\varphi,\mu,\lambda)$. The first-order conditions for $t \geq 1$ are

$$(15)\qquad F_L(L(s^t),K(s^{t-1}),s^t,t) = \frac{-W_L(c(s^t),L(s^t);\,\varphi,\mu,\lambda)}{W_c(c(s^t),L(s^t);\,\varphi,\mu,\lambda)},$$

$$W_c(c(s^t),L(s^t);\ \varphi,\mu,\lambda) = \beta \sum_{s_{t+1}} W_c(c(s^{t+1}),L(s^{t+1});\ \varphi,\mu,\lambda) R^*(s^{t+1}) \Pr(s_{t+1}|s^t), \tag{16}$$

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. When a lump-sum tax is available, the first-order condition with respect to T implies

$$\sum_{i \in I} \mu^i \pi^i = 0, \tag{17}$$

so that the term involving T always vanishes in (14). The first-order conditions with respect to the market weights φ will not be needed in what follows, so I omit them.

Optimal Tax Rates. Dividing equation (10) by (15) and using equation (6) for $w(s^t)$ gives $\tau(s^t) = \tau^*(c(s^t),L(s^t);\ \varphi,\mu,\lambda)$ for $t \geq 1$, where

$$\tau^*(c,L;\ \varphi,\mu,\lambda) \equiv 1 - \frac{U_L^m(c,L;\ \varphi)}{W_L(c,L;\ \varphi,\mu,\lambda)} \frac{W_c(c,L;\ \varphi,\mu,\lambda)}{U_c^m(c,L;\ \varphi)}. \tag{18}$$

The labor-income tax rate is a function of current aggregate consumption and labor only.

Using equilibrium prices (11) in the no-arbitrage condition (3) gives

$$U_c^m(c(s^t),L(s^t);\ \varphi) = \beta \sum_{s_{t+1}} U_c^m(c(s^{t+1}),L(s^{t+1});\ \varphi) R(s^{t+1}) \Pr(s_{t+1}|s^t). \tag{19}$$

In general, there are several $R(s^{t+1})$ that ensure that equations (16) and (19) are compatible. One choice that suits our purposes is

$$R(s^{t+1}) = R^*(s^{t+1}) \frac{U_c^m(c(s^t),L(s^t);\ \varphi)}{W_c(c(s^t),L(s^t);\ \varphi,\mu,\lambda)} \frac{W_c(c(s^{t+1}),L(s^{t+1});\ \varphi,\mu,\lambda)}{U_c^m(c(s^{t+1}),L(s^{t+1});\ \varphi)}. \tag{20}$$

For example, if the ratio $W_c(c,L;\ \varphi,\mu,\lambda)/U_c^m(c,L;\ \varphi)$ is constant, then the capital tax can be set to zero so that $R(s^{t+1}) = R^*(s^{t+1})$ for $t \geq 1$. This formula reveals a version of the celebrated Chamley–Judd result: if the economy settles down to a deterministic steady state, with $c(s^{t+1}) = c(s^t)$ and $L(s^{t+1}) = L(s^t)$, then the tax on capital income can be set to zero and $R(s^{t+1}) = R^*(s^{t+1})$.[9]

The form of the Lagrangian (14) as a discounted sum of the pseudo-utility function W and the tax-rate formulas (18) and (20) provide a first methodological link with the primal approach often used in representative-agent Ramsey analyses [see e.g., Chari, Christiano, and Kehoe 1994; Atkeson, Chari, and Kehoe 1999]. That is, the derivation applies with or without either worker heterogeneity or a lump-sum tax. Section IV.C provides an even tighter connection for two common preference specifications.

III.E. Initial Taxation

I allow unrestricted initial wealth taxation as my benchmark, requiring only that gross returns on capital and bonds not be negative. Tighter restrictions on initial wealth taxation are hard to justify because, as is well-known, a combination of consumption and labor-income taxes can replicate their effect. That is, ignoring consumption taxes, as I have done here, is without loss in generality if and only if initial wealth taxation is unrestricted.

The first-order condition for κ_0, corresponding to that for $R_0 \in [0,\infty)$, gives

$$\sum_{i \in I} \mu^i k_0^i \pi^i = 0 \quad \text{or} \quad R_0 = 0. \tag{21}$$

Similarly, the first-order condition for $\kappa_B(s_0) \in (-\infty,1]$ gives

$$\sum_{i \in I} \mu^i B^i(s_0) \pi^i = 0 \quad \text{or} \quad \kappa_B(s_0) = 1. \tag{22}$$

Together conditions (21) and (22) imply that the first-order conditions (15) and (16) derived for $t \geq 1$ also apply now for $t = 0$, extending the conclusion for tax rates to $\tau(s_0)$ and $\kappa(s_1)$.

In some cases initial wealth taxation is unnecessary. If all workers start with the same capital holdings, so that k_0^i is independent of $i \in I$, then the effect of the initial capital levy κ_0 is equivalent to a lump-sum tax. If a lump-sum tax is already available, then (21) is implied by (17) and any κ_0 is optimal; in particular a zero tax $\kappa_0 = 0$ is optimal. Similarly, if initial bond

9. In different ways, Chamley [1986], Judd [1985], and Chari and Kehoe [1999] consider heterogeneous agents, but not lump-sum taxation, in long-run capital-taxation results.

holdings are equal, so that $B^i(s_0)$ is independent of $i \in I$, then $\kappa_B(s_0) = 0$ is optimal. Equality of wealth corresponds to the canonical optimal-taxation scenario where skill differences are the primordial source of all heterogeneity.

In contrast, in representative-agent Ramsey analyses, just as the lump-sum tax is arbitrarily ruled-out, restrictions on the taxation of consumption and initial wealth are imposed. If some taxation of initial wealth is permitted, it is always optimal to use initial levies on capital and bonds κ_0 and $\kappa_B(s_0)$ (or consumption taxes) to the full extent allowable to imitate the missing lump-sum tax. With *ad hoc* restrictions initial wealth $R_0 k_0^i + (1 - \kappa_B(s_0))B^i(s_0)$ does not drop out of the first-order conditions for $c(s_0)$ and $L(s_0)$, which are thus different from (15) and (16), leading to different conditions for initial tax rates $\tau(s_0)$ and $\kappa(s_1)$.

IV. Two Cases Solved

It is now straightforward to apply the general analysis and formulas laid out in the previous section to any particular case of interest by simply computing the U^m and W functions. In this section, I explore heterogeneity arising from skill differences and consider two classes of utility functions: (i) a separable and isoelastic specification; and (ii) a nonseparable balanced-growth specification. The last subsection discusses the partial-equivalence result with the representative-agent Ramsey model.

IV.A. Separable Isoelastic Utility: Perfect Tax Smoothing

I first consider the case where the underlying utility function is separable and isoelastic:

$$U^i(c,L) = u(c) - v(L/\theta^i), \tag{23}$$
$$\text{where} \quad u(c) \equiv \frac{c^{1-\sigma}}{1-\sigma} \quad \text{and} \quad v(n) \equiv \alpha\frac{n^\gamma}{\gamma},$$

with $\sigma, \alpha > 0$ and $\gamma > 1$.

With these preferences, individual consumption and labor are proportional to their aggregates: $c^i = h^{i,c}(c,L) = \omega_c^i c$, and $L^i = h^{i,L}(c,L) = \omega_L^i L$ with

$$\omega_c^i = \frac{(\varphi^i)^{1/\sigma}}{\sum_{i\in I}(\varphi^i)^{1/\sigma}\pi^i},$$

and

Errata: missing minus sign added.

$$\omega_L^i = \frac{(\theta^i)^{\frac{\gamma}{\gamma-1}}\,(\varphi^i)^{\frac{-1}{\gamma-1}}}{\sum_{i\in I}(\theta^i)^{\frac{\gamma}{\gamma-1}}\,(\varphi^i)^{\frac{-1}{\gamma-1}}\pi^i}.$$

Moreover, the functions U^m and W inherit the separable and isoelastic form of the utility function:

$$(24)\quad U^m = \Phi_u^m u(c) - \Phi_v^m v(L) \quad \text{and} \quad W = \Phi_u^W u(c) - \Phi_v^W v(L),$$

where the constants Φ_u^m, Φ_v^m, Φ_u^W, and Φ_v^W are functions of φ, μ, and λ (see Appendix).

Note that whenever $\Phi_v^m/\Phi_u^m \neq \Phi_v^W/\Phi_u^W$, the functions U^m and W put different weight on consumption versus labor. Applying formula (18) gives

$$(25)\qquad \tau^*(c,L) = \bar{\tau} \equiv 1 - \frac{\Phi_v^m}{\Phi_u^m}\frac{\Phi_u^W}{\Phi_v^W},$$

so that labor-income tax rates are constant over time and across histories, $\tau(s^t) = \bar{\tau}$; Section V explores this formula for the tax rate further. Note that, although the tax rate remains constant across realizations of uncertainty, the stochastic processes for government expenditure and technology itself does generally affect the optimal constant level $\bar{\tau}$. In other words, the tax rate is not necessarily invariant to comparative-static exercises on these processes. Finally, since W_c is proportional to U_c^m, equation (20) implies that the tax on capital can be set to zero.[10]

PROPOSITION 2. When preferences are separable and isoelastic as in (23), (a) perfect tax smoothing is optimal: $\tau(s^t) = \bar{\tau}$ given by equation (25); and (b) zero capital tax rates $\kappa(s^t) = 0$ for $t \geq 1$ are optimal. These results hold with or without a lump-sum tax T.

My model nests the representative-agent Ramsey case, which obtains by setting $\theta^i = 1$ for all $i \in I$ and restricting the lump-sum tax to zero. Applied to this special case, Proposition 2 echoes Kingston's [1991] and Zhu's [1992] representative-agent Ramsey results.

One intuition for the optimality of zero capital taxes is based on the well-known uniform taxation principles due to Diamond

10. The isoelastic specification for the disutility of labor is not needed for this last result: a zero capital tax is optimal as long as utility is separable and isoelastic in consumption $c^{1-\sigma}/(1-\sigma) - v(n)$ for any v function.

and Mirrlees [1971]: since preferences in (23) are homothetic over consumption paths and separable from labor, consumption at different dates should be taxed uniformly, which is equivalent to a zero capital tax.

The intuition for the tax-smoothing result is best conveyed by the natural case that allows for lump-sum taxation. Distortionary taxation is then a redistribution mechanism: a positive tax rate makes high-skilled, rich workers pay more taxes than low-skilled, poor workers. The optimal tax rate at any point in time balances distributional concerns against efficiency. Tax smoothing emerges because the determinants of inequality are constant over time and invariant to government expenditure or aggregate technology shocks.

In representative-agent settings, tax-smoothing results are often explained by the following informal argument: in order to minimize the total cost from distortions it is optimal to equate the marginal cost of distortions over time, which requires equating taxes over time [Barro 1979]. The result derived here refines this intuition: at any point in time, the marginal cost from distortions should be equated to the marginal benefit from redistribution. If the latter is constant over time and invariant to shocks, then the marginal cost from distortions should be equated over time, implying the same for the tax rate.

When lump-sum taxes are ruled out, the only difference is that the overall level of taxation is, by necessity, driven by budgetary needs instead of by distributional concerns. However, the timing of taxes is still affected by distributional concerns. Tax-smoothing is optimal because the skill distribution is constant over time and invariant to shocks.

Some Extensions. I now provide some extensions that do not affect the conclusions for optimal taxes. Proposition 2 still applies if the utility function is generalized to

$$\sum_{t,s^t} \beta^t \left(\chi_t^u(s^t) \cdot u(c^i(s^t)) - \chi_t^v(s^t) \cdot v\left(\frac{L^i(s^t)}{\theta^i}\right)\right) \Pr(s^t). \tag{26}$$

The functions $\chi_t^u(s^t)$ and $\chi_t^v(s^t)$ capture shocks to the marginal rate of substitution between consumption and labor through $\chi_t^v(s^t)/\chi_t^u(s^t)$, equivalent to the "wedges" emphasized in the business-cycle literature [e.g., Chari, Kehoe, and McGrattan 2006]. They also affect the intertemporal marginal rate of substitution, or stochastic discount factor, $(\chi_t^u(s^{t+1})/$

$\chi_t^u(s^t))\beta u'(c(s^{t+1}))/u'(c(s^t))$, which determines asset pricing. Thus, the ratio $\chi_t^u(s^{t+1})/\chi_t^u(s^t)$ could be used to ensure that the model is consistent with asset-returns data.

Although government expenditures do not enter the production function explicitly, the history of states s^t is allowed to affect it in a general way. This implicitly captures any effect that the history $g^t(s^t) \equiv (g_0(s_0), g_1(s^1), \ldots, g_t(s^t))$ of government expenditures—which, after all, is simply a function of the history of states s^t—may have on production possibilities. For example, the stock of public infrastructure, a function of current and past government investments, may affect private production possibilities. By the same reasoning, all the results extend to the case where government expenditures are valued according to the utility function $\chi_t^u(g^t,s^t)u(c_t^i(s^t)) - \chi_t^v(g^t,s^t)v(n_t^i(s^t)) + \chi_t(g^t,s^t)$. The additive term $\chi_t(g^t,s^t)$ plays no role, while the multiplicative factors in the utility function (26) already implicitly capture government expenditures (since, again, $g^t(s^t)$ is a function of s^t).[11] Finally, if government expenditures are endogenous, then the problem studied here is the taxation subproblem that takes as given the solution $g_t^*(s^t)$ for government expenditures.

Lastly, the assumption that labor types are perfectly substitutable can be relaxed. One can show that Proposition 2 holds as long as labor is weakly separable from capital so that production is given by $F(k(s^{t-1}), g(\{L^i(s^t)\}), s^t, t)$ for any aggregator function $g(\{L^i\})$ that is homogeneous of degree one.

IV.B. Balanced Growth Preferences

When utility is not isoelastic or is nonseparable, optimal tax rates do change over time and do respond to shocks. However, I now argue that the previous result provides a useful benchmark by considering the balanced-growth specification chosen by Chari, Christiano, and Kehoe [1994] for their quantitative Ramsey analysis:

$$U^i(c,L) \equiv U(c, L/\theta^i), \tag{27}$$

$$\text{where } U(c,n) \equiv \frac{1}{1-\sigma}\,(c^{\alpha}(1-n)^{1-\alpha})^{1-\sigma} \quad \text{for } \sigma \neq 1,$$

and $U(c,n) = \alpha \log(c) + (1-\alpha)\log(1-n)$ for $\sigma = 1$.

11. The argument can be generalized to make government expenditure a vector: some elements may primarily affect production while others affect utility.

With these preferences, individual consumption and leisure are proportional to their aggregates: $c^i(s^t) = h^{i,c}(c,L) = \omega_c^i c(s^t)$ and $1 - L^i(s^t)/\theta^i = 1 - h^{i,L}(c,L)/\theta^i = \omega_L^i(1 - L(s^t))$, for some fixed weights $\{\omega_c^i, \omega_L^i\}$ determined by φ; it follows that $U^m(c,L)$ is proportional to $U(c,L)$. Also,

$$W(c,L) = \Phi_U^W U(c,L) + \Phi_{U_L}^W U_L(c,L), \tag{28}$$

for some constants Φ_U and Φ_{U_L} determined by φ and μ. Formula (18) implies

$$\tau^*(L) = \frac{1}{(1-L)\Phi_U^W/\Phi_{U_L}^W + \sigma(1-\alpha) + \alpha}, \tag{29}$$

so that the tax rate only depends on current labor L. Using equation (20) and (29) gives

$$\frac{R(s^{t+1})}{R^*(s^{t+1})} = \frac{1 - L(s^t)}{1 - L(s^{t+1})} \cdot \frac{\tau^*(L(s^{t+1}))^{-1} - 1}{\tau^*(L(s^t))^{-1} - 1}. \tag{30}$$

For the logarithmic utility case with $\sigma = 1$, equations (29) and (30) imply that $R(s^{t+1}) = R^*(s^{t+1})$, the tax on capital is zero, and $\kappa(s^{t+1}) = 0$. For other values of σ, these equations reveal that $\kappa(s^{t+1})$ typically takes on both signs, with the magnitude of its fluctuations around zero depending on the magnitude of changes in labor.

PROPOSITION 3. With balanced-growth preferences as in (27) the optimal labor-income tax rate is a function of current labor $\tau^*(L)$ given by equation (29), and its sensitivity is

$$L\tau^{*\prime}(L) = -\frac{L}{1-L}\tau^*(L)(1 - \tau^*(L)(\sigma(1-\alpha) + \alpha)). \tag{31}$$

It is optimal to set the capital-income tax rate so that the after tax rate of return on capital is given by equation (30).

Proof of Proposition 3. Equation (31) follows by differentiating $\tau^*(L)$ in equation (29) with respect to labor L and using equation (29) to substitute out the ratio $\Phi_U^W/\Phi_{U_L}^W$ for the tax rate $\tau^*(L)$. QED

The semi-elasticity $L\tau^{*\prime}(L)$ provides an estimate of the magnitude of likely variations in tax rates

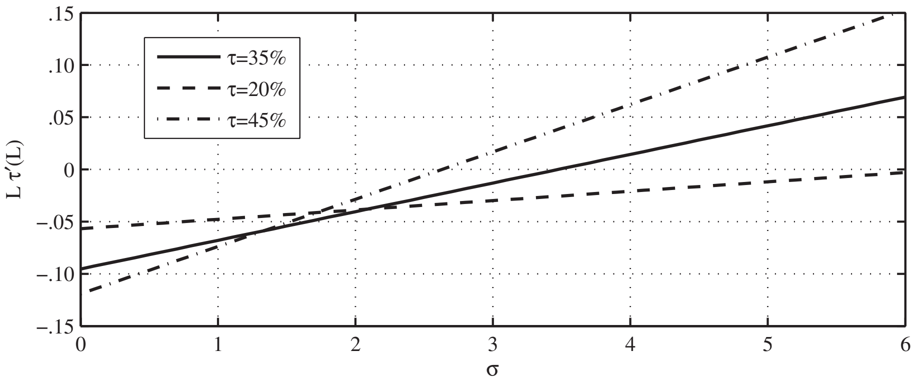

FIGURE I
Sensitivity of Labor-Income Tax Rate with Respect to Labor for Balanced Growth Preferences

$$\tau(s^t) - \tau(\bar{L}) \approx \bar{L}\tau'(\bar{L}) \cdot \frac{L(s^t) - \bar{L}}{\bar{L}} \Rightarrow$$

$$\mathrm{Std}(\tau(s^t)) \approx \bar{L}\tau^{*\prime}(\bar{L}) \cdot \mathrm{Std}(L(s^t)/\bar{L}),$$

for some average value of labor $\bar{L}$.

To get a sense of the magnitudes, consider an example. Suppose a tax rate of $\tau(\bar{L}) = 0.35$, $\bar{L} = 0.23$ and that utility is logarithmic ($\sigma = 1$), then a 1 percent increase in labor changes the tax rate by $\bar{L}\tau'(\bar{L}) \approx -0.074$ of a percentage point, so that the tax rate drops from 35 percent to 34.936 percent. Figure I plots $L\tau^{*\prime}(L)$ as a function of σ using $\alpha = 0.25$ and $\bar{L} = 0.23$ (as in Chari et al.'s [1994] calibration) for three values the tax rate $\tau(L) = 0.20$, 0.35, and 0.45. For the magnitude of business-cycle fluctuations in labor, these calculations suggest small movements in optimal tax rates. Indeed, Chari, Christiano, and Kehoe [1994] found minuscule variations for a calibrated representative-agent Ramsey model—equation (31) explains their findings and extends them to the case with heterogeneity and lump-sum taxation. Finally, as Figure I illustrates, condition (31) implies that perfect tax-smoothing may hold.

COROLLARY 1. If for some level of labor $\bar{L}$ the labor-income tax rate is such that

$$\tau^*(\bar{L}) = \frac{1}{\sigma(1-\alpha) + \alpha}$$

or $\tau^*(\bar{L}) = 0$, then the labor-income tax rate is constant $\tau(L) = \tau(\bar{L})$ for all L, and perfect tax-smoothing $\tau(s^t) = \tau(\bar{L})$ is optimal.

IV.C. Equivalence with Ramsey

The previous analysis actually uncovers a partial-equivalence result between the general model, with heterogeneity and lump-sum taxation, and the representative-agent Ramsey model that rules out lump-sum taxes. This equivalence can be used as a foundation or reinterpretation for some aspects of Ramsey analyses.

The point is that for both preference classes the difference between the functions $W(c,L)$ and $U^m(c,L)$, which determines tax rates, is indexed by a one-dimensional variable: for the separable-isoelastic case, it is the ratio $(\Phi_v^m \Phi_u^W)/(\Phi_u^m \Phi_v^W)$, while for the balanced-growth case, it is $\Phi_U^W/\Phi_{U_L}^W$. Whatever the primitives are—the skill distribution, the initial capital and debt distribution, the availability of lump-sum taxation or initial wealth levies, etc.—it all comes down to the value of this ratio. In particular, the model with heterogeneous workers and lump-sum taxation and the representative-agent Ramsey model, which rules out lump-sum taxation, both deliver the same tax rates for $t \geq 1$, if their ratios coincide. Only differences in the first period remain due to the different assumptions regarding restrictions on initial wealth taxation (see the discussion in Section III.E).

Another way to see this, which provides a closer link to the Ramsey methodology, is that for both preference classes one can show that $W(c,L)$ is proportional to

$$U^m(c,L) + \hat{\mu}(U_c^m(c,L) \cdot c + U_L^m(c,L) \cdot L).$$

This expression is equivalent to that of $W(c,L)$ for a representative-agent Ramsey economy with preferences $U^m(c,L)$. The scalar $\hat{\mu}$ is a transformation of the ratios discussed earlier and provides an equivalent metric of the difference between W and U^m.

PROPOSITION 4. Assume that preferences are either separable and isoelastic, as in (23), or are of the balanced-growth class (27). Optimal tax rates can be expressed as a function of the allocation as in equations (18) and (20) that belong to a class indexed by a one-dimensional parameter $\hat{\mu}$ that summarizes the model's primitives. In particular, this is true for both the model with heterogeneity and lump-sum taxation and the representative-agent Ramsey model that rules out lump-sum taxation.

In the full model, with skill inequality and lump-sum taxation, an important determinant of $\hat{\mu}$ is the degree of skill inequality, or the desire for redistribution captured by the Pareto weights λ. For example, a higher weight on low-skilled workers leads to a higher $\hat{\mu}$, implying higher tax rates and higher transfers $-T$. In the representative-agent Ramsey model an important determinant of $\hat{\mu}$ is the initial level of debt $B(s_0)$. Indeed, all feasible values of $\hat{\mu}$ can be spanned by varying initial debt $B(s_0)$. The first-best is attained if $B(s_0)$ is sufficiently negative, while more indebted governments set higher tax rates to finance the servicing of the debt.

Suppose one solves the planning problem for an economy with heterogeneous workers and a lump-sum tax. Among other things, this yields a pseudo-utility function $U^m(c,L)$ and a tax policy expressed as a function of the allocation. Now, consider solving a representative-agent economy where preferences are given by the $U^m(c,L)$ obtained from the previous exercise, with the same specification of uncertainty and technology and some initial level of debt.[12] Then there exists some initial level of debt for which the tax policy that comes out of both exercises is identical. Moreover, the first-order conditions characterizing the allocation are also identical.

This provides a connection between initial debt $B(s_0)$ in the representative-agent Ramsey model and the chosen level of transfers $-T$ in the model with heterogeneous agents and lump-sum taxation. Interestingly, Chari, Christiano, and Kehoe [1994] calibrate their representative-agent Ramsey economy with a fictitiously high level of debt to capture the important transfers present in the United States tax system (around 12 percent of gross national product in 1985), but absent in their model. The present discussion provides a justification for such a shortcut.

12. Recall that with balanced-growth preferences U^m was equivalent to U, up to an irrelevant constant of proportionality. With separable-isoelastic preferences, U^m places a different weight than U on the disutility of labor.

The equivalence result is useful to reinterpret previous theoretical and quantitative work using the representative-agent Ramsey framework. For example, the simulated dynamics for the optimal allocation and tax rates reported in Chari, Christiano, and Kehoe [1994], using a representative-agent Ramsey model, can be directly adjudicated to my model, with skill heterogeneity and lump sum taxation. On the other hand, things are different regarding initial capital taxation, time-inconsistency of policy, and debt management. I discuss these issues in Section VI.

V. Shocks to the Distribution of Skills

To bring out the importance of distributional concerns in determining the optimal tax rate, I now extend the model to allow skills to vary over time or with the state of the economy: $\theta_t^i(s_t)$ for a worker of type $i \in I$. This can capture, for example, increases in inequality that do not change the ranking of worker types, as well as idiosyncratic shocks to skills that affect workers' rankings without necessarily affecting the cross-sectional distribution of skills. These changes in the distribution may be the result of shocks (e.g., if inequality rises during recessions) or deterministic trends (such as the rise of wage inequality in the United States during the 1980s).

Fortunately, the general analysis from Section III is virtually unaffected by this extension. The only difference is that utility $U^i(c,L;\, s_t,t) = U(c,L/\theta_t^i(s_t))$ now depends on the state s_t and the period t, which induces the same in the functions h^i, U^m, and W. With this small change, all the analysis from Section III extends.

For the rest of this section, to focus on the impact on tax rates from changes in the distribution, I adopt the separable-isoelastic utility specification (23). The functions U^m and W are now:

$$U^m = \Phi_u^m u(c) - \Phi_{v,t}^m(s_t)v(L) \quad \text{and} \quad W = \Phi_u^W u(c) - \Phi_{v,t}^W(s_t)v(L)$$

for coefficients Φ_u^m, $\Phi_{v,t}^m(s_t)$, Φ_u^W and $\Phi_{v,t}^W(s_t)$, that vary with the state s_t and the period t solely through their effect on the distribution of skills $\{\theta_t^i(s_t)\}$ (see Appendix I). Applying formula (18) gives

$$\tau(s^t) = \bar{\tau}_t(s_t) \equiv 1 - \frac{\Phi^m_{v,t}(s_t)}{\Phi^m_u}\frac{\Phi^W_u}{\Phi^W_{v,t}(s_t)} \tag{32}$$
$$= 1 - \frac{\Sigma_{i\in I}\ \omega^i_c\left(\frac{\lambda^i}{\varphi^i} + (1-\sigma)\mu^i\right)\pi^i}{\Sigma_{i\in I}\ \omega^i_{L,t}(s_t)\left(\frac{\lambda^i}{\varphi^i} + \gamma\mu^i\right)\pi^i},$$

where $\omega^i_c = (\varphi^i)^{1/\sigma}/(\Sigma_{i\in I}\ (\varphi^i)^{1/\sigma}\pi^i)$ and

$$\omega^i_{L,t}(s_t) \equiv \frac{(\theta^i_t(s_t))^{\frac{\gamma}{\gamma-1}}(\varphi^i)^{\frac{-1}{\gamma-1}}}{\Sigma_{i\in I}\ (\theta^i_t(s_t))^{\frac{\gamma}{\gamma-1}}\ (\varphi^i)^{\frac{-1}{\gamma-1}}\ \pi^i}. \tag{33}$$

The share of labor $\omega^i_{L,t}(s_t)$ and the tax rate $\bar{\tau}_t(s_t)$ vary only with the skill distribution $\{\theta_t(s_t)\}$.

PROPOSITION 5. With shocks to the distribution of skill and preferences given by (23), (a) the optimal tax rate on labor income is $\tau(s^t) = \bar{\tau}_t(s_t)$ given by equation (32), which varies only with the distribution of skills $\{\theta^i_t(s_t)\}_{i\in I}$; and (b) a zero capital tax rate $\kappa(s^t) = 0$ is optimal for $t \geq 1$. These results hold with or without a lump-sum tax T.

The tax rate is unresponsive to shocks affecting government expenditures or aggregate technology. That movements in the distribution of relative skills are the only source for tax rate fluctuations underscores the point made earlier that distributional concerns are a crucial determinant of the level of labor-income tax rates. Indeed, as discussed earlier when a lump-sum tax is available, distributional concerns are the main determinant of the overall level of tax rates. Proposition 5 generalizes this comparative-static notion by showing that fluctuations in the distribution of skills also lead to fluctuating tax rates over time.

To see the link between inequality and taxes more clearly, consider the case where a lump-sum tax is available. Using the first-order condition (17), equation (32) becomes

$$\bar{\tau}_t(s_t) = 1 - \frac{\tilde{\mathbb{E}}\left[\frac{\lambda^i}{\varphi^i}\right] + \widetilde{\mathrm{cov}}\left(\omega^i_c, \frac{\lambda^i}{\varphi^i} + (1-\sigma)\mu^i\right)}{\tilde{\mathbb{E}}\left[\frac{\lambda^i}{\varphi^i}\right] + \widetilde{\mathrm{cov}}\left(\omega^i_L(s_t), \frac{\lambda^i}{\varphi^i} + \gamma\mu^i\right)}, \tag{34}$$

where $\tilde{\mathbb{E}}[x^i] \equiv \Sigma_{i\in I}\ x^i\pi^i$ and $\widetilde{\mathrm{cov}}\ (x^i, y^i) \equiv \tilde{\mathbb{E}}[x^i y^i] - \tilde{\mathbb{E}}[x^i]\tilde{\mathbb{E}}[y^i]$

represent special cross-sectional expectations and covariance operators that add across worker types i using population fractions $\{\pi^i\}$ as probabilities. This version of the tax-smoothing formula highlights the central role that the dispersion in labor income across workers can play. To be concrete, suppose that (as in the example below) the second term in the denominator's covariance, $(\lambda^i/\varphi^i) + \gamma\mu^i$, increases with the worker's skill type i. Suppose further that the share of labor earnings, $\omega_L^i(s_t)$, is also increasing in the worker's skill type. The denominator's covariance is then positive, and a rise in the dispersion of labor increases the covariance, making the tax rate rise. The greater the dispersion in labor income, the more effective the tax as a redistributive device.

Recall the intuition that, with a lump-sum tax, the marginal cost from distortions should equal the marginal benefit from increased redistribution in each period. As long as the skill distribution does not vary, the marginal benefit from redistribution is unchanging so that the marginal cost from distortions should be equated over time, leading to a constant tax rate. However, when the distribution of skills does shift, the marginal benefit from redistribution shifts with it, and the marginal cost from distortions should not be equated over time. As a result, the optimal tax rate responds to shifts in the skill distribution.

Tax rates vary with the distribution of skills even without the lump-sum tax. Although in this case distributional concerns cannot affect the overall level of tax rates, they can shape their timing.

An Example. To illustrate some features of the tax rate's dependence on the relative skill distribution, I turn to an example. I adopt the Brock–Mirman specification, with logarithmic utility from consumption ($\sigma = 1$), Cobb–Douglas production function and full depreciation $\delta = 1$. This permits an almost-closed-form solution (see Appendix I for details). The example abstracts from government expenditure and technology shocks to focus on skill-distribution shocks.

Workers are split into two equally populated types, $I = \{L,H\}$ with $\pi^H = \pi^L = 1/2$, with low and high productivity, $\theta^H(s_t) \geq 1 \geq \theta^L(s_t) = 2 - \theta^H(s_t)$. The possible values for relative skills $\theta_H(s_t)/\theta_L(s_t)$ live on an equally-spaced grid, with ten points with the lowest value equal to 1 (no inequality) and the highest

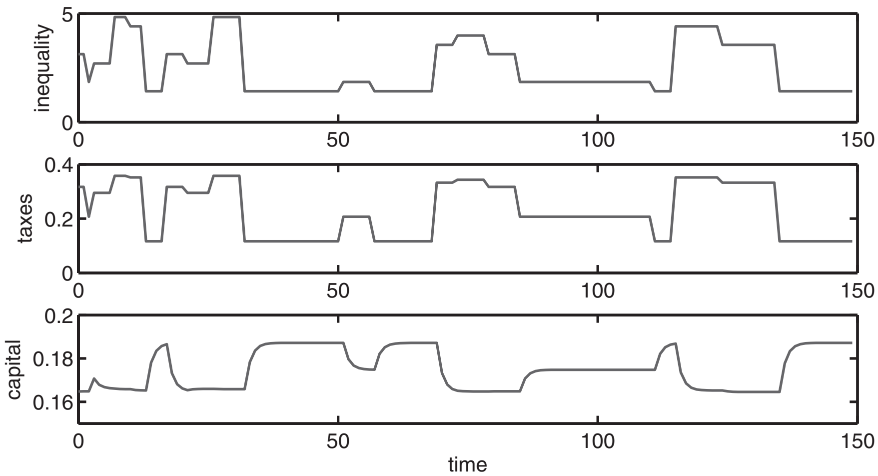

FIGURE II
Simulated Sample Path for Inequality, Taxes, and Capital

value around 4.8. The state s_t is a simple Markov chain: each period it changes with some constant probability, and its new value is drawn with equal probability from the grid. Finally, the example assumes no initial inequality in financial wealth and solves the Utilitarian case with equal weights on both groups, $\lambda^L = \lambda^H = 1$.[13]

Figure II shows a simulated sample path from the solution. Tax rates only adjust when the distribution of skills changes, illustrating Proposition 5. The dynamics for capital are smoother; output and consumption are not shown but behave similarly.

Figure III shows the optimal relationship between skill inequality $\theta_H(s_t)/\theta_L(s_t)$ and the tax rate $\tau(s_t)$ that holds at any point in time along the equilibrium. Tax rates increase with the current skill inequality, ranging from -4.2 up to 35.7 percent. Figure IV shows the allocation for effective labor L^i as a function of the state, which is independent of capital in this Brock–Mirman specification. When the relative skill of the high type increases, the high type's effective labor supply increases, while the low type's decreases.

For the two-type case, formula (34) implies that optimal tax

13. The other parameters are: $\beta = 0.95$; $g_t(s_t) = 0$; $\gamma = 2$; $\alpha = 1$; and $F(L,K) = K^{\rho}L^{1-\rho}$ with the share of capital set to $\rho = 0.4$; the probability of not changing states in the Markov chain was chosen near .9 but adjusted so that there was a state with no inequality in labor-income: it came out to be .8999119. The initial state of the economy was taken to be the middle grid point.

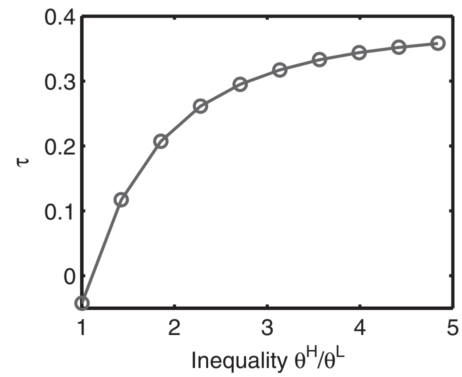

FIGURE III
Relative Inequality and Tax Rates

rates are solely a function of the difference $\omega_L^H - \omega_L^L$. Figures III and IV show that the optimal tax rate increases with this difference, as expected. For sufficiently low skill-inequality, the tax rate may even become negative. This occurs because the ranking of labor-income flips when $\theta^H/\theta^L = 1$, so that low-skilled workers earn more than high-skilled workers due to an income effect: leisure is a normal good and low-skilled workers are poorer, since their productivity is never higher but sometimes strictly lower than that of high-skilled workers (this shows up as a lower ω_c^i in equation (33)). A flip in the ranking of earnings changes the sign of the covariance term in the denominator of formula (34) and can lead to a negative optimal tax rate.

Figures III and IV show that the tax rate is still positive

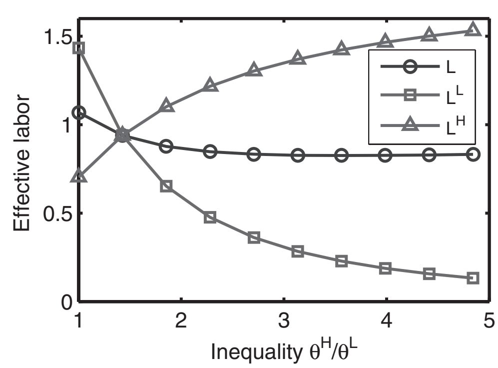

FIGURE IV
Relative Inequality and Labor

when both types of workers earn the same labor income, which occurs at the second-lowest state of inequality. Equation (34) with $\sigma = 1$ and $\lambda^i = 1$ confirms that the tax is nonzero whenever $\omega^i_L(s_t) = 1$ for all $i \in I$ as long as there is consumption inequality, so that ω^i_c are not all equal to 1. This may seem counterintuitive: if there is no difference in labor income to redistribute, why tax it? But if the tax rate were zero, workers would intertemporally substitute work towards these periods with no inequality, away from periods with positive tax rates. This reallocation of labor is inefficient—it does not reflect differences in productivity—and a positive tax helps mitigate it. When the utility from consumption is linear ($\sigma = 0$) labor supply in each period is a function of the current net wage, so there is no intertemporal substitution effect. Formula (34) confirms that the optimal tax rate is zero in this case whenever there is no labor-income inequality.[14]

It is important to recall that the tax changes pictured in these figures represent the equilibrium response to changes in inequality over time, which is different from a comparative-static exercise. How do these changes in tax rates along the equilibrium path differ from a comparative-static exercise which changes some fixed skill distribution? I performed some numerical simulations and found that the comparison is ambiguous. Whenever the changes in inequality considered were small enough, the equilibrium path responses to shocks were smaller than the corresponding comparative-static ones. However, for large enough changes in inequality, the reverse is possible.[15]

VI. Discussion: Three Differences with Ramsey

In this section I discuss some further implications of my model, focusing on three issues that differ sharply with representative-agent Ramsey settings.

14. This follows since $\mu^i = -\lambda^i + \sum_{i \in I} \lambda^i \pi^i$ and $\varphi^i = 1$ in this case. Of course, for the problem to be interesting one requires $\lambda^i \neq \lambda^j$ for some $i, j \in I$; otherwise there is no reason to redistribute and $\mu^i = 0$ for all $i \in I$, so that tax rates are always zero.

15. For example, suppose that $\sigma = 1$, $\lambda^i = 1$, and that there are two possibilities: no skill inequality or some positive skill inequality. Furthermore, suppose the probability of changing states along the equilibrium path is near 0. Then, when the initial state has no inequality, taxes are zero in both states. In contrast, when the initial state has positive inequality, taxes are positive there, but can turn negative when the state switches to no inequality (this is illustrated by the example worked out in the next subsection). Thus, when the initial state has positive inequality, the change along the equilibrium path is larger than the change obtained from the comparative-static exercise.

VI.A. Capital Taxation 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.[16] This tension, between long-run and short-run tax prescriptions, makes government policy time inconsistent.

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 are already directly available. However, as discussed in Section III.E, capital levies are no longer neutral if initial asset holdings are unequal. For example, consider a simple two-type case $I = \{L,H\}$ where higher skilled workers are also wealthier, so that $\theta^L < \theta^H$ and $k^L < k^H$. A tax on initial wealth then acts as an ideal redistributive device, taking more from the rich, as income taxation does, but without introducing distortions. The taxation of initial wealth is desirable since it shifts out the frontier of attainable utilities.

In a nutshell, the Ramsey framework 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 can always be accomplished by using the lump-sum tax. Rather, it is the distributional concern regarding from *whom* the government is extracting resources. Instead of redistribution from private to public sector, what is crucial is redistribution *within* the private sector.

These differences regarding capital levies affect the issue of time inconsistency of policy. Indeed, unlike in the representative-agent Ramsey setting, the optimum may, in some cases, be time consistent. As an example, consider a deterministic economy that finds itself initially at the steady-state level of capital, given the optimal policy. Suppose further that each worker owns the same amount of capital, i.e., $k_0^i = k_0^j$. Then, capital levies simply

16. To avoid the first-best, most analyses impose arbitrary upper bounds on the capital levy. In contrast, here a nontrivial tax problem remains without such restrictions.

replicate lump-sum taxes, which are already directly available. As a result, capital levies can be set to zero. Since the economy is at a steady state, this situation simply repeats itself over time, implying that the optimal tax policy is time consistent.

This simple example makes the point that when lump-sum taxes are available, policy may be time consistent. However, in general, time inconsistency problems may emerge. For example, if along the equilibrium more productive workers tend to accumulate more assets over time, then, *ex post,* a tax on capital, combined with a reduction of the tax on labor, may create a Pareto improvement. Hence, optimal policy is not necessarily *ex post* Pareto efficient. The reason is precisely that, *ex post,* the capital levy no longer imitates a lump-sum tax: in this example, it falls more heavily on richer, more productive workers. That is, the incidence of the capital levy matters. This discussion emphasizes that the mechanism for time inconsistency is different and that it suggests new issues—regarding the distribution of assets, and its evolution, within the private sector—that cannot be addressed by a representative-agent model.

VI.B. Debt Management

Since Barro [1979], second-best tax problems have been used to avoid the neutrality results implied by Ricardian equivalence and determine an optimal debt management policy. Barro argued that tax rates should be smoothed over time, that taxes should be set with an eye towards permanent government spending, as opposed to current spending. Government debt is key to smoothing tax rates; it should be used to buffer any resulting deficits and surpluses. By allowing state-contingent debt, Lucas and Stokey [1983] extended this argument and found that tax rates should also be smoothed across states of the world, as well as time. In both models the solution to the tax problem determines a debt management policy. This is the case because, with proportional taxation, average and marginal taxes coincide.

However, this tight link between average and marginal tax rates is broken when lump-sum taxes are available. Ricardian equivalence is recovered, rendering the mix between debt and lump-sum taxes indeterminate. Indeed, at one extreme, the government could refrain from using debt altogether and balance its budget each period by using the lump-sum tax.[17] However, this does not imply that the asset market is unimportant: even in this case, workers may need to borrow and save, or to provide insurance to each other; in general, it cannot be dispensed with.

Simple extensions of the model that overcome Ricardian equivalence may provide a determinate theory of debt management. One possibility is to model some individuals as having limited participation in asset markets. As an extreme example, suppose that a particular type of worker $i \in I$ lives hand-to-mouth: with no initial assets and no access to asset markets whatsoever, these workers simply consume, in each period, their entire labor income net of taxes. The desire to smooth their consumption then pins down the optimal lump-sum tax, and with it, public debt.[18]

VI.C. Can Taxes on Capital Replicate Complete Markets?

In representative-agent Ramsey settings, Zhu [1992] and Chari, Christiano, and Kehoe [1994] showed that capital taxation can help implement complete-market outcomes, even if state-contingent bonds are unavailable. Roughly, the argument is based on the fact that distortionary effects from taxation are determined by some average (namely, the marginal utility weighted average) of next period's capital tax rate. Then, the tax rate's dependence on next period's state acts as a state-contingent source of revenue, which can be exploited to replicate revenue from state-contingent bonds.[19]

However, these ideas depend heavily on the representative-agent Ramsey framework. Firstly, they are based on imitating, *ex post,* a missing lump-sum tax to provide a nondistortive source of state-contingent revenues. In contrast, when a lump-sum tax is available it already provides a nondistortive source of state-contingent revenues. Thus, if capital levies simply replicated lump-sum taxes, as they do in the representative-agent Ramsey case, then they would be completely irrelevant.

Secondly, with heterogeneous workers both the role of complete markets and the effect of capital levies when markets are incomplete are more involved. Complete markets are more

17. A related indeterminacy arises in Bassetto and Kocherlakota [2004] in the context of a model that allows taxes to be a function of past earnings.

18. For the simple case described here, the planner can actually replicate the allocation that is optimal with full participation in asset markets; whether or not doing so is optimal remains an open question.

19. A similar idea applies to monetary models regarding the price level, i.e., surprise inflations.

than just a source for state-contingent revenue for the government. They also provide insurance between different types of workers. That is, since workers are heterogeneous, in general, they trade with each other in the asset market. Replicating complete markets requires replicating the transfers between workers obtained from these contractual arrangements. Capital levies do not simply imitate lump-sum taxes; they also redistribute within the private sector. While proportional taxes on capital may make up for some of the state-contingent transfers across workers that are needed, in general, they are too coarse an instrument to do the job.

VII. Mirrleesian Taxation: Constrained Efficiency

In this section I treat the Mirrleesian scenario, where no restrictions are placed on tax instruments—allowing, for example, nonlinear taxation of labor and capital income—so that the economy achieves efficient allocations that are constrained only by asymmetric information. Alternatively, instead of stressing informational frictions, the analysis applies without modification if one simply assumes that the only restriction on the government is that taxes, for some reason, cannot depend directly on a worker's type $i \in I$.

Naturally, tax schemes that implement constrained-efficient allocations are more involved than the linear taxes considered in previous sections. I first focus on characterizing the implicit marginal tax rates implied by these allocations. I then provide a tax scheme based on income averaging that implements constrained-efficient allocations as part of a competitive equilibrium.

The contribution here is to study a dynamic economy with aggregate uncertainty and characterize the response to shocks. In contrast, previous work within a Mirrleesian setting has studied static models or focused on idiosyncratic uncertainty. The focus on the response to aggregate fluctuations and uncertainty—an integral part of the Ramsey literature—has not been explored in a Mirrleesian setting.

VII.A. The Planning Problem

I assume the additively-separable utility specification (23), except that the assumption that the utility function $u(c)$ is isoelastic is no longer required; the assumption that the disutility function $v(n)$ is isoelastic, on the other hand, is required for the tax-smoothing result derived below.

I apply the Revelation Principle and consider a direct truth-telling mechanism, where workers report their skill type and receive an allocation as a function of this report. The incentive-compatibility constraints ensure that truthful reporting is optimal:

$$\sum_{t,s^t} \beta^t \left(u(c^i(s^t)) - v\left(\frac{L^i(s^t)}{\theta_t^i(s_t)}\right)\right) \Pr(s^t) \geq \sum_{t,s^t} \beta^t \left(u(c^j(s^t)) - v\left(\frac{L^j(s^t)}{\theta_t^i(s_t)}\right)\right) \Pr(s^t) \tag{35}$$

for all types $i \in I$ and reports $j \in I$.

To characterize all Pareto constrained-efficient allocations, consider the planning problem that maximizes the weighted sum of utilities

$$\sum_{t,s^t,i\in I} \beta^t \lambda^i \left(u(c^i(s^t)) - v\left(\frac{L^i(s^t)}{\theta_t^i(s_t)}\right)\right) \Pr(s^t)\pi^i$$

subject to the resource constraints (2) and the incentive constraints (35).[20]

Let the multiplier on the incentive constraint for worker type $i \in I$ reporting to be $j \in I$ be $\psi^{i,j}\pi^i$. Exploiting the fact that $v(L/\theta^j) = (\theta^i/\theta^j)^\gamma v(L/\theta^i)$, one can write the Lagrangian that incorporates these constraints as

$$\sum_{i,t,s^t} \beta^t \left(\phi_c^i u(c^i(s^t)) - \phi_{L,t}^i(s_t) v\left(\frac{L^i(s^t)}{\theta_t^i(s_t)}\right)\right) \Pr(s^t)\pi^i,$$

where

$$\begin{aligned} \phi_c^i &\equiv \lambda^i + \sum_j \left(\psi^{i,j} - \frac{\pi_j}{\pi_i}\psi^{j,i}\right) \\ \phi_{L,t}^i(s_t) &\equiv \lambda^i + \sum_j \left(\psi^{i,j} - \frac{\pi_j}{\pi_i}\left(\frac{\theta_t^i(s_t)}{\theta_t^j(s_t)}\right)^\gamma \psi^{j,i}\right). \end{aligned} \tag{36}$$

20. Unlike the case with linear taxes, one can consider the maximization of this objective without loss in generality, instead of simply using it to derive first-order conditions. This follows because the planning problem is convex after the following change in variables: from $c^i(s^t)$ and $L^i(s^t)$, to $u^i(s^t) = u(c^i(s^t))$ and $v^i(s^t) = v(L^i(s^t))$. This convexity implies that it is without loss in generality to consider the maximization of the weighted sum of utilities: varying $\{\lambda^i\}$ traces out the utility frontier and characterizes all Pareto constrained-efficient allocations.

VII.B. Tax Smoothing with Nonlinear Taxation

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

$$\phi_c^i u'(c^i(s^t)) = \eta(s^t), \tag{37}$$

$$\phi_{L,t}^i(s_t)\,\frac{1}{\theta_t^i(s_t)}\,v'\!\left(\frac{L^i(s^t)}{\theta_t^i(s_t)}\right) = \eta(s^t)F_L(L(s^t),K(s^{t-1}),s^t,t),\ \text{and} \tag{38}$$

$$\eta(s^t) = \beta \sum_{s_{t+1}} \eta(s^{t+1})R^*(s^{t+1})\ \Pr(s_{t+1}|s^t), \tag{39}$$

where $\beta^t\eta(s^t)\ \Pr(s^t)$ is the multiplier on the resource constraint (2) in period t with history s^t.

Combining the first-order conditions (37) and (39) gives the standard intertemporal consumption Euler equation

$$u'(c^i(s^t)) = \beta \sum_{s_{t+1}} u'(c^i(s^t,s_{t+1}))R^*(s^t,s_{t+1})\ \Pr(s_{t+1}|s^t) \tag{40}$$

for all $i \in I$, so that, in this sense, saving decisions are not distorted.

Define the implicit marginal tax on labor $\tau^i(s^t)$ as the solution to

$$\frac{1}{\theta_t^i(s_t)}\,\frac{v'(L^i(s^t)/\theta_t^i(s_t))}{u'(c^i(s^t))} = F_L(L(s^t),K(s^{t-1}),s^t,t)(1-\tau^i(s^t)).$$

Combining the first-order conditions (37) and (38) gives

$$\tau^i(s^t) = 1 - \frac{\phi_c^i}{\phi_{L,t}^i(s_t)} \equiv \bar{\tau}_t^i(s_t). \tag{41}$$

Note that $\phi_{L,t}^i(s_t)$, defined by (36), depends on period t and state s_t only through the effect these may have on the distribution of skills $\{\theta_t^i(s_t)\}_{i\in I}$; equation (41) implies that this property is inherited by the marginal tax on labor income.

PROPOSITION 6. At any constrained-efficient allocation (a) the intertemporal consumption Euler equation (40) holds, and (b) each worker type $i \in I$ faces an implicit marginal tax on labor income $\tau^i(s^t) = \bar{\tau}_t^i(s_t)$ that depends only on the current skill distribution $\{\theta_t^i(s_t)\}_{i\in I}$.

When the skill distribution is fixed, Proposition 6 provides a benchmark for zero capital taxation and for perfect tax smoothing in a setting with nonlinear Mirrleesian taxation.[21] As in Proposition 2, the marginal tax on labor faced by each worker is then constant across time and states. However, unlike the linear tax case, different workers face different implicit marginal tax rates.[22] In the next subsection I investigate what these results on *implicit* marginal tax rates may imply for *explicit* tax systems.

VII.C. Income Tax Averaging: Implementing Tax Smoothing

To derive properties of constrained-efficient allocations, the last two subsections used the abstract tool of a direct mechanism, where workers made an initial report on their type that determined their allocation thereafter. In particular, no markets were involved, and the only choice made by workers was their report. In contrast, in the case of linear taxation studied earlier, the tax implementation was an integral part of the analysis in that the optimal allocation was derived jointly with a tax policy and market equilibrium that sustained it. The idea of this subsection is to place the nonlinear Mirrleesian scenario on equal footing: incorporating markets, prices, and an explicit tax system that implements constrained-efficient allocations.

The results regarding implicit marginal tax rates obtained in the previous subsection identify properties that explicit tax systems that do allow workers to make savings and labor choices at market-determined prices and wages should have to achieve efficient outcomes. For the case where the skill distribution is fixed, the results suggest taxation based on income averages since this ensures that each worker faces a constant marginal tax but allows this marginal tax to vary across workers. I now formalize this idea by proving that such a tax system implements constrained-efficient allocations characterized in the previous subsections as part of a competitive equilibrium.

The implementation works as follows. The government

21. Distorting savings would be optimal if there were ensuing privately-observed individual skill shocks [Diamond and Mirrlees 1978; Rogerson 1985; Golosov, Kocherlakota, and Tsyvinski 2003]. Farhi and Werning [2005] explore the importance, in terms of welfare gains, of distorting savings in such environments.

22. As in the linear case, for nonseparable utility or for disutility functions $v(n)$ that are nonisoelastic, tax rates generally do vary over time and with shocks, even with a fixed distribution of skills. An open question is whether the magnitude of these changes is significant. It may be possible to obtain estimates of the sensitivity of tax rates to these shocks, as Section IV.B did for the linear taxation case.

places a nonlinear tax on the present value of labor earnings and does not tax capital income. A competitive equilibrium in this context is a tax function Ψ, a sequence of prices $\{p(s^t), r(s^t), w(s^t)\}$ that satisfy the no-arbitrage condition (3), and nonnegative quantities $\{c^i(s^t), L^i(s^t), K(s^t)\}$, such that (i) workers maximize utility

$$\sum_{t,s^t} \beta^t \left(u(c^i(s^t)) - v\left(\frac{L^i(s^t)}{\theta^i} \right) \right) \Pr(s^t) \tag{42}$$

subject to the budget constraint

$$\sum_{t,s^t} p(s^t) c^i(s^t) \leq \sum_{t,s^t} p(s^t) w(s^t) L^i(s^t) - \Psi\left(\sum_{t,s^t} p(s^t) w(s^t) L^i(s^t) \right); \tag{43}$$

(ii) firms maximize profits so that $w(s^t)$ and $r(s^t)$ are given by (5) and (6); and (iii) markets clear so that the resource constraints (2) hold for all periods t and histories s^t.

To see why this might work, take any tax function Ψ that is differentiable. Then the first-order conditions for the problem of a worker of type $i \in I$ yield

$$\frac{1}{\theta^i} \frac{v'(L^i(s^t)/\theta^i)}{u'(c^i(s^t))} = \left(1 - \Psi'\left(\sum_{t,s^t} p(s^t) w(s^t) L^i(s^t) \right) \right) w(s^t)$$

for all $t = 0, 1, \ldots$ and s^t. Thus, the implicit marginal tax on each worker is constant, but not necessarily the same across workers (since Ψ may be nonlinear), exactly the property shared by constrained-efficient allocations. This suggests that an appropriately chosen Ψ might implement the optimum. In Appendix II, I prove that this is indeed the case: there exists a tax function Ψ and prices $\{p(s^t), w(s^t), r(s^t)\}$ that implement the constrained-efficient allocation as part of a competitive equilibrium.

PROPOSITION 7. Suppose that the skill distribution is constant over time and does not vary with the state s_t, i.e., $\theta^i_t(s_t) = \theta^i$. Then any constrained-efficient allocation can be implemented by a competitive equilibrium with no tax on capital income and a nonlinear tax $\Psi(\sum_{t,s^t} p(s^t) w(s^t) L^i(s^t))$ on the present value of labor income. The tax function Ψ can be chosen to be continuous and piecewise differentiable.

Variations on this implementation are possible. As an example, Appendix III provides a sequential variant for a deterministic version of the economy. Instead of taxing workers once and for all as a function of the present value of future earnings, workers pay taxes in all periods as a function of past earnings.

VIII. Conclusions

The model developed here provides a flexible and tractable framework to address optimal taxation issues in dynamic economies. Like the representative-agent Ramsey framework, the model can handle rich dynamic environments; unlike the representative-agent Ramsey framework, distortive taxes arise naturally from distributional concerns. The analysis provided a bridge between the Ramsey and Mirrleesian approaches to dynamic issues in optimal taxation.

Two extensions of the model, not explored here, may be of interest for future work. One is to relax the assumption that asset markets are complete in favor of some incomplete-market alternative. Another is to explore overlapping-generations demographics instead of the infinitely-lived dynastic framework. It is hoped that the model in this paper may provide a useful benchmark or starting point for approaching these and other extensions.

Appendix I: U^m and W in Isoelastic Separable Case

I treat the general case where the distribution of skills $\{\theta_t^i(s_t)\}_{i\in I}$ may vary with the state s_t or period t. Given the expressions for ω_c^i and $\omega_{L,t}^i(s_t)$ in the text, one finds that $U^m = \Phi_u^m u(c) - \Phi_{v,t}^m(s_t)v(L)$, with

$$\Phi_u^m \equiv \left(\sum_{i\in I} (\varphi^i)^{\frac{1}{\sigma}} \pi^i\right)^{\sigma}$$

$$\Phi_{v,t}^m(s_t) \equiv \left(\sum_{i\in I} (\theta_t^i(s_t))^{\frac{\gamma}{\gamma-1}} (\varphi^i)^{\frac{-1}{\gamma-1}} \pi^i\right)^{1-\gamma}.$$

One can verify that $(\omega_c^i)^{1-\sigma} = \Phi_u^m \omega_c^i/\varphi^i$ and $(\omega_L^i(s_t)/\theta^i(s_t))^{\gamma} = \Phi_{v,t}^m(s_t)\omega_L^i(s_t)/\varphi^i$. Using these expressions and the fact that $u'(c)c = (1-\sigma)u(c)$ and $v'(L)L = \gamma v(L)$ then gives $W = \Phi_u^W u(c) - \Phi_{v,t}^W(s_t)v(L)$ with

$$\Phi_u^W \equiv \Phi_u^m \sum_{i\in I} \omega_c^i\left(\frac{\lambda^i}{\varphi^i} + (1-\sigma)\mu^i\right)\pi^i$$

$$\Phi_v^W(s_t) \equiv \Phi_v^m(s_t) \sum_{i \in I} \omega_{L,t}^i(s_t)\left(\frac{\lambda^i}{\varphi^i} + \gamma\mu^i\right)\pi^i.$$

Appendix II: Solution Procedure for the Example from Section V

For the logarithmic case, i.e., $\sigma = 1$, the implementability condition becomes

$$\sum_{t,s^t} \beta^t \left(1 - \gamma\left(\frac{\omega_{L,t}^i(s_t)}{\theta_t^i(s_t)}\right)^\gamma v(L(s^t))\right) \Pr(s^t) = (\omega_c^i C_0)^{-1}((1 - \kappa_B(s_0))B^i(s_0) + R_0 k_0^i - T).$$

Equating the lump-sum T and using (33) to substitute for $\omega_{L,t}^L(s_t)$ reduces the two implementability conditions for $i = L, H$ to the single constraint (recall that the example assumes no inequality in initial wealth, $(1 - \kappa_B(s_0))B^i(s_0) + R_0 k_0^i$):

$$\sum_{t,s^t} \beta^t \Pr(s^t)(\omega_c^L - \omega_c^H - \gamma((\omega_c^L)^{-1/(\gamma-1)}\theta_t^L(s_t)^{\gamma/(\gamma-1)} - (\omega_c^H)^{-1/(\gamma-1)}\theta_t^H(s_t)^{\gamma/(\gamma-1)})(\Phi_{v,t}^m(s_t))^{\frac{\gamma}{\gamma-1}} v(L(s^t))) = 0. \tag{44}$$

Using equations (33) and (44), the planning problem can be written as (recall that the example assumes the Utilitarian specification with $\lambda^i = 1$) maximizing

$$\sum_{t,s^t} \beta^t(\log(c(s^t)) - d_t(s_t; \{\omega_c^i\},\tilde{\mu})v(L(s^t)) + \kappa(\{\omega_c^i\},\tilde{\mu})) \Pr(s^t), \tag{45}$$

subject to the resource constraint (2) and $\omega_c^L\pi^L + \omega_c^H\pi^H = 1$; here $\tilde{\mu}$ is the multiplier on equation (44), while

$$d_t(s_t; \{\omega_c^i\},\tilde{\mu}) \equiv \left(\sum_i \pi_i\left(\frac{\theta_t^i(s_t)}{\omega_c^i}\right)^{\gamma/(\gamma-1)} + \tilde{\mu}\gamma((\omega_c^L)^{-1/(\gamma-1)}\theta_t^L(s_t)^{\gamma/(\gamma-1)} - (\omega_c^H)^{-1/(\gamma-1)}\theta_t^H(s_t)^{\gamma/(\gamma-1)})\right)(\Phi_{v,t}^m(s_t))^{\frac{\gamma}{\gamma-1}}$$

$$\kappa(\{\omega_c^i\},\tilde{\mu}) \equiv \sum_i \pi_i \log(\omega_c^i) + \tilde{\mu}(\omega_c^L - \omega_c^H).$$

For the purpose of optimizing over the aggregates variables, $K(s^t)$, $c(s^t)$, and $L(s^t)$, the term involving $\kappa(\{\omega_c^i\},\tilde{\mu})$ is simply a constant that can be ignored. Now define

$$N(s^t) \equiv L(s^t)d_t(s_t;\ \{\omega_c^i\},\tilde{\mu})^{1/\gamma} \quad \text{and} \quad z(s^t) \equiv Ad_t(s_t;\ \{\omega_c^i\},\tilde{\mu})^{(\rho-1)/\gamma},$$

where the production function is the Cobb–Douglas specification $F(L,K) = AK^{\rho}L^{1-\rho}$. Using the resource constraint (2) to substitute consumption $c(s^t)$ out, reduces the problem of optimizing over aggregates to

$$\max_{N(s^t),K(s^t)} \sum_{t,s^t} \beta^t(\log\,(z(s^t)K(s^{t-1})^{\rho}N(s^t)^{1-\rho} - K(s^t)) - v(N(s^t)))\ \Pr(s^t),$$

which is exactly the Brock–Mirman problem, with stochastic (pseudo) technology shocks, $z(s^t)$, and (pseudo) labor supply, $N(s^t)$. The solution is well known (recalling that the example sets $\alpha = 1$):

$$N(s^t) = \bar{N} \equiv \left(\frac{1-\rho}{1-\rho\beta}\right)^{1/\gamma} \quad \text{and} \quad K_{t+1} = \rho\beta z_t K_t^{\rho} N^{1-\rho},$$

aggregate labor is then $L(s^t) = \bar{N}d_t(s_t;\ \{\omega_c^i\},\tilde{\mu})^{-1/\gamma}$.

Replacing the solution for aggregates into the objective function (45) gives (ignoring the term involving k_0, which does not involve $\{\omega_c^i\}$):

$$\max_{\{\omega_c^i\}} \left(\left(1 + \frac{\beta}{1-\beta\rho}\right) \sum_{t,s^t} \beta^t \mathbb{E}[\log\,(z(s^t))] + \frac{\kappa(\{\omega_c^i\},\tilde{\mu})}{1-\beta}\right), \tag{46}$$

subject to $\omega_c^L\pi^L + \omega_c^H\pi^H = 1$. By substituting out $\omega_c^H = (1 - \omega_c^L\pi^L)/\pi^H$, this becomes a one-dimensional optimization problem over ω_c^L. For each value of $\tilde{\mu}$, the solution to (46) yields an optimal value for ω_c^L (note that both ω_c^i and $\tilde{\mu}$ enter the definition of $z(s^t)$).

One then has a relation between $\tilde{\mu}$ and ω_c^L; the other relation needed to pin down the solution is given by the implementability condition (44), which can be rewritten as

$$\sum_{t,s^t} \beta^t(\omega_c^L - \omega_c^H - \psi_t(s_t))\ \Pr(s^t) = 0, \tag{47}$$

where

$$\psi_t(s_t) \equiv d_t(s_t;\ \{\omega_c^i\},\tilde{\mu})^{-1} \times ((\omega_c^L)^{-1/(\gamma-1)}\theta_t^L(s_t)^{\gamma/(\gamma-1)} - (\omega_c^H)^{-1/(\gamma-1)}\theta_t^H(s_t)^{\gamma/(\gamma-1)})\Phi_t(s_t)^{-\gamma}\bar{N}^{\gamma}.$$

Now, for each value of $\tilde{\mu}$, (47) can be seen as defining the value of ω_c^L, which is consistent with the implementability condition (44) (note that ω_c^L and $\tilde{\mu}$ enter the definition of $\psi_t(s_t)$).

The optimization in (46) and constraint (47) provide the two conditions that pin down the solution for $\tilde{\mu}$ and ω_c^L. The entire allocation can then be computed, and the tax rate is given by

$$\tau(s^t) = \bar{\tau}_t(s_t)$$

$$\equiv 1 - \frac{1}{(\pi_1 + \tilde{\mu}\gamma\omega_c^L)(\omega_{L,t}^L(s_t))/(\omega_c^L) + (\pi_2 - \tilde{\mu}\gamma\omega_c^H)(\omega_{L,t}^H(s_t))/(\omega_c^H)},$$

which is just an analog of formula (32).

Appendix III

Proof of Proposition 7. Consider any constrained-efficient allocation $\{c^{i*}(s^t), L^{i*}(s^t), K^*(s^t)\}$ with its associated multipliers $\{\eta^*(s^t)\}$ and $\{\phi_c^{i*}, \phi_L^{i*}\}$. The resource constraints (2) hold, so market clearing is guaranteed. Setting factor prices to their marginal products,

$$w(s^t) = F_L(L^*(s^t), K^*(s^{t-1}), s^t, t) \tag{48}$$

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

ensures firm maximization. Let

$$p(s^t) = \eta^*(s^t) \tag{49}$$

so that the no arbitrage condition (3) is then implied by (39). The only requirement for a competitive equilibrium left to be verified is that the constrained-efficient allocation solves the worker problem in (42) and (43).

It is useful to split the worker problem in (42) and (43), for any tax function Ψ and prices $\{p(s^t), w(s^t)\}$, into two stages. In the second stage, the worker solves the subproblem of choosing consumption and labor for any given present value of these variables. That is, they maximize (42) over $\{c^i(s^t), L^i(s^t)\}$ subject to

$$\sum_{t,s^t} p(s^t)c^i(s^t) = c_{PV}^i \quad \text{and} \quad \sum_{t,s^t} p(s^t)w(s^t)L^i(s^t) = L_{PV}^i.$$

for some given present-value levels of consumption and labor (c_{PV}^i, L_{PV}^i); let $V^i(c_{PV}^i, L_{PV}^i)$ denote the maximal utility attained. In the first stage the worker chooses the optimal pair (c_{PV}^i, L_{PV}^i) constrained by the budget constraint (43). That is, they solve

$$\max_{c_{PV}^i, L_{PV}^i} V^i(c_{PV}^i, L_{PV}^i) \quad \text{subject to } c_{PV}^i \leq L_{PV}^i - \Psi(L_{PV}^i).$$

Our goal is to construct a tax function Ψ so that, given the proposed prices, the constrained-efficient allocation solves these two stages for each worker type $i \in I$.

Equations (37) and (38) hold at the constrained-efficient allocation $\{c^{i*}(s^t), L^{i*}(s^t)\}$. Given the proposed prices (48) and (49), these are precisely the necessary and sufficient first-order conditions for the second stage of the worker's problem with $(c^{i*}_{PV}, L^{i*}_{PV})$ defined by

$$c^{i*}_{PV} \equiv \sum_{t,s^t} p(s^t)c^{i*}(s^t) \quad \text{and} \quad L^{i*}_{PV} \equiv \sum_{t,s^t} p(s^t)w(s^t)L^{i*}(s^t).$$

Thus, if one can ensure that $(c^{i*}_{PV}, L^{i*}_{PV})$ is chosen in the first stage by type $i \in I$ workers for some tax function Ψ then one can guarantee that the constrained-efficient allocation $\{c^{i*}(s^t), L^{i*}(s^t), K^*(s^t)\}$ is implemented. Turning to this first stage, set the tax function Ψ so that

$$\Psi(L^{i*}_{PV}) = L^{i*}_{PV} - c^{i*}_{PV} \quad \text{for all } i \in I, \tag{50}$$

and for values $L \neq L^{i*}_{PV}$ set $\Psi(L)$ in any way so that $\Psi(L) \geq \underline{\Psi}(L)$ where

$$\underline{\Psi}(L_{PV}) \equiv L_{PV} - \max_{c_{PV}} \cap \{c_{PV} : V^i(c_{PV}, L_{PV}) \leq V^i(c^{i*}_{PV}, L^{i*}_{PV})\}.$$

That is, take the set of points (c_{PV}, L_{PV}) that are not preferred by any worker to their corresponding $(c^{i*}_{PV}, L^{i*}_{PV})$. Then $\underline{\Psi}(L)$ is constructed from the frontier of this set and represents the most any worker is willing to pay (in present value) to produce L_{PV} and give up $(c^{i*}_{PV}, L^{i*}_{PV})$. Equivalently, take the set of indifference curves for $V^i(c^i_{PV}, L^i_{PV})$ for each worker type $i \in I$ that corresponds to $(c^{i*}_{PV}, L^{i*}_{PV})$ and write these in terms of c_{PV} as a function of L_{PV}; then $\underline{\Psi}$ is the lower envelope of these functions. Thus, it is continuous and piecewise differentiable.

Take any Ψ such that (50) holds and $\Psi(L_{PV}) \geq \underline{\Psi}(L_{PV})$ for $L_{PV} \neq L^{i*}_{PV}$. In particular, the function $\underline{\Psi}(L_{PV})$ will work. Then, since the original allocation is incentive compatible and the tax function $\Psi(L_{PV})$ only offers each worker further alternatives that are no better, the pair $(c^{i*}_{PV}, L^{i*}_{PV})$ solves the first stage of the problem of worker of type $i \in I$. Thus, given the constructed prices and taxes, the constrained-efficient allocation solves the worker problem in (42) and (43). This completes the proof.

APPENDIX IV: SEQUENTIAL IMPLEMENTATION EXAMPLE

Consider a deterministic version of the economy. The implementation builds on that in Section VII.C except that instead of taxing workers once and for all as a function of the present value of future earnings, workers pay taxes in all periods as a function of past earnings. Equivalently, given prices and wages $\{p_t,w_t\}$, taxes Ψ_t can be written as a function of effective labor choices. The budget constraint faced by workers is then

$$(51)\quad \sum_{t=0}^{\infty} p_t c_t^i \leq \sum_{t=0}^{\infty} p_t w_t L_t^i - \sum_{t=0}^{\infty} p_t \Psi_t(L_0^i, L_1^i, \ldots, L_t^i; \{p_t, w_t\}).$$

Now, take any continuous tax function Ψ and equilibrium prices $\{p_t,w_t\}$ that implement a constrained-efficient allocation, as guaranteed by Proposition 7. Then, for tax functions $\{\Psi_t\}$ satisfying

$$(52)\quad \sum_{t=0}^{\infty} p_t \Psi_t(L_0^i, L_0^i, \ldots, L_t^i; \{p_t, w_t\}) = \Psi\left(\sum_{t=0}^{\infty} p_t w_t L_t^i\right) \quad \text{for all nonnegative } \{L_t^i\},$$

the budget constraint (51) is equivalent to the deterministic version of (43) from the previous implementation; Proposition 7 then implies that workers will choose the constrained-efficient allocation. For example, setting $\Psi_0(L_0^i; \{p_t,w_t\}) = \Psi(p_0 w_0 L_0^i)/p_0$ and

$$\Psi_t(L_0^i, L_1^i, \ldots, L_t^i; \{p_t, w_t\}) = \frac{1}{p_t}\left(\Psi\left(\sum_{s=0}^{t} p_s w_s L_s^i\right) - \Psi\left(\sum_{s=0}^{t-1} p_s w_s L_s^i\right)\right) \quad \text{for } t = 1, \ldots$$

will satisfy condition (52) and ensure the sequential implementation of the constrained-efficient allocation as part of a competitive equilibrium.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

Aiyagari, S. Rao, Albert Marcet, Thomas J. Sargent, and Juha Soppälä, "Optimal Taxation without State-Contingent Debt," *Journal of Political Economy,* CX (2002), 1220–1254.

Albanesi, Stefania, and Christopher Sleet, “Dynamic Optimal Taxation with Private Information,” *Review of Economic Studies,* LXXIII (2006), 1–30.
Angeletos, George-Marios, “Fiscal Policy with Non-contingent Debt and the Optimal Maturity Structure,” *Quarterly Journal of Economics,* CXVII (2002), 1105–1131.
Atkeson, Andrew, V. V. Chari, and Patrick J. Kehoe, “Taxing Capital Income: A Bad Idea,” *Federal Reserve Bank of Minneapolis Quarterly Review,* XXIII (1999), 3–17.
Barro, Robert, “On the Determination of the Public Debt,” *Journal of Political Economy,* LXXXVII (1979), 940–71.
Bassetto, Marco, and Narayana Kocherlakota, “On the Irrelevance of Government Debt when Taxes are Distortionary,” *Journal of Monetary Economics,* LI (2004), 299–304.
Buera, Francisco J., and Juan Pablo Nicolini, “Optimal Maturity of Government Debt without State Contingent Bonds,” *Journal of Monetary Economics,* IV (2004), 531–554.
Chamley, Christophe, “Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives,” *Econometrica,* LIV (1986), 607–622.
Chari, V. V., Lawrence J. Christiano, and Patrick J. Kehoe, “Optimal Fiscal Policy in a Business Cycle Model,” *Journal of Political Economy,* CII (1994), 617–652.
Chari, V. V., and Patrick J. Kehoe, “Optimal Fiscal and Monetary Policy,” in *Handbook of Macroeconomics,* J. B. Taylor and M. Woodford, eds. (New York: Elsevier, 1999).
Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan, “Business Cycle Accounting,” Staff Report 328, Federal Reserve Bank of Minneapolis, 2006.
Diamond, Peter A., and James A. Mirrlees, “Optimal Taxation and Public Production II: Tax Rules,” *American Economic Review,* LXI (1971), 261–78.
——, “A Model of Social Insurance with Variable Retirement,” *Journal of Public Economics,* X (1978), 295–336.
Farhi, Emmanuel, “Capital Taxation and Ownership When Markets are Incomplete,” Mimeograph, 2005.
Farhi, Emmanuel, and Iván Werning, “Capital Taxation: Quantitative Explorations of the Inverse Euler Equation,” Mimeograph, M.I.T., 2005.
Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski, “Optimal Indirect and Capital Taxation,” *Review of Economic Studies,* LXX (2003), 569–587.
Hellwig, Martin, “The Optimal Linear Income Tax Revisited,” *Journal of Public Economics,* XXXI (1986), 163–179.
Hugget, Mark, Gustavo Ventura, and Amir Yaron, “Sources of Lifetime Inequality,” Mimeograph, Georgetown University, 2006.
Judd, Kenneth L., “Redistributive Taxation in a Perfect Foresight Model,” *Journal of Public Economics,* XXVIII (1985), 59–83.
Keane, Michael, and Kenneth Wolpin, “The Career Decisions of Young Men,” *Journal of Political Economy,* CV (1997), 473–522.
Kingston, Geoffrey, “Should Marginal Tax Rates be Equalized Over Time?” *Quarterly Journal of Economics,* CVI (1991), 911–924.
Lucas, Robert E., Jr., and Nancy L. Stokey, “Optimal Fiscal and Monetary Policy in an Economy without Capital,” *Journal of Monetary Economics,* XII (1983), 55–93.
Mirrlees, James A., “An Exploration in the Theory of Optimum Income Taxation,” *Review of Economic Studies,* XXXVIII (1971), 175–208.
Rogerson, William P., “Repeated Moral Hazard,” *Econometrica,* LIII (1985), 69–76.
Sheshinski, Eytan, “The Optimal Linear Income-Tax,” *Review of Economic Studies,* XXXIX (1972), 297–302.
Storesletten, Kjetil, Christopher Telmer, and Amir Yaron, “Consumption and Risk Sharing Over the Life Cycle,” *Journal of Monetary Economics,* LI (2004), 609–633.
Vickrey, William S., *Agenda for Progressive Taxation* (New York: Ronald, 1947).
Werning, Iván, “Optimal Dynamic Taxation,” Ph.D. thesis, University of Chicago, 2002.
——, “Tax Smoothing with Incomplete Markets,” Mimeograph, M.I.T., (2005).
Zhu, Xiaodong, “Optimal Fiscal Policy in a Stochastic Growth Model,” *Journal of Economic Theory,* LVIII (1992), 250–89.