

Optimal Government Policies in Models with Heterogeneous Agents

Radim Boháček and Michal Kejak*
CERGE-EI, Prague, Czech Republic

October 15, 2010

Abstract

In this paper we develop a new approach for finding optimal government policies in economies with heterogeneous agents. Using the calculus of variations, we present three classes of equilibrium conditions from government's and individual agent's optimization problems: 1) the first order conditions: the government's Lagrange-Euler equation and the individual agent's Euler equation; 2) the stationarity condition on the distribution function; and, 3) the aggregate market clearing conditions. These conditions form a system of functional equations which we solve numerically. The solution takes into account simultaneously the effect of the government policy on individual allocations and the resulting optimal distribution of agents in the steady state. This approach is applicable to a wide class of general equilibrium, Bewley type economies where the government looks for an optimal nonlinear, second-best fiscal or monetary policy. We illustrate it on a steady state Ramsey problem with heterogeneous agents, finding the optimal tax schedule.

JEL Keywords: Optimal macroeconomic policy, optimal taxation, computational techniques, heterogeneous agents, distribution of wealth and income

*Contact: CERGE-EI, Politických veznu 7, 111 21 Prague 1, Czech Republic. Email: radim.bohacek@cerge-ei.cz, michal.kejak@cerge-ei.cz. First version: September 2002. For helpful comments we thank Jim Costain, Mark Gertler, Max Gillman, Boyan Jovanovic, Marek Kapička, Dirk Krueger, Josep Pijoan-Mas, Thomas Sargent, Harald Uhlig, Gianluca Violante, Galyna Vereshchagina, and the participants at the CNB/CERGE-EI Macro Workshop 2004, SED 2004 conference, University of Stockholm, the Macroeconomic Seminar at the Federal Reserve Bank of Minneapolis, SED 2007 conference, Cardiff Business School, and the Macroeconomic Seminar at the Stern School of Business of NYU. We are especially grateful to Michele Boldrin, Tim Kehoe and Stan Zin for their support and advice. Anton Tyutin, Jozef Zubrický and Tomáš Leško provided excellent research assistance. All errors are our own.

1 Introduction

This paper provides a new approach for computing equilibria in which the stationary distribution of agents is a part of an optimal nonlinear, second-best in a general equilibrium, Bewley type economy with heterogeneous agents. We formulate the optimal government policy problem as a calculus of variations problem where the government maximizes an objective functional subject to a system of operator constraints: 1) the first order conditions from the individual agent's problem; 2) the stationarity condition on the distribution function; and, 3) the aggregate market clearing conditions. The first order necessary conditions of the government functional problem given by the Euler-Lagrange equation (with transversality conditions) form altogether a system of functional equations in individual agents' and government's policies and in the distribution function over agents' individual state variables. We solve this system numerically by the standard projection method.

It should be emphasized that our approach does not use any additional restrictions or assumptions on the equilibrium allocations but is strictly derived from the first order and envelope conditions and from the stationarity of the endogenous distribution in the steady state. Our main contribution is in the formulation of the Euler-Lagrange equation for the government problem and for the stationary distribution over individual state variables. In this way, we are able to solve *simultaneously* for the *government optimal policy*, for the *optimal individual allocations*, and for the (from a government's point of view) *optimal distribution* of agents in the steady state. Additionally, the derived first-order transversality conditions for the boundary agents allow for a qualitative analysis of the shape of the optimal government policy function. To our knowledge, this paper is the first one that provides a solution method for this kind of problem.

Our approach can be applied to a wide range of government optimal fiscal or monetary policy problems. We illustrate this general methodology on a steady state Ramsey problem with heterogeneous agents. We recast the original Ramsey (1927) and Lucas (1990) normative question for an economy with heterogeneous agents: What choice of a tax schedule will lead to maximal social welfare in the steady state, consistent with given government consumption and with market determination of quantities and prices? What is the welfare differential with respect to social welfare resulting from the existing progressive tax schedule in the U.S. economy and as well as from the usual flat-tax reform?

The tradeoff between efficiency and income or wealth distribution plays a central role in analyzing tax policies. In dynamic, general equilibrium models with household heterogeneity from uninsurable idiosyncratic risk, a tax schedule provides incentives for agents to accumulate wealth. The optimal tax schedule seeks to arrive at such a steady state where the distribution of agents is optimal with respect to the aggregate welfare in the economy.

The steady state wealth distribution is perhaps the most important part of the resulting equilibrium: it determines equilibrium prices and fractions of poor and rich agents, i.e. all elements that affect agents' ability to self-insure against idiosyncratic shocks through the accumulated buffer stock of savings. In our example, we find a welfare maximizing tax schedule on the total income from labor and capital which takes into account simultaneously its effects on agents' allocations and on the stationary distribution of agents in a steady state. Previous models analyzing the effects of government policies in this class of models were limited to sub-optimal policy reforms exogenously imposed on the model. Within the context of optimal taxation, several papers have analyzed the steady state implications (and transition paths) resulting from an *ad hoc* flat-tax reform or from an *ad hoc* removal of double taxation of capital income.

In this paper, we solve for the *optimal* tax schedule on the total income that maximizes aggregate welfare in a steady state of a standard neoclassical, general equilibrium, full information and full commitment economy with heterogeneous agents and incomplete markets. In order to evaluate the benefits of the optimal tax schedule, we compare the steady state aggregate levels, welfare, efficiency and distribution of resources associated with this optimal tax schedule to a simulated steady state of the U.S. economy with the existing progressive tax schedule and to a steady state resulting from a standard flat-tax reform.

The optimal tax schedule we find is a function that is neither progressive nor monotone. It is a positive, U-shaped function, taxing the lowest income at 45%, decreasing to a minimum of 19% and rising to 62% at the highest level of total income. It provides incentives for agents to accumulate assets while preserving the equality measures in the economy. Its impact on aggregate levels and welfare is large. Compared to the progressive tax schedule steady state, average welfare increases by 4.4%, capital stock by 49%, output by 15.8%, and consumption by 5.8%. Relative to the flat-tax steady state, welfare goes up by 0.8%, capital stock by 15%, output by 4.5%, and consumption by 1.1%. The marginal tax rate is also a U-shaped function, but almost flat at the low incomes, reaching negative levels around the average income, and then rising to positive levels.

The efficiency and distributional effects of the optimal tax are the main mechanisms behind these large changes. Related to the efficiency are the general equilibrium effects: a higher stock of capital increases productivity of labor and, therefore, the income of poor agents. For the steady state distributional effects, the optimal tax schedule concentrates the agents around the mean at high levels of wealth, something what a social planner with an access to the first-best, lump-sum transfers would do. The high tax rate at low income levels provides incentives for these agents to save more for precautionary reasons in the long

run steady state. The even higher tax rate on high incomes discourages additional savings by the wealthiest agents. In the middle of the total income levels, the tax rate is lower than the one found for the flat-tax reform. In this way, the optimal tax schedule solves the tradeoff between efficiency and equality. For comparison, the *ad hoc* flat tax reform also increases aggregate levels but does not take into account the distribution of agents. On the other hand, the progressive tax schedule provides too much short-run insurance at the cost of the long-run average levels.

Finally, in order to evaluate the short run costs of the optimal tax reform, we compute transitions from the progressive and the flat-rate tax schedule steady states to the steady state of the optimal tax schedule. We find that a majority of population, 73%, would benefit from a reform that replaces the progressive tax schedule by the optimal tax schedule. On the other hand, only one third of agents would support the reform starting from the steady state with flat tax. These results as well as a detailed efficiency and distributional analysis are described in a great detail in the following sections.

We limit our example to the optimal tax schedule on the total income from labor and capital that is needed to raise a given fraction of GDP. There are several reasons why we choose this setup. First, the tax on the total income enables us to study a tax system with a non-degenerate distribution of agents in a steady state. If the government had an access to a lump-sum, first best taxation the model would collapse to a representative agent one. Second, to a large extent the current U.S. tax code does not distinguish between the sources of taxable income. The last reason for a simple tax on the total income is the complexity of the problem we solve.

By focusing on a steady state analysis and by imposing a single tax rate on labor and capital income we tried to isolate the shape of the optimal tax schedule on total income. The closest paper to ours is Conesa and Krueger (2006), who compute the optimal *progressivity* of the income tax code in an overlapping generations economy. They search in a class of monotone tax functions to find a welfare maximizing tax schedule. We show in this paper that limiting the analysis to monotone functions seems rather restrictive with respect to welfare maximization.

In this paper, we do not address the following important issues related to optimal taxation: the issue of time-consistency, the issue of the optimal capital income tax rate in the steady state, and the distortionary effects of taxation on labor supply. Our government is fully and credibly committed, the tax schedule is constant over time.¹ Aiyagari (1995) showed that for our class of models with incomplete insurance markets and borrowing constraints, the optimal tax rate on capital income is positive even in the long run (see

¹For the time-consistency problem see Kydland and Prescott (1977) and Klein and Rios-Rull (2004).

also a recent paper by Conesa, Kitao, and Krueger (2009)).² Due to the complexity of our work, we study the simplest utility maximization problem on the consumption-investment margin. However, we would like to stress that our methodology can be applied to different aggregate welfare criteria and a wide variety of optimal government policy problems, including those with endogenous labor supply, separate taxation of labor and capital incomes as well as public goods, population growth, or a life-cycle earnings process as in Ventura (1999). Finally, in future research we also plan to analyze a much more difficult problem, that of a stationary competitive equilibrium which is the limit of the optimal dynamic tax schedule.

The paper is organized as follows. The following section describes the economy with heterogeneous agents, defines the stationary recursive competitive equilibrium and the stationary Ramsey problem. Section 3 specifies the equilibrium as a system of functional equations and defines the operator stationary Ramsey problem. Section 4 formulates the Ramsey problem by means of the calculus of variations. The first-order necessary conditions for the optimal government policy schedule expressed in the form of a generalized Euler-Lagrange equation including the transversality conditions and related analytical results are described in Section 5. Section 6 illustrates our approach by an example of the optimal income tax schedule. Section 7 presents the numerical solution and Section 8 concludes. Appendix contains the proofs and analytical results.

2 The Economy

The economy is populated by a continuum of infinitely lived agents on a unit interval. Each agent has preferences over consumption in period $t \geq 0$, c_t , given by a utility function

$$E \sum_{t=0}^{\infty} \beta^t U(c_t), \quad 0 < \beta < 1,$$

where $U : \mathbb{R}_+ \rightarrow \mathbb{R}$ is a twice continuously differentiable, strictly increasing and strictly concave function. We assume that the utility function satisfies the Inada conditions.

At all $t \geq 0$, each agent is identified by an endogenous state variable, the accumulated stock of capital, $k_t \in B = [\underline{k}, \infty)$ with $\underline{k} = 0$, and by an exogenous labor productivity shock $z_t \in Z = \{z_1, z_2, \dots, z_J\}$. The shock represents labor efficiency units and follows a first-order Markov chain with a transition function $Q(z, z') = \text{Prob}(z_{t+1} = z' | z_t = z)$. We assume that Q is monotone, satisfies the Feller property and the mixing condition defined in Stokey, Lucas, and Prescott (1989). As the labor productivity shock is independent

²For the optimal capital tax with two types of agents see Chamley (1986) and Judd (1985).

across agents there is no uncertainty at the aggregate level. We preserve the heterogeneity in the economy by assuming incomplete markets.

In each period, agents supply labor and accumulated capital stock to a representative firm with a production function $F(K_t, L_t)$, where $K_t \in B$ is the aggregate capital stock, $L_t \in \mathbb{R}_+$ is the aggregate effective labor. The production function is concave, twice continuously differentiable, increasing in both arguments, and displays constant returns to scale. Profit maximization implies the following factor prices

$$r_t = F_K(K_t, L_t) - \delta \quad \text{and} \quad w_t = F_L(K_t, L_t), \quad (1)$$

where $\delta \in (0, 1)$ is the depreciation rate of capital.

Finally, there is a government that finances its expenditures by taxing the agents in the economy. We assume the government is fully committed to a sequence of tax schedules $\pi = \{\pi_t\}_{t=0}^\infty$ to finance its expenditures. We assume that these expenditures equal a constant fraction of output, g , in each period, that they are not returned to the agents, and that the government cannot use the first best, lump-sum taxation.³ The policy π is applied to a broadly defined taxable activity of each agent, $x_t \in \mathbb{R}_+$. We will assume that $x_t = x(z_t, k_t)$ where $x : Z \times B \rightarrow \mathbb{R}_+$ and $x_z, x_k > 0$. Thus in each period, the policy schedule is a function $\pi_t : \mathbb{R}_+ \rightarrow \mathbb{R}$, so that an agent with a total income from labor and capital, $y_t \in \mathbb{R}_+$, $y_t = y(k_t, z_t) = r_t k_t + w_t z_t$, and a taxable activity $x_t = x(k_t, z_t)$ pays taxes $\pi_t(x_t)x_t$ and is left with an after-tax income $y_t - \pi_t(x_t)x_t$. In our example in Section 6, we illustrate this policy by a proportional taxation of this total income from labor and capital, i.e. when $x = y$.

The economy's state is characterized by the sequence of government policies π , and by a distribution of agents over capital and shock in each period, $\lambda = \{\lambda_t\}_{t=0}^\infty$. The latter is in each period a probability measure defined on subsets of the state space, describing the heterogeneity of agents over their individual state $(z, k) \in Z \times B$. Let (B, \mathcal{B}) and (Z, \mathcal{Z}) be measurable spaces, where \mathcal{B} denotes the Borel sets that are subsets of B and \mathcal{Z} is the set of all subsets of Z .

2.1 Stationary Recursive Competitive Equilibrium

We will analyze the economy in a stationary recursive competitive equilibrium in which the government policy schedule and the distribution of agents are time-invariant. Given the equilibrium prices and the time-invariant government policy π , an agent (z, k) solves

³Our analysis equally applies to the case when government finances any expenditures $\{G_t\}_{t=0}^\infty$ and corresponding revenue neutral reforms.

the following dynamic programming problem

$$v(z, k) = \max_{c, k^+} \left\{ u(c) + \beta \sum_{z^+} v(z^+, k^+) Q(z, z^+) \right\}, \quad (2)$$

subject to a budget constraint

$$c(z, k) + k^+(z, k) \leq y + k - \pi(x(z, k))x(z, k), \quad (3)$$

with taxable activity $x(z, k)$, total income $y = rk + wz$, a borrowing constraint,

$$k^+(z, k) \geq \underline{k}. \quad (4)$$

Definition 1 (Stationary Recursive Competitive Equilibrium) *For a given share of government expenditures g and a time-invariant government policy schedule π , a stationary recursive competitive equilibrium is a set of functions (v, c, y, x, k^+) , aggregate levels (K, L) , prices (r, w) , and a probability measure λ , such that*

1. *given prices and the government policy, the policy functions solve each agent's optimization problem (2);*
2. *firms maximize profit (1);*
3. *the probability measure is time invariant,*

$$\lambda(z^+, B^+) = \sum_z \int_{\{(z, k) \in Z \times B: k^+(z, k) \in B^+\}} Q(z, z^+) \lambda(z, k) dk, \quad (5)$$

for all $(z^+, B^+) \in \mathcal{Z} \times \mathcal{B}$;

4. *the aggregate conditions hold,*

$$K = \sum_z \int k^+(z, k) \lambda(z, k) dk, \quad (6)$$

$$L = \sum_z \int z \lambda(z, k) dk; \quad (7)$$

5. *the government budget constraint holds at equality,*

$$g = \sum_z \int \pi(x(z, k))x(z, k) \lambda(z, k) dk / F(K, L); \quad (8)$$

6. *and the allocations are feasible,*

$$\sum_z \int [c(z, k) + k^+(z, k)] \lambda(z, k) dk + gF(K, L) = F(K, L) + (1 - \delta)K. \quad (9)$$

Agents have rational expectations, take the behavior of prices as given by a predetermined function that depends on aggregate variables in equation (1). As we look for optimal government policy schedule the goal of our paper is to solve the following stationary Ramsey problem.

Definition 2 (Stationary Ramsey Problem) *A solution to the Ramsey problem for a stationary economy with heterogeneous agents is a time-invariant government policy schedule $\pi : \mathbb{R}_+ \rightarrow \mathbb{R}$ that maximizes social welfare in the steady state,*

$$\max_{\pi} \sum_z \int v(z, k; \pi) \lambda(z, k; \pi) dk,$$

consistent with a given government consumption and with allocations satisfying the definition of the stationary recursive competitive equilibrium, where $v : Z \times B \rightarrow \mathbb{R}$ is the value function of individual agents and $\lambda : Z \times B \rightarrow [0, 1]$ is the stationary distribution.

Our notation indicates that the value and distribution functions depend on the government policy, π , i.e. $v(z, k; \pi)$ and $\lambda(z, k; \pi)$, respectively.

It is easy to show that the solution to the Stationary Ramsey Problem is equivalent to that of maximizing the average current period utility,

$$\max_{\pi} \sum_z \int u(c(z, k; \pi)) \lambda(z, k; \pi) dk.$$

In the following Sections, we will characterize the optimal government policy schedule using this latter specification.⁴

3 The Operator Stationary Ramsey Problem

Since the problem is to find an optimal, welfare maximizing time-invariant function $\pi : \mathbb{R}_+ \rightarrow \mathbb{R}$, the Stationary Ramsey Problem can be transformed into an operator form. In order to express the stationary recursive competitive equilibrium in this form, we define two operators: an operator on the Euler equation, \mathcal{F} , and that on the stationary distribution, \mathcal{L} .

For a given government policy schedule $\pi : \mathbb{R}_+ \rightarrow \mathbb{R}$ the Euler equation operator is defined on the savings function $h : Z \times B \rightarrow B$, and the stationary distribution operator is defined on the distribution function $\lambda : Z \times B \rightarrow [0, 1]$, and the savings function h . We

⁴Our analysis of optimal government policies can also be applied to other types of welfare functions.

will assume that these functions are square integrable functions on some closed domain⁵: $h, \lambda \in L^2(Z \times B)$ where $L^2(Z \times B)$ is a Hilbert space with the inner product $(u, v) = \int_{Z \times B} u(t)v(t)dt$. The operator $\mathcal{F} : C^1(Z \times B) \subset L^2(Z \times B) \rightarrow C^1(Z \times B) \subset L^2(Z \times B)$ is the mapping from a space of continuously differentiable functions into a space of continuously differentiable functions; and the operator $\mathcal{L} : C^1(Z \times B) \times C^1(Z \times B) \rightarrow C^1(Z \times B) \subset L^2(Z \times B)$.

Operator \mathcal{F} on the Euler Equation An individual agent's allocations are characterized by the Euler equation from the optimization problem (2)-(4). For all $(z, k) \in Z \times B$,

$$u'(y - \pi(x)x + k - k^+) \geq \beta \sum_{z^+} u'(y^+ - \pi(x^+)x^+ + k^+ - k^{++}) \cdot \{1 + y_k^+ [\pi(x^+) + \pi'(x^+)x^+] x_k^+\} Q(z, z^+), \quad (10)$$

where

$$\begin{aligned} y &= rk + wz, \\ x &= x(z, k), \\ y^+ &= rk^+ + wz^+, \\ x^+ &= x(z^+, k^+), \\ x_k^+ &= x_k(z^+, k^+), \\ y_k^+ &= r \frac{\partial k^+}{\partial k}. \end{aligned}$$

Clearly, $y - \pi(x)x$ is disposable income and $\{1 + y_k^+ - [\pi(x^+) + \pi'(x^+)x^+]x_k^+\}$ is the next period 'after-policy' marginal return to capital where π' is the marginal government policy schedule.

The solution of the Euler equation is a time invariant savings function $h : Z \times B \rightarrow B$. In the stationary equilibrium, prices, as functions of the aggregate capital K , are constant. When the government searches for the optimal policy schedule it needs to take into account the effect of its policy on prices. For this reason it will be advantageous to introduce aggregate capital, K , as an explicit variable in the specification of the equilibrium. In order to make this effect more transparent, we will write the equilibrium prices as $r(K)$ and $w(K)$. As the savings function depends on the aggregate capital stock (through equilibrium prices) and on the government policy, π , we denote $k^+ = h(z, k; K, \pi)$ and $k^{++} = h(z^+, k^+; K, \pi) = h(z^+, h(z, k; K, \pi); K, \pi)$. Since $K = \sum_z \int k \lambda(z, k; \pi) dk$, the optimal savings function is affected through prices by the distribution function λ .

⁵In more precise terms we actually assume that the functions are from the subspace $W^{1,2}(Z \times B)$ which contains $L^2(Z \times B)$ -functions which have weak derivatives of order one.

In the text below, we present only the case of the unconstrained agents (i.e. those for whom the Euler equation holds with equality and $k^+(z, k; K, \pi) > \underline{k}$). The case of the borrowing constrained agents is discussed in Appendix A.

The operator on Euler equation \mathcal{F} is defined by,

$$\mathcal{F}(h; \pi) \equiv u'(c) - \beta \sum_{z^+} u'(c^+) \{1 + y_k^+ - [\pi(x^+) + \pi'(x^+)x^+] x_k^+\} Q(z, z^+), \quad (11)$$

where

$$\begin{aligned} c &= y(z, k; K) - \pi(x(z, k; K))x(z, k; K) + k - h(z, k; K, \pi), \\ y(z, k; K) &= r(K)k + w(K)z, \\ c^+ &= y(z^+, h(z, k; K, \pi); K) - \pi(x(z^+, h(z, k; K, \pi); K))x(z^+, h(z, k; K, \pi); K) \\ &\quad + h(z, k; K, \pi) - h(z^+, h(z, k; K, \pi); K, \pi), \end{aligned}$$

and the terms $y_k^+ = y_k(z^+, h(z, k; K, \pi); K)$ and $x_k^+ = x_k(z^+, h(z, k; K, \pi); K)$ are the marginal effects of individual savings on income y and taxable activity x in the next period. The operator equation is simply $\mathcal{F}(h) = 0$.

Operator \mathcal{L} on the Stationary Distribution In order to guarantee a unique stationary distribution of agents in the steady state we need to assume that the government policy function π is such that the resulting individual agents' behavior does not display pathological features (for example, that wealthy agents save less than poor agents). The following assumption is used in virtually all models with heterogeneous agents as a part of conditions for the existence of a unique equilibrium and is innocuous.

Assumption 1 *Assume that for all $z \in Z$ and for a given (K, π) , the individual savings function $h : Z \times B \rightarrow B$ is a monotone function of k over the whole interval $[\underline{k}(z), \infty]$.*

Thus for all $z \in Z$ and for a given pair (K, π) , there exists an inverse function h^{-1} assigning a current value of capital k to savings k^+ according to $k = h^{-1}(z, k^+; K, \pi)$. In Section 6, we show in Theorem 2 that our example economy is consistent with this assumption.

In the stationary recursive competitive equilibrium, under the Assumption 1 above, the operator \mathcal{L} for equation (5) is

$$\mathcal{L}(\lambda, h; \pi) \equiv \lambda(z^+, k^+; K, \pi) - \sum_z \lambda[z, h^{-1}(z, k^+; K, \pi); K, \pi] Q(z, z^+), \quad (12)$$

for all $(z^+, k^+) \in Z \times [h(\underline{k}(z), z), \bar{k}(z)]$.

The related operator equation is $\mathcal{L}(\lambda, h) = 0$.

Definition 3 (Operator Stationary Recursive Competitive Equilibrium) *Given a share of government expenditures g and a time-invariant government policy schedule π , an operator stationary recursive competitive equilibrium is a set of operators $(\mathcal{F}, \mathcal{L})$, prices (r, w) , a savings function h , a probability measure λ , and aggregate levels (K, L) such that*

1. *given prices and government policy π , the policy function h is the solution to each agent's optimization problem in the operator equation*

$$\mathcal{F}(h; \pi) = 0, \tag{13}$$

2. *firms maximize profit (1),*
3. *the time-invariant probability measure is the solution to the operator equation*

$$\mathcal{L}(h, \lambda; \pi) = 0, \tag{14}$$

4. *the capital and labor markets clear, (6)-(7),*
5. *the government budget constraint (8) holds at equality,*
6. *and the allocations are feasible, (9).*

We can now specify an operator version of the Stationary Ramsey Problem.

Definition 4 (Operator Stationary Ramsey Problem) *A solution to the Operator Stationary Ramsey Problem for an economy in a stationary recursive competitive equilibrium in Definition 1 is a time-invariant government policy π that maximizes social welfare in the steady state,*

$$\arg \max_{\pi} \sum_z \int_{\underline{k}}^{\bar{k}} W(v(z, k; K, \pi)) \lambda(z, k; K, \pi) dk, \tag{15}$$

subject to a system of operator equations (13)-(14), consistent with equilibrium prices (1) and the market clearing conditions (6)-(7) in Definition 1, where $v : Z \times B \rightarrow \mathbb{R}$ is the value function of individual agents, $\lambda : Z \times B \rightarrow [0, 1]$ is the stationary distribution, W is a positive linear social aggregator function, \underline{k} is the exogenously given lower bound on capital holding, and \bar{k} is the endogenous upper bound on individual savings, i.e. $\bar{k} = h(\bar{z}, \bar{k}; K, \pi)$.

4 Ramsey Problem as Calculus of Variations Problem

The first-order conditions for the solution π to the Operator Stationary Ramsey Problem in Definition 4 are best formulated in the calculus of variations.⁶ As it is standard, the government policy function π and its derivative π' are treated as two independent functions. Additionally, since we are looking for the government policy as a function of individual activity x rather than capital k , we now reformulate the problem with activity x as an independent variable.

The social welfare function in equation (15) at the new coordinates x is

$$\sum_z \int_{\underline{k}}^{\bar{k}} W[u(c(z, k; K, \pi, \pi'))] \lambda(z, k; K, \pi, \pi') dk = \sum_z \int_{\underline{x}(z)}^{\bar{x}(z)} \mathcal{W}[z, x; K, \pi(x), \pi'(x)] dx$$

with

$$\mathcal{W}[z, x; K, \pi(x), \pi'(x)] \equiv W[u(c(z, k(z, x; K); K, \pi, \pi'))] \lambda[z, k(z, x; K); K, \pi, \pi'] k_x(z, x; K), \quad (16)$$

where $k_x[z, x; K] = [x_k(z, k(z, x; K); K)]^{-1}$ is the inverse function to the marginal effect of individual savings on the taxable activity x_k .

Before we proceed further, we have to clarify two important aspects of this dynamic optimal problem. First, observing that K , which determines the equilibrium prices, is one of the arguments in the objective function above, we have to use the condition on the aggregate capital stock. Writing the condition properly using the fact that K also depends on the government policy π and its derivative π' , we get at the new coordinates

$$K = \sum_z \int_{\underline{k}}^{\bar{k}} k \lambda(z, k; K, \pi, \pi') dk = \sum_z \int_{\underline{x}(z)}^{\bar{x}(z)} \mathcal{K}[z, x; K, \pi(x), \pi'(x)] dx,$$

where

$$\mathcal{K}[z, x; K, \pi(x), \pi'(x)] \equiv k(z, x; K) \lambda[z, k(z, x; K); K, \pi, \pi'] k_x(z, x; K).$$

Second, the bounds on taxable activity, $\underline{x}(z)$ and $\bar{x}(z)$, for each $z \in Z$, are endogenous functions of a chosen government policy. The lower bound $\underline{x}(z) = \underline{x}(z, \underline{k}; K)$ depends on z , on the exogenously given lower bound on capital, \underline{k} , and on the equilibrium prices

⁶Mirlees (1976) also uses the calculus of variations to derive the first-order conditions for the optimum income tax schedule in an economy with heterogenous agents. However, while his problem is a static one with an exogenously imposed distribution of agents' abilities we have a dynamic problem with an endogenous distribution of agents.

determined by K . The upper bound $\bar{x}(z) = \bar{x}(z, \bar{k}; K)$ depends on z , on the endogenous upper bound of capital, \bar{k} , and also on K .⁷

Finally, we also need to reformulate the side condition of the problem given by the government budget constraint in equation (8), i.e.

$$\begin{aligned} & \sum_z \int_{\underline{k}}^{\bar{k}} [\pi(x(z, k; K))x(z, k; K) - g y(z, k; K)] \lambda(z, k; K, \pi, \pi') dk \\ = & \sum_z \int_{\underline{x}(z)}^{\bar{x}(z)} \mathcal{G}[z, x; K, \pi(x), \pi'(x)] dx = 0, \end{aligned} \quad (17)$$

where

$$\mathcal{G}[z, x; K, \pi(x), \pi'(x)] \equiv [\pi(x)x - g y(z, k(z, x; K); K, \pi, \pi')] \lambda[z, k(z, x; K); K, \pi, \pi'] k_x(z, x; K).$$

Definition 5 (Calculus of Variation Ramsey Problem) *The Ramsey problem in the calculus of variation is formulated as the following generalized isoperimetric problem,*

$$\max_z \sum_z \int_{\underline{x}(z)}^{\bar{x}(z)} \mathcal{W}[z, x; K, \pi(x), \pi'(x)] dx, \quad (18)$$

subject to

$$\sum_z \int_{\underline{x}(z)}^{\bar{x}(z)} \mathcal{G}[z, x; K, \pi(x), \pi'(x)] dx = 0, \quad (19)$$

with the definition of the aggregate capital stock in equation (17), the individual policy function h given implicitly by the operator Euler equation (13), the distribution function, λ , given implicitly by the operator equation (14), the endogenously determined bounds of taxable activity, $\underline{x}(z)$ and $\bar{x}(z)$ for all values of $z \in Z$, and the free values of the government policy at the extreme lower and upper bounds, $\pi(\underline{x}(\underline{z}))$, and $\pi(\bar{x}(\bar{z}))$.

Note that since \bar{k} is endogenous, the endpoint $\bar{x}(\bar{z})$ is equality constrained.

⁷Clearly, the maximal interval is $[\underline{x}(\underline{z}), \bar{x}(\bar{z})]$ where $\underline{x}(\underline{z})$ is the lower bound of the lowest shock, \underline{z} , and $\bar{x}(\bar{z})$ is the upper bound of the highest shock, \bar{z} . So any taxable activity interval associated with a shock $z \in Z$ is a subinterval of the maximal interval, $[\underline{x}(z), \bar{x}(z)] \subset [\underline{x}(\underline{z}), \bar{x}(\bar{z})]$.

5 Necessary Conditions for the Optimal Government Policy Schedule

In order to derive the first-order conditions, we define the Lagrange function \mathbf{L} for the Calculus of Variations Ramsey Problem in Definition 5 as

$$\mathbf{L}(z, x) = \begin{cases} 0 & \text{for } x \in [\underline{x}(z), \underline{x}(z)], \\ \mathcal{W}(z, x) + \mu\mathcal{G}(z, x) & \text{for } x \in [\underline{x}(z), \bar{x}(z)], \\ 0 & \text{for } x \in (\bar{x}(z), \bar{x}(z)], \end{cases} \quad (20)$$

for each $z \in Z = \{z, \bar{z}\}$.

Note that the social welfare function in (16) is the sum of integrands $\mathcal{W}(z, x) = \mathcal{W}[z, x; K, \pi(x), \pi'(x)]$ integrated on the intervals $[\underline{x}(z), \bar{x}(z)]$ for each $z \in Z$. The same is true for integrands $\mathbf{L}(z, x)$ in equation (20).

Interestingly, as we show in Theorem 1 below, where we derive the first-order conditions, the relevant Lagrange function that emerges from the solution to the maximization problem is one amended by a term which captures the effect of the distribution of capital on the social welfare: $\tilde{\mathbf{L}}(z, x) = \mathbf{L}(z, x) + \Psi\mathcal{K}(z, x)$, where Ψ is the marginal effect of the aggregate capital on social welfare.⁸ In other words, the term $\Psi\mathcal{K}(z, x)$ takes into account the effect of prices (determined by the aggregate capital) on social welfare of agents characterized by (z, x) .

Theorem 1 (First Order Necessary Conditions) *Using the modified Lagrange function $\tilde{\mathbf{L}}$ for the Calculus of Variations Ramsey Problem in Definition 5,*

$$\tilde{\mathbf{L}}(z, x) = \begin{cases} 0 & \text{for } x \in [\underline{x}(z), \underline{x}(z)], \\ \mathcal{W}(z, x) + \mu\mathcal{G}(z, x) + \Psi\mathcal{K}(z, x) & \text{for } x \in [\underline{x}(z), \bar{x}(z)], \\ 0 & \text{for } x \in (\bar{x}(z), \bar{x}(z)], \end{cases} \quad (21)$$

for each $z \in Z = \{z, \bar{z}\}$, the first order necessary conditions for the Ramsey problem are

1. the Euler-Lagrange condition,

$$\sum_z \left(\tilde{\mathbf{L}}_\pi(z, x) - \frac{d}{dx} \tilde{\mathbf{L}}_{\pi'}(z, x) \right) = 0; \quad (22)$$

2. the transversality condition on the free boundary value, $\pi(\bar{x}(\bar{z}))$, at the equality constrained endpoint, $\bar{x}(\bar{z})$,

$$\left[\tilde{\mathbf{L}}(\bar{z}, x) - \left(\pi'(x) - \frac{k_x(\bar{z}, x)}{\omega_\pi(\bar{z}, x)} \right) \tilde{\mathbf{L}}_{\pi'}(\bar{z}, x) \right]_{x=\bar{x}(\bar{z})} = 0; \quad (23)$$

⁸Note that Ψ is the effect of variation in K on the variation in \mathbf{L} , i.e. $\delta\mathbf{L}/\delta K$.

3. the transversality condition on the free boundary value $\pi(\underline{x}(\underline{z}))$ at $\underline{x}(\underline{z})$

$$\left[\tilde{\mathbf{L}}_{\pi'}(\underline{z}, x) \right]_{x=\underline{x}(\underline{z})} = 0; \quad (24)$$

4. and the condition on the Langrange multiplier, μ , at which (19) is satisfied.

The marginal effect Ψ of the aggregate capital stock on social welfare is

$$\begin{aligned} \Psi \equiv \Psi_K \sum_z \left\{ \int_{\underline{x}(z)}^{\bar{x}(z)} \mathbf{L}_K(z, x) dx + [(\mathbf{L}(z, x) - \pi'(x)\mathbf{L}_{\pi'}(z, x)) x_K]_{x=\underline{x}(z)} \right. \\ \left. + [(\mathbf{L}(z, x) - \pi'(x)\mathbf{L}_{\pi'}(z, x)) (x_K + x_k \omega_K)]_{x=\bar{x}(z)} \right\}, \end{aligned}$$

with

$$\begin{aligned} \Psi_K^{-1} \equiv 1 - \sum_z \left\{ \int_{\underline{x}(z)}^{\bar{x}(z)} \mathcal{K}_K(z, x) dx + [(\mathcal{K}(z, x) - \pi'(x)\mathcal{K}_{\pi'}(z, x)) x_K]_{x=\underline{x}(z)} \right. \\ \left. + [(\mathcal{K}(z, x) - \pi'(x)\mathcal{K}_{\pi'}(z, x)) (x_K + x_k \omega_K)]_{x=\bar{x}(z)} \right\}. \end{aligned} \quad (25)$$

Proof For the proof and more detailed specifications of all terms see the Appendix.

Inspecting the first order conditions in Theorem 1, we see that the condition (22) forms a functional equation in the unknown government policy function, π , with the side conditions (23)-(24) and the condition on the value of the Lagrangean multiplier, μ . From the setup of the problem it is clear that the only free boundary values of the government policies are the values at the lower and upper bounds, $\underline{x}(\underline{z})$ and $\bar{x}(\bar{z})$. Additionally, the more detailed first-order conditions in the Appendix contain savings and distribution functions, h and λ , and their derivatives with respect to π , π' , and K , i.e. h_π , $h_{\pi'}$, h_K , $h_{\pi'\pi'}$, $h_{\pi'\pi}$, λ_π , $\lambda_{\pi'}$, λ_K , $\lambda_{\pi'\pi'}$, and $\lambda_{\pi'\pi}$, respectively.

If we knew how agents' saving policies h and *simultaneously* how the distribution λ depend on the government policy schedule, i.e. if we could solve at equilibrium prices for the optimal policy π which is a function of the distribution and prices which in turn are determined by $h(\pi, \pi')$ which is itself a function of the optimal policy and prices, the task of the derivation of the first order conditions for this dynamic optimization would be straightforward. However, not only we have to solve for these functions simultaneously but also we are in a much more difficult situation since these functions are specified only implicitly as operator equations. The proper specification of the derivative of the savings and distribution functions with respect to the government policy functions requires a generalized concept of derivative, the so called Frechet derivative.

5.1 The Effects of Government Policy on Stationary Recursive Equilibrium

For any government policy schedule π , an agents' saving policy and the distribution functions are known only implicitly as a solution to the two operator equations (13)-(14) together with the aggregate conditions specifying the equilibrium stationary prices. In order to derive the first order conditions in the calculus of variation, we need to specify the derivatives of the integrand functions, \mathcal{W} and \mathcal{G} , with respect to the marginal changes in government policy, π and π' . For this purpose, we use the concept of generalized derivatives on mappings between two Banach spaces (B-spaces), the Fréchet derivatives.⁹

Definition 6 (The Fréchet Derivative) *Given a nonlinear operator $\mathcal{N}(u)$ on function u , its Fréchet differential is defined by*

$$\mathcal{N}_u \Delta \equiv \left. \frac{\partial \mathcal{N}(u + \varepsilon \Delta)}{\partial \varepsilon} \right|_{\varepsilon=0},$$

where \mathcal{N}_u is the Fréchet derivative.

We now formulate functional equations for the unknown F-derivatives h_π , $h_{\pi'}$, h_K , $h_{\pi'\pi'}$, $h_{\pi'\pi}$, λ_π , $\lambda_{\pi'}$, λ_K , $\lambda_{\pi'\pi'}$, and $\lambda_{\pi'\pi}$. The following Lemma derives the effects of the government policy function π on the operator Euler equation by specifying five unknown “sensitivity” functions $h_\pi : Z \times K \rightarrow \mathbb{R}^+$, $h_{\pi'} : Z \times K \rightarrow \mathbb{R}^+$, $h_K : Z \times K \rightarrow \mathbb{R}^+$, $h_{\pi',\pi} : Z \times K \rightarrow \mathbb{R}^+$, and $h_{\pi'\pi'} : Z \times K \rightarrow \mathbb{R}^+$.

Lemma 1 (The Effects of π , π' , and K on the Euler Equation) *The total F-derivatives of the operator Euler equation \mathcal{F} with respect to the government policy function π , to its derivative π' , and to the aggregate capital stock K are, respectively,*

$$\mathcal{F}_\pi = u''(c)c_\pi - \beta \sum_{z^+} \{u''(c^+)c_\pi^+ R^+ + u'(c^+)R_\pi^+\} Q(z, z^+) = 0, \quad (26)$$

$$\mathcal{F}_{\pi'} = u''(c)c_{\pi'} - \beta \sum_{z^+} \{u''(c^+)c_{\pi'}^+ R^+ + u'(c^+)R_{\pi'}^+\} Q(z, z^+) = 0, \quad (27)$$

$$\mathcal{F}_K = u''(c)c_K - \beta \sum_{z^+} \{u''(c^+)c_K^+ R^+ + u'(c^+)R_K^+\} Q(z, z^+) = 0, \quad (28)$$

⁹The compliance of the Fréchet derivatives (also called the F-derivatives) with the derivations of the first order conditions in the calculus of variation is reflected by the fact that the F-differential is identical to the variation. Our derivations are more complicated than the standard Fréchet derivative because our operators are recursive.

and further

$$\begin{aligned}\mathcal{F}_{\pi'\pi'} &= u'''(c) [c_{\pi'}]^2 + u''(c) c_{\pi'\pi'} \\ &- \beta \sum_{z^+} \left\{ \left[u'''(c^+) [c_{\pi'}^+]^2 + u''(c^+) c_{\pi'\pi'}^+ \right] R^+ + 2u''(c^+) c_{\pi'}^+ R_{\pi'}^+ \right. \\ &+ \left. u'(c^+) R_{\pi'\pi'}^+ \right\} Q(z, z^+) = 0,\end{aligned}\tag{29}$$

and

$$\begin{aligned}\mathcal{F}_{\pi'\pi} &= u'''(c) c_{\pi'} c_{\pi} + u''(c) c_{\pi'\pi} \\ &- \beta \sum_{z^+} \left\{ \left[u'''(c^+) c_{\pi'}^+ c_{\pi}^+ + u''(c^+) c_{\pi'\pi}^+ \right] R^+ + u''(c^+) [c_{\pi'}^+ R_{\pi}^+ + c_{\pi}^+ R_{\pi'}^+] \right. \\ &+ \left. u'(c^+) R_{\pi'\pi}^+ \right\} Q(z, z^+) = 0.\end{aligned}\tag{30}$$

Proof For the proof and the definition of terms see the Appendix.

Similarly, we derive the effects of the government policy on the shape of the distribution function λ by specifying functional equations which implicitly determine the unknown “sensitivity” functions $\lambda_{\pi} : Z \times K \rightarrow \mathbb{R}^+$, $\lambda_{\pi'} : Z \times K \rightarrow \mathbb{R}^+$, $\lambda_K : Z \times K \rightarrow \mathbb{R}^+$, $\lambda_{\pi'\pi'} : Z \times K \rightarrow \mathbb{R}^+$ and $\lambda_{\pi'\pi} : Z \times K \rightarrow \mathbb{R}^+$.

Lemma 2 (The Effects of π , π' , and K on the Stationary Distribution Function)

The total F -derivative of the operator stationary distribution function \mathcal{L} with respect to the government policy function π , to its derivative π' and to the aggregate capital stock K are, respectively,

$$\begin{aligned}\mathcal{L}_{\pi} &= \lambda_{\pi}(z^+, k^+) - \sum_z \left\{ \lambda_k [z, h^{-1}(z, k^+)] h_{\pi}^{-1}(z, k^+) + \lambda_{\pi} [z, h^{-1}(z, k^+)] \right\} Q(z, z^+) = 0, \\ \mathcal{L}_{\pi'} &= \lambda_{\pi'}(z^+, k^+) - \sum_z \left\{ \lambda_k [z, h^{-1}(z, k^+)] h_{\pi'}^{-1}(z, k^+) + \lambda_{\pi'} [z, h^{-1}(z, k^+)] \right\} Q(z, z^+) = 0, \\ \mathcal{L}_K &= \lambda_K(z^+, k^+) - \sum_z \left\{ \lambda_k [z, h^{-1}(z, k^+)] h_K^{-1}(z, k^+) + \lambda_K [z, h^{-1}(z, k^+)] \right\} Q(z, z^+) = 0,\end{aligned}$$

and further

$$\begin{aligned}\mathcal{L}_{\pi'\pi'} &= \lambda_{\pi'\pi'}(z^+, k^+) - \sum_z \left\{ \lambda_{\pi'\pi'} [z, h^{-1}(z, k^+)] + \lambda_{\pi'k} [z, h^{-1}(z, k^+)] h_{\pi'}^{-1}(z, k^+) \right. \\ &+ \left. [\lambda_{\pi'k} [z, h^{-1}(z, k^+)] + \lambda_{kk} [z, h^{-1}(z, k^+)] h_{\pi'}^{-1}(z, k^+)] h_{\pi'}^{-1}(z, k^+) \right. \\ &+ \left. \lambda_k [z, h^{-1}(z, k^+)] h_{\pi'\pi'}^{-1}(z, k^+) \right\} Q(z, z^+) = 0,\end{aligned}\tag{31}$$

and

$$\begin{aligned}\mathcal{L}_{\pi'\pi} &= \lambda_{\pi'\pi}(z^+, k^+) - \sum_z \left\{ \lambda_{\pi'\pi} [z, h^{-1}(z, k^+)] + \lambda_{\pi'k} [z, h^{-1}(z, k^+)] h_{\pi}^{-1}(z, k^+) \right. \\ &+ \left. [\lambda_{\pi k} [z, h^{-1}(z, k^+)] + \lambda_{kk} [z, h^{-1}(z, k^+)] h_{\pi}^{-1}(z, k^+)] h_{\pi'}^{-1}(z, k^+) \right. \\ &+ \left. \lambda_k [z, h^{-1}(z, k^+)] h_{\pi'\pi}^{-1}(z, k^+) \right\} Q(z, z^+) = 0.\end{aligned}\tag{32}$$

Proof For the proof and the definition of terms see the Appendix.

In this way we obtain functional equations (11), (26)-(30), (12), and (31)-(32) in the unknown functions h , h_π , $h_{\pi'}$, h_K , $h_{\pi'\pi'}$, $h_{\pi'\pi}$, λ , λ_π , $\lambda_{\pi'}$, λ_K , $\lambda_{\pi'\pi'}$, and $\lambda_{\pi'\pi}$, respectively. Finally, by adding the first-order conditions from Theorem 1, the problem of finding the optimal government policy π is a system of thirteen functional equations in thirteen unknown functions with two side conditions and one condition on the Lagrange multiplier.

5.2 Analytical Results

Despite the complexity of the problem we are able to derive several analytical results. It will be useful to compare the optimal government policy and its effects to those of a flat-tax economy in which the government also consumes a fraction g of the total output. Note that the term¹⁰ $gy(z, x)$ in equation (19) captures the fraction of government expenditures “related” to an agent (z, x) . This fraction of government expenditures $gy(z, x)$ is identical to the amount of taxes hypothetically paid by the agent (z, x) under a flat tax g on total income. We will further use this concept as a useful benchmark in the following analysis.

Since the first-order conditions in Theorem 1 specify the explicit transversality conditions on agents with the lowest and the highest taxable activity, they can be used for qualitative results on the behavior of the policy schedule at these boundary points.

In the following Proposition we relate the tax payment of the “poorest” agent under the optimal tax policy to the tax payment in the economy with a flat tax g . This poorest agent is constrained at the lowest value of capital \underline{k} and is hit by the lowest shock \underline{z} .

Proposition 1 (Optimal Tax at Lowest Taxable Activity) *The transversality condition in equation (24) in Theorem 1 implies that for an agent with the lowest taxable activity level $\underline{x}(\underline{z})$, the optimal difference between taxes paid $\pi(\underline{x}(\underline{z}))\underline{x}(\underline{z})$ and the income taxes paid under a flat tax regime, $g\underline{y}(\underline{z}, \underline{x}(\underline{z}))$, is proportional to the sum of the agent’s utility and the marginal contribution of his savings to the aggregate welfare,*

$$\pi(\underline{x}(\underline{z}))\underline{x}(\underline{z}) - g\underline{y}(\underline{z}, \underline{x}(\underline{z})) = -\frac{W[u(\underline{c}(\underline{z}, \underline{x}(\underline{z}))) + \Psi\underline{k}]}{\mu},$$

where $\underline{c}(\underline{z}, \underline{x}(\underline{z})) = \underline{y}(\underline{z}, \underline{x}(\underline{z})) - \pi(\underline{x}(\underline{z}))\underline{x}(\underline{z}) - \underline{k}$ is consumption, $\underline{y}(\underline{z}, \underline{x}(\underline{z})) = r\underline{k} + w\underline{z}$ is before-tax income, $\mu < 0$ is the shadow price of government expenditures, and Ψ is the marginal effect of the aggregate capital stock on the aggregate welfare defined in Theorem 1.

Proof See the Appendix.

¹⁰To shorten the notation we write $gy(z, x) = gy(z, k(z, x))$.

The following corollary states the conditions for which the amount of taxes paid by the poorest agent under the optimal tax policy is larger than that under a flat income tax, i.e. $\pi(\underline{x}(\underline{z}))\underline{x}(\underline{z}) > g\underline{y}(\underline{z}, \underline{x}(\underline{z}))$.

Corollary 1 *If the savings contribution of the poorest agent to aggregate welfare is non-negative, $\Psi \underline{k} \geq 0$, then at the lowest taxable activity level $\underline{x}(\underline{z})$ the amount of taxes paid under the optimal tax policy $\pi(\underline{x}(\underline{z}))\underline{x}(\underline{z})$ is larger than that under a flat income tax, $g\underline{y}(\underline{z}, \underline{x}(\underline{z}))$.*

The poorest agent pays more taxes than in the flat-tax regime if one of these three conditions are satisfied: $\Psi = 0$, or if $\underline{k} = 0$, or if simultaneously $\Psi > 0$ and $\underline{k} > 0$). The Corollary illustrates the incentives to the agent to accumulate more assets.¹¹ We discuss these incentives in the following Section.

In a similar way, the Proposition below specifies the optimal tax payment of the “richest” agent relative to his tax burden at the flat tax equal to g . The richest agent’s current and future savings are at the endogenous upper bound $\bar{k}(\bar{z}) = \bar{k}$ and he receives the highest shock \bar{z} .

Proposition 2 (Optimal Tax at Highest Taxable Activity) *Assuming $\pi'(\bar{x}(\bar{z})) \neq \left[\frac{\delta x}{\delta \pi}(\bar{x}(\bar{z}))\right]^{-1}$, the transversality condition in equation (23) in Theorem 1 implies that for an agent with the highest taxable activity $\bar{x}(\bar{z})$, the optimal difference between taxes paid $\pi(\bar{x}(\bar{z}))\bar{x}(\bar{z})$ and the taxes paid under the flat tax regime $g\bar{y}(\bar{z}, \bar{x}(\bar{z}))$, is proportional to the sum of the agent’s utility and the marginal contribution of his savings to the aggregate welfare,*

$$\pi(\bar{x}(\bar{z}))\bar{x}(\bar{z}) - g\bar{y}(\bar{z}, \bar{x}(\bar{z})) = -\frac{W[u(\bar{c}(\bar{z}, \bar{x}(\bar{z})))] + \Psi \bar{k}}{\mu},$$

where $\bar{c}(\bar{z}, \bar{x}(\bar{z})) = \bar{y}(\bar{z}, \bar{x}(\bar{z})) - \pi(\bar{x}(\bar{z}))\bar{x}(\bar{z}) - \bar{k}(\bar{z})$ is consumption, $\bar{y}(\bar{z}) = r\bar{k} + w\bar{z}$ is before-tax income, $\mu < 0$ is the shadow price of government expenditures, and Ψ is the marginal effect of the aggregate capital stock on the aggregate welfare defined in Theorem 1.

Proof See the Appendix.

Notice that as $\omega_\pi = \frac{h_\pi}{1-h_k} < 0$, $\left[\frac{\delta x}{\delta \pi}(\bar{x}(\bar{z}))\right]^{-1} = \frac{1}{x_k(\bar{z}, \bar{k})\omega_\pi(\bar{z}, \bar{k})} < 0$ implies that the assumption on the slope of the tax schedule at $\bar{x}(\bar{z})$ is satisfied whenever it is non-decreasing, i.e. $\pi'(\bar{x}(\bar{z})) \geq 0$, and that $\Psi \geq 0$.

The following corollary states the conditions for which the amount of taxes paid by the richest agent under the optimal tax schedule is larger than that under a flat tax, i.e. $\pi(\bar{x}(\bar{z}))\bar{x}(\bar{z}) > g\bar{y}(\bar{z}, \bar{x}(\bar{z}))$.

¹¹If a substantial debt can be accumulated, $\underline{k} < 0$, the ability of the government to tax capital stock becomes limited. Finally, it is easy to show that the equilibrium might not exist if $\Psi < 0$.

Corollary 2 *If the marginal contribution of the aggregate capital stock to the aggregate welfare is nonnegative $\Psi \geq 0$, and $\underline{k} \leq 0$ then at the highest taxable activity level the optimal tax contribution $\pi(\bar{x}(\bar{z}))\bar{x}(\bar{z})$ is larger than that under a flat tax regime $g\bar{y}(\bar{z}, \bar{x}(\bar{z}))$,*

$$\pi(\bar{x}(\bar{z}))\bar{x}(\bar{z}) - g\bar{y}(\bar{z}, \bar{x}(\bar{z})) = -\frac{W(u(\bar{c}(\bar{z}, \bar{x}(\bar{z}))) + \Psi\bar{k}}{\mu} > 0.$$

Additionally,

$$\frac{\pi(\bar{x}(\bar{z}))\bar{x}(\bar{z}) - g\bar{y}(\bar{z}, \bar{x}(\bar{z}))}{\pi(\underline{x}(\underline{z}))\underline{x}(\underline{z}) - g\underline{y}(\underline{z}, \underline{x}(\underline{z}))} \geq \frac{W(u(\bar{c}(\bar{z}, \bar{x}(\bar{z})))}{W(u(\underline{c}(\underline{z}, \underline{x}(\underline{z})))} \left(\frac{\Psi\bar{k}}{W(u(\bar{c}(\bar{z}, \bar{x}(\bar{z})))} + 1 \right)$$

implies that the amount of taxes paid by the richest agent is greater than that of the poorest agent, i.e.

$$\pi(\bar{x}(\bar{z}))\bar{x}(\bar{z}) > \pi(\underline{x}(\underline{z}))\underline{x}(\underline{z}).$$

These Corollaries show that the socially optimal amount of the tax revenues paid by the ‘boundary’ agents depends only on these agents’ individual characteristics (on their taxable activity, x , income, y , and consumption, c) and the two aggregate shadow prices: the shadow price of government spending, μ , and the shadow price of the aggregate capital, Ψ , both expressed in terms of social welfare.

Together, Corollaries 1 and 2 can say a lot about the shape of the optimal government policy schedule. If $\underline{k} = 0$, then both ends of the tax schedule are above the flat-tax level and the implied tax schedule is a “U-shape” function. Its bottom lies under the flat-tax level to clear the government budget constraint proportional to g . This is also the case if $\Psi = 0$ and if both $\Psi > 0$ and $\underline{k} > 0$. Finally, if $\Psi > 0$ and $\underline{k} < 0$, the optimal tax schedule could either be “U-shaped” or an increasing function of income. These qualitative assessments are confirmed by the numerical results in our example in the following Section.

6 An Example: The Optimal Income Tax Schedule

In this section we demonstrate our method by finding the optimal government policy π defined as a tax schedule on total income from capital and labor. Therefore, the taxable activity is

$$x(z, k; K) = y(z, k; K) = rk + wz,$$

and the individual budget constraint is

$$c \leq (1 - \pi(x))x + k - k^+.$$

We assume that the borrowing constraint is

$$\underline{k} = 0,$$

and that there are only two levels of shock, $Z = \{z, \bar{z}\}$. The total tax revenues are equal to a fraction g of the total output.

Thus the Euler equation (10) for a (z, k) -agent's optimal savings function $k^+(z, k)$ for all $(z, k) \in Z \times B$ is now

$$u'(c) \geq \beta \sum_{z^+} u'(c^+) [(1 - \pi(x^+) - \pi'(x^+)x^+)r + 1] Q(z, z^+),$$

where $c^+ = (1 - \pi(x^+))x^+ + k^+ - k^{++}$, $x^+ = rk^+ + wz^+$, and $k^{++} = k^+(k^+(z, k), z^+)$. Note that for this specification $x_k = y_k = r$ and $k_x = 1/x_k = 1/r$.

For this concrete example we first analyze the conditions for the existence of the stationary recursive competitive equilibrium.

6.1 The Existence of Stationary Recursive Competitive Equilibrium

Because the tax schedule is an arbitrary function, we must ensure that the first order approach is valid.¹² In order to characterize the admissible tax functions and to prove the Schauder Theorem for economies with distortions, we follow the notation in Stokey, Lucas, and Prescott (1989), Chapter 18. For each agent $(z, k) \in B \times Z$, denote the after-tax gross income as

$$\psi(z, k) \equiv (1 - \pi(x(z, k)))x(z, k) + k.$$

Using $\psi(z, k)$, rewrite the Euler equation as

$$u'(\psi(z, k) - k^+(z, k)) = \beta \sum_{z^+} u'(\psi(k^+(z, k), z^+) - k^+(k^+(z, k), z^+)) \psi_1(k^+(z, k), z^+) Q(z, z^+),$$

where

$$\psi_1(k^+(z, k), z^+) = (1 - \pi(x(k^+(z, k), z^+)) - \pi'(x(k^+(z, k), z^+))x(k^+(z, k), z^+))r + 1$$

is the marginal after-tax return on an extra unit of investment. In the following Theorem we establish the validity of the first order approach and the existence of the competitive equilibrium.

¹²Again, we analyze the case of borrowing constrained agents in the Appendix.

Theorem 2 For a given tax schedule $\pi : \mathbb{R}_+ \rightarrow \mathbb{R}$, if for each $(z, k) \in B \times Z$

1. $\psi_1(z, k) > 0$, and
2. ψ is quasi-concave,

then the solution to each agent's maximization problem and the stationary recursive competitive equilibrium exist.

Proof See the Appendix.

The following corollary characterizes the set of admissible tax schedules that satisfy the conditions of Theorem 2. We have earlier introduced the notation for the endogenous upper bound on capital for any agent, \bar{k} . Let $\{\underline{w}, \bar{w}\}$ and $\{\underline{r}, \bar{r}\}$ denote some arbitrary, non-binding lower and upper bounds for equilibrium wage and interest rate, respectively. Finally, $\varepsilon_x^\pi(x) \equiv \frac{\pi'(x)}{\pi(x)}x$ is the elasticity of the tax rate to the taxable income.

Corollary 3 (Admissible Tax Schedule Functions) Let $C^2(\mathbb{R}_+)$ be a set of continuously differentiable functions from \mathbb{R}_+ to \mathbb{R} . If a tax schedule function $\pi \in C^2(\mathbb{R}_+)$ belongs to the set of admissible tax schedules Υ ,

$$\Upsilon = \left\{ \pi \in C^2(\mathbb{R}_+) : \pi(x) \left(1 + \varepsilon_x^\pi \right) < 1 + \frac{1}{\bar{r}} \right\}$$

for all $x \in [\underline{rk} + \underline{wz}, \bar{r}\bar{k} + \bar{wz}]$, then it satisfies the conditions of Theorem 2.

The above statement follows directly from the fact that $\psi_1(z, k) > 0$ and that ψ is quasi-concave. The corollary implies that there exists an upper bound on the marginal tax rate, $\pi(x) (1 + \varepsilon_x^\pi)$. This upper bound is not likely to bind for a very wide range of tax schedules.¹³

6.2 The Shape of the Optimal Income Tax Schedule

To obtain the first order conditions for the optimal income schedule we plug the example-specific terms into the general conditions in Theorem 1 and Lemmata 1 and 2. These terms are listed in the Appendix C.

Adapting Propositions 1 and 2 and the related Corollaries 1 and 2 to our example economy we obtain the following results.

¹³As an example, for a realistic equilibrium interest rate $r = 0.05$, the upper bound on the marginal tax rate is equal $1/(1+r) \simeq 21$. Therefore, for a high level of tax rate, say $\pi(x) = 0.42$, the elasticity of the tax schedule at that level of income would have to be $\varepsilon_x^\pi = 49$ in order to violate the admissibility condition. When numerically solving for the optimal tax schedule in the next Section we do not impose any of these exogenous bounds but we check the admissibility of the optimal tax schedule ex post.

Corollary 4 (Optimal Tax Rates at Lowest and Highest Income) *If π is the optimal tax schedule on the total income, then*

1. *the optimal tax rate at the lowest income level $\underline{x}(\underline{z})$ is strictly greater than a flat tax g at that income level,*

$$\pi(\underline{x}(\underline{z})) > g;$$

2. *and, provided that $\Psi \geq 0$, the optimal tax rate at the highest income level $\bar{x}(\bar{z})$ is strictly greater than the optimal tax rate at the lowest income level,*

$$\pi(\bar{x}(\bar{z})) > \pi(\underline{x}(\underline{z})) > g.$$

Proof The results follow directly from Corollaries 1 and 2.

Thus both the poorest and richest agents pay higher taxes than in a corresponding flat-tax economy. The government budget constraint and the continuity of the optimal tax function imply that there must exist a measure of agents at intermediate income levels who face lower tax rates than the flat tax rate. Thus the optimal tax schedule must be a “U-shape” function.

We interpret these results from the point of view of the social planner who could use the first best, lump sum taxes. To insure agents against idiosyncratic shocks, he would set such a tax schedule on each pair of individual states (z, k) to arrive at a stationary equilibrium in which all agents would accumulate the same amount of capital K and consume the same amount of goods. In other words, the distribution would consist of a mass of agents at two points, (\underline{z}, K) and (\bar{z}, K) . A U-shape transfer (tax) system is the way to induce agents to arrive at this optimal outcome.

Although our government is constrained from using the first best policy, it strives to accomplish the same outcome. By imposing higher than average taxes on the poorest and richest agents, the government tries to provide incentives for agents to save towards the average capital stock. The poorest agents are motivated by high but decreasing tax schedule, $\left(\frac{d(\pi(x)x)}{dx}\right)_{\underline{x}(\underline{z})} < 0$. Thus, according to the Euler equation, their expected return on capital increases. On the other hand, the richest agents are discouraged from further savings by the increasing tax schedule at high levels of capital. Finally, agents with savings around the average capital stock are motivated by lower than average taxes to keep their savings at the current level. The U-shape tax function provides the incentives for agents to eventually arrive and stay at the individual capital stock $k = K$.

7 Numerical Solution

In this Section we solve for the optimal tax schedule and compare the associated steady state allocations to those resulting from the existing progressive tax schedule in the U.S. economy and from the usual flat-tax reform. In order to evaluate the welfare implications of these tax reforms, we conduct the transition analysis.

The solution to the problem of finding the optimal policy in the sense of the Operator Steady State Ramsey Problem in our example leads, as in the general case, to solving the functional system of the first-order conditions given by Theorem 1 together with the functional equations specifying the stationary equilibrium (expressed by h and λ) and the sensitivity functions given by F-derivatives h_π , $h_{\pi'}$, h_K , $h_{\pi'\pi'}$, $h_{\pi'\pi}$, λ_π , $\lambda_{\pi'}$, λ_K , $\lambda_{\pi'\pi'}$, and $\lambda_{\pi'\pi}$. Thus we obtain the system of thirteen functional equations in thirteen unknown functions with two side conditions and one condition on the Lagrange multiplier. We solve this complex functional equation problem by the least squares projection method. Its application to our problem and the approximation of the optimal tax schedule can be found in Appendix D.¹⁴

7.1 Parameterization

The uninsurable idiosyncratic shock to labor productivity follows a two-state, first order Markov chain. We use the results of Heaton and Lucas (1996) who, using the PSID labor market data, estimate the household annual labor income process between 1969 and 1984 by a first-order autoregression of the form

$$\log(\eta_t) = \bar{\eta} + \rho \log(\eta_{t-1}) + \epsilon_t,$$

with $\epsilon \sim N(0, \sigma_\epsilon^2)$. They find that $\rho = 0.53$ and $\sigma_\epsilon^2 = 0.063$. Tauchen and Hussey (1991) approximation procedure for a two-state Markov chain implies $z_L = 0.665$, $z_H = 1.335$ and $Q(z_L, z_L) = Q(z_H, z_H) = 0.74$. These values imply an aggregate effective labor supply equal to one with agents evenly split over the two shocks.¹⁵ We set the discount factor at $\beta = 0.95$. The rest of the parameters is taken from Prescott (1986), in particular $\alpha = 0.36$, $\delta = 0.1$, and the preference parameter $\sigma = 1$.

¹⁴For a more detailed explanation of the use of the projection methods to stationary equilibria in economies with a continuum of heterogenous agents see Bohacek and Kejak (2002).

¹⁵Similar parameterization is in Storesletten, Telmer, and Yaron (2007) with $z_L = 0.73$, $z_H = 1.27$ and $Q(z_L, z_L) = Q(z_H, z_H) = 0.82$. Diaz-Jimenez, Quadrini, and Rios-Rull (1997) use $z_L = 0.5$, $z_H = 3.0$ and $Q(z_L, z_L) = 0.9811$, $Q(z_H, z_H) = 0.9261$. In future research we plan to add life cycle features to the model as in Ventura (1999).

Finally, for all steady states we consider a Ramsey problem in which government is required to raise tax revenues equal to 20% of the total output, i.e. $g = 0.2$.

7.2 The U.S. Progressive Tax Schedule

We model the progressive tax schedule as Ventura (1999), the closest model analyzing a flat-tax reform in an economy with heterogeneous agents.¹⁶ An agent's budget constraint is

$$c + k' \leq rk + zw + k - T,$$

where T represents the amount of tax paid by the agent according to the progressive tax schedule. The amount of tax is determined according to which tax bracket the total taxable income, $I = rk + \max\{0, zw - I^*\}$, falls in, with a labor-income tax deductible amount $I^* \geq 0$. There are M brackets with associated tax rates, $\tau_m, m = 1, \dots, M$, defined on intervals between brackets' bounds I_0, \dots, I_{M-1} . For $M = 5$ the tax rates are $\tau_m \in \{0.15, 0.28, 0.31, 0.36, 0.396\}$ and the tax brackets, expressed as a multiple of the average income, $I_{m-1} \in \{0, 0.85, 2.06, 3.24, 5.79\}$. In addition, capital income, rk , is taxed at a flat rate $\tau_k = 0.25$.

For income $I \in (I_{m-1} - I_m]$, the total tax is then

$$T = \tau_1(I_1 - I_0) + \tau_2(I_2 - I_1) + \dots + \tau_m(I - I_{m-1}) + \tau_k rk.$$

The government budget constraint is cleared by finding an equilibrium value of the tax exemption level I^* . Aggregate statistics of the steady state are shown in the left column of Table 1.

7.3 A Flat-Tax Reform

The flat-tax reform consists of replacing the progressive tax schedule with a single flat tax τ on the total income from labor and capital. The budget constraint of each agent becomes

$$c + k' \leq (1 - \tau)(rk + zw) + k.$$

Note that the flat tax reform, like in Ventura (1999), does not eliminate taxation of capital income. We find that the equilibrium flat tax rate is $\tau = 0.254$.

INSERT TABLE 1 ABOUT HERE

¹⁶Compared to his model, our agents are infinitely lived, so we omit the life-cycle variables, accidental bequests, government transfers, and social security tax and benefits. Except for capital depreciation, we do not consider tax deductions.

The middle column in Table 1 describes the steady state results. Relative to the progressive tax schedule steady state, the flat-tax reform increases the steady state levels by similar magnitudes found in the literature: capital stock increases by 30%, output by 10.8%, consumption by 4.6%, and welfare by 3.9%. As in Ventura (1999), the flat-tax reform increases inequality: Gini income coefficients rise from 0.22 to 0.31 before tax and from 0.21 to 0.32 after tax.¹⁷

7.4 The Optimal Tax Schedule

We use our methodology described in the previous Sections to solve for the optimal tax schedule that maximizes average steady state welfare¹⁸. The right column in Table 1 summarizes the optimal tax schedule steady state. The impact of the optimal tax schedule is very large. Steady state average welfare increases by 4.4%. Aggregate capital stock rises by 49%, output by 15.8%, and consumption by 5.8%. Inequality increases too but not as much as in the flat-tax reform: Gini income coefficients are 0.28 before and 0.27 after tax, respectively. General equilibrium effects cause the interest rate to drop by almost one half and the wage to increase due to a higher productivity of labor used in production with such a high capital stock. Compared to the flat-tax steady state, capital stock increases by 15%, output by 4.5%, consumption by 1.1%, and welfare by 0.8%.

INSERT FIGURE 2 ABOUT HERE

Figure 2 shows the optimal tax schedule and the marginal tax rate function. The average tax rate is a U-shaped function taxing the lowest total income at a 45%, decreasing to a minimum of 19% and rising to 62% at the highest level of total income. Although the whole shape of the tax function is important for the resulting allocations, a majority of agents face the decreasing or the flat part of the tax schedule. The marginal tax rate is also U-shaped, almost flat and close to zero for low incomes, falling to negative levels around the average total income and then rising at high income levels. Note that the maximal marginal rate is 2.5 and that the optimal tax schedule easily satisfies the admissibility condition from Corollary 3 (the interest rate implies an upper bound equal to 19.7).

The optimal tax function τ is strictly positive and very nonlinear. Both results are different from Mirrlees (1971) static model with fixed distribution of skills, where a welfare

¹⁷Elimination of capital income tax in Lucas (1990) increases capital stock by 30-34% and consumption by 6.7%. A flat-rate reform with heterogeneous agents in Ventura (1999) increases the total capital stock by one third, output by 15%. Without a well calibrated life-cycle earnings process we are not able to match well the inequality coefficients, especially that of wealth.

¹⁸The residual errors of the functional equations associated to the main approximated functions are reported in Table ??.

maximizing tax schedule is close to a linear, non-decreasing function. In his model, the marginal tax rate is between zero and one, and zero at both ends of the distribution. Compared to our model with insurance and savings incentives, Mirrlees (1971) results follow from labor incentives related to the distribution of skills and consumption-leisure preferences.¹⁹

We want to emphasize that our stationary distribution is endogenous and there are no restrictions on the optimal tax schedule to be positive or to be of any particular shape. Conesa and Krueger (2006), also in a general equilibrium framework but with added life cycle features, searched for the optimal progressivity of a tax schedule, limiting their class of tax schedules to monotone functions as in Gouveia and Strauss (1994).²⁰ In this class of functions, the optimal tax schedule is basically a flat tax with a fixed deduction, delivering a welfare gain of 1.7% compared to the existing tax progressive tax in the United States. The class of monotone functions seems rather restrictive for the optimal tax schedule. Our class of admissible functions includes all progressive tax schedules but these were found significantly inferior with respect to the welfare criterion.

7.5 The Tradeoff Between Efficiency and Distribution

Apart from the general equilibrium effects, the huge welfare impact of the optimal tax schedule arises from the distributional effects. The stationary distributions of capital in the three steady states are shown in Figure 3. Although both the flat and the optimal tax schedules increase the aggregate levels, the difference between them is that the ad hoc flat tax schedule does not take into account the distribution of agents. The flat tax reform helps more the agents with high incomes: the mean wealth increases much more than the median so that the median/mean ratio falls to 0.77. In the flat-tax steady state the aggregate levels increase but from “the optimal distribution” point of view the mass of agents moves too much to the left while wealthy agents emerge at the right tail of the distribution. The progressive tax schedule has the lowest inequality measures because the high taxes on rich agents narrows the distribution towards the mean. However, the low tax

¹⁹Building on the Mirrlees (1971) and Mirrlees (1976) seminal work, Kocherlakota (2005), Golosov, Kocherlakota, and Tsyvinski (2003) or Albanesi and Sleet (2003) study optimal social planner policies with asymmetric information. In such an environment, positive capital income taxes are optimal despite the associated efficiency loss and informational frictions are necessary for a characterization of the optimal policies.

²⁰The simple search method cannot be used for computation of a stationary competitive equilibrium which is the limit of the optimal dynamic tax schedule. We show in Bohacek and Kejak (2004) that under some parameterization the first order conditions of the general dynamic Ramsey problem can simplify to the first order conditions of the steady state Ramsey problem analyzed in this paper.

rates on low incomes do not provide incentives for the poor households to save and move to higher income levels. In other words, it provides too much short-run insurance at the cost of the long-run average levels.

INSERT FIGURE 3 ABOUT HERE

This is exactly what the optimal income tax schedule improves. The main mechanism behind the large increase in the aggregate levels is the incentive effect of the optimal tax schedule. The U-shaped function in the top panel of Figure 2 effectively concentrates the agents around the mean, something what a social planner with an access to lump-sum transfers would do.²¹ The high tax rate at low income levels provides incentives for these agents to save more and move to higher income levels. On the other hand, the even higher tax rate on high income discourages further savings by the wealthiest agents. In the middle of the total income levels, the tax rate is lower than that found for the flat-tax reform. The optimal tax schedule preserves the median/mean wealth ratio of the progressive tax schedule by increasing the median by 47% and the mean by 49%. The support of the invariant distribution becomes wider but inequality measures do not increase as much as in the flat-tax reform.²²

To further analyze the tradeoff between efficiency and distribution, we adopt the approach in Domeij and Heathcote (2004) to distinguish the efficiency gain from distributional gains. The efficiency gain for an individual agent is the percentage of the original consumption that would allow the agent to consume the same fraction of the aggregate consumption after the reform as he or she was consuming in the original steady state. In the case of logarithmic utility, the gain is the same for all agents (see Domeij and Heathcote (2004) for a simple proof and other details). The distributional gain is the difference between the individual welfare gain and the efficiency gain.²³

Table 1 displays the average efficiency and distributional gains of the optimal steady state relative to the other two steady states. It is apparent that the steady state associated with the optimal, U-shaped tax function is welfare and efficiency superior to the other

²¹In many countries, marginal taxes are favorable to middle income groups. In practice, high rates on rich can break large fortunes while on poor they provide a floor for poverty. The result is a more equal distribution. Saez (2002) studies the optimal progressivity of capital income tax in a partial equilibrium model with exogenous labor and distribution. He finds that a progressive tax is a powerful tool to redistribute accumulated wealth.

²²Table 1 also shows the fraction of agents constrained in their borrowing: only 1.16% of agents is constrained in the progressive tax schedule steady state. The flat tax schedule increases this number to 1.88%, while the optimal tax steady state it is 1.42 (Domeij and Heathcote (2004) obtained similar results).

²³The individual welfare gain is the percentage of the original consumption level that would make an agent as well off as in the optimal tax steady state.

two steady states: both average welfare and efficiency measures are positive, and naturally greater for the comparison with the progressive steady state. As it was noted before, the optimal tax schedule obtains an average distributional loss relative to the progressive tax (-0.57%) but a gain relative to the flat tax steady state (2.86%).

INSERT FIGURE 4 ABOUT HERE

The individual gains, for agents with high and low labor productivity shocks, are shown in Figure 4. They are monotonically decreasing functions for all agents at all asset and labor income levels (with some exceptions). Most of the asset-poor agents have both welfare and distributional gains while the rich have losses relative to the steady state with the flat tax schedule. There are two forces present: first is the tax rate (especially for the rich agents in the flat tax steady state) and general equilibrium effects. The huge welfare gains (5-20%) for poor agents are mostly due to the higher wage in the optimal steady state. Note that the big efficiency gain from the optimal tax schedule is not sufficient to compensate all agents for the more unequal distribution (compared to the progressive tax steady state, an agent with a low productivity shock has always a distributional loss in Figure 4, top panel).

INSERT TABLE 2 ABOUT HERE

Table 2 shows the distribution of resources for quintiles of the wealth distribution. Because of the high tax rate on incomes in the bottom quintile, these agents in the optimal tax schedule steady state consume 6.5% less than those of the progressive tax schedule. However, in all the other quintiles the optimal tax schedule steady state, agents consume on average more than in the other two steady states. Dividing these levels by the average consumption in each steady state, we can calculate average quintile consumption relative to the steady state average. Under the optimal tax schedule, the bottom quintile consumes 73% of the average consumption, in the flat tax it is 77% , in the progressive it is 82% . This shows that the savings incentives of the optimal tax schedule overweight the insurance aspects (i.e., redistribution) in both the progressive and the flat tax schedules.

The distribution of capital reveals that the incentives contained in the optimal-tax schedule move the distribution to higher capital levels. The poorest quintile owns on average 17% more assets than in the progressive steady state. This increase is even larger for the other quintiles (40% on the top). Again, the flat-rate steady state leads to lower level of savings by the bottom two quintiles. These levels are reflected in the shares of the total capital stock. For all steady states the bottom quintile owns only around 5% of the total stock while the top quintile around one third (43% in the flat-tax steady state).

The investment-to-income ratios reveal the agents in the bottom quintile of the optimal schedule invest much more than similar agents in the other two steady states. Agents in the optimal tax schedule steady state invest 30% of their income, more than those in the flat-tax (27%) and progressive (22%) steady states. The investment is also more evenly distributed over the quintiles. Note also that the flat-rate tax schedule favors capital accumulation by the top quintile.

The income and after-tax income distribution show the differences between the three tax schedules. The progressive tax helps the bottom quintile while the flat tax helps the top quintile. The U-shape of the optimal tax provides the right incentives at the cost of lowest after-tax income for the poor agents. Finally, the optimal tax actually equalizes tax contribution share of total tax revenues across the quintiles. Both the flat-tax and progressive-tax steady states put more relative burden on the higher income quintiles.

INSERT FIGURE 5 ABOUT HERE

Finally, Figure 5 shows the sensitivity functions h_π and λ_π . The top panel shows the effect of a change in the optimal tax schedule on the savings decision of agents. For the low shock it is close to zero, for the high shock it is negative and monotonically decreasing. The bottom panel displays the same effects on the probability density function of the stationary distribution λ , again for each shock. We know from the stationarity condition of the distribution that the integral of these functions must be zero.²⁴

7.6 Transition to the Optimal Tax Schedule Steady State

Pure welfare steady-state comparisons could be misleading because tax changes imply substantial redistribution in the short run. In Domeij and Heathcote (2004) model of capital tax cuts, the expected discounted present value of welfare losses during transition are so large that they overturn the steady state welfare improvement. The short-run cost in the form of higher labor taxes is too heavy a price to pay for all except for the wealth-richest households.²⁵

INSERT TABLE 3 ABOUT HERE

²⁴Our numerical solution is only very close to zero due to approximation errors.

²⁵This is similar to Garcia-Mila, Marcet, and Ventura (1995) and Auerbach and Kotlikoff (1987) who find that reducing capital income taxation shifts the tax burden away from households who receive a large fraction of their income from capital and towards those who receive a disproportionate fraction from labor. Transition costs in Lucas (1990) reduce the welfare gains from zero capital tax reform to 0.75-1.25 percent of average consumption in the initial steady state.

Table 3 shows the results for our tax reform experiment. It compares the expected present discounted value from an unanticipated optimal tax reform of the progressive and flat-tax steady state. In each case the optimal tax schedule is imposed on the stationary distribution of the initial steady state.²⁶ We guess a sufficiently large number of convergence periods and iterate on paths of equilibrium interest rates and wages to clear markets in each period of the transition, returning possible excess tax revenues to all agents in each period. The convergence is relatively fast lasting around thirty periods.

INSERT FIGURE 6 ABOUT HERE

Contrary to Domeij and Heathcote (2004), we find that the reform makes both the mean and the median agents in the progressive tax schedule economy better off. Their welfare gains are positive but smaller than in the pure steady-state comparison (3.44% and 3.86%, respectively, measured as per period consumption transfers as a percentage of the initial steady state average consumption). The top panel in Figure 6 shows the expected present discounted values in the progressive-rate steady state and at the moment of the unanticipated reform to the optimal tax schedule. While 73% of the population is better off from the reform, it is not Pareto improving as the poorest 27% of all households are worse off (they are hit by high tax rates from the optimal schedule).

On the other hand, a transition from the flat-tax steady state would not be supported by the mean nor by the median agent (they loose 1.81% and 1.97%, respectively). The poor and now also the wealthy, for whom the tax increases dramatically, are worse off during the transition. The bottom panel in Figure 6 shows the expected present discounted values of the flat-rate steady state and of the transition to the optimal tax schedule. Political support is not sufficient, equal only to 33% of the population. We do not know whether an *optimal* transition would be welfare improving from this steady state.

As usual, this transition exercise shows that a tax reform is not Pareto improving for all agents. However, the gains from the optimal tax reform of the existing progressive tax schedule are so large that they are supported by majority of agents despite their transitional costs. Conesa and Krueger (2006) also find that a majority of the population would benefit from their optimal tax reform. However, in their case the poor and rich benefit, while it is the middle class (38%) who would be against the reform.

INSERT FIGURE 7 ABOUT HERE

²⁶Of course, it is not the *optimal* transition to the optimal tax schedule steady state. An optimal reform would implement a time-specific optimal tax schedule at each period of the transition

Finally, Figure 7 shows the efficiency and distributional individual gains from transition.²⁷ Relative to the steady state analysis, the averages for the progressive steady state reform decline: while the average welfare and efficiency gains remain still positive the distributional loss reaches negative 7%. A reform from the flat rate steady state delivers average welfare and efficiency losses but improves the distribution. Note that due to sizeable general equilibrium effects, the functions for poor agents are still positive and monotonically declining.

8 Conclusions

Quah (2003) shows that average levels are of the first order importance for economic growth and welfare, much more important than inequality. Government policies focusing on aggregate levels, including obviously optimal fiscal policy and taxation, are essential. However, it is the distribution of agents that delivers these aggregate levels. This paper shows that it is crucial to think of policies that target the distribution of agents. Only in this way the high aggregate levels and welfare improvements can be achieved.

To our knowledge, this paper is the first one that provides a solution method for such optimal government policies in heterogeneous agent economies. We think of these policies as optimal because they take into account their effects on the distribution of agents. As an example, we find the optimal tax schedule for a steady state Ramsey problem in an economy with heterogeneous agents. The optimal tax schedule is U-shaped, it increases all aggregate levels by providing the right incentives for the agents to accumulate high aggregate levels but not at the cost of increased inequality. Welfare gain in the steady state is large: it is positive for both mean and median agent as well as in a transition following an unanticipated optimal tax reform of the progressive tax schedule steady state.

The approach developed in this paper can be applied to any optimal government policy. Within the field of optimal taxation, in our future research we plan to study the optimal tax schedule with elastic labor supply and realistic life-cycle income profiles. An endogenous labor-leisure decision might affect the shape of the optimal tax schedule, the aggregate labor supply and the distribution of labor hours. We would also like to explore different (Rawlsian) welfare functions. Another topic that has received a lot of attention is the optimal capital taxation in models with heterogeneous agents (see Aiyagari (1995) for the initial contribution). Finally, we plan to use this methodology to analyze optimal dynamic taxation.

²⁷These gains are defined in the same way as in the steady state. A gain from transition is a constant, per-period percentage of consumption in the original steady state that equalizes its corresponding expected present discounted value from the whole transition. For details, see in Domeij and Heathcote (2004).

A Appendix: Analysis of the Borrowing Constrained Agents

In general, for all $z \in Z$ there exists a current minimal accumulated asset level $\underline{k}(z)$ above which agents are not borrowing constrained. For these agents, i.e. for those with $(k, z) \in [\underline{k}, \underline{k}(z)) \times Z$ and the next period savings k^+ equal to \underline{k} , the Euler equation is satisfied in the form of inequality

$$u'(\underline{c}) > \beta \sum_{z^+} u'(\underline{c}^+) \left[1 + \underline{y}_k^+ - (\pi(\underline{x}^+) + \pi'(\underline{x}^+) \underline{x}^+) \underline{x}_k^+ \right] Q(z, z^+),$$

where

$$\begin{aligned} \underline{c} &= (1 - \pi(x))x + k - \underline{k}, \\ \underline{c}^+ &= (1 - \pi(\underline{x}^+))\underline{x}^+ + \underline{k} - h(z^+, \underline{k}; K, \pi, \pi'), \\ \underline{y}_k^+ &= r(K), \\ \underline{x}^+ &= x(z^+, \underline{k}; K), \\ \underline{x}_k^+ &= x_k(z^+, \underline{k}; K) \end{aligned}$$

implying $h = \underline{k}$.

Taking this into account we can define the extended Euler equation operator,

$$\tilde{\mathcal{F}}(h) \equiv \begin{cases} \mathcal{F}(h) & \text{for } (z, k) \in Z \times [\underline{k}(z), \bar{k}(\bar{z})], \\ h & \text{for } (z, k) \in Z \times [\underline{k}, \underline{k}(z)), \end{cases}$$

and thus the operator equation in the form $\tilde{\mathcal{F}}(\tilde{h}) = 0$ determines the savings function with the segment of constrained savings, \tilde{h} .

For the sake of brevity we present here only the effect of the borrowing constrained agents at the lowest shock, \underline{z} , with the next period capital \underline{k} . The stationarity of the distribution functions implies

$$\lambda(z^+, \underline{k}) = \int_{\underline{k}}^{\underline{k}(\underline{z})} \lambda(\underline{z}, k) Q(\underline{z}, z^+) dk + \sum_{z \neq \underline{z}} \lambda(z, h^{-1}(z, k^+; K, \pi, \pi')) Q(z, z^+),$$

where $\lambda(z^+, \underline{k})$ is the mass of agents with the next period capital \underline{k} and the next-period shock z^+ . Let us note such amended distribution function by $\tilde{\lambda}$.

Clearly, it means that the amended distribution function has a discontinuity at $\tilde{\lambda}(\cdot, \underline{k})$ in the sense that

$$\tilde{\lambda}(z, \underline{k}) > \lim_{k \downarrow \underline{k}} \tilde{\lambda}(z, k),$$

for all $z \in Z$. However, since we assume here that functions are integrable in the Lebesgue sense, it follows that

$$\int_{\underline{k}}^{\bar{k}(\bar{z})} \tilde{\lambda}(z, k) dk = \int_{\underline{k}}^{\bar{k}(\bar{z})} \lambda(z, k) dk,$$

for any $z \in Z$ and the distribution functions $\tilde{\lambda}$ and λ are equivalent. So we can simply consider only the distribution function λ given by \mathcal{L} in (12) and h given by (11).

B Appendix: Proofs

B.1 Proof of Theorem 1

For the first order conditions for the Ramsey problem in (18)-(19) we define

$$\begin{aligned}
J(\varepsilon) &= \int_{\underline{x}(z)-\varepsilon\delta x}^{\bar{x}(z)} \left\{ \mathcal{W} [z, x; \tilde{K}, \tilde{\pi}(x), \tilde{\pi}'(x)] + \mu \mathcal{G} [z, x; \tilde{K}, \tilde{\pi}(x), \tilde{\pi}'(x)] \right\} dx \\
&+ \sum_{z \in Z \setminus \{z, \bar{z}\}} \int_{\underline{x}(z)}^{\bar{x}(z)} \left\{ \mathcal{W} [z, x; \tilde{K}, \tilde{\pi}(x), \tilde{\pi}'(x)] + \mu \mathcal{G} [z, x; \tilde{K}, \tilde{\pi}(x), \tilde{\pi}'(x)] \right\} dx \\
&+ \int_{\underline{x}(\bar{z})}^{\bar{x}(\bar{z})+\varepsilon\delta\bar{x}} \left\{ \mathcal{W} [z, x; \tilde{K}, \tilde{\pi}(x), \tilde{\pi}'(x)] + \mu \mathcal{G} [z, x; \tilde{K}, \tilde{\pi}(x), \tilde{\pi}'(x)] \right\} dx,
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\pi}(x) &\equiv \pi^*(x) + \varepsilon\delta\pi(x), \\
\tilde{\pi}'(x) &\equiv \pi^{*'}(x) + \varepsilon\delta\pi'(x), \\
\tilde{K} &\equiv K^* + \varepsilon\delta K.
\end{aligned}$$

The dependence of the bounds on the value of shocks $z \in Z$ makes our problem a little harder than the standard calculus of variation problem. However, as Theorem 1 states we construct the variations (the perturbation of functions from the optimum) being zero at all (interior) bounds – see Figure 8. Therefore, only the values of the government policy at the boundaries of the maximal interval, $\pi(\underline{x}(z))$ and $\pi(\bar{x}(\bar{z}))$, are free while all other interior bounds are fixed. Then the condition

$$J'(0) = \lim_{\varepsilon \rightarrow 0} \frac{dJ(\varepsilon)}{d\varepsilon} = 0$$

gives us the first-order conditions.

$$\begin{aligned}
J'(0) &= \sum_{z \in Z} \int_{\underline{x}(z)}^{\bar{x}(z)} \left\{ \mathcal{W}_\pi [z, x; K^*, \pi^*(x), \pi^{*'}(x)] + \mu \mathcal{G}_\pi [z, x; K^*, \pi^*(x), \pi^{*'}(x)] \right\} \delta\pi(x) \\
&+ \left\{ \mathcal{W}_{\pi'} [z, x; K^*, \pi^*(x), \pi^{*'}(x)] + \mu \mathcal{G}_{\pi'} [z, x; K^*, \pi^*(x), \pi^{*'}(x)] \right\} \delta\pi'(x) \\
&+ \left\{ \mathcal{W}_K [z, x; K^*, \pi^*(x), \pi^{*'}(x)] + \mu \mathcal{G}_K [z, x; K^*, \pi^*(x), \pi^{*'}(x)] \right\} \delta K \\
&- \left\{ \mathcal{W} [z, x; K^*, \pi^*(x), \pi^{*'}(x)] \Big|_{\underline{x}(z)} + \mu \mathcal{G} [z, x; K^*, \pi^*(x), \pi^{*'}(x)] \Big|_{\underline{x}(z)} \right\} (-\delta\underline{x}) \\
&+ \left\{ \mathcal{W} [\bar{z}, x; K^*, \pi^*(x), \pi^{*'}(x)] \Big|_{\bar{x}(\bar{z})} + \mu \mathcal{G} [\bar{z}, x; K^*, \pi^*(x), \pi^{*'}(x)] \Big|_{\bar{x}(\bar{z})} \right\} \delta\bar{x},
\end{aligned}$$

where the last two lines come from the fact that the lower and upper bounds are free.

INSERT FIGURE 8 ABOUT HERE

Using the following notation

$$\mathcal{W}(z, x) \equiv \mathcal{W} [z, x; K^*, \pi^*(x), \pi^{*'}(x)]$$

and

$$\mathbf{L}(z, x) \equiv \mathcal{W}(z, x) + \mu \mathcal{G}(z, x),$$

integration by parts delivers

$$\int_{\underline{x}(z)}^{\bar{x}(z)} \mathbf{L}_{\pi'}(z, x) \delta \pi'(x) dx = [\mathbf{L}_{\pi'}(z, x) \delta \pi(x)]_{\underline{x}(z)}^{\bar{x}(z)} - \int_{\underline{x}(z)}^{\bar{x}(z)} \frac{d}{dx} \mathbf{L}_{\pi'}(z, x) \delta \pi(x) dx.$$

Thus we can rewrite the formula above in a more compact form as

$$\begin{aligned} J'(0) &= \sum_{z \in Z} \int_{\underline{x}(z)}^{\bar{x}(z)} \left\{ \left[\mathbf{L}_{\pi}(z, x) - \frac{d}{dx} \mathbf{L}_{\pi'}(z, x) \right] \delta \pi(x) + \mathbf{L}_K(z, x) \delta K \right\} dx \\ &+ [\mathbf{L}_{\pi'}(z, x) \delta \pi(x)]_{\underline{x}(z)}^{\bar{x}(z)} - \mathbf{L}(z, x) |_{\underline{x}(z)} (-\delta \underline{x}) + \mathbf{L}(z, x) |_{\bar{x}(z)} \delta \bar{x}. \end{aligned} \quad (33)$$

At the free upper bound, the variation at the end-value of the policy function, $\delta \bar{\pi}$, can be expressed as

$$\delta \bar{\pi} \equiv \pi(\bar{x} + \delta \bar{x}) - \pi^*(\bar{x}) = \pi(\bar{x}) + \pi^{*'}(\bar{x}) \delta \bar{x} - \pi^*(\bar{x}),$$

and

$$\delta \pi(\bar{x}) = \pi(\bar{x}) - \pi^{*'}(\bar{x}) \delta \bar{x}.$$

This implies that

$$\delta \pi(\bar{x}) = \delta \bar{\pi} - \pi^{*'}(\bar{x}) \delta \bar{x}, \quad (34)$$

i.e. the variance of the policy function at the upper bound can be expressed as a function of the variance at the end-value of policy function, $\delta \bar{\pi}$, and the variance at the end-value of the taxable activity value, $\delta \bar{x}$.

We can similarly specify the variation of the start-value of the policy function,

$$\delta \pi(\underline{x}) = -\delta \underline{\pi} + \pi^{*'}(\underline{x}) \delta \underline{x}. \quad (35)$$

The situation is more complicated at the equality constrained endpoint $\bar{x}(\bar{z})$, because the upper bound for capital, \bar{k} , is implicitly given by the saving function, $\bar{k} = h(\bar{z}, \bar{k}; K, \pi)$. The total variance differential is

$$\delta \bar{k} [h_k(\bar{z}, \bar{k}; K, \pi) - 1] + h_K(\bar{z}, \bar{k}; K, \pi) \delta K + h_{\pi}(\bar{z}, \bar{k}; K, \pi) \delta \bar{\pi} = 0$$

and thus

$$\delta \bar{k} = [1 - h_k(\bar{z}, \bar{k}; K, \pi)]^{-1} [h_K(\bar{z}, \bar{k}; K, \pi) \delta K + h_{\pi}(\bar{z}, \bar{k}; K, \pi) \delta \bar{\pi}]. \quad (36)$$

As $\bar{x} = x[\bar{z}, \bar{k}(\bar{z}; K, \pi); K]$, we determine the variation

$$\delta \bar{x} = x_k[\bar{z}, \bar{k}(\bar{z}; K, \pi); K] \delta \bar{k} + x_K[\bar{z}, \bar{k}(\bar{z}; K, \pi); K] \delta K,$$

where we can further substitute for $\delta \bar{k}$ from (36)

$$\delta \bar{x} = \{x_k(\bar{z}, \bar{k}; K) \omega_K + x_K(\bar{z}, \bar{k}; K)\} \delta K + x_k(\bar{z}, \bar{k}; K) \omega_{\pi} \delta \bar{\pi}, \quad (37)$$

where

$$\begin{aligned}\omega_K &\equiv \frac{\delta \bar{k}}{\delta K} = \frac{h_K(\bar{z}, \bar{k}; K, \pi)}{1 - h_k(\bar{z}, \bar{k}; K, \pi)}, \\ \omega_\pi &\equiv \frac{\delta \bar{k}}{\delta \pi} = \frac{h_\pi(\bar{z}, \bar{k}; K, \pi)}{1 - h_k(\bar{z}, \bar{k}; K, \pi)}.\end{aligned}$$

Going back to (33) and using (34) and (35) we obtain

$$\begin{aligned}J'(0) &= \sum_{z \in Z} \int_{\underline{x}(z)}^{\bar{x}(z)} \left\{ \left[\mathbf{L}_\pi(z, x) - \frac{d}{dx} \mathbf{L}_{\pi'}(z, x) \right] \delta \pi(x) + \mathbf{L}_K(z, x) \delta K \right\} dx \quad (38) \\ &+ \mathbf{L}_{\pi'}(\underline{z}, x) \Big|_{\underline{x}(\underline{z})} \delta \underline{\pi} - [\pi'(x) \mathbf{L}_{\pi'}(\underline{z}, x) - \mathbf{L}(\underline{z}, x)]_{\underline{x}(\underline{z})} \delta \underline{x} \\ &+ \mathbf{L}_{\pi'}(\bar{z}, x) \Big|_{\bar{x}(\bar{z})} \delta \bar{\pi} - [\pi'(x) \mathbf{L}_{\pi'}(\bar{z}, x) - \mathbf{L}(\bar{z}, x)]_{\bar{x}(\bar{z})} \delta \bar{x}.\end{aligned}$$

Since the upper bound is equality constrained, $\delta \bar{x}$ is not independent and we need to use (37) to obtain

$$\begin{aligned}J'(0) &= \sum_{z \in Z} \int_{\underline{x}(z)}^{\bar{x}(z)} \left\{ \left[\mathbf{L}_\pi(z, x) - \frac{d}{dx} \mathbf{L}_{\pi'}(z, x) \right] \delta \pi(x) + \mathbf{L}_K(z, x) \delta K \right\} dx \quad (39) \\ &+ \mathbf{L}_{\pi'}(\underline{z}, x) \Big|_{\underline{x}(\underline{z})} \delta \underline{\pi} + [\mathbf{L}(\underline{z}, x) - \pi'(x) \mathbf{L}_{\pi'}(\underline{z}, x)]_{\underline{x}(\underline{z})} \underline{x}_K \delta K \\ &+ \mathbf{L}_{\pi'}(\bar{z}, x) \Big|_{\bar{x}(\bar{z})} \delta \bar{\pi} + [\mathbf{L}(\bar{z}, x) - \pi'(x) \mathbf{L}_{\pi'}(\bar{z}, x)]_{\bar{x}(\bar{z})} \{(\bar{x}_K + \bar{x}_k \bar{\omega}_K) \delta K + \bar{x}_k \bar{\omega}_\pi \delta \bar{\pi}\}.\end{aligned}$$

For the variation of the aggregate capital δK we use equation (17)

$$\begin{aligned}\delta K &= \sum_{z \in Z} \int_{\underline{x}(z)}^{\bar{x}(z)} \{ \mathcal{K}_\pi(z, x) \delta \pi(x) + \mathcal{K}_{\pi'}(z, x) \delta \pi'(x) + \mathcal{K}_K(z, x) \delta K \} dx \\ &- \mathcal{K}(\underline{z}, x) \Big|_{\underline{x}(\underline{z})} (-\delta \underline{x}) + \mathcal{K}(\bar{z}, x) \Big|_{\bar{x}(\bar{z})} \delta \bar{x},\end{aligned}$$

again using integration by parts and the conditions for the free boundary points,

$$\begin{aligned}\delta K &= \sum_{z \in Z} \int_{\underline{x}(z)}^{\bar{x}(z)} \left\{ \left[\mathcal{K}_\pi(z, x) - \frac{d}{dx} \mathcal{K}_{\pi'}(z, x) \right] \delta \pi(x) + \mathcal{K}_K(z, x) \delta K \right\} dx \\ &+ \mathcal{K}_{\pi'}(\underline{z}, x) \Big|_{\underline{x}(\underline{z})} \delta \underline{\pi} + [\mathcal{K}(\underline{z}, x) - \pi'(x) \mathcal{K}_{\pi'}(\underline{z}, x)]_{\underline{x}(\underline{z})} \underline{x}_K \delta K \\ &+ \mathcal{K}_{\pi'}(\bar{z}, x) \Big|_{\bar{x}(\bar{z})} \delta \bar{\pi} + [\mathcal{K}(\bar{z}, x) - \pi'(x) \mathcal{K}_{\pi'}(\bar{z}, x)]_{\bar{x}(\bar{z})} \{(\bar{x}_K + \bar{x}_k \bar{\omega}_K) \delta K + \bar{x}_k \bar{\omega}_\pi \delta \bar{\pi}\}.\end{aligned}$$

The variation of the aggregate capital is

$$\begin{aligned}\delta K &= \Psi_K \left\{ \sum_{z \in Z} \int_{\underline{x}(z)}^{\bar{x}(z)} \left\{ \left[\mathcal{K}_\pi(z, x) - \frac{d}{dx} \mathcal{K}_{\pi'}(z, x) \right] \delta \pi(x) \right\} dx \right. \\ &+ \mathcal{K}_{\pi'}(\underline{z}, x) \Big|_{\underline{x}(\underline{z})} \delta \underline{\pi} + \left. \left\{ \mathcal{K}_{\pi'}(\bar{z}, x) \Big|_{\bar{x}(\bar{z})} + [\mathcal{K}(\bar{z}, x) - \pi'(x) \mathcal{K}_{\pi'}(\bar{z}, x)]_{\bar{x}(\bar{z})} \bar{x}_k \bar{\omega}_\pi \right\} \delta \bar{\pi} \right\},\end{aligned}$$

where

$$\begin{aligned}\Psi_K^{-1} &\equiv 1 - \sum_{z \in Z} \int_{\underline{x}(z)}^{\bar{x}(z)} \mathcal{K}_K(z, x) dx - [\mathcal{K}(\underline{z}, x) - \pi'(x) \mathcal{K}_{\pi'}(\underline{z}, x)]_{\underline{x}(\underline{z})} \underline{x}_K \\ &- [\mathcal{K}(\bar{z}, x) - \pi'(x) \mathcal{K}_{\pi'}(\bar{z}, x)]_{\bar{x}(\bar{z})} (\bar{x}_K + \bar{x}_k \bar{\omega}_K).\end{aligned}$$

Substituting the formula for δK into (39) we get

$$\begin{aligned} \delta J &= \sum_{z \in Z} \int_{\underline{x}(z)}^{\bar{x}(z)} \left[\mathbf{L}_\pi(z, x) + \Psi \mathcal{K}_\pi(z, x) - \frac{d}{dx} (\mathbf{L}_{\pi'}(z, x) + \Psi \mathcal{K}_{\pi'}(z, x)) \right] \delta \pi(x) dx \\ &+ [\mathbf{L}_{\pi'}(\underline{z}, x) + \Psi \mathcal{K}_{\pi'}(\underline{z}, x)]|_{\underline{x}(\underline{z})} \delta \underline{\pi} + \{ [\mathbf{L}_{\pi'}(\bar{z}, x) + \Psi \mathcal{K}_{\pi'}(\bar{z}, x)]|_{\bar{x}(\bar{z})} \\ &+ [\mathbf{L}(\bar{z}, x) + \Psi \mathcal{K}(\bar{z}, x) - \pi'(x) (\mathbf{L}_{\pi'}(\bar{z}, x) + \Psi \mathcal{K}_{\pi'}(\bar{z}, x))]|_{\bar{x}(\bar{z})} \bar{x}_k \bar{\omega}_\pi \} \delta \bar{\pi}, \end{aligned}$$

with

$$\begin{aligned} \Psi &\equiv \frac{\delta \mathbf{L}}{\delta K} = \Psi_K \left\{ \sum_{z \in Z} \int_{\underline{x}(z)}^{\bar{x}(z)} \mathbf{L}_K(z, x) dx + [\mathbf{L}(\underline{z}, x) - \pi'(x) \mathbf{L}_{\pi'}(\underline{z}, x)]|_{\underline{x}(\underline{z})} \underline{x}_K \right. \\ &\left. + [\mathbf{L}(\bar{z}, x) - \pi'(x) \mathbf{L}_{\pi'}(\bar{z}, x)]|_{\bar{x}(\bar{z})} (\bar{x}_K + \bar{x}_k \bar{\omega}_K) \right\}. \end{aligned}$$

Now, in order to get FOCs we assign $\delta J = 0$. Since the first term is zero for any $\delta \pi(x)$ for all x , the following terms must be zero

$$\begin{aligned} \sum_{z \in Z} \left[\mathbf{L}_\pi(z, x) + \Psi \mathcal{K}_\pi(z, x) - \frac{d}{dx} (\mathbf{L}_{\pi'}(z, x) + \Psi \mathcal{K}_{\pi'}(z, x)) \right] &= 0, \\ [\mathbf{L}_{\pi'}(\underline{z}, x) + \Psi \mathcal{K}_{\pi'}(\underline{z}, x)]|_{\underline{x}(\underline{z})} &= 0, \\ [\mathbf{L}(\bar{z}, x) + \Psi \mathcal{K}(\bar{z}, x)]|_{\bar{x}(\bar{z})} + \left[\frac{1}{\bar{x}_k \bar{\omega}_\pi} - \pi'(x) \right] [\mathbf{L}_{\pi'}(\bar{z}, x) + \Psi \mathcal{K}_{\pi'}(\bar{z}, x)]|_{\bar{x}(\bar{z})} &= 0. \end{aligned}$$

Note that

$$\left[\frac{\delta x}{\delta \pi} \right]^{-1} = \frac{1}{\bar{x}_k \bar{\omega}_K}.$$

If we denote the modified Lagrange function by $\tilde{\mathbf{L}}(z, x) \equiv \mathbf{L}(z, x) + \Psi \mathcal{K}(z, x)$, then clearly the derived FOCs are those in (22)-(25). Q.E.D.

B.1.1 Definition of Terms in Theorem 1

Recall that according to (16)

$$\mathcal{W}[z, x; K, \pi(x), \pi'(x)] = W[u(c(z, k(z, x; K); K, \pi, \pi'))] \lambda[z, k(z, x; K); K, \pi, \pi'] k_x(z, x; K),$$

then

$$\mathcal{W}_\pi(z, x) \delta \pi(x) = \lim_{\varepsilon \rightarrow 0} \frac{d\mathcal{W}[z, x; K, \tilde{\pi}(x), \pi'(x)]}{d\varepsilon},$$

with

$$\begin{aligned} c(z, k(z, x; K); K, \tilde{\pi}, \pi') &= y(z, k(z, x; K); K) - \tilde{\pi}(x)x + k(z, x; K) - h(z, k(z, x; K); K, \tilde{\pi}, \pi'), \\ y(z, k(z, x; K); K) &= r(K)k(z, x; K) + w(K)z, \end{aligned}$$

and $\tilde{\pi}(x) \equiv \pi(x) + \varepsilon \delta \pi(x)$.

To simplify notation we will omit the obvious arguments, writing $c = c(z, k(z, x; K))$, $c_\pi = c_\pi(z, k(z, x; K))$, $\lambda = \lambda(z, k(z, x; K))$ and so on. Thus,

$$\mathcal{W}_\pi = \{W'[u(c)]u'(c)c_\pi\lambda + W[u(c)]\lambda_\pi\}k_x,$$

where $c_\pi = -x - h_\pi$, and the Frechet derivatives of the savings policy and the distribution function, h_π and λ_π , are given by Lemmas 1 and 2, respectively.

Similarly,

$$\mathcal{W}_{\pi'} = \{W'[u(c)]u'(c)c_{\pi'}\lambda + W[u(c)]\lambda_{\pi'}\}k_x, \quad (40)$$

where $c_{\pi'} = -h_{\pi'}$, for the Frechet derivatives $h_{\pi'}$ and $\lambda_{\pi'}$. Finally, for h_K and λ_K ,

$$\mathcal{W}_K = \{W'[u(c)]u'(c)c_K\lambda + W[u(c)][\lambda_K + \lambda_k k_K]\}k_x + W[u(c)]\lambda k_{xK},$$

where

$$\begin{aligned} r_K &= F_{11}(K, \bar{L}), \\ w_K &= F_{21}(K, \bar{L}), \\ c_K &= r_K k + w_K z + [1 + r - h_k]k_K - h_K. \end{aligned}$$

For the equation

$$\frac{d}{dx}\mathcal{W}_{\pi'}(z, x) = \mathcal{W}_{\pi'\pi'}(z, x)\pi''(x) + \mathcal{W}_{\pi'\pi}(z, x)\pi'(x) + \mathcal{W}_{\pi'x}(z, x),$$

we obtain, using $W = W[u(c)]$, $W' = W'[u(c)]u'(c)$, and $W'' = W''[u(c)][u'(c)]^2 + W'[u(c)]u''(c)$,

$$\begin{aligned} \mathcal{W}_{\pi'\pi'}(z, x) &= \{W''[c_{\pi'}]^2\lambda + W'(c_{\pi'}\lambda + 2c_{\pi'}\lambda_{\pi'}) + W\lambda_{\pi'\pi'}\}k_x, \\ \mathcal{W}_{\pi'\pi}(z, x) &= \{W''c_{\pi'}c_\pi\lambda + W'[c_{\pi'}\pi\lambda + c_{\pi'}\lambda_\pi + c_\pi\lambda_{\pi'}] + W\lambda_{\pi'\pi}\}k_x, \\ \mathcal{W}_{\pi'x}(z, x) &= \{W''c_{\pi'}c_x\lambda + W'[c_{\pi'}x\lambda + c_{\pi'}\lambda_x + c_x\lambda_{\pi'}] + W\lambda_{\pi'x}\}k_x + \{W'c_{\pi'}\lambda + W\lambda_{\pi'}\}k_{xx}, \end{aligned}$$

with $c_{\pi'\pi'} = -h_{\pi'\pi'}$, $c_{\pi'\pi} = -h_{\pi'\pi}$, $c_{\pi'x} = -h_{\pi'k}k_x(z, x)$, and $\lambda_{\pi'x} = \lambda_{\pi'k}k_x(z, x)$.

According to (17),

$$\mathcal{G}[z, x; K, \pi(x), \pi'(x)] = [\pi(x)x - g y(z, k(z, x; K); K)]\lambda[z, k(z, x; K); K, \pi, \pi']k_x(z, x; K),$$

and therefore,

$$\begin{aligned} \mathcal{G}_\pi(z, x) &= \{x\lambda + [\pi(x)x - g y]\lambda_\pi\}k_x, \\ \mathcal{G}_{\pi'}(z, x) &= [\pi(x)x - g y]\lambda_{\pi'}k_x, \\ \mathcal{G}_K(z, x) &= \{-g[r_K k + r k_K + w_K z]\lambda + [\pi(x)x - g y][\lambda_K + \lambda_k k_K]\}k_x + [\pi(x)x - g y]\lambda k_{xK}, \end{aligned} \quad (41)$$

and

$$\begin{aligned} \frac{d}{dx}\mathcal{G}_{\pi'}(z, x) &= \mathcal{G}_{\pi'\pi'}(z, x)\pi''(x) + \mathcal{G}_{\pi'\pi}(z, x)\pi'(x) + \mathcal{G}_{\pi'x}(z, x), \\ \mathcal{G}_{\pi'\pi'}(z, x) &= \{[\pi(x)x - g y]\lambda_{\pi'\pi'}\}k_x, \\ \mathcal{G}_{\pi'\pi}(z, x) &= \{x\lambda_\pi + [\pi(x)x - g y]\lambda_{\pi'\pi}\}k_x, \\ \mathcal{G}_{\pi'x}(z, x) &= \{[\pi(x)x - g y]k_x\lambda_{\pi'} + [\pi(x)x - g y]\lambda_{\pi'x}\}k_x + [\pi(x)x - g y]\lambda_{\pi'}k_{xx}. \end{aligned}$$

Lastly, according to (17)

$$\mathcal{K}[z, x; K, \pi(x), \pi'(x)] = k(z, x; K) \lambda[z, k(z, x; K); K, \pi, \pi'] k_x(z, x; K),$$

and thus

$$\begin{aligned} \mathcal{K}_\pi(z, x) &= k \lambda_\pi k_x, \\ \mathcal{K}_{\pi'}(z, x) &= k \lambda_{\pi'} k_x, \\ \mathcal{K}_K(z, x) &= \{k_K \lambda + k [\lambda_k k_K + \lambda_K]\} k_x + k \lambda k_{xK}, \end{aligned} \tag{42}$$

and

$$\begin{aligned} \frac{d}{dx} \mathcal{K}_{\pi'}(z, x) &= \mathcal{K}_{\pi'\pi'}(z, x) \pi''(x) + \mathcal{K}_{\pi'\pi}(z, x) \pi'(x) + \mathcal{K}_{\pi'x}(z, x), \\ \mathcal{K}_{\pi'\pi'}(z, x) &= k \lambda_{\pi'\pi'} k_x, \\ \mathcal{K}_{\pi'\pi}(z, x) &= k \lambda_{\pi'\pi} k_x, \\ \mathcal{K}_{\pi'x}(z, x) &= \{k_x^2 + k k_{xx}\} \lambda_{\pi'} + k \lambda_{\pi'x} k_x. \end{aligned}$$

B.2 Proof of Lemma 1

From equation (16) the Euler equation operator with the variation $\tilde{\pi} \equiv \pi + \varepsilon \delta \pi$ is

$$\begin{aligned} &\mathcal{F}(h, \pi)(z, k; K, \tilde{\pi}(x), \pi'(x)) \equiv u'(c(z, k; K, \tilde{\pi}, \pi')) \\ &- \beta \sum_{z^+} u'(c^+(z^+, h(z, k; K, \tilde{\pi}, \pi'); K, \tilde{\pi}, \pi')) R^+(z^+, h(z, k; K, \tilde{\pi}, \pi'); K, \tilde{\pi}, \pi') Q(z, z^+), \end{aligned}$$

and

$$\mathcal{F}_\pi(z, k) \delta \pi = \lim_{\varepsilon \rightarrow 0} \frac{d\mathcal{F}[z, k; K, \tilde{\pi}(x), \pi'(x)]}{d\varepsilon},$$

Using abbreviated notation $x = x(z, k)$, $c = c(z, k)$, $h = h(z, k)$, $h^+ = h(z^+, h)$, $x^+ = x(z^+, h)$, $y^+ = y(z^+, h)$, $\pi^+ = \pi(z^+, h)$, $c^+ = c^+(z, z^+, k)$, and $R^+ = R^+(z, z^+, k)$,

$$\mathcal{F}_\pi(z, k) = u''(c) c_\pi - \beta \sum_{z^+} \{u''(c^+) c_\pi^+ R^+ + u'(c^+) R_\pi^+\} Q(z, z^+) = 0,$$

where

$$\begin{aligned} c &= y - \tilde{\pi}(x)x + k - h, \\ y &= rk + wz, \\ c^+ &= y^+ - \tilde{\pi}(x^+)x^+ + h - h^+, \\ R^+ &= 1 + y_k^+ - [\tilde{\pi}(x^+) + \pi'(x^+)x^+] x_k^+. \end{aligned}$$

Terms for \mathcal{F}_π above and in equation (26) are

$$\begin{aligned} c_\pi &= -x - h_\pi, \\ c_\pi^+ &= [1 + r - h_k^+ - [\pi'^+ x^+ + \pi^+] x_k^+] h_\pi - h_\pi^+ - x^+, \\ R_\pi^+ &= r h_{k\pi} - [1 + [2\pi'(x^+) + \pi''(x^+)x^+] x_k^+ h_\pi] x_k^+ - [\pi'^+ x^+ + \pi^+] x_{kk}^+ h_\pi. \end{aligned}$$

For

$$\mathcal{F}_{\pi'}(z, k) = u''(c) c_{\pi'} - \beta \sum_{z^+} \{u''(c^+) c_{\pi'}^+ R^+ + u'(c^+) R_{\pi'}^+\} Q(z, z^+) = 0,$$

we use terms

$$\begin{aligned} c_{\pi'} &= -h_{\pi'}, \\ c_{\pi'}^+ &= \{1 + r - h_k^+ - [\pi'^+ x + \pi^+] x_k^+\} h_{\pi'} - h_{\pi'}^+, \\ R_{\pi'}^+ &= r h_{k\pi'} - \{x^+ + [2\pi'(x^+) + \pi''(x^+) x^+] x_k^+ h_{\pi'}\} x_k^+ - [\pi'^+ x^+ + \pi^+] x_{kk}^+ h_{\pi'}. \end{aligned}$$

For

$$\mathcal{F}_K(z, k) = u''(c) c_K - \beta \sum_{z^+} \{u''(c^+) c_K^+ R^+ + u'(c^+) R_K^+\} Q(z, z^+) = 0,$$

the terms are, as well as for equation (28),

$$\begin{aligned} c_K &= r_K k + w_K z - [\pi'(x) x + \pi(x)] x_K - h_K, \\ c_K^+ &= r_K h + w_K z^+ - [\pi'^+ x^+ + \pi^+] [x_K^+ + x_k^+ h_K] - h_K^+, \\ R_K^+ &= r_K h_k + r h_{kK} - [2\pi'(x^+) + \pi''(x^+) x^+] x_k^+ [x_K^+ + x_k^+ h_K] - [\pi'^+ x^+ + \pi^+] [x_{kK}^+ + x_{kk}^+ h_K]. \end{aligned}$$

For

$$\begin{aligned} \mathcal{F}_{\pi'\pi'}(z, k) &= u'''(c) c_{\pi'}^2 + u''(c) c_{\pi'\pi'} \\ &- \beta \sum_{z^+} \left\{ [u'''(c^+) [c_{\pi'}^+]^2 + u''(c^+) c_{\pi'\pi'}^+] R^+ + 2u''(c^+) c_{\pi'}^+ R_{\pi'}^+ + u'(c^+) R_{\pi'\pi'}^+ \right\} Q(z, z^+) = 0, \end{aligned}$$

we use, as well in equation (27), the terms

$$\begin{aligned} c_{\pi'\pi'} &= -h_{\pi'\pi'}, \\ c_{\pi'\pi'}^+ &= [1 + r - (h_{k\pi'}^+ + h_{kk}^+ h_{\pi'})] h_{\pi'} + (1 + r - h_k^+) h_{\pi'\pi'} - (h_{\pi'k}^+ h_{\pi'} + h_{\pi'\pi'}^+) \\ &- [x^+ + (\pi''(x^+) x^+ + 2\pi'(x^+)) x_k^+ h_{\pi'}] x_k^+ h_{\pi'} \\ &- [\pi'(x^+) x^+ + \pi(x^+)] (x_{kk}^+ (h_{\pi'})^2 + x_k^+ h_{\pi'\pi'}), \\ R_{\pi'\pi'}^+ &= r h_{k\pi'\pi'} - [3(x_k^+)^2 + x^+ x_{kk}^+] h_{\pi'} - [3\pi''(x^+) + \pi'''(x^+) x^+] (x_k^+)^3 (h_{\pi'})^2 \\ &- [2\pi'(x^+) + \pi''(x^+) x^+] [3x_k^+ x_{kk}^+ (h_{\pi'})^2 + (x_k^+)^2 h_{\pi'\pi'}] \\ &- [\pi'(x^+) x^+ + \pi(x^+)] [x_{kkk}^+ (h_{\pi'})^2 + x_{kk}^+ h_{\pi'\pi'}]. \end{aligned}$$

Finally,

$$\begin{aligned} \mathcal{F}_{\pi'\pi}(z, k) &= u'''(c) c_{\pi'} c_{\pi} + u''(c) c_{\pi'\pi} - \beta \sum_{z^+} \{ [u'''(c^+) c_{\pi'}^+ c_{\pi}^+ + u''(c^+) c_{\pi'\pi}^+] R^+ \\ &+ u''(c^+) [c_{\pi'}^+ R_{\pi}^+ + c_{\pi}^+ R_{\pi'}^+] + u'(c^+) R_{\pi'\pi}^+ \} Q(z, z^+) = 0. \end{aligned}$$

Terms for $\mathcal{F}_{\pi'\pi'}$ above and in equation (27) are

$$\begin{aligned}
c_{\pi'\pi} &= -h_{\pi'\pi}, \\
c_{\pi'\pi'}^+ &= -(h_{k\pi}^+ + h_{kk}^+ h_\pi) h_{\pi'} + (1 + r - h_k^+) h_{\pi'\pi} - (h_{\pi'k}^+ h_\pi + h_{\pi'\pi}^+) \\
&\quad - [1 + (\pi''(x^+) x^+ + 2\pi'(x^+)) x_k^+ h_\pi] x_k^+ h_{\pi'} \\
&\quad - [\pi'(x^+) x^+ + \pi(x^+)] (x_{kk}^+ h_{\pi'} h_\pi + x_k^+ h_{\pi'\pi}), \\
R_{\pi'\pi'}^+ &= r h_{k\pi'\pi} - [(x_k^+)^2 + x^+ x_{kk}^+] h_\pi - [3\pi''(x^+) + \pi'''(x^+) x^+] (x_k^+)^3 h_{\pi'} h_\pi \\
&\quad - [2\pi'(x^+) + \pi''(x^+) x^+] [3x_k^+ x_{kk}^+ h_{\pi'} h_\pi + (x_k^+)^2 h_{\pi'\pi}] \\
&\quad - [\pi'(x^+) x^+ + \pi(x^+)] [x_{kkk}^+ h_{\pi'} h_\pi + x_{kk}^+ h_{\pi'\pi}].
\end{aligned}$$

Q.E.D.

B.3 Proof of Lemma 2

Using equation (16), the stationary distribution operator is

$$\mathcal{L}(h, \lambda, \pi) \equiv \lambda(z^+, k^+; K, \pi, \pi') - \sum_z \lambda[z, h^{-1}(z, k^+; K, \pi, \pi'); K, \pi, \pi'] Q(z, z^+),$$

and so

$$\mathcal{L}_\pi(z^+, k^+) \delta\pi = \lim_{\varepsilon \rightarrow 0} \frac{d\mathcal{L}[z^+, k^+; K, \tilde{\pi}(x), \pi'(x)]}{d\varepsilon},$$

with $\tilde{\pi} \equiv \pi + \varepsilon \delta\pi$. Therefore, abbreviating for (K, π, π') ,

$$\mathcal{L}_\pi(z^+, k^+) = \lambda_\pi(z^+, k^+) - \sum_z [\lambda_\pi(z, h^{-1}(z, k^+)) + \lambda_k(z, h^{-1}(z, k^+)) h_\pi^{-1}(z, k^+)] = 0$$

with $h_\pi^{-1}(z, k^+) = h_\pi(z, h^{-1}(z, k^+)) / h_k(z, h^{-1}(z, k^+))$.

Similarly, we can derive the total F-derivative of the Euler equation operator with respect to the derivative of government policy function,

$$\mathcal{L}_{\pi'}(z^+, k^+) = \lambda_{\pi'}(z^+, k^+) - \sum_z [\lambda_{\pi'}(z, h^{-1}(z, k^+)) + \lambda_k(z, h^{-1}(z, k^+)) h_{\pi'}^{-1}(z, k^+)] = 0$$

with $h_{\pi'}^{-1}(z, k^+) = h_{\pi'}(z, h^{-1}(z, k^+)) / h_k(z, h^{-1}(z, k^+))$, and

$$\mathcal{L}_K(z^+, k^+) = \lambda_K(z^+, k^+) - \sum_z [\lambda_K(z, h^{-1}(z, k^+)) + \lambda_k(z, h^{-1}(z, k^+)) h_K^{-1}(z, k^+)] = 0,$$

with $h_K^{-1}(z, k^+) = h_K(z, h^{-1}(z, k^+)) / h_k(z, h^{-1}(z, k^+))$.

Further, we can derive the following total F-derivative of the Euler equation operator

$$\begin{aligned}
\mathcal{L}_{\pi'\pi'}(z^+, k^+) &= \lambda_{\pi'\pi'}(z^+, k^+) - \sum_z \{ \lambda_{\pi'\pi'}(z, h^{-1}(z, k^+)) + \lambda_{\pi'k}(z, h^{-1}(z, k^+)) h_{\pi'}^{-1}(z, k^+) \\
&\quad + [\lambda_{\pi'k}(z, h^{-1}(z, k^+)) + \lambda_{kk}(z, h^{-1}(z, k^+))] [h_{\pi'}^{-1}(z, k^+)]^2 \\
&\quad + \lambda_k(z, h^{-1}(z, k^+)) h_{\pi'\pi'}^{-1}(z, k^+) \} = 0
\end{aligned}$$

where $h_{\pi'\pi}^{-1}(z, k^+) = h_{\pi'\pi'}(z, h^{-1}(z, k^+)) / h_{kk}(z, h^{-1}(z, k^+))$.

Finally,

$$\begin{aligned}\mathcal{L}_{\pi'\pi}(z^+, k^+) &= \lambda_{\pi'\pi}(z^+, k^+) - \sum_z \{ \lambda_{\pi'\pi}(z, h^{-1}(z, k^+)) + \lambda_{\pi'k}(z, h^{-1}(z, k^+)) h_{\pi}^{-1}(z, k^+) \\ &+ [\lambda_{\pi k}(z, h^{-1}(z, k^+)) + \lambda_{kk}(z, h^{-1}(z, k^+))] [h_{\pi}^{-1}(z, k^+)]^2 \\ &+ \lambda_k(z, h^{-1}(z, k^+)) h_{\pi'\pi}^{-1}(z, k^+) \} = 0\end{aligned}$$

where $h_{\pi'\pi}^{-1}(z, k^+) = h_{\pi'\pi}(z, h^{-1}(z, k^+)) h_{kk}(z, h^{-1}(z, k^+))$.

Q.E.D.

B.4 Proof of Proposition 1

Using the definition for $\tilde{\mathbf{L}}$ in equation (21), the boundary first-order condition in (24) can be expressed as

$$\tilde{\mathbf{L}}_{\pi'}(\underline{z}, \underline{x}(\underline{z})) = \mathcal{W}_{\pi'}(\underline{z}, \underline{x}(\underline{z})) + \mu \mathcal{G}_{\pi'}(\underline{z}, \underline{x}(\underline{z})) + \Psi \mathcal{K}_{\pi'}(\underline{z}, \underline{x}(\underline{z})) = 0.$$

The term $\mathcal{W}_{\pi'}$ in (40) evaluated at $(\underline{z}, \underline{x}(\underline{z}))$ gives

$$\begin{aligned}\mathcal{W}_{\pi'}(\underline{z}, \underline{x}(\underline{z})) &= \{W'[u(c(\underline{z}, \underline{k}))] u'(c(\underline{z}, \underline{k})) c_{\pi'}(\underline{z}, \underline{k}) \lambda(\underline{z}, \underline{k}) \\ &+ W[u(c(\underline{z}, \underline{k}))] \lambda_{\pi'}(\underline{z}, \underline{k})\} k_x(\underline{z}, \underline{x}(\underline{z})).\end{aligned}\quad (43)$$

From $c_{\pi'} = -h_{\pi'}$ it follows that

$$c_{\pi'}(\underline{z}, \underline{k}) = -h_{\pi'}(\underline{z}, \underline{k}) = 0,$$

since the ‘variation’ of the savings policy function related to the lowest shock, $h(\underline{z}, \cdot)$, with respect to the slope of the government policy function, $h_{\pi'}$, at $(\underline{z}, \underline{k})$ is clearly zero. Note that the saving function of the borrowing-constrained agents is flat and equal to zero. It implies that the first term in (43) is also equal to zero. Using $\mathcal{G}_{\pi'}$, and $\mathcal{K}_{\pi'}$ from (41) and (42), respectively, we get

$$W[u(c(\underline{z}, \underline{k}))] + \mu [\pi(\underline{x}(\underline{z})) \underline{x}(\underline{z}) - g y(\underline{z}, \underline{k})] + \Psi \underline{k} = 0,$$

where we used the fact that both $\lambda_{\pi'}(\underline{z}, \underline{k})$ and $k_x(\underline{z}, \underline{x}(\underline{z}))$ are non-zero. The result of the Proposition follows. Q.E.D.

B.5 Proof of Proposition 2

We see that terms $\tilde{\mathbf{L}}$ and $\tilde{\mathbf{L}}_{\pi'}$ appear in the boundary first-order condition in (23). Using the definition of $\tilde{\mathbf{L}}$ in (21), we get

$$\tilde{\mathbf{L}}(\bar{z}, \bar{x}(\bar{z})) = \{W[u(c(\bar{z}, \bar{k}))] + \mu [\pi(\bar{x}(\bar{z})) \bar{x}(\bar{z}) - g y(\bar{z}, \bar{k})] + \Psi \bar{k}\} \lambda(\bar{z}, \bar{k}) = 0,$$

where we have used (20), (17), and (17) for \mathcal{W} , \mathcal{G} , and \mathcal{K} , respectively, and also the fact that at the upper endogenous limit on capital, \bar{k} , $\lambda(\bar{z}, \bar{k}) = 0$. Thus the first-order boundary condition (23) becomes

$$\left(\pi'(\bar{x}(\bar{z})) - \frac{k_x(\bar{z}, \bar{x}(\bar{z}))}{\omega_\pi(\bar{z}, \bar{x}(\bar{z}))} \right) \tilde{\mathbf{L}}_{\pi'}(\bar{z}, \bar{x}(\bar{z})) = 0.$$

If we assume that $\pi'(\bar{x}(\bar{z})) - \frac{k_x(\bar{z}, \bar{x}(\bar{z}))}{\omega_\pi(\bar{z}, \bar{x}(\bar{z}))} > 0$, then the condition above is satisfied if the second term $\tilde{\mathbf{L}}_{\pi'}(\bar{z}, \bar{x}(\bar{z}))$ is equal to zero. By the inspection of $\mathcal{W}_{\pi'}$ in (40) we see that since $\lambda(\bar{z}, \bar{k}) = 0$ then again the first term in (40) is zero and

$$W[u(c(\bar{z}, \bar{k}))] + \mu[\pi(\bar{x}(\bar{z}))\bar{x}(\bar{z}) - gy(\bar{z}, \bar{k})] + \Psi\bar{k} = 0,$$

where we again used the fact that both $\lambda_{\pi'}(\bar{z}, \bar{k})$ and $k_x(\bar{z}, \bar{x}(\bar{z}))$ are non-zero. This implies the result of the Proposition. Q.E.D.

B.6 Proof of Theorem 2

We need to show that the first order approach to each agent's maximization problem is valid. First, agents maximize over a quasi-convex set: $\Psi = \{x \in B : 0 \leq x \leq \psi(k, z) \text{ for all } (k, z) \in B \times Z\}$. If the function ψ is increasing and quasi-concave, then the set Ψ is quasi-convex. Further, we need to satisfy Assumptions 18.1 in Stokey, Lucas, and Prescott (1989), particularly that (i) $\beta \in (0, 1)$; (ii) utility function $u : \mathbb{R}_+ \rightarrow \mathbb{R}$ is twice continuously differentiable, strictly increasing and strictly concave function; (iii) for some $\bar{k}(z) > 0$, $\psi(k, z) - k$ is strictly positive on $[0, \bar{k}(z))$ and strictly negative for $k > \bar{k}(z)$, where the value \bar{k} , the maximum sustainable capital stock out of after-tax income for any agent, is defined as $\bar{k} = \max\{\bar{k}(z_1), \dots, \bar{k}(z_J)\}$; and, (iv) given the tax-schedule function the right-hand side of the Euler equation is strictly positive

$$\beta \sum_{z^+} u'(\psi(k^+(k, z), z^+) - k^+(k^+(k, z), z^+)) \psi_1(k^+(k, z), z^+) Q(z, z^+) > 0,$$

where

$$\psi_1(k^+(k, z), z^+) Q(z, z^+) = (1 - \tau(y(k^+(k, z), z^+)) - \tau'(y(k^+(k, z), z^+)) y(k^+(k, z), z^+)) r + 1.$$

It can be easily checked that the assumptions (i)-(iii) are satisfied from our previous assumptions and the model. The assumption (iv) follows directly from the fact that ψ is increasing in k , i.e. $\psi_1 > 0$.

The other assumptions needed for proving the existence of a stationary recursive competitive equilibrium (see Assumption 18.2 in Stokey, Lucas, and Prescott (1989)) are satisfied : (i) the equilibrium marginal return on capital for any $k \in B$ is finite (in our case the interest rate r); and (ii) that $\lim_{c \rightarrow 0} u'(c) = \infty$.

Then to prove the Schauder's Theorem, let $C(B, Z)$ be the set of continuous bounded functions $h : B \times Z \rightarrow B$ and define a subset $F = \{h \in C(B, Z)\}$ where the function h satisfies $0 \leq h(k, z) \leq \psi(k, z)$, all $(k, z) \in B \times Z$, and h and $\psi - h$ are nondecreasing. Note

that $B \times Z$ is a bounded subset of \mathbb{R}^2 and that the family of functions F is nonempty, closed, bounded, and convex. Define an operator T on F

$$u'(\psi(k, z) - (Th)(k, z)) = \beta \sum_{z^+} u'(\psi((Th)(k, z), z^+) - h[(Th)(k, z), z^+]) \cdot [(1 - \tau(y(\cdot)) - \tau'(y(\cdot))y(\cdot))r + 1] Q(z, z^+),$$

where $y(\cdot) = y((Th)(k, z), z^+)$.

Then it is easy to prove that T is well defined, continuous and that $T : F \rightarrow F$. From the conditions on function h and finite return on capital, it follows that F is an equicontinuous family. That the operator T has a fixed point in F follows from the Schauder's Theorem (see e.g. Theorem 17.4 in Stokey, Lucas, and Prescott (1989)).

The existence of the stationary recursive competitive equilibrium is standard from the monotonicity, Feller and mixing property of Q and the non-decreasing policy functions (see Chapter 12 in Stokey, Lucas, and Prescott (1989)). Q.E.D.

C Terms for the Example

The terms from Theorem 1 for Example are those in B.1.1 together with

$$\begin{aligned} y(z, k(z, x; K)) &= x, \\ k_x(z, x; K) &= \frac{1}{r(K)}, \\ k_K(z, x; K) &= -\frac{w_K(K)zr(K) + [x - w(K)z]r_K(K)}{[r(K)]^2}, \end{aligned}$$

and

$$k_{xK}(z, x; K) = -\frac{r_K}{[r(K)]^2}.$$

D Appendix: The Least Squares Projection Method

The optimal income tax policy, π , is a solution of the following system of operator equations:

1. FOC for π given by the Euler-Lagrange condition in (22);
2. the Euler equation (11) capturing the individual optimal behavior h ;
3. five operator equations (26)-(30) for F-derivatives of h based on the Euler equation, h_π , $h_{\pi'}$, h_K , $h_{\pi'\pi'}$, and $h_{\pi'\pi}$;
4. the operator equation for distribution function, λ , in (12); and
5. five operator equations (31)-(32) for F-derivatives of λ based on the operator equation for distribution function: λ_π , $\lambda_{\pi'}$, λ_K , $\lambda_{\pi'\pi'}$, and $\lambda_{\pi'\pi}$.

In order to solve the problem numerically, we first approximate all the unknown functions by combinations of polynomials from a polynomial base. Approximated solutions are specified by unknown parameters transforming the original infinitely dimensional problem into a finite dimensional one. After substituting the approximated functions into the original operator equations we construct the *residual equations*. Ideally, the residual functions should be uniformly equal to zero. In practical situations, however, this is not achievable and we limit the problem to a finite number of conditions, the so called *projections*, whose satisfaction guarantees a reasonably good approximation. There are many possibilities how to define the projections.²⁸ We have chosen the least squares projection method for its good convergence properties and advantage in solving systems of nonlinear operator equations. We search for parameters approximating the functional equations that minimize the squared residual functions.

As we specified above, in the system of operator equations given by (11), (12), (22), (26)-(30), and (31)-(32), there are thirteen unknown classes of functions $\{\pi, h, h_\pi, h_{\pi'}, h_K, h_{\pi'\pi'}, h_{\pi'\pi}, \lambda, \lambda_\pi, \lambda_{\pi'}, \lambda_K, \lambda_{\pi'\pi'}, \lambda_{\pi'\pi}\}$. Since we assume that the shocks are discrete, $z \in Z = \{z_1, z_2, \dots, z_J\}$ and $J > 1$, we define the following family of policy and distribution functions, and their derivatives $\{h^i(k), \lambda^i(k), h_\pi^i(k), h_{\pi'}^i(k), h_K^i(k), h_{\pi'\pi'}^i(k), h_{\pi'\pi}^i(k), \lambda_\pi^i(k), \lambda_{\pi'}^i(k), \lambda_K^i(k), \lambda_{\pi'\pi'}^i(k), \lambda_{\pi'\pi}^i(k)\}_{i=1}^J$, for each shock value z_1, z_2, \dots, z_J . We interpret the policy function h^i as the next-period capital function of an agent who was hit by a shock level z_i . Analogously, the distribution function λ^i is the distribution of agents with the shock z_i , etc. Similarly, we assign the Euler and distribution function operators to every shock level, \mathcal{F}^i and \mathcal{L}^i , respectively. We approximate all unknown functions by the orthogonal Chebyshev polynomial base $\{T_i(x)\}_{i=0}^\infty$ defined for $x \in [-1, 1]$.

As we have to define our approximation on a finite interval, we set the highest capital level to a value \widehat{k} , greater than the endogenous upper bound on the stationary distribution. Let the interval of approximation be $[\underline{k}, \widehat{k}]$ and the degrees of approximation for $\{h^i(k), h_\pi^i(k), h_{\pi'}^i(k), h_K^i(k), h_{\pi'\pi'}^i(k), h_{\pi'\pi}^i(k), \lambda^i(k), \lambda_\pi^i(k), \lambda_{\pi'}^i(k), \lambda_K^i(k), \lambda_{\pi'\pi'}^i(k), \lambda_{\pi'\pi}^i(k)\}$ be $M, M_\pi, M_{\pi'}, M_K, M_{\pi'\pi'}, M_{\pi'\pi}, N, N_\pi, N_{\pi'}, N_K, N_{\pi'\pi'}, N_{\pi'\pi}, P \geq 2$, respectively.²⁹

Thus, we obtain

$$\begin{aligned}\widehat{h}_m^i(k; a_m^i) &\equiv \sum_{j=1}^{M_m} a_{m,j}^i \phi_j(k), \\ \widehat{\lambda}_m^i(k; b_m^i) &\equiv \sum_{j=1}^{N_m} b_{m,j}^i \phi_j(k),\end{aligned}$$

²⁸For an excellent survey and description of these methods see Chapter 11 in Judd (1998).

²⁹The details on Chebyshev polynomials can be found in Judd (1992), Judd (1998) or in any book on numerical mathematics. The linear transformation $\xi : [\underline{k}, \widehat{k}] \rightarrow [-1, 1]$ is necessary if we want to use the Chebyshev polynomials on the proper domain. It is straightforward to show that $\xi(k) = 2(k - \underline{k}) / (\widehat{k} - \underline{k}) - 1$.

with $i \in \{1, 2, \dots, J\}$ and $m \in \{\emptyset, \pi, \pi', K, \pi'\pi', \pi'\pi\}$, and

$$\widehat{\pi}(x; \mathbf{c}) \equiv \sum_{j=1}^P c_j \phi_j(x),$$

for any $k \in [\underline{k}, \overline{k}]$ and $x \in [r\underline{k} + w\underline{z}, r\overline{k} + w\overline{z}]$, where $\phi_j(k) \equiv T_{j-1}(\xi(k))$, and a 's, b 's, and c 's are the unknown parameters.

Now we have to define *residual functions* as approximations to the original operator functions (11), (12), (22), (26)-(30), and (31)-(32). Substituting the above approximations for the unknown functions,

$$R^{\tilde{\mathbf{L}}}(x; \mathbf{p}) = \sum_{z_j} \left[\tilde{\mathbf{L}}_\pi(\widehat{\mathbf{h}}, \widehat{\Lambda}, \widehat{\pi}) + \frac{d}{dx} \tilde{\mathbf{L}}'_\pi(\widehat{\mathbf{h}}, \widehat{\Lambda}, \widehat{\pi}) \right], \quad (44)$$

$$R_m^{\mathcal{F}^i}(k; \mathbf{p}) = \mathcal{F}_m^i(\widehat{\mathbf{h}}, \widehat{\pi}), \quad (45)$$

$$R_m^{\mathcal{L}^i}(k; \mathbf{p}) = \mathcal{L}_m^i(\widehat{\mathbf{h}}, \widehat{\Lambda}, \widehat{\pi}), \quad (46)$$

with $i = 1, \dots, J$ and $m \in \{\emptyset, \pi, \pi', K, \pi'\pi', \pi'\pi\}$ where

$$\begin{aligned} \mathbf{p} &\equiv (\mathbf{a}, \mathbf{a}_\pi, \mathbf{a}_{\pi'}, \mathbf{a}_K, \mathbf{a}_{\pi'\pi'}, \mathbf{a}_{\pi'\pi}, \mathbf{b}, \mathbf{b}_\pi, \mathbf{b}_{\pi'}, \mathbf{b}_K, \mathbf{b}_{\pi'\pi'}, \mathbf{b}_{\pi'\pi}, \mathbf{c}), \\ \mathbf{a}_m &\equiv (a_m^1, a_m^2, \dots, a_m^J), \\ \mathbf{b}_m &\equiv (b_m^1, b_m^2, \dots, b_m^J), \end{aligned}$$

and \mathbf{p} is of a size $S = J \times (\sum_m (M_m + N_m)) + P$,

$$\begin{aligned} \widehat{\mathbf{h}} &\equiv (\widehat{h}, \widehat{h}_\pi, \widehat{h}_{\pi'}, \widehat{h}_K, \widehat{h}_{\pi'\pi'}, \widehat{h}_{\pi'\pi}), \\ \widehat{\mathbf{h}}_m &\equiv (\widehat{h}_m^1, \dots, \widehat{h}_m^J), \\ \widehat{\Lambda} &\equiv (\widehat{\lambda}, \widehat{\lambda}_\pi, \widehat{\lambda}_{\pi'}, \widehat{\lambda}_K, \widehat{\lambda}_{\pi'\pi'}, \widehat{\lambda}_{\pi'\pi}), \\ \widehat{\lambda}_m &\equiv (\widehat{\lambda}_m^1, \dots, \widehat{\lambda}_m^J), \end{aligned}$$

for any $i = 1, \dots, J$.

The least squares projection method searches for a vector of parameters \mathbf{p} that minimizes the sum of weighted residuals,

$$\sum_{i=1}^J \int_{\underline{k}}^{\overline{k}} \left\{ \sum_m \left([R_m^{\mathcal{F}^i}(k; \mathbf{p})]^2 + [R_m^{\mathcal{L}^i}(k; \mathbf{p})]^2 \right) \right\} w(k) dk + \int_{\underline{k}}^{\overline{k}} [R^{\tilde{\mathbf{L}}}(x(k); \mathbf{p})]^2 w(k) dk,$$

with the weighting function given by $w(k) \equiv \left(1 - \left(2 \frac{k-\underline{k}}{\overline{k}-\underline{k}} \right)^2 \right)^{-1/2}$ and $i = 1, \dots, J$. After approximating the integrals by the Gauss-Chebyshev quadrature, we obtain a minimization problem

$$\min_{\mathbf{p} \in \mathbb{R}^S} \sum_{\underline{k}} \left\{ \sum_{i=1}^J \sum_m \left([R_m^{\mathcal{F}^i}(k; \mathbf{p})]^2 + [R_m^{\mathcal{L}^i}(k; \mathbf{p})]^2 \right) + [R^{\tilde{\mathbf{L}}}(x(k); \mathbf{p})]^2 \right\}, \quad (47)$$

with \check{k} 's being the zeros of the polynomial ϕ of a degree greater than the biggest degree of the polynomial approximations, i.e. $\max\{M, M_\pi, M_{\pi'}, M_K, M_{\pi'\pi'}, M_{\pi'\pi}, N, N_\pi, N_{\pi'}, N_K, N_{\pi'\pi'}, N_{\pi'\pi}, P\}$.

Since the least squares projection method sets up an optimization problem we can use standard methods of numerical optimization, e.g. the Gauss-Newton or the Levenberg-Marquardt methods. Again, the discussion of these methods is not the aim of our paper. However, we found that these traditional methods did not work in our high-dimensional problem mainly due to possible multiple local solutions. We tried several other methods (simulated annealing or genetic algorithm with quantization, for example) and finally succeeded with a genetic algorithm with multiple populations and local search. The used degrees of polynomial approximation for the optimal individual policy functions h , distribution functions, λ , the related sensitivity functions h_π , λ_π , and the optimal government policy function, π , were 4, 12, 3, 3, and 4, respectively. The residuals of the related functional equations were of the order 10^{-3} or 10^{-4} with the exception of h_π which was of the order 10^{-2} .

References

- Aiyagari, R. S. (1995). Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting. *Journal of Political Economy* 103, 1158–1175.
- Albanesi, S. and C. Sleet (2003). Dynamic optimal taxation with private information. *Mimeo, University of Iowa*.
- Auerbach, A. J. and L. J. Kotlikoff (1987). *Dynamic Fiscal Policy*. New York, N. Y.: Cambridge University Press.
- Bohacek, R. and M. Kejak (2002). Projection methods for economies with heterogeneous agents. *CERGE-EI Working Paper*.
- Bohacek, R. and M. Kejak (2004). On the optimal tax schedule. *CERGE-EI Working Paper*.
- Chamley, C. (1986). Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica* 54(3), 607–622.
- Conesa, J. C., S. Kitao, and D. Krueger (2009). Taxing capital? not a bad idea after all! *American Economic Review* 99, 25–48.
- Conesa, J. C. and D. Krueger (2006). On the optimal progressivity of the income tax code. *Journal of Monetary Economics* 53, 1425–1450.
- Diaz-Jimenez, J., V. Quadrini, and J. V. Rios-Rull (1997, Spring). Dimensions of inequality: Facts on the u.s. distribution of earnings, income, and wealth. *Federal Reserve Bank of Minneapolis Quarterly Review*.
- Domeij, D. and J. Heathcote (2004). On the distributional effects of reducing capital taxes. *International Economic Review* 45(2), 523–554.
- Garcia-Mila, T., A. Marcet, and E. Ventura (1995). Supply side interventions and redistributions. *Working Paper UPF* 115.
- Golosov, M., N. R. Kocherlakota, and O. Tsyvinski (2003). Optimal indirect and capital taxation. *Review of Economic Studies* 70, 569–587.
- Gouveia, M. and R. P. Strauss (1994). Effective federal individual income tax functions: An exploratory empirical analysis. *National Tax Journal* (47), 317–339.
- Heaton, J. and D. Lucas (1996). Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy* 104(3), 443–487.
- Judd, K. L. (1985). Redistributive taxation in a simple perfect foresight model. *Journal of Public Economics* 28, 59–83.
- Judd, K. L. (1992). Projection methods for solving aggregate growth models. *Journal of Economic Theory* 58, 410–452.

- Judd, K. L. (1998). *Numerical Methods in Economics*. Cambridge, MA: MIT Press.
- Klein, P. and J.-V. Rios-Rull (2004). Time-consistent optimal fiscal policy. *International Economic Review* 44(4), 1217–1245.
- Kocherlakota, N. R. (2005). Zero expected wealth taxes: A mirrless approach to dynamic optimal taxation. *Econometrica* 73(5), 1587–1621.
- Kydland, F. and E. C. Prescott (1977). Rules rather than discretion: the inconsistency of optimal plans. *Journal of Political Economy* 84, 473–491.
- Lucas, Jr., R. E. (1990). Supply-side economics: An analytical review. *Oxford Economic Papers* (42), 293–316.
- Mirrlees, J. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies* 38, 175–208.
- Mirrlees, J. (1976). Optimal tax theory: A synthesis. *Journal of Public Economics* 6, 327–358.
- Prescott, E. C. (1986). Theory ahead of business cycle measurement. *Carnegie-Rochester Conference Series on Public Policy* 25, 11–44.
- Quah, D. (2003). One third of the world’s growth and inequality. In T. Eicher and S. Turnovsky (Eds.), *Growth and Inequality: Issues and Policy Implications*, Cambridge. MIT Press.
- Ramsey, F. P. (1927). A contribution to the theory of taxation. *The Economic Journal* 37, 47–61.
- Saez, E. (2002). Optimal progressive capital income taxes in the infinite horizon model. *NBER Working Paper* (9046).
- Stokey, N. L., R. E. Lucas, Jr., and E. C. Prescott (1989). *Recursive Methods in Economic Dynamics*. Cambridge: Harvard University Press.
- Storesletten, K., C. Telmer, and A. Yaron (2007). Asset pricing with idiosyncratic risk and overlapping generations. *Review of Economic Dynamics* 10, 519–548.
- Tauchen, G. and R. Hussey (1991). Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica* 59, 371–396.
- Ventura, G. (1999). Flat tax reform: A quantitative exploration. *Journal of Economic Dynamics and Control* 23, 1425–1458.

Steady State Results			
Average	Tax Schedule		
	Progressive	Flat Tax	Optimal
Capital	2.54	3.29	3.80
Output	1.40	1.54	1.62
Consumption	0.86	0.90	0.91
Median Wealth	2.45	2.55	3.60
Median/Mean Wealth Ratio	0.97	0.77	0.95
Interest Rate (%)	9.87	6.77	5.33
Wage	0.894	0.983	1.035
Gini Total Income	0.22	0.31	0.27
Gini Total After Tax Income	0.21	0.32	0.28
Constrained agents (%)	1.16	1.88	1.42
Steady State Welfare Gains from the Optimal Tax Schedule			
Welfare Gain (%)	4.39	0.84	—
Efficiency Gain (%)	5.67	1.27	—
Distributional Gain (%)	-0.57	2.86	—

Notes: Wealth in terms of accumulated capital stock. Constrained agents is a fraction of agents whose wealth equals the exogenous lower bound on capital. Welfare measured as consumption level corresponding to the average utility. Average Welfare Gain measured as percentage of the average consumption each agent would have to receive in the progressive and the flat-rate tax steady state so that the average welfare equals that in the optimal tax steady state. Average Efficiency and Distributional Gains defined in the text.

Table 1: Steady State Results.

Steady State Distribution of Resources					
Tax Schedule	Quintile				
	1st	2nd	3rd	4th	5th
Average Consumption Level					
Optimal	0.6673	0.8272	0.9171	1.0076	1.1489
Flat	0.6990	0.8189	0.8904	0.9685	1.1213
Progressive	0.7120	0.8113	0.8642	0.9237	1.0148
Average Asset Level					
Optimal	0.9166	2.4051	3.6545	4.9934	7.0381
Flat	0.6336	1.7708	2.8488	4.1933	6.9921
Progressive	0.7705	1.8195	2.5666	3.2509	4.2725
Average Investment/Income Ratio					
Optimal	0.1222	0.2148	0.2935	0.3692	0.4643
Flat	0.0869	0.1630	0.2363	0.3197	0.4624
Progressive	0.1058	0.1774	0.2252	0.2592	0.2964
Average Income Level					
Optimal	1.0141	1.1520	1.2353	1.3221	1.4620
Flat	0.9644	1.1009	1.1881	1.2883	1.4884
Progressive	0.9158	1.0565	1.1410	1.2337	1.3739
Average After-Tax Income Level					
Optimal	0.6996	0.8341	0.9142	0.9964	1.1240
Flat	0.7194	0.8213	0.8863	0.9611	1.1103
Progressive	0.7319	0.8168	0.8644	0.9184	0.9948
Average Tax Contribution Share					
Optimal	0.1944	0.1957	0.1995	0.2014	0.2090
Flat	0.1588	0.1824	0.1982	0.2133	0.2473
Progressive	0.1311	0.1702	0.1988	0.2258	0.2741

Table 2: Distribution of Resources in Steady States.

Tax Reform		
Transition to the Optimal Tax Schedule Steady State		
	Transition From Steady State	
	Progressive	Flat Tax
Average		
Welfare Gain (%)	3.44	-1.81
Efficiency Gain (%)	4.14	-1.35
Distributional Gain (%)	-7.29	3.25
Median		
Welfare Gain (%)	3.86	-1.97
Political Support		
% of Population	72.9	33.2

Notes: Average Welfare Gain from transition is a percentage of the average consumption each agent would have to receive in each period of transition so that the average welfare from transition equals in expected present discounted value that of the initial steady state. Average Efficiency and Distributional Gains from transition defined in the text.

Table 3: Transition to the Optimal Tax Schedule Steady State.

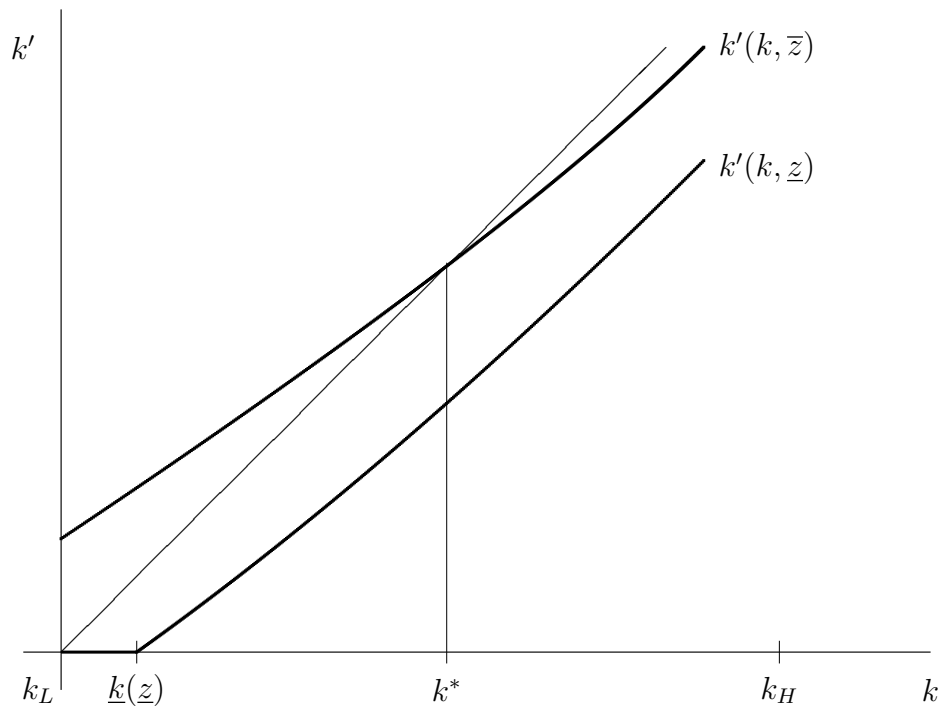


Figure 1: Policy functions for the next period capital stock. An example with two productivity shocks $\bar{z} > \underline{z}$. There is an exogenous lower bound k_L and an endogenous upper bound $k^* < k_H$. The stationary distribution has a unique ergodic set $E = [k_L, k^*]$. Agents with shock \underline{z} and capital stock $k < \underline{k}(\underline{z})$ are borrowing constrained.

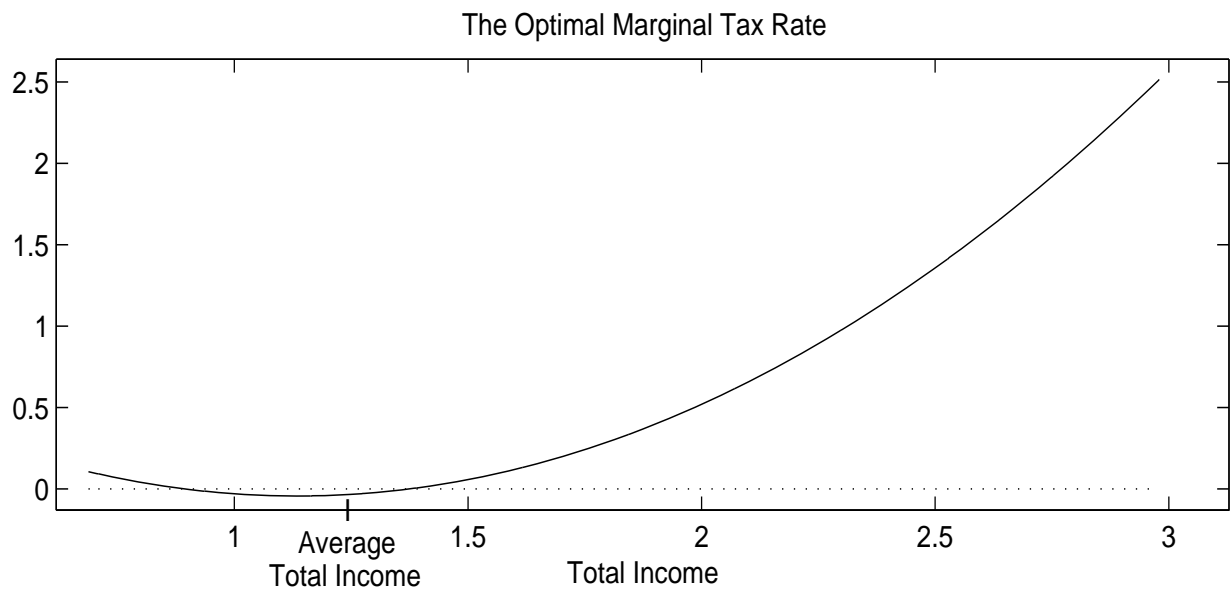
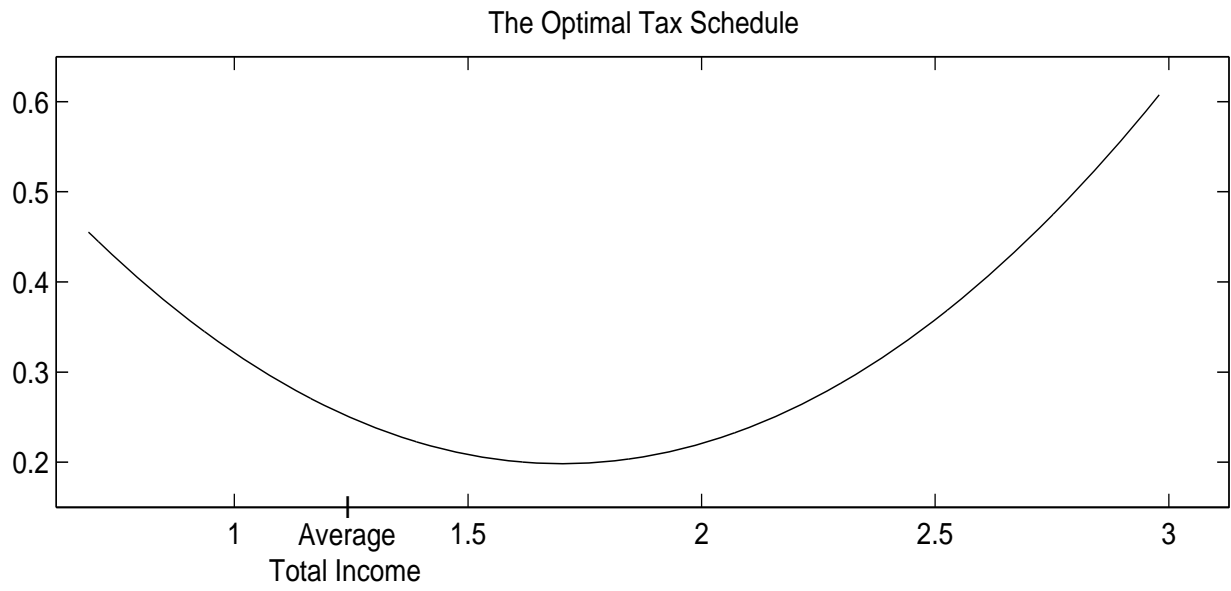


Figure 2: The Optimal Tax Schedule and The Optimal Marginal Tax.

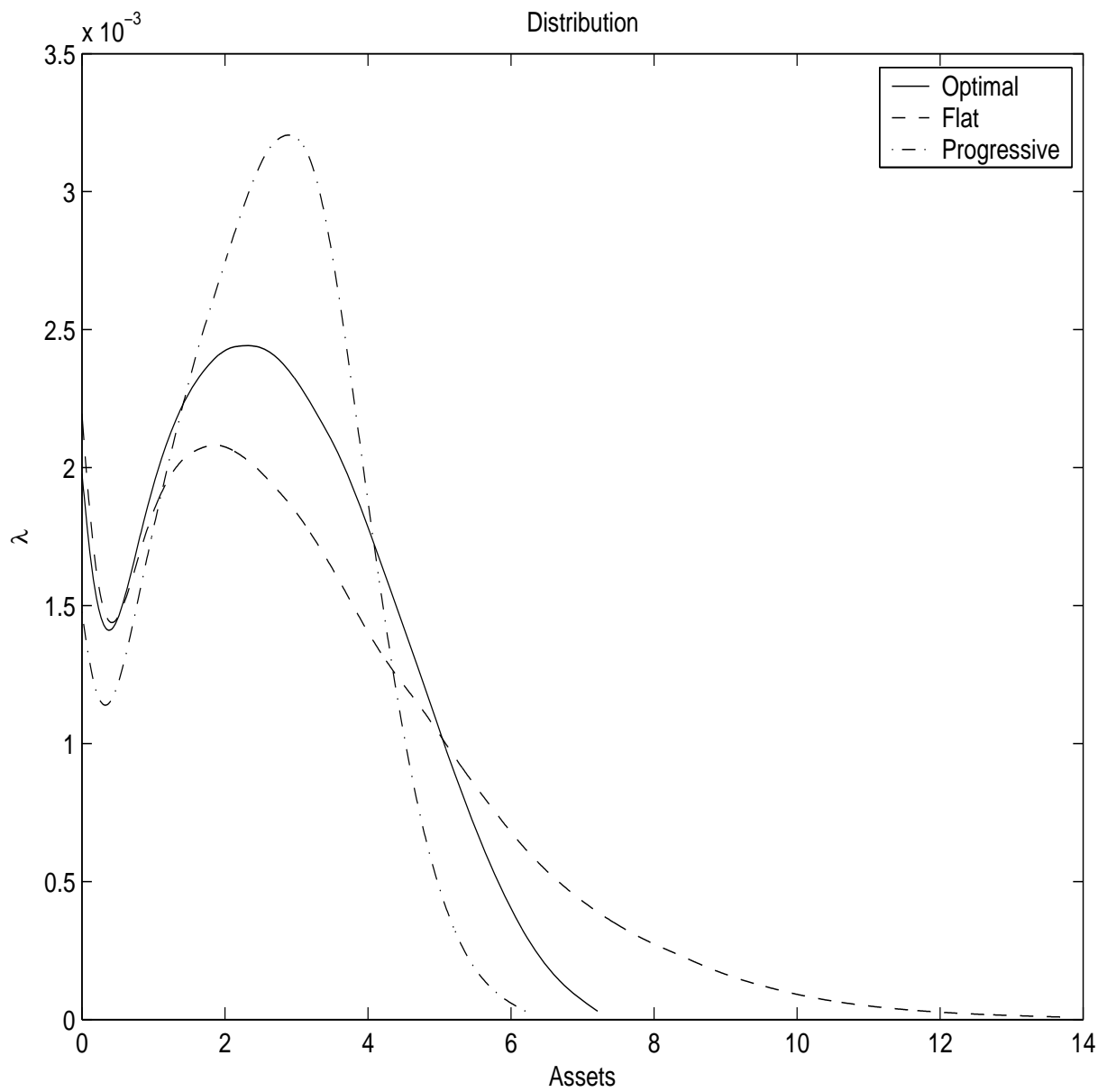


Figure 3: Stationary Distribution of Agents Over Assets in the Optimal, Flat, and Progressive Tax Schedule Steady States.

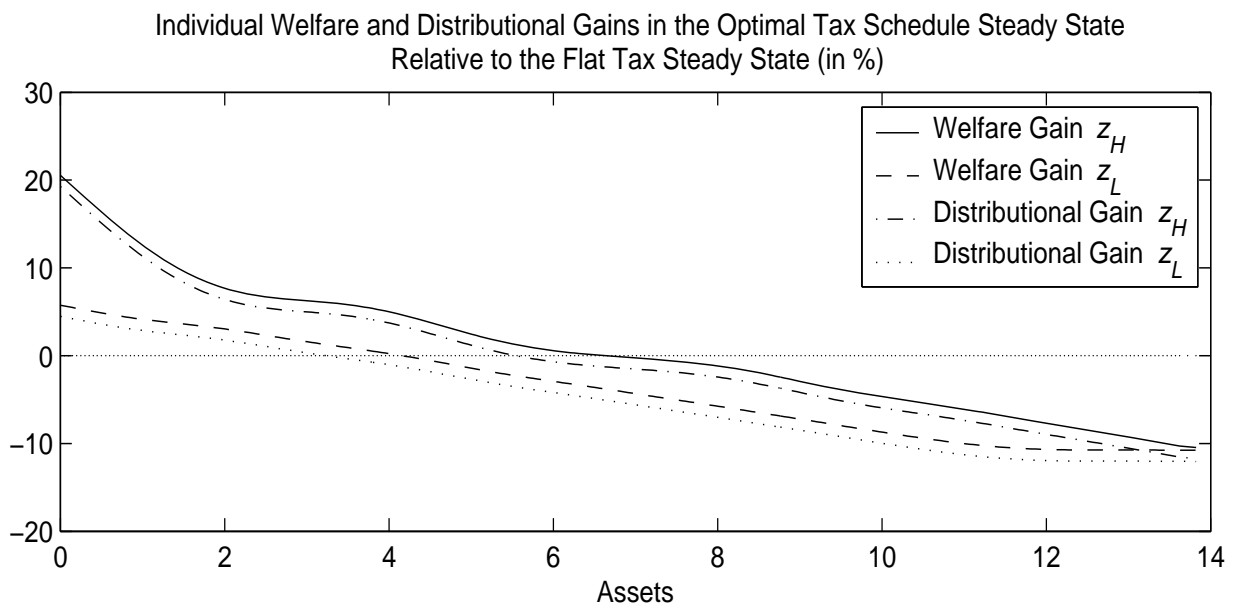
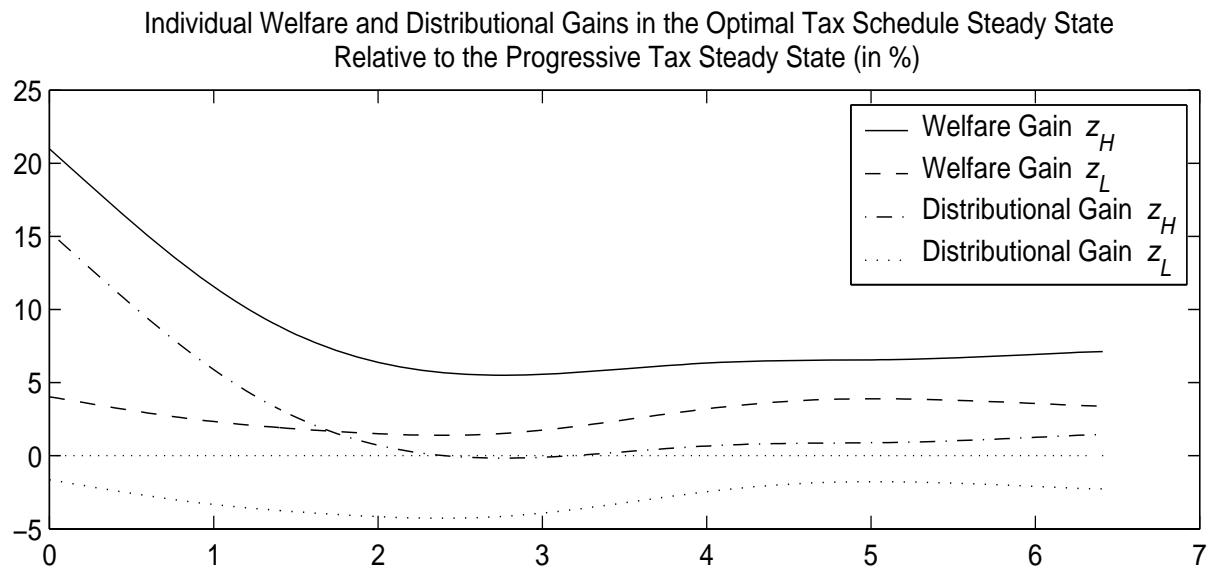


Figure 4: Individual Gains in the Optimal Tax Schedule Steady State.

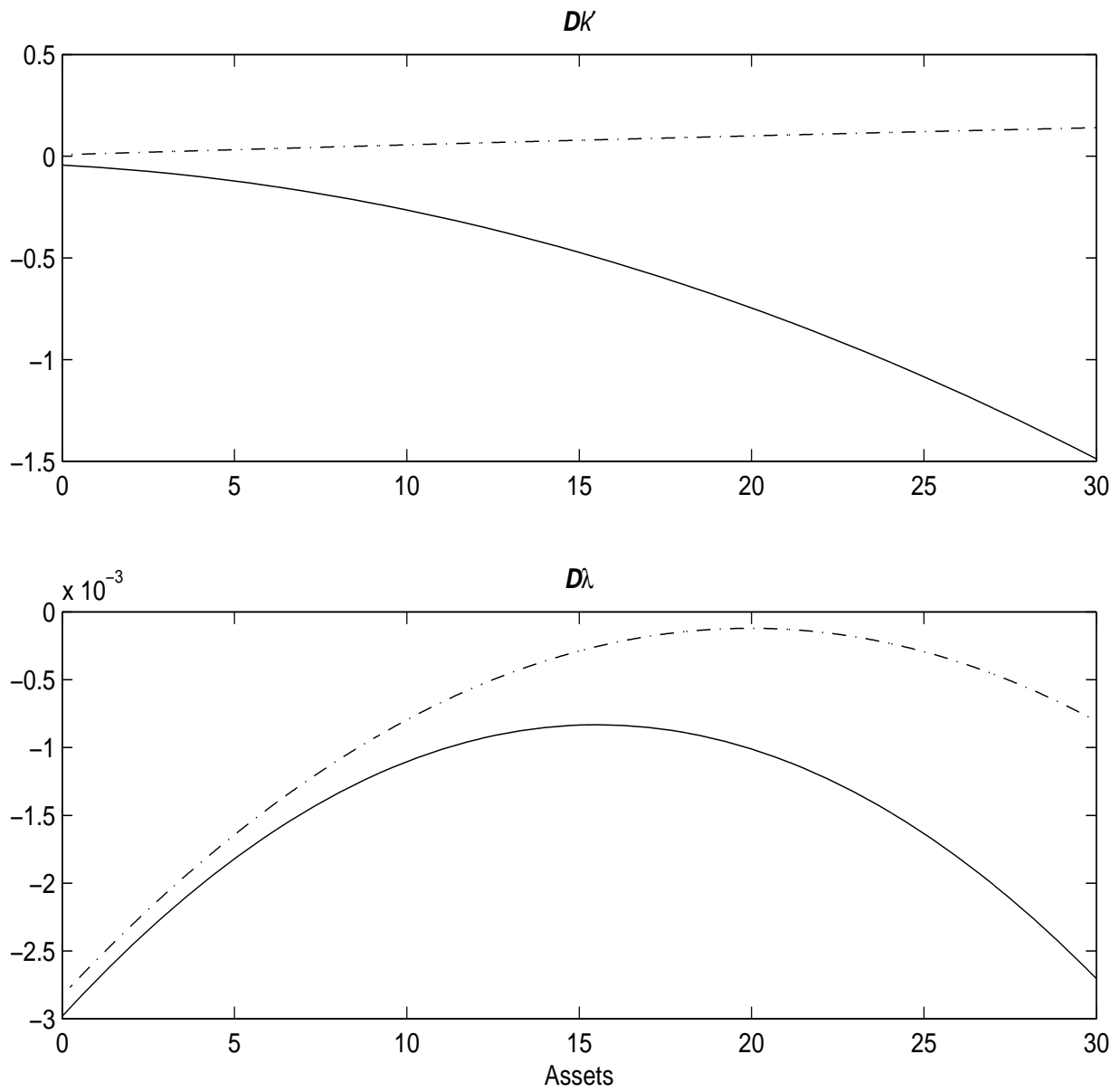


Figure 5: Savings and Distribution Sensitivity Functions for High (-) and Low (-.-) Productivity Shocks.

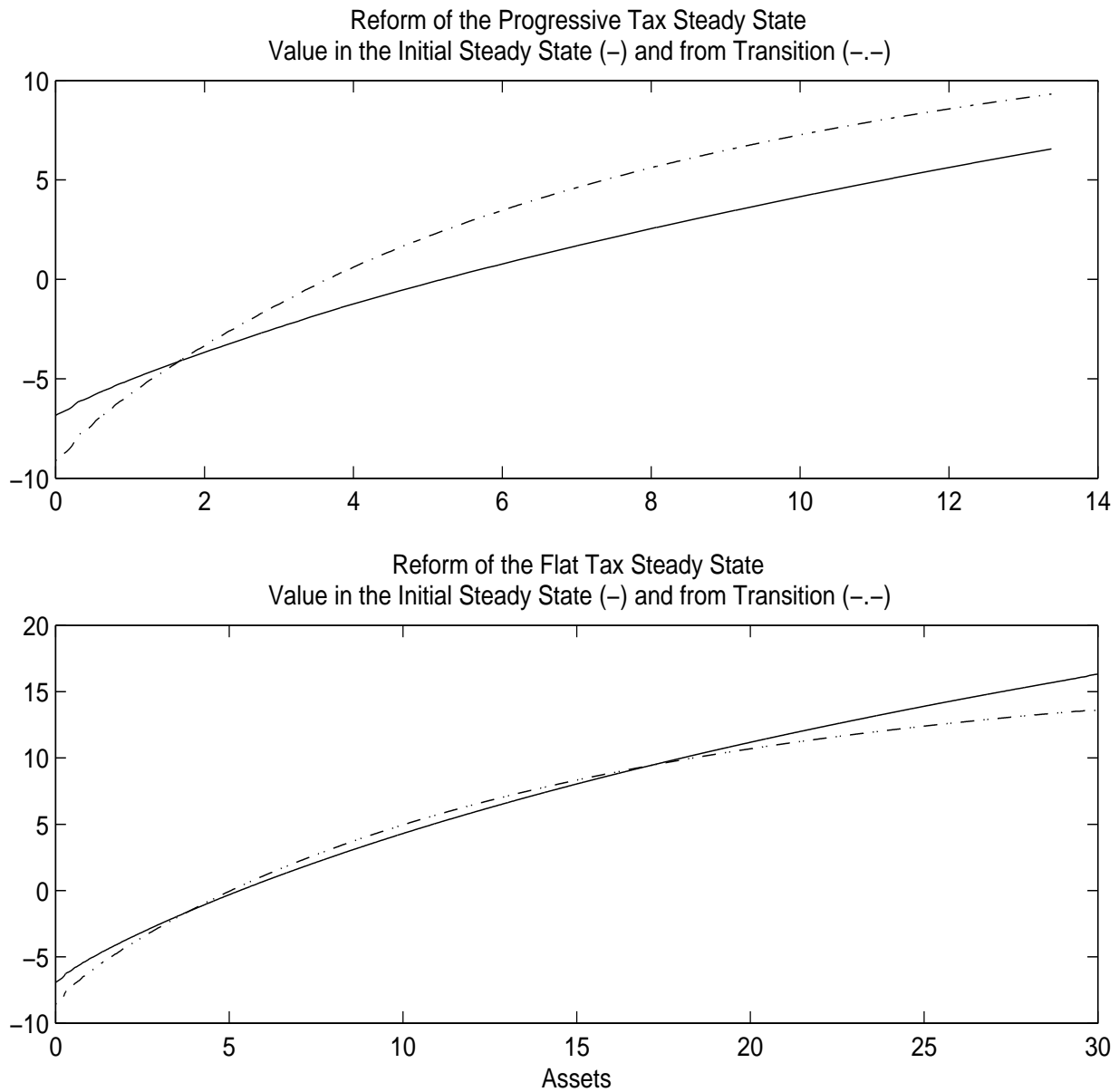


Figure 6: Welfare Gains from a Tax Reform. Transition from the Progressive and the Flat Tax Steady State to the Optimal Tax Schedule Steady State.

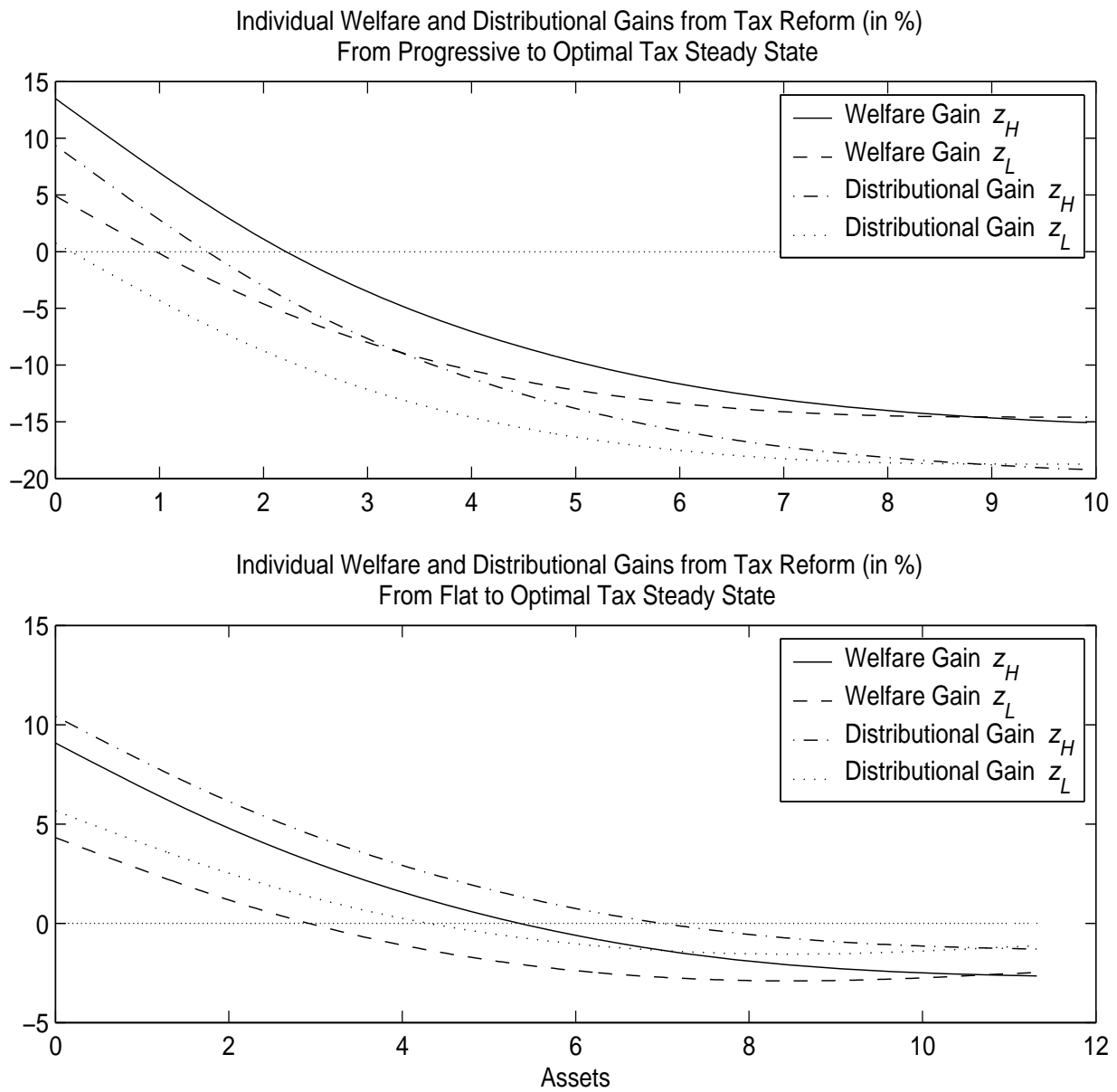


Figure 7: Individual Gains from a Tax Reform.

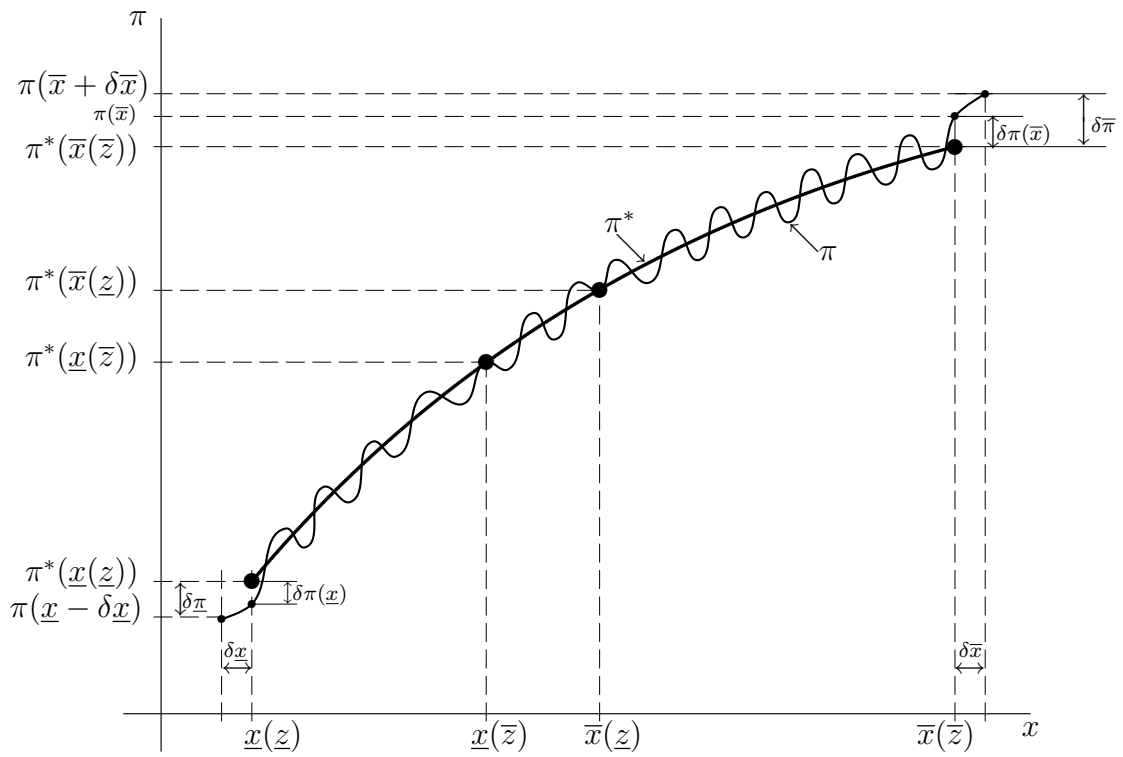


Figure 8: Variations Around Optimal Tax Schedule