Optimal Government Policies in Models with Heterogeneous Agents

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

August 9, 2016

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, the resulting optimal distribution of agents in the steady state and, therefore, equilibrium prices. We illustrate the methodology on a Ramsey problem with heterogeneous agents, finding the optimal limiting tax on total income.

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

*Contact: CERGE-EI, Politickych 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 the editor and the referees, Jim Costain, Mark Gertler, Max Gillman, Christian Hellwig, Boyan Jovanovic, Marek Kapička, Dirk Krueger, Per Krusell, Michal Pakos, Josep Pijoan-Mas, Thomas Sargent, Jakub Steiner, 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 Stern School of Business. We are especially grateful to Michele Boldrin, Tim Kehoe and Stan Zin for their support and advice. Anton Tyutin, Jozef Zubrický and Tomáś Leśko provided excellent research assistance. The financial support of the Czech Science Foundation project No. P402/12/G097 (DYME Dynamic Models in Economics) is gratefully acknowledged. 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 government problem in a general equilibrium, Bewley type economy with heterogenous 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 condition for 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 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.

Our main contribution is the derived Euler-Lagrange equation for the government problem and the operator formulation of the individual agent's Euler equation and of the endogenous stationary distribution. 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. The first and second order conditions, in the form of the Euler-Lagrange equation and a modified Legendre condition, respectively, represent the necessary and sufficient conditions for concavity and a unique maximum attained by the government policy function. Additionally, the first-order transversality conditions for the boundary agents allow for a qualitative analysis of the shape of the optimal government policy function. There are two restrictions we must impose on the solution: the government cannot use taxes that are state-contingent (to preserve incomplete markets and heterogeneity in the economy) and, because of the variational approach, the tax function belongs to the class of continuously differentiable functions. We do not impose additional assumptions on the shape of the government policy function. The optimal policy is derived from the first order and envelope conditions and from the stationarity of the endogenous distribution in the steady state. To our knowledge, this paper is the first one that provides a solution method for this kind of optimal government problem in an economy with heterogeneous agents. The main innovation of our approach lies in its consideration of the general equilibrium effects arising from the endogeneity of the distribution function attained by the optimal policy function.

We formulate the government problem as a modified Golden Rule. That is, we solve for an optimal government policy under an assumption that the economy converges to a steady state. The optimal limiting government policy is a long-run optimal outcome that takes into account intertemporal discounting and the convergence to the steady state. In a related paper, Davila, Hong, Krusell, and Rios-Rull (2012) consider a social planner with an ability to change the aggregate capital stock in the steady state and focus on pecuniary externality arising from general equilibrium effects. They derive a functional first-order condition that is a necessary condition for the constrained-efficiency planning problem and use variations approach to solve it. Compared to Davila, Hong, Krusell, and Rios-Rull (2012), the contribution of our paper is in the formulation of the Euler-Lagrange conditions for an implementable government policy.

We illustrate this methodology on a Ramsey problem, solving for the optimal limiting tax schedule on total income that maximizes aggregate welfare in a steady state of a standard neoclassical, dynamic general equilibrium model with heterogeneous agents and incomplete markets. We compare the steady state aggregate levels, welfare, efficiency and distribution of resources associated with this optimal limiting tax schedule to a simulated steady state of an economy with a progressive tax schedule approximated from the U.S. data based on Heathcote, Storesletten, and Violante (2014), to an optimal progressive tax reform, and to a steady state resulting from a standard flat-tax reform.

The optimal limiting tax schedule we find is a U-shaped function, taxing the lowest income at 37%, decreasing to a minimum of 25% and rising to 48% at the highest level of total income. The marginal tax rate is also U-shaped, negative at low incomes and increasing to positive levels at higher incomes. Compared to the approximated progressive tax steady state, average welfare increases by 1.02% while relative to the flat-tax steady state, welfare goes up by 1.14%. Similar to Davila, Hong, Krusell, and Rios-Rull (2012), we find that the aggregate capital stock corresponding to the optimal limiting tax policy is higher than in the other steady states in order to increase wages of poor agents with labor-intensive incomes. An important implication is that optimal limiting taxation leads to much higher capital accumulation than standard precautionary savings models including, in particular, the flat tax-reform.[1]

The efficiency and distributional effects of the optimal tax are the main mechanisms behind these large changes. Efficiency gains arise from 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 negative marginal tax rate at low income levels induces agents to save more to escape relative poverty and to secure better insurance against idiosyncratic risk, while the higher tax rate on high incomes provides

[1] We also compute transitions to the steady state associated with the optimal limiting tax schedule and find that only around one fifth of the population is better off from the reform. However, the transition analysis is far from an optimal one. For example, a possibility to issue debt repaid from the efficiency gains in the optimal steady state might provide resources to compensate agents and obtain political support for the reform (Gottardi et al. (2011)).

resources for redistribution. At the aggregate level, these incentives lead to pecuniary externality from general equilibrium effects. Altogether, the implementable optimal limiting tax function not only increases the aggregate capital stock but also significantly reduces inequality, close to what a social planner with an access to the first-best, lump-sum transfers would do. In this way, the optimal limiting tax function helps to resolve the tradeoff between efficiency and equality.

The literature on optimal taxation has provided important insights due to restrictions imposed either on the functional form of the tax function (Diamond and Mirrlees (1971)) or on information available to the government (Mirrlees (1971) and Golosov, Kocherlakota, and Tsyvinski (2003)). In our example, we study a full-information economy without any prior restriction on the shape of the continuously differentiable tax function. In partial equilibrium models, Diamond (1998) shows that optimal marginal taxes are U-shaped if the distribution of skills is single-peaked with a Pareto distribution above the peak. Saez (2001) generalizes these results for preferences with income effects. In a static model calibrated to empirical cross-sectional distribution of labor income and empirical tax rates, he finds that optimal marginal taxes are U-shaped and rise up to 50–70 percent. Diamond and Saez (2011), Mirrlees (1971), Kanbur and Tuomala (1994), and Mankiw et al. (2009) all note that the shape of the marginal tax schedule is very sensitive to the exogenous distribution of skill or income.[2] Compared with the Mirrleesian literature, in our approach it is the tax schedule that seeks to attain a steady state where the endogenous distribution of agents is optimal with respect to the aggregate welfare in the economy.

In general equilibrium models, the tax schedule is usually restricted to a specific, usually linear or monotone, functional forms. Several important papers have analyzed the steady state implications (and transition paths) resulting from an *ad hoc* flat-tax reform or from a removal of double taxation of capital income. Heathcote, Storesletten, and Violante (2014) and Conesa and Krueger (2006) compute the optimal progressivity of the income tax code. Our paper shows that narrowing the analysis to monotone functions may be restrictive with respect to welfare maximization. An important step towards quantitatively studying dynamic optimal taxation has been made in Kapicka (2013) and Golosov, Tsyvinski, and Troshkin (2010). In a related paper, Golosov, Tsyvinski, and Werquin (2014) also use variational approach for their analysis of optimal tax systems. In a partial equilibrium framework with exogenous distribution of agents, they compute Gateaux differentials of local tax perturbations and look for a globally optimal tax function that cannot be locally improved within a restricted class of tax functions. For many realistic parameters the optimal marginal tax rates are also U-shaped. Using a similar variations methods, Davila

[2] In general, the information-constrained optimal marginal tax rates increase in the tail ratio of the skill distribution and in the desired degree of redistribution; they decrease when labor elasticity is high.

et al. (2012) derive a functional first-order condition for the constrained-efficiency planning problem without characterizing the implementation by tax policies.

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 output. There are several reasons why we choose this setup. First, the tax on the total income preserves incomplete markets 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. In this paper, we therefore do not address many important issues related to optimal taxation: the issue of time-consistency, the issue of the optimal taxation of capital income, the distortionary effects of taxation on labor supply, or a more detailed model with life-cycle features. We also abstract from government debt that could be particularly important during a transition analysis. Due to the complexity of our work, we study the simplest utility maximization problem on the consumption-investment margin. In our future research we will apply our methodology to these issues as well as to other than utilitarian welfare criteria.

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 formulates the limiting Ramsey problem in the calculus of variations. The first-order necessary and the second-order sufficient conditions for the optimal government policy expressed in the form of a generalized Euler-Lagrange equations and Legendre condition, respectively, with related analytical results are described in Section 4. Section 5 illustrates our approach by an example of the optimal income tax schedule with numerical simulation presented in Section 6. Section 7 concludes. Appendices contain proofs, analytical results, and details on the simulated optimal tax schedule.

2 The Economy

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

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

where $U : \mathbb{R}_+ \to \mathbb{R}$ 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, $k_t \in B = [\underline{k}, \overline{k}]$, and by a discrete, exogenous labor productivity shock $z_t \in Z = \{\underline{z}, \ldots, \overline{z}\}$. We assume that there is a borrowing constraint that prevents the individual savings from being negative. The lower bound could be motivated by solvency constraints or by an explicit borrowing constraint. As is standard, the upper bound $\overline{k}$ is set very high and verified not to be binding in equilibrium. 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). The labor productivity shock is independent across agents and we preserve the heterogeneity in the economy by assuming incomplete markets: Agents do not have access to state-contingent contracts but can only accumulate the risk-free capital stock.

In each period, agents inelastically 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

$$r_t = F_K(K_t, L_t) - \delta \quad \text{and} \quad w_t = F_L(K_t, L_t), \tag{2}$$

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

Finally, there is a government that finances its expenditures by taxation. In order to preserve incomplete markets and, therefore, heterogeneity in the economy, we impose that the government cannot use state-contingent taxes. We assume that the government cannot issue debt and is fully committed to a sequence of tax functions $\{\pi_t\}_{t=0}^{\infty}$ to finance its expenditures equal to a fraction of output, g, not returned back to the agents.[3] The tax schedule policy 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 $x : Z \times B \to \mathbb{R}_+$ and $x_z, x_k > 0$. Thus in each period, the policy schedule is a twice continuously differentiable function $\pi_t : \mathbb{R}_+ \to \mathbb{R}$, 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 $\pi_t(x_t)$ and is left with an after-tax income $y_t - \pi_t(x_t)$.[4] The individual budget constraint in each period is then

$$c_t + k_{t+1} \leq r_t k_t + w_t z_t - \pi_t(x_t) + k_t.$$

The economy's aggregate state is characterized by the sequences of government policies

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

[4]In Section 5, we compute an economy with a tax on total income, i.e. when $x = y$.

$\{\pi_t\}_{t=0}^{\infty}$ and the distribution of agents over capital and productivity shock in each period, $\{\lambda_t\}_{t=0}^{\infty}$. 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. Agents have rational expectations and take prices as given by equation (2). In order to determine prices, agents also have to know the evolution of the distribution function from an initial distribution λ_0, for each sequence of government policies $\{\pi_t\}_{t=0}^{\infty}$.

The objective of the government is to choose a sequence of $\{\pi_t\}_{t=0}^{\infty}$ to maximize

$$\sum_{t=0}^{\infty} \beta^t \sum_z \int u(c_t(z,k))\lambda_t(z,k)dk, \tag{3}$$

subject to agents' optimal allocations in each period, equilibrium prices determined by the aggregate capital stock and labor,

$$K_t = \sum_z \int k\lambda_t(z,k)dk, \quad \text{and} \quad L_t = \int \sum_z z\lambda_t(z,k)dk, \tag{4}$$

the government budget constraint

$$g\, F(K_t, L_t) = \sum_z \int \pi_t(x_t(z,k))\, \lambda_t(z,k)\, dk; \tag{5}$$

and a law of motion of the distribution,

$$\lambda_{t+1}(z', B') = \sum_z \int_{\{(z,k)\in Z\times B:\, h_t(z,k)\in B'\}} Q(z,z')\, \lambda_t(z,k)\, dk, \tag{6}$$

given an initial distribution λ_0,

The government problem assigns equal weights to all agents. This utilitarian approach is chosen for two main reasons. We prefer to start the economy from an initial distribution λ_0 where all agents have identical wealth and labor productivity. Second, assigning equal weights to all agents allows us to treat identical agents identically and derive properties for the long-run equilibrium associated with the optimal limiting government policy function.[5]

The goal of our paper is to solve the following stationary Ramsey problem.

Definition 1 (Stationary Ramsey Problem) *A solution to the Stationary Ramsey Problem is a time-invariant limiting government tax policy function $\pi : \mathbb{R}_+ \to \mathbb{R}$ such that $\pi = \lim_{t\to\infty} \pi_t$ maximizes the government problem in (3)-(6).*

Note that our analysis is not a pure steady state utility maximization. The optimal limiting government policy is a long run optimal outcome that takes into account intertemporal

[5]See Davila et al. (2012) for a similar discussion. The initial distribution is arbitrary.

discounting and the convergence to the steady state. In other words, we study a steady state under a modified Golden Rule. That is, we study a steady state of an economy for which the optimal limiting government policy implies a convergence to that steady state.

2.1 Recursive Formulation

Define the value function of each agent as $v : Z \times B \to \mathbb{R}$ and the savings function as $h : Z \times B \to B$. Since the government searches for a limiting optimal policy schedule, it needs to take into account the effect of its policy on equilibrium prices determined by the aggregate capital stock.[6] We make this relationship explicit in policy functions. Given the equilibrium prices and the limiting optimal government policy π, an agent (z, k) solves the following dynamic programming problem

$$v(z,k;K) = \max_{c,h} \left\{ u(c(z,k;K)) + \beta \sum_{z'} v(z', h(z,k;K); K')\, Q(z,z') \right\}, \tag{7}$$

subject to a budget constraint

$$c(z,k;K) + h(z,k;K) \leq y(z,k;K) + k - \pi(x(z,k;K)), \tag{8}$$

with a taxable activity $x(z,k;K)$, total income $y(z,k;K) = r(K)k + w(K)z$, and a borrowing constraint,

$$h(z,k;K) \geq \underline{k}. \tag{9}$$

Definition 2 (Recursive Competitive Equilibrium) *For a given share of government expenditures g and the government policy π on a taxable activity x, a recursive competitive equilibrium is a set of functions (v, c, h), aggregate levels (K, L), prices (r, w), and a probability measure $\lambda : Z \times B \to [0,1]$, such that for given prices and government policies,*

1. *the policy functions solve each agent's optimization problem (7);*
2. *firms maximize profit (2);*
3. *the probability measure evolves according to a law of motion,*

$$\lambda'(z', B') = \sum_{z} \int_{\{(z,k)\in Z\times B:\, h(z,k;K)\in B'\}} Q(z,z')\, \lambda(z,k)\, dk, \tag{10}$$

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

4. *the aggregation conditions hold,*

$$K = \sum_{z} \int k\, \lambda(z,k)\, dk \quad \text{and} \quad L = \sum_{z} \int z\, \lambda(z,k)\, dk; \tag{11}$$

[6] With exogenous labor supply, the aggregate labor converges deterministically to a constant due to the law of large numbers. In the following exposition, we normalize the aggregate labor supply and write the equilibrium prices as functions of the aggregate capital K only.

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

$$g\,F(K,L) = \sum_z \int \pi(x(z,k;K))\,\lambda(z,k)\,dk; \tag{12}$$

6. *and the allocations are feasible,*

$$\sum_z \int \left[c(z,k;K) + h(z,k;K)\right]\,\lambda(z,k)\,dk + gF(K,L) = F(K,L) + (1-\delta)K. \tag{13}$$

In the recursive formulation of the optimal limiting government policy problem, the government maximizes

$$W(\lambda) = \max_{\pi} \int \sum_z u(c(z,k;K))dk + \beta W(\lambda'), \tag{14}$$

subject to allocations satisfying the conditions in Definition 2.

The steady state of the economy corresponding to the limiting optimal government policy π is characterized by a time-invariant distribution λ. That the optimal limiting government policy π allows for a convergence to the steady state requires a regularity condition on its properties. Denote the interval of individual savings at which an agent with a productivity shock z is borrowing constrained as $[\underline{k}, \underline{k}(z)]$. For future reference also denote $\overline{k}(z)$ as the highest savings by an agent with a current productivity shock z.

Assumption 1 (Regularity Condition) *The government policy function π is such that for each $z \in Z$, the individual savings function $h : Z \times B \to B$ is a strictly increasing function for $k > \underline{k}(z)$ and is constant $h(k,z) = \underline{k}$ for $k \in [\underline{k}, \underline{k}(z)]$.*

A similar condition is generally required for the existence of a unique stationary recursive equilibrium in all models with heterogeneous agents (see Stokey, Lucas, and Prescott (1989)). It implies that the savings function does not display pathological features (for example, that wealthy agents save less than poor agents) and, therefore, that the stationary distribution has a unique ergodic set. We want to make explicit here that this assumption is completely innocuous.[7]

[7] The Regularity Condition guarantees that for all $z \in Z$, the government policy function π is such that for given prices determined by K, there exists an inverse function h^{-1} assigning a current value of capital k to savings h according to $k = h^{-1}(z,h;K,\pi)$. The Regularity Condition is used only in the law of motion for the distribution λ in the operator for the distribution function in equation (16). Davila et al. (2012) make a similar assumption (an increasing savings function h, in Appendix). In their case the savings policy is not distorted by any government policy function.

3 Solution to the Stationary Ramsey Problem as a Calculus of Variations Problem

Since the problem is to find an optimal limiting, welfare maximizing continuous function $\pi \in \mathcal{C}^2(\mathbb{R}_+, \mathbb{R})$, we transform the Stationary Ramsey Problem into an operator form and solve it by the calculus of variations.[8] The calculus of variations is much more suitable for solving our problem with a complicated set of functional constraints and complex boundary conditions than dynamic programming or optimal control methods.

In order to express the stationary recursive competitive equilibrium in this form, we define two operators: on the Euler equation and on the stationary distribution. For a given government policy function π, the Euler equation operator $\mathcal{F}$ is defined on the savings function $h : Z \times B \to B$. The distribution operator $\mathcal{L}$ is defined on the probability measure $\lambda : Z \times B \to [0, 1]$ as well as on the savings function h. We assume that these functions are square integrable functions on some closed domain[9]: $h, \lambda \in L^2(Z \times B)$ 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 $\mathcal{F} : C^1(Z \times B) \subset L^2(Z \times B) \to C^1(Z \times B) \subset L^2(Z \times B)$ is the mapping from a space of continuously differentiable functions into a space of continuously differentiable functions; and the operator $\mathcal{L} : C^1(Z \times B) \times C^1(Z \times B) \to C^1(Z \times B) \subset L^2(Z \times B)$. All functions in the calculus of variations depend on the government policy π and its derivative π_x.

Operator $\mathcal{F}$ on the Euler Equation An individual agent's optimization problem in (7)-(9) is characterized by the Euler equation with an operator[10]

$$\mathcal{F}(h) \equiv u_c(c) - \beta \sum_{z'} u_c(c') \left[1 + y'_k - \pi_x(x')x'_k\right] Q(z, z'), \tag{15}$$

where the next period variables are $c' = y(z', h(z, k; K); K') - \pi(x(z', h(z, k; K)); K') + h(z, k; K) - h(z', h(z, k; K); K')$, $y' = r(K')h(z, k; K) + w(K')z'$, and $x' = x(z', h(z, k; K))$. The term $[1 + y'_k - \pi_x(x')x'_k]$ is the next-period after-tax marginal return to capital where $\pi_x(x')$ is the next period marginal government policy function. Finally, $y'_k = y_k(z', h(z, k; K); K') = r(K')$ is the marginal effect of individual savings on the total income y' next period, and $x'_k = x_k(z', h(z, k; K); K')$ is the marginal effect of individual

[8] Davila, Hong, Krusell, and Rios-Rull (2012) use a similar approach. Mirrlees (1976) also uses the calculus of variations to derive the first-order conditions for the optimal income tax schedule.

[9] In 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.

[10] In the text below, we present only the case of the unconstrained agents (for whom the Euler equation holds with equality and $h(z, k; K) > \underline{k}$). The case of borrowing constrained agents is in Appendix A.

savings on the taxable activity x' next period.[11] The operator equation is $\mathcal{F}(h) = 0$.

Operator $\mathcal{L}$ on the Distribution Function Under the Regularity Condition, the operator $\mathcal{L}$ for the distribution function in equation (10) is

$$\mathcal{L}(\lambda, \lambda', h) \equiv \lambda'(z', k') - \sum_{z} \lambda[z, h^{-1}(z, k')]\, Q(z, z'), \tag{16}$$

for all $(z', k') \in Z \times [h(z, \underline{k}(z)), \overline{k}(z)]$. The operator equation is $\mathcal{L}(\lambda, \lambda', h) = 0$.

Ramsey Problem as Calculus of Variations Given the operator equations $(\mathcal{F}, \mathcal{L})$ and a current distribution λ, the limiting optimal tax policy must maximize the government problem in equation (14),

$$\sum_{z} \int \left\{ u(c(z,k)) + \beta \sum_{z'} u(c(z', h(z,k)))\, Q(z, z') \right\} \lambda(z,k)\, dk, \tag{17}$$

where

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

The next-period prices are determined by

$$K' = \sum_{z'} \int k' \sum_{z} \lambda[z, h^{-1}(z, k'; K')]\, Q(z, z')\, dk' = \sum_{z} \int h(z,k;K')\, \lambda(z,k)\, dk. \tag{18}$$

The objective is to derive the first-order conditions for a limiting solution to the Ramsey Problem in the calculus of variations. Since we are looking for the government policy as a function of the taxable activity x, we reformulate the Ramsey Problem with the taxable activity x as the independent variable. Given the distribution λ and equilibrium prices in the current period, the choice of government policy for the next period, π', directly influences the savings function h and taxable activity in the next period. The social welfare function in equation (17) at new coordinates x is

$$\sum_{z} \int_{\underline{x}(z)}^{\overline{x}(z)} \mathcal{W}[z, x; \pi', \pi'_x, K']\, dx, \tag{19}$$

[11] The Euler equation for an individual agent is standard. When the taxable activity equals total income,

$$\mathcal{F}(h) \equiv u_c(c) - \beta \sum_{z'} u_c(c')\, [1 + r(K) - \pi_x(x') r(K)]\, Q(z, z').$$

where

$$\mathcal{W}[z,x;\pi',\pi'_x,K'] \equiv \{u(c(z,k(z,x);\pi',\pi'_x,K')) \\ +\beta \sum_{z'} u(c(z',h(z,k(z,x);\pi',\pi'_x,K');\pi',\pi'_x,K'))Q(z,z')\}\lambda[z,k(z,x);\pi',\pi'_x,K']k_x(z,x),$$

with $k_x(z,x) = [x_k(z,k(z,x))]^{-1}$ as the inverse function to the marginal effect of individual savings on the taxable activity x_k. To make clear that the government policy function affects equilibrium outcomes in the next period, we add the arguments (π', π'_x, K').

The bounds on taxable activity, $\underline{x}(z)$ and $\overline{x}(z)$, 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 determined by K. Similar arguments apply to the upper bound $\overline{x}(z) = \overline{x}(z, \overline{k}; K)$.[12]

The aggregate capital stock rewritten at the new coordinates is

$$K' \;=\; \sum_z \int_{\underline{x}(z)}^{\overline{x}(z)} \mathcal{K}[z,x;\pi',\pi'_x,K']\,dx, \tag{20}$$

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

Finally, the side condition on the government budget constraint in equation (12) is,

$$\sum_z \int_{\underline{x}(z)}^{\overline{x}(z)} \mathcal{G}[z,x;\pi',\pi'_x,K']\,dx = 0, \tag{21}$$

where

$$\mathcal{G}[z,x;\pi',\pi'_x,K'] \equiv \sum_{z'} [\pi'(x(z',h(z,x(z,k);\pi',\pi'_x,K'))) - gy(z',h(z,x(z,k);\pi',\pi'_x,K');K')] \\ \cdot Q(z,z')\,\lambda[z,k(z,x);\pi',\pi'_x,K']\,k_x(z,x),$$

Definition 3 (Calculus of Variations Stationary Ramsey Problem) *The Stationary Ramsey Problem in the calculus of variations is a generalized isoperimetric problem*

$$\max_{\pi} \left\{ \sum_z \int_{\underline{x}(z)}^{\overline{x}(z)} \mathcal{W}[z,x;\pi',\pi'_x,K']\,dx \right\}, \tag{22}$$

subject to the government budget constraint (21), the definition of the aggregate capital stock (20), the individual policy function h given implicitly by the operator Euler equation $\mathcal{F}(h) = 0$, the law of motion for the distribution function, λ, given implicitly by the operator equation $\mathcal{L}(\lambda, \lambda', h) = 0$, the endogenously determined bounds of taxable activity, $\underline{x}(z)$ and

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

$\overline{x}(z)$, for all values of $z \in Z$, and the free values of the government policy at the extreme lower and upper bounds.

Note that since the upper bound $\overline{k}(z)$ are endogenous, the endpoints $\overline{x}(z)$ are equality constrained. This might be also true for the lower bounds $\underline{k}(z)$ and their endpoints $\underline{x}(z)$.

4 Necessary and Sufficient Conditions for the Stationary Government Policy Function

In this Section we derive the first-order necessary and second order sufficient conditions for the optimal limiting government policy function. In order to derive these conditions in the calculus of variations, we need to specify the derivatives of the functionals $\mathcal{W}$ and $\mathcal{G}$ with respect to marginal changes in government policy, π' and π'_x. For this purpose, we use the concept of generalized derivatives on mappings between two Banach spaces (B-spaces), the Fréchet derivatives. The Fréchet derivative is a generalization of the concept of a derivative on functional and operator spaces (see Luenberger (1969) or Ok (2007)).[13]

Definition 4 (Fréchet Derivative) *Given a nonlinear operator $\mathcal{N}(u)$ on function u, the Fréchet differential $\delta\mathcal{N}(u;\delta h) = \mathcal{N}_u \delta h$ is*

$$\lim_{\|\delta h\| \to 0} \frac{\|\mathcal{N}(u+\delta h) - \mathcal{N}(u) - \mathcal{N}_u \delta h\|}{\|\delta h\|} = 0,$$

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

Define the Lagrange function $\mathbf{L}$ for the Calculus of Variations Ramsey Problem in Definition 3, for each $z \in Z$ as

$$\mathbf{L}(z,x) \;=\; \begin{cases} 0 & \text{for } x \in [\underline{x}(\underline{z}), \underline{x}(z)), \\ \mathcal{W}(z,x) + \mu\mathcal{G}(z,x) & \text{for } x \in [\underline{x}(z), \overline{x}(z)], \\ 0 & \text{for } x \in (\overline{x}(z), \overline{x}(\overline{z})]. \end{cases} \tag{23}$$

Note that the social welfare function is the sum of integrands $\mathcal{W}(z,x) = \mathcal{W}[z,x;\pi',\pi'_x,K']$ and $\mathcal{G}(z,x) = \mathcal{G}[z,x;\pi',\pi'_x,K']$ integrated on intervals $[\underline{x}(z), \overline{x}(z)]$ for each $z \in Z$. The same is true for integrands $\mathbf{L}(z,x)$ in equation (23).

Interestingly, as we show in Theorem 1 below, the relevant Lagrange function that emerges from the solution to the maximization problem is one amended by a term which

[13] The compliance of the Fréchet derivatives (also called the F-derivatives) with the derivations of the first order conditions in the calculus of variations 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 functional equations are recursive. Practically, the Fréchet derivative can be obtained using a weaker concept of the Gateaux derivative $\mathcal{N}_u = \lim_{\varepsilon \to 0} \frac{\mathcal{N}(u+\varepsilon\delta h)}{\varepsilon}$ when the obtained derivative is continuous.

captures the effect of the aggregate capital on social welfare: $\widetilde{\mathbf{L}}(z,x) = \mathbf{L}(z,x) + \Psi\mathcal{K}(z,x)$, where Ψ represents the marginal effect of the aggregate capital on social welfare.[14] In other words, the term $\Psi\mathcal{K}(z,x)$ takes into account the general equilibrium effects on social welfare by individual allocations of agents characterized by (z,x).

Theorem 1 (First Order Necessary Conditions) *Using the modified Lagrange function $\widetilde{\mathbf{L}}$ for the Calculus of Variations Stationary Ramsey Problem in Definition 3,*

$$\widetilde{\mathbf{L}}(z,x) \quad = \quad \begin{cases} 0 & \textit{for } x \in [\underline{x}(\underline{z}), \underline{x}(z)), \\ \mathcal{W}(z,x) + \mu\mathcal{G}(z,x) + \Psi\mathcal{K}(z,x) & \textit{for } x \in [\underline{x}(z), \overline{x}(z)], \\ 0 & \textit{for } x \in (\overline{x}(z), \overline{x}(\overline{z})], \end{cases} \tag{24}$$

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

1. *the Euler-Lagrange condition,*

$$\sum_z \left(\widetilde{\mathbf{L}}_{\pi'}(z,x) - \frac{d}{dx}\widetilde{\mathbf{L}}_{\pi'_x}(z,x) \right) \quad = \quad 0; \tag{25}$$

2. *the transversality condition on the free boundary value $\overline{x}(\overline{z})$,*

$$\left[\widetilde{\mathbf{L}}(\overline{z},x) - \left(\pi_x(x) - \frac{k_x(\overline{z},x)}{\omega_\pi(\overline{z},x)} \right) \widetilde{\mathbf{L}}_{\pi'_x}(\overline{z},x) \right]_{x=\overline{x}(\overline{z})} \quad = \quad 0; \tag{26}$$

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

$$\left[\widetilde{\mathbf{L}}(\underline{z},x) - \left(\pi_x(x) - \frac{k_x(\underline{z},x)}{\omega_\pi(\underline{z},x)} \right) \widetilde{\mathbf{L}}_{\pi'_x}(\underline{z},x) \right]_{x=\underline{x}(\underline{z})} \quad = \quad 0; \tag{27}$$

4. *and the condition on the Lagrange multiplier, μ, at which (21) is satisfied.*

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

$$\Psi \equiv \frac{\delta \mathbf{L}}{\delta K'} \quad = \quad \Psi_{K'} \sum_z \left\{ \int_{\underline{x}(z)}^{\overline{x}(z)} \mathbf{L}_{K'}(z,x)dx + \mathbf{L}_{K'}(\underline{z}, \underline{x}(\underline{z})) + \mathbf{L}_{K'}(\overline{z}, \overline{x}(\overline{z})) \right\}, \tag{28}$$

where

$$\Psi_{K'}^{-1} \quad \equiv \quad 1 - \sum_z \left\{ \int_{\underline{x}(z)}^{\overline{x}(z)} \mathcal{K}_{K'}(z,x)dx + \mathcal{K}_{K'}(\underline{z}, \underline{x}(\underline{z})) + \mathcal{K}_{K'}(\overline{z}, \overline{x}(\overline{z})) \right\}. \tag{29}$$

The marginal effects at the lower and upper bounds are defined for $\mathbf{I} = \{\mathbf{L}, \mathcal{K}\}$ as

$$\begin{aligned} \mathbf{I}_{K'}(\underline{z}, \underline{x}(\underline{z})) \quad &\equiv \quad \left[\left(\mathbf{I}(\underline{z},x) - \pi_x(x)\mathbf{I}_{\pi'_x}(\underline{z},x) \right) (x_{K'} + x_k\omega_{K'}(\underline{z},x)) \right]_{x=\underline{x}(\underline{z})}. \\ \mathbf{I}_{K'}(\overline{z}, \overline{x}(\overline{z})) \quad &\equiv \quad \left[\left(\mathbf{I}(\overline{z},x) - \pi_x(x)\mathbf{I}_{\pi'_x}(\overline{z},x) \right) (x_{K'} + x_k\omega_{K'}(\overline{z},x)) \right]_{x=\overline{x}(\overline{z})}. \end{aligned}$$

[14] Note that Ψ is a functional derivative of $\mathbf{L}$ with respect to K', i.e. $\delta\mathbf{L}(\pi', \pi'_x, K')/\delta K'$. Klein, Krusell, and Rios-Rull (2008) derive a similar effect of government policy on aggregate capital in their generalized Euler equation.

Proof At the endogenous upper bound the endpoint is equality constrained. If the extreme lower bound is exogenous, then the condition 3 simplifies to $\widetilde{\mathbf{L}}_{\pi'_x}(\underline{z},x)|_{x=\underline{x}(\underline{z})} = 0$. For the proof and more detailed specifications of all terms see the Appendix.

The Lagrange-Euler conditions in Theorem 1 capture the tradeoff between the level of tax, π', and its curvature captured by the marginal tax rate, π'_x. At the optimum, the marginal effect of a change in the level tax schedule π' on the social welfare, $\tilde{\mathbf{L}}_{\pi'}$, has to be equal to the marginal effect of an implied change in the tax schedule curvature expressed by the derivative of the implied change, $\frac{d}{dx}\tilde{\mathbf{L}}_{\pi'_x}$, due to the changing marginal tax rate, π'_x.[15]

The social welfare function has three components: First, the weighted marginal utilities $\mathcal{W}$ from the sequence of optimal government tax schedules evaluated at the optimal limiting tax policy schedule that prevails in the long-run. $\mathcal{W}$ is composed from marginal utilities from two successive periods. The second term captures the value of the government balanced budget in every period, with μ as its shadow price. The third component captures the intertemporal effect of adjusted savings for the next-period aggregate capital. It affects social welfare through general equilibrium effects (in a representative-agent model this term would be zero). This term represents the additional term in the first-order condition in Davila et al. (2012). The optimal limiting government policy balances these three effects in the tradeoff between intertemporal efficiency and equality in the resulting stationary equilibrium. That is, the weighted sum of these variations at the optimum has to be equal to zero

$$\sum_{z \in Z} \Delta \tilde{\mathbf{L}}_{\pi'} = \sum_{z \in Z} \left(\Delta \mathcal{W}_{\pi'} + \mu \Delta \mathcal{G}_{\pi'} + \Psi \Delta \mathcal{K}_{\pi'} \right) = 0,$$

for any taxable activity x. Similar to Davila et al. (2012), the pecuniary externality operating through the general equilibrium effects is positive if a higher aggregate stock of capital would rise total income of agents with labor-intensive total income.

Theorem 2 (Second Order Sufficient Conditions) *A tax schedule π' satisfying the first-order conditions in Theorem 1 attains a strict maximum if and only if (i) the Lagrange function defined in equation (23) satisfies the second-order Legendre condition*

$$\sum_{z \in Z} \boldsymbol{L}_{\pi'_x \pi'_x}(z,x) < 0 \quad \textit{for all } x \in \left[\underline{x}(\underline{z}), \overline{x}(\overline{z})\right],$$

and (ii) the interval $\left[\underline{x}(\underline{z}), \overline{x}(\overline{z})\right]$ contains no points conjugate to $\underline{x}(\underline{z})$.

Proof For the proof see the Appendix.

[15]If we restrict the tax policy to be only a flat tax and the marginal tax rate is constant, the first-order conditions degenerate to the standard optimization problem $\tilde{\mathbf{L}}_{\pi'}(z,x) = 0$.

The Euler-Lagrange equation in Theorem 1 and the modified Legendre condition in Theorem 2 represent necessary and sufficient conditions for concavity and a unique maximum of the Calculus of Variations Stationary Ramsey Problem.

4.1 The Effects of Government Policy on the Equilibrium

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 which is a function of the distribution and prices which in turn are determined by $h(\cdot)$ 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 for any government policy schedule, agents' saving policy and the distribution functions are known only implicitly as a solution to the two operator equations ($\mathcal{F}(h) = 0$ and $\mathcal{L}(\lambda, \lambda', h) = 0$) and the aggregate equilibrium condition for equilibrium prices.

The next Lemma derives the effects of the government policy function π on the operator Euler equation by specifying five unknown "sensitivity" functions $h_\pi : Z \times B \longrightarrow \mathbb{R}$, $h_{\pi_x} : Z \times B \longrightarrow \mathbb{R}$, $h_K : Z \times B \longrightarrow \mathbb{R}$, $h_{\pi_x \pi} : Z \times B \longrightarrow \mathbb{R}$, and $h_{\pi_x \pi_x} : Z \times B \longrightarrow \mathbb{R}$.[16] Denoting the next-period after-tax marginal return to capital from equation (15) as $R' = 1 + y'_k - \pi_x(x')\, x'_k$, we obtain

Lemma 1 (The Effects of π, π_x, 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_x, and to the aggregate capital stock K, are,*

$$\mathcal{F}_i = u_{cc}(c)c_i - \beta \sum_{z'} \left\{ u_{cc}(c')c'_i R' + u_c(c') R'_i \right\} Q(z, z') = 0, \tag{30}$$

where the subscript denotes derivatives with respect to $i \in \{\pi, \pi_x, K\}$, and

$$\begin{aligned} \mathcal{F}_{ij} \;=\; & u_{ccc}(c)c_i c_j + u_{cc}(c)c_{ij} - \\ \beta \sum_{z'} & \left\{ \left[u_{ccc}(c')c'_i c'_j + u_{cc}(c')c'_{ij} \right] R' + u_{cc}(c') \left[c'_i R'_j + c'_j R'_i \right] + u'(c') R'_{ij} \right\} Q(z, z') = 0. \end{aligned} \tag{31}$$

where the subscripts denote derivatives with respect to $ij \in \{\pi_x \pi_x, \pi_x \pi\}$.

Proof For the proof and a full 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

[16] The proper specification of the derivative of the savings and distribution functions with respect to the government policy function again requires the generalized concept of the Fréchet derivative.

"sensitivity" functions $\lambda_\pi : Z \times B \longrightarrow \mathbb{R}$, $\lambda_{\pi_x} : Z \times B \longrightarrow \mathbb{R}$, $\lambda_K : Z \times B \longrightarrow \mathbb{R}$, $\lambda_{\pi_x\pi_x} : Z \times B \longrightarrow \mathbb{R}$ and $\lambda_{\pi_x\pi} : Z \times B \longrightarrow \mathbb{R}$.

Lemma 2 (The Effects of π, π_x, and K on the Distribution Function) *The total F-derivative of the operator distribution function $\mathcal{L}$ with respect to the government policy function π, to its derivative π_x, and to the aggregate capital stock K are,*

$$\mathcal{L}_i = \lambda_i(z', k') - \sum_z \left\{\lambda_k\left[z, h^{-1}(z,k')\right] h_i^{-1}(z,k') + \lambda_i\left[z, h^{-1}(z,k')\right]\right\} Q(z,z') = 0, \quad (32)$$

where the subscript denotes derivatives with respect to $i \in \{\pi, \pi_x, K\}$, and

$$\begin{aligned} \mathcal{L}_{ij} &= \lambda_{ij}(z',k') - \sum_z \left\{\lambda_{ij}\left[z, h^{-1}(z,k')\right] + \lambda_{ik}\left[z, h^{-1}(z,k')\right] h_j^{-1}(z,k') \right. \\ &+ \left[\lambda_{jk}\left[z, h^{-1}(z,k')\right] + \lambda_{kk}\left[z, h^{-1}(z,k')\right] h_j^{-1}(z,k')\right] h_i^{-1}(z,k') \\ &+ \left. \lambda_k\left[z, h^{-1}(z,k')\right] h_{ij}^{-1}(z,k')\right\} Q(z,z') = 0. \qquad (33) \end{aligned}$$

where the subscripts denote derivatives with respect to $ij \in \{\pi_x\pi_x, \pi_x\pi\}$.

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

In this way we obtain functional equations (15), (30)-(31), (16), and (32)-(33) in the unknown functions h, h_π, h_{π_x}, h_K, $h_{\pi_x\pi_x}$, $h_{\pi_x\pi}$, λ, λ_π, λ_{π_x}, λ_K, $\lambda_{\pi_x\pi_x}$, and $\lambda_{\pi_x\pi}$, 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.

4.2 Analytical Results: The U-Shaped Tax Function

As the first-order conditions in Theorem 1 specify explicit transversality conditions on agents with the lowest and the highest taxable activity, they can be used for qualitative results on the properties of the optimal limiting government policy at these boundary points. In equation (21), the term $gy(z,x)$ raises the same amount of revenues as a flat tax g on total income. We will use this concept as a useful benchmark.

First, we relate the tax payment of the "poorest" agent with $(\underline{k}, \underline{z})$ under the optimal tax policy π to the tax payment in the economy with a flat tax g.

Proposition 1 (Optimal Tax at Lowest Taxable Activity) *Assuming $\pi_x(\overline{x}(\overline{z})) \neq \left[\frac{\delta x}{\delta \pi}(\overline{x}(\overline{z}))\right]^{-1}$, the transversality condition in equation (27) in Theorem 1 implies that for an agent with the lowest taxable activity level $\underline{x}(\underline{z})$, the difference between optimal taxes paid $\pi(\underline{x}(\underline{z}))$ and 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 agent's savings to the aggregate welfare,*

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

where $\underline{c}(\underline{z}, \underline{x}(\underline{z})) = \underline{y}(\underline{z}, \underline{x}(\underline{z})) - \pi(\underline{x}(\underline{z})) - \underline{k}$ *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 aggregate capital on aggregate welfare defined in Theorem 1.*

Proof See the Appendix.

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 tax.

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 is larger than that under a flat income tax,* $\pi(\underline{x}(\underline{z})) > g\underline{y}(\underline{z}, \underline{x}(\underline{z}))$.

The poorest agent pays more taxes than in the flat-tax regime if one of these three conditions is satisfied: $\Psi = 0$, or if $\underline{k} = 0$ (as assumed in this paper), or if simultaneously $\Psi > 0$ and $\underline{k} > 0$. The Corollary illustrates the incentives to the agent to accumulate more assets.[17] 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 under a flat tax equal to g. The richest agent's current and future savings are at the endogenous upper bound $\overline{k}(\overline{z}) = \overline{k}$.

Proposition 2 (Optimal Tax at Highest Taxable Activity) *Assuming* $\pi_x(\overline{x}(\overline{z})) \neq \left[\frac{\delta x}{\delta \pi}(\overline{x}(\overline{z}))\right]^{-1}$, *the transversality condition in equation (26) in Theorem 1 implies that for an agent with the highest taxable activity* $\overline{x}(\overline{z})$, *the difference between optimal taxes paid* $\pi(\overline{x}(\overline{z}))$ *and the taxes paid under the flat tax regime* $g\,\overline{y}(\overline{z}, \overline{x}(\overline{z}))$, *is proportional to the sum of the agent's utility and the marginal contribution of agent's savings to aggregate welfare,*

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

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

Proof See the Appendix.

[17] If 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 $\Psi < 0$.

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

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.

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

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

Additionally,

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

implies that the amount of taxes paid by the richest agent is greater than that of the poorest agent, i.e., $\pi(\overline{x}(\overline{z})) > \pi(\underline{x}(\underline{z}))$*.*

These Corollaries show that the socially optimal tax payments of the 'boundary' agents depend only on these agents' individual characteristics 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 reveal 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 as its trough must lie under the flat-tax level to clear the government budget constraint proportional to g.[18]

The marginal effect Ψ of the aggregate capital stock on social welfare reflects the importance of the optimal wealth distribution in analyzing tax policies. The shape of the distribution of the individual state variable is behind the U-shape marginal tax rates in Diamond (1998) and Saez (2001) as well as in the dynamic models of Kapicka (2013) and Golosov, Tsyvinski, and Werquin (2014). These qualitative assessments are confirmed by the numerical results in our example in the following Section.

[18]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.

5 An Example: The Optimal Income Tax Schedule

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

$$x(z, k; K) = y(z, k; K) = r(K)k + w(K)z,$$

and the individual budget constraint is

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

As before, there is a borrowing constraint $\underline{k} = 0$ and the total tax revenues are equal to a fraction g of the total output. The Euler equation (15) for a (z, k)-agent's optimal savings function $h(z, k; K)$ is now

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

where $c' = x' - \pi(x') + h - h'$, $x' = r(K')h + w(K')z'$, and $h' = h(z', h(z, k; K); K')$. Note that for this specification $x_k = y_k = r(K)$ and $k_x = 1/x_k = 1/r(K)$.

5.1 Admissible Tax Functions

Because the tax schedule is an arbitrary continuous function, we must ensure that the first-order approach is valid and that the stationary recursive equilibrium exists.[19] 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). For each agent $(z, k) \in B \times Z$, denote the after-tax gross income as

$$\psi(z, k; K) \equiv x(z, k; K) - \pi(x(z, k; K)) + k,$$

and rewrite the Euler equation as

$$\begin{aligned} u'(\psi(z, k; K) \; - \; h(z, k; K)) = \\ \beta \sum_{z'} u'(\psi(z', h(z, k; K); K') - h(z', h(z, k; K); K'))\psi_1(z', h(z, k; K); K')Q(z, z'), \end{aligned}$$

where $\psi_1(z', h(z, k; K); K') = 1 + r(K') - \pi_x(x(z', h(z, k; K)); K')r(K')$ is the marginal after-tax return of investment.

Theorem 3 *For a given tax schedule* $\pi : \mathbb{R}_+ \to \mathbb{R}$, *if for each* $(z, k) \in B \times Z$

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

[19]Again, we analyze the case of borrowing constrained agents in the Appendix.

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 functions.

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

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

for all $x \in [r(K)\underline{k} + w(K)\underline{z}, r(K)\overline{k} + w(K)\overline{z}]$, then it satisfies the conditions of Theorem 3.

The above statement follows directly from the fact that $\psi_1(z, k; K) > 0$ and that ψ is quasi-concave. The corollary implies that the marginal tax rate must be smaller than $1 + 1/r(K)$. This upper bound is not likely to bind for a very wide range of tax schedules.[20] Application of Theorem 1 and Theorem 2 then implies necessary and sufficient conditions for the unique maximum attained by the tax schedule.

5.2 The Shape of the Optimal Income Tax Schedule

To obtain the first order conditions for the optimal income schedule we insert the terms from our example into the general conditions in Theorem 1 and Lemma 1 and Lemma 2 (see the Appendix C). Then, adapting Propositions 1 and 2 and the related Corollaries 1 and 2 to our example economy we obtain the following analytical results.

Corollary 4 (Optimal Tax Rates at Lowest and Highest Income) *If π is the optimal limiting tax function 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,*
$$\frac{\pi(\underline{x}(\underline{z}))}{\underline{x}(\underline{z})} > g;$$
2. *and, provided that $\Psi \geq 0$, the optimal tax rate at the highest income level $\overline{x}(\overline{z})$ is strictly greater than the optimal tax rate at the lowest income level,*
$$\frac{\pi(\overline{x}(\overline{z}))}{\overline{x}(\overline{z})} > \frac{\pi(\underline{x}(\underline{z}))}{\underline{x}(\underline{z})} > g.$$

[20] For a realistic equilibrium interest rate $r = 0.03$, the upper bound on the marginal tax rate is equal $1 + 1/r \simeq 33$. When numerically solving for the optimal tax schedule in the next Section we do not impose any bounds on the marginal tax rate but we check the admissibility of the optimal tax schedule ex post.

Proof The results follow directly from Corollaries 1 and 2.

Both the poorest and the richest agents face higher tax rates 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 a social planner who could use the first best, lump sum taxes. In order to fully insure the agents against idiosyncratic risk, the social planner would arrive at a stationary equilibrium in which all agents accumulate the same amount of capital K and consume the same amount of goods. Although our government is constrained from using such a first best policy, it strives to attain a similar distributional 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 optimal aggregate capital stock. The poor agents are motivated by high but decreasing tax schedule: In our simulated economy in the next Section, the optimal limiting marginal tax function is negative at low income levels. On the other hand, the rich 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 same level.

Similar incentives for agents in dynamic models of optimal taxation are present in Kapicka (2013), Golosov, Tsyvinski, and Troshkin (2010), or Golosov, Tsyvinski, and Werquin (2014). Davila, Hong, Krusell, and Rios-Rull (2012) also consider a change of the aggregate capital stock in the steady state that provides for a better insurance against idiosyncratic shocks through general equilibrium effects. We illustrate the optimal policy by numerical simulations in the following Section.

6 Numerical Solution

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

The solution to the problem of finding the optimal limiting policy in terms of the Stationary 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

Parameters			
$\beta = 0.976$	$\sigma = 2.0$	$\alpha = 0.36$	$\delta = 0.08$
Earnings Process			
$z \in Z = \{0.78, 1.0, 1.27\}$		$Q(z,z') = \begin{bmatrix} 0.66 & 0.28 & 0.07 \\ 0.27 & 0.44 & 0.27 \\ 0.07 & 0.28 & 0.66 \end{bmatrix}$	

Table 1: Parameters of the Benchmark Economy.

functions given by the F-derivatives. We obtain a system of functional equations with two side conditions and one condition on the Lagrange multiplier. We solve this complex functional equations problem by the least squares projection method. Its application to our problem, all definitions of the functional equations, and the approximation of the optimal tax schedule can be found in Appendix D.[21]

6.1 Parameterization

We follow the calibration described in Diaz-Jimenez et al. (2003) and Castaneda et al. (2003). The uninsurable idiosyncratic shock to labor productivity follows a three-state, first order Markov chain. Using the PSID data, the household annual labor income process is estimated by a first-order autoregression a persistence parameter 0.6 and a volatility 0.2. Tauchen and Hussey (1991) approximation procedure for a three-state Markov chain is displayed in Table 1.[22]

We set the discount factor at $\beta = 0.976$ to arrive at a capital-output ratio around 3. The rest of the parameters is standard, $\alpha = 0.36$, $\delta = 0.08$, and the risk aversion parameter $\sigma = 2$ (with the intertemporal elasticity of substitution 1/2). We provide a sensitivity analysis for different values of σ in the Appendix E.

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

[21]The least squares projection method is an efficient and well-behaved method for functional equations problems. For 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).

[22]These values imply an aggregate effective labor supply $L = 1.0169$ and a stationary distribution of labor productivity $Q^* = \{0.337, 0.326, 0.337\}$. A similar parameterization is used by Storesletten et al. (2004) and Diaz-Jimenez et al. (2003). We also experimented with a labor income process from Heaton and Lucas (1996) with $\rho = 0.53$ and $\sigma^2_\epsilon = 0.063$, obtaining similar results.

Steady State Results				
	Tax Schedule			
	Progressive		Flat	Optimal
Average	US Data	Optimal		
τ	0.151	0.094	0.268	—
Capital	5.76	5.95	6.25	6.56
Output	1.89	1.92	1.95	1.99
Capital-output ratio	3.04	3.10	3.20	3.30
Consumption	1.06	1.06	1.07	1.07
Median wealth	5.36	5.71	5.27	6.64
Median/mean wealth ratio	0.93	0.96	0.84	1.01
Interest rate (%)	3.86	3.62	3.26	2.92
Wage	1.19	1.21	1.23	1.25
Gini wealth	0.33	0.31	0.39	0.22
Constrained agents (%)	0.26	0.19	0.38	0.08

Notes: Constrained agents is a fraction of agents whose wealth equals the exogenous lower bound on capital.

Table 2: Steady State Results.

6.2 Simulated Economies

We model the progressive tax schedule as Heathcote, Storesletten, and Violante (2014) by specifying a tax on the total income as a function

$$\pi^{PT}(y(z,k)) = y(z,k) - \chi y(z,k)^{1-\tau},$$

where τ is a parameter of the rate of progressivity and χ is a level parameter that clears the government budget constraint. The tax function is progressive if $0 < \tau < 1$ with strictly increasing marginal tax rates. Heathcote et al. (2014) estimate from the PSID data (years 2000-2006 for individuals aged 20 to 60) that the rate of progressivity in the United States is $\tau^{US} = 0.151$. We use this rate to simulate a steady state of our economy to approximate the existing tax on total income in the United States. Next we apply our methodology to finding an *optimal* progressivity (or regressivity) of the tax schedule by constraining the tax function $\pi(x)$ to only monotone functions. We find that an optimal limiting monotone tax function is less progressive with $\tau^{PT*} = 0.094$. A simple flat-tax reform consists of applying a single proportional tax $\tau^{FT} = 0.268$ on the total income. Finally, we use our methodology described in the previous Sections to solve for the optimal limiting tax schedule.

6.3 Steady State Results

Steady states outcomes of these four economies are shown in Table 2. Relative to the approximated progressive tax schedule, the optimal progressive tax schedule increases capital stock by 3.3% and delivers a welfare gain of 0.41% in terms of consumption equivalent units.[23] The flat tax reform further increases the aggregate capital stock and improves steady state welfare by 0.57% and 0.16%, relative to the US data and optimal progressive steady states, respectively. As in Ventura (1999), the flat-tax reform increases inequality: Gini coefficients of wealth rises to 0.39 and the fraction of borrowing constrained agents becomes 0.38%.

The last column in Table 2 summarizes the optimal tax schedule steady state. The impact of the optimal limiting tax schedule is large. Steady state welfare increases by 1.02% and 0.98% relative to the progressive taxation and by 1.14% relative to the flat tax. Aggregate capital stock rises by 13.9%, 10.3%, and 5%, respectively. Output levels increase 5.3%, 3.6%, and 2.1%, respectively. Inequality falls significantly, with a fraction of constrained agents at 0.08%. The median agent accumulates more assets than the mean. General equilibrium effects cause the interest rate to drop below 3 percent while the wage increases.

Figure 1 shows the average and the marginal optimal limiting tax functions. The average tax rate is a U-shaped function taxing the lowest total income at 37%, decreasing to a minimum of 25% and rising to 48% at the highest level of total income. The marginal tax rate is negative for low incomes and rises at higher levels of accumulated savings.[24]

As in Davila et al. (2012), we find that the capital stock corresponding to the optimal limiting tax policy is higher than in the other steady states (although not as high as in their paper). Because the poor agents have labor intensive income, the optimal policy delivers a larger stock of capital in order to increase their wages. The optimal limiting tax function leads to *higher* capital accumulation than standard precautionary savings models including, in particular, the flat tax reform. However, in Davila et al. (2012), when the social planner directly manipulates agents' first order conditions, the resulting distribution of agents in the steady state stays very similar to the market economy. In our approach, the

[23]For comparison, Heathcote et al. (2014) find an optimal progressivity applied to labor income at $\tau^* = 0.065$ with a welfare gain of 0.5%.

[24]Note that the optimal limiting tax schedule easily satisfies the admissibility condition from Corollary 3 (the equilibrium interest rate implies an upper bound equal to 34). Optimal marginal tax rates with a similar shape are also found in dynamic models of Kapicka (2013) or Golosov, Tsyvinski, and Troshkin (2010) and in static models by Diamond (1998) and Saez (2001). In the original contribution of Mirrlees (1971), the welfare maximizing tax schedule is close to a linear non-decreasing function with the marginal tax rate between zero and one, and zero at both ends of the distribution due to labor incentives related to the distribution of skills and consumption-leisure preferences. Building on Mirrlees (1971) and Mirrlees (1976) seminal work, Kocherlakota (2005), Golosov et al. (2003) or Albanesi and Sleet (2006) study optimal social planner policies with asymmetric information needed for characterization of optimal policies.

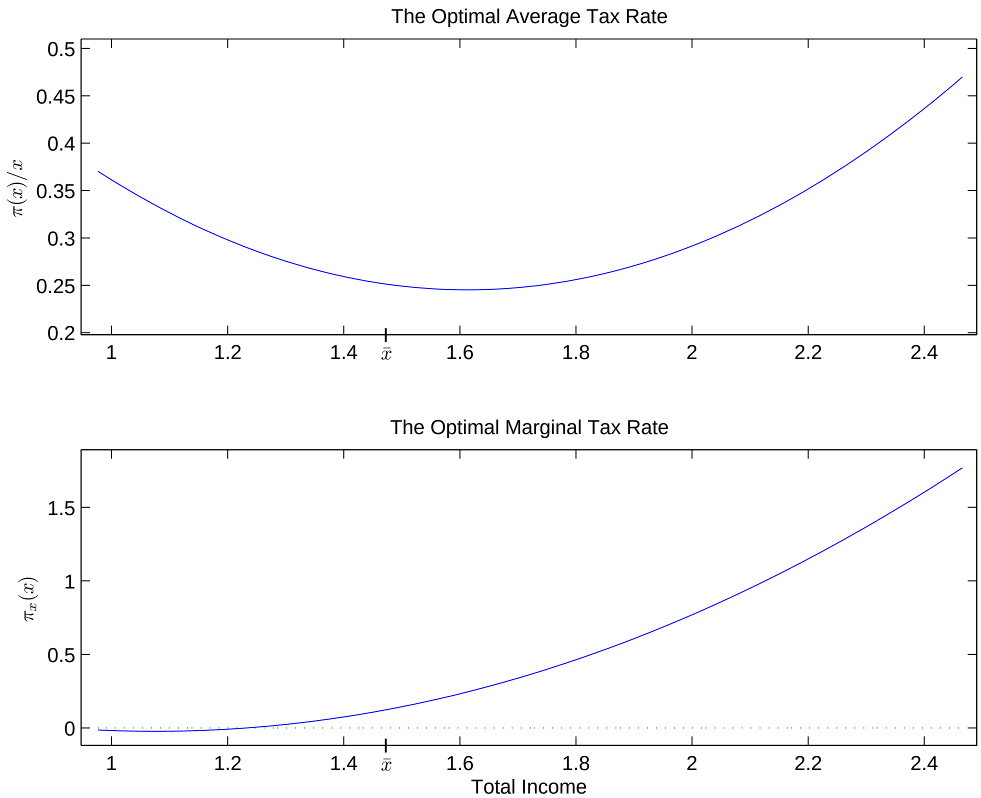

Figure 1: The optimal average tax rate and the optimal marginal tax rate.

implementable optimal limiting tax function not only increases the aggregate capital stock but also significantly reduces inequality. An important implication is that the optimal limiting tax function can resolve the tradeoff between efficiency and equality.

We want to emphasize that our stationary distribution is endogenous and there are no restrictions on the continuous optimal tax function to be positive or to be of any particular shape. Our class of admissible functions includes all progressive tax schedules but these were found significantly inferior with respect to the welfare criterion.[25] The comparison to the parameterization by Heathcote et al. (2014) shows that limiting the class of tax schedules to monotone functions could be rather restrictive.

6.4 Incentives Provided by the Optimal Limiting Tax Function

The large welfare effects of the optimal limiting tax function arise from the distributional effects. The stationary distributions of agents over assets in the four steady states are shown in Figure 2. Although all reforms increase the aggregate levels, the progressive and flat tax reforms do not take into account the distribution of agents and the associated general equilibrium effects. The flat-tax reform helps more the agents with high incomes

[25] Examples of monotone tax functions are Conesa and Krueger (2006) and Gouveia and Strauss (1994).

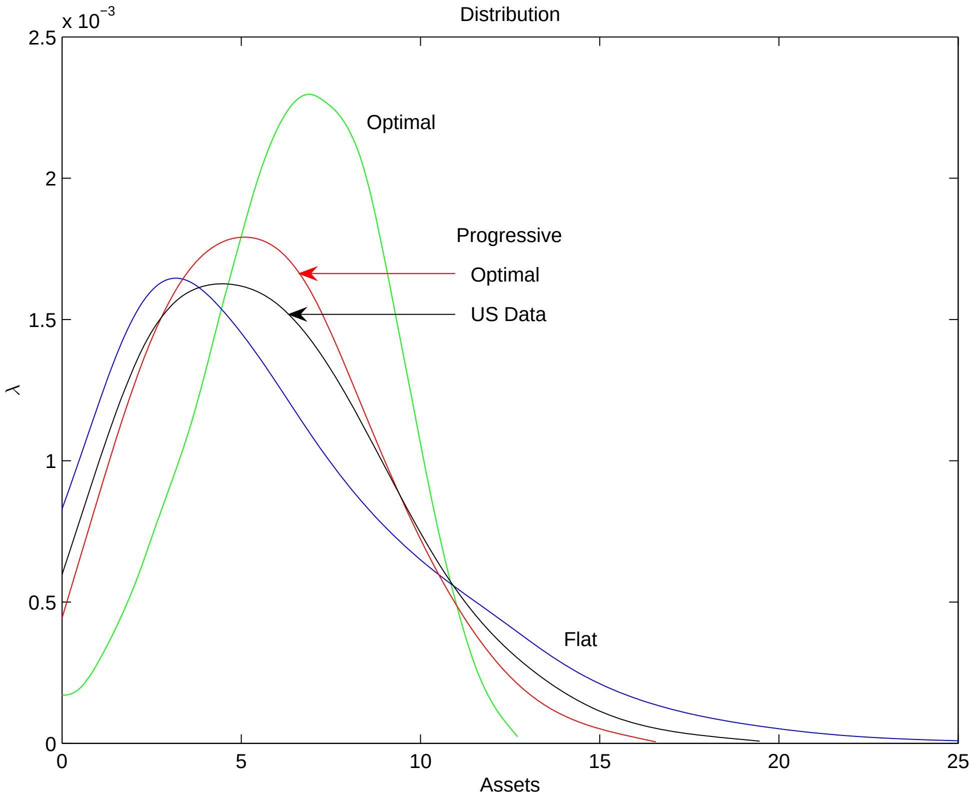

Figure 2: Stationary distribution of agents over assets in the progressive (US data and optimal), flat, and the optimal tax steady states.

with the median/mean wealth ratio falling to 0.84: 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. On the other hand, the progressive tax schedules do not provide incentives for poor households to save and move to higher steady-state income levels. In other words, progressive taxation provides too much short-run insurance at the cost of the long-run average levels.

This is exactly what the optimal limiting income tax function improves. The main mechanism are the incentive effects provided by the shape of the optimal tax schedule. The U-shaped function in the top panel of Figure 1 effectively motivates the agents to concentrate around the desired aggregate capital stock. The negative marginal tax rate at low income levels provides incentives to save more. The increasing tax rate on high incomes delivers tax revenues for redistribution. Around the median income level, the tax rate is lower than that found in the flat-tax reform.

To further analyze the tradeoff between efficiency and the distribution of agents, we adopt the approach of Domeij and Heathcote (2004) who distinguish the efficiency gain from distributional gains. The efficiency gain for an individual agent is the percentage of

Gains from the Optimal Tax Schedule (in %)			
	Relative to the Steady State with		
	Progressive Tax		Flat Tax
Average	US Data	Optimal	
Steady State Comparison			
Welfare Gain	1.02	0.88	1.14
Efficiency Gain	0.55	0.48	0.31
Distributional Gain	0.47	0.40	0.83
Including the Cost of the Transition			
Welfare Gain	-1.08	-0.91	-0.67
Efficiency Gain	-0.99	-0.84	-0.64
Distributional Gain	-0.09	-0.07	-0.03
Median Welfare Gain	-0.73	-0.72	-0.33
Political Support	19.6	20.3	22.1

Notes: Steady state and transition gains defined in the text.

Table 3: Steady state and transition gains from the optimal tax schedule.

the original consumption that would allow the agent to consume the same fraction of the aggregate consumption after the reform as in the original steady state. The distributional gain is the difference between the individual welfare gain and the efficiency gain.[26] The top panel of Table 3 shows that the steady state associated with the optimal tax function is superior in efficiency and distribution gains to the other steady states (the latter especially with respect to the flat-tax steady state).

We present a detailed analysis of individual welfare and distributional gains in the Appendix D. Most of these results can be summarized by the shares of assets owned and taxes paid by each quintile of the stationary distribution in Table 4. Recall that the steady state distribution is much more concentrated in the optimal tax steady state. In the progressive and flat tax steady states, the bottom quintile owns only around 5% of the total capital stock, while in the optimal tax steady state it is more than 8% (in terms of levels, it is almost twice as much). Vice versa, the top quintile in the other economies hold around 40% while it is only 30% in the optimal steady state (in levels, it is 25 percentage points less than in the flat-tax economy). Overall, in the optimal tax steady state the previously relatively poorer bottom three quintiles hold much more assets and at the same time receive higher wages. As a consequence, the distribution of the tax burden leads to more equitable outcomes (shares between 18.9 and 21.6%). Redistribution in a dynamic sense is a very important part of the optimal limiting tax schedule. Compared to Davila

[26]The 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. 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).

Steady State Distribution of Assets and Tax Contributions					
	Share of Each Quintile (in %)				
Tax Schedule	1st	2nd	3rd	4th	5th
			Assets		
Progressive US Data	5.31	12.32	18.61	25.45	38.30
Progressive Optimal	5.95	13.20	19.22	25.35	36.28
Flat Tax	4.26	10.57	17.01	25.32	42.84
Optimal	8.93	16.00	20.31	24.35	30.41
			Tax Contributions		
Progressive US Data	16.36	18.48	19.85	21.33	23.99
Progressive Optimal	16.79	18.53	19.74	21.10	23.85
Flat Tax	17.24	18.67	19.78	20.88	23.42
Optimal	18.88	19.41	19.83	20.30	21.58

Table 4: Steady State Distribution of Resources: Shares in Quintiles.

et al. (2012), the changes in savings are not mostly executed by the rich but by the poor agents. The optimal limiting tax function induces the poor agents to save more to escape relative poverty and to secure better insurance against idiosyncratic risk. At the aggregate level, these incentives lead to the desired pecuniary externality from general equilibrium effects.

6.5 Transition to the Optimal Tax Schedule Steady State

Simple steady-state comparisons could be misleading for welfare analysis because tax reforms imply substantial redistribution in the short run. The bottom part of Table 3 compares the expected present discounted value from an unanticipated optimal tax reform of the progressive and flat-tax steady states. In each case the optimal tax schedule is imposed on the stationary distribution of the initial steady state. We guess a sufficiently large number of convergence periods and iterate on paths of equilibrium prices to clear markets in each period of the transition, returning possible excess tax revenues to all agents in each period. The convergence lasts around thirty periods.

As in Domeij and Heathcote (2004) and many other papers, the expected discounted present value of welfare losses during a transition are so large that they overturn the steady state welfare improvement. The welfare losses are 1.08%, 0.91%, and 0.67%, respectively (measured in terms of per-period consumption transfers as a percentage of the initial steady state consumption). In total, only around one fifth of the population is better off from the reform (see Appendix D for more details). However, a possibility to issue debt repaid from the efficiency gains in the optimal steady state might provide resources to compensate agents and obtain political support for the reform.[27] Obviously, this transition is very far from an optimal transition during which the government would want to change the shape and level of the optimal tax policy in each period.

6.6 Sensitivity Analysis

Finally, we discuss the effects of different parameter values on the optimal tax schedule. In particular, we analyze steady states and transitions with alternative specifications of the coefficient of intertemporal substitution, $\theta = 1/\sigma$. Detailed results and tables with steady state and transition allocations can be found in Appendix E. The U-shape of the optimal limiting tax schedule is preserved in all simulated economies. Similar to Imrohoroglu (1998), with higher risk aversion, the agents have their own incentives to accumulate more capital as a buffer stock against the idiosyncratic risk. Consequently, the optimal tax schedules are less steep at low incomes. The optimal marginal tax functions are still U-shaped and negative at low levels of income. The aggregate capital stock increases in σ for precautionary reasons. These simulations show that the methodology developed in the paper delivers consistent results.

7 Conclusions

In this paper, we provide a solution method for implementing optimal limiting government policies in a general equilibrium economy with incomplete markets. We think of these policies as optimal because they take simultaneously into account their effects on the distribution of agents and equilibrium prices. As an example, we find the optimal limiting tax on total income in a stationary Ramsey problem. As in Diamond (1998), Saez (2001), or Golosov et al. (2010), the optimal tax schedule is U-shaped, it increases aggregate levels by providing the right incentives but not at the cost of increased inequality. Welfare gains in the steady state are large and compensation schemes or optimal transition policies could be designed to alleviate welfare losses from the transition to the optimal steady state. As Davila et al. (2012), we find that the capital stock corresponding to the optimal limiting tax policy is higher than in standard precautionary savings models, including the flat-tax reform. However, an important result is that the implemented optimal tax can resolve the tradeoff between efficiency and equality.

The present paper considers a very basic optimization problem with an inelastic labor supply and a simple process for individual income risk. We believe that it is a first step for analyzing more realistic models with important policy implications. Within the field

[27] See Gottardi et al. (2011) for a recent analysis.

of optimal taxation, in our future research we plan to study the optimal tax schedule with endogenous labor supply and realistic life-cycle income profiles as in Golosov et al. (2009) and Farhi and Werning (2012). Another topic that has received a lot of attention is the optimal capital taxation in models with heterogeneous agents. With respect to the class of optimal government policies, an important extension is in relaxing the assumption on continuity of the tax function. Finally, we would also like to explore different (Rawlsian) welfare functions and the role of government debt.

References

Albanesi, S. and C. Sleet (2006). Dynamic optimal taxation with private information. *Review of Economic Studies 73*, 1–30.

Bohacek, R. and M. Kejak (2002). Projection methods for economies with heterogeneous agents. *CERGE-EI Working Paper*.

Castaneda, A., J. Diaz-Jimenez, and J.-V. Rios-Rull (2003). Accounting for the u.s. earnings and wealth inequality. *Journal of Political Economy 111*(4), 818–857.

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

Davila, J., J. H. Hong, P. Krusell, and J.-V. Rios-Rull (2012). Constrained efficiency in the neoclassical growth model with uninsurable idiosyncratic shocks. *Econometrica 80*(6), 2431–2467.

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

Diamond, P. and J. Mirrlees (1971). Optimal taxation and public production II: Tax rules. *American Economic Review 61*(3), 261–278.

Diamond, P. and E. Saez (2011). The case for a progressive tax: From basic research to policy recommendations. *Journal of Economic Perspectives 25*(4), 165–190.

Diaz-Jimenez, J., J. Pijoan-Mas, and J. V. Rios-Rull (2003). Habit formation: Implications for the wealth distribution. *Journal of Monetary Economics 50*(6), 1257–1291.

Domeij, D. and J. Heathcote (2004). On the distributional effects of reducing capital taxes. *International Economic Review 45*(2), 523–554.

Farhi, E. and I. Werning (2012). Insurance and taxation ove the life cycle. *Review of Economic Studies 80*(2), 596–635.

Gelfand, I. M. and S. V. Fomin (2000). *Calculus of Variations*. Courier Corporation.

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

Golosov, M., A. Tsyvinski, and M. Troshkin (2009). A quantitative exploration in the theory of dynamic optimal taxation. *Working paper*.

Golosov, M., A. Tsyvinski, and M. Troshkin (2010). Optimal dynamic taxes. *Working paper*.

Golosov, M., A. Tsyvinski, and N. Werquin (2014). A variational approach to the analysis of tax systems. *Working paper*.

Gottardi, P., A. Kajii, and T. Nakajima (2011). Optimal taxation and constrained inefficiency in an infinite-horizon economy with incomplete markets. *CESifo Working Paper 3560*.

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

Heathcote, J., K. Storesletten, and G. Violante (2014). Optimal tax progressivity: An analytical framework. *NBER Working Paper 19899*.

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.

Imrohoroglu, S. (1998). A quantitative analysis of capital income taxation. *ier 39*(2), 307–328.

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.

Kanbur, R. and M. Tuomala (1994). Inherent inequality and the optimal graduation of marginal tax rates. *Scandinavian Journal of Economics 96*(2), 275–282.

Kapicka, M. (2013). Efficient allocations in dynamic private information economies with persistent shocks: A first-order approach. *Review of Economic Studies 80*(3), 1027–1054.

Klein, P., P. Krusell, and J.-V. Rios-Rull (2008). Time-consistent public policy. *Review of Economic Studies 75*(3), 789–808.

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

Liberzon, D. (2011). *Calculus of variations and optimal control theory: a concise introduction*. Princeton: Princeton University Press.

Luenberger, D. G. (1969). *Optimization by vector space methods*. New York: John Wiley and Sons.

Mankiw, N. G., M. C. Weinzierl, and D. Yagan (2009). Optimal taxation in theory and practice. *Journal of Economic Perspectives 23*(4), 147–174.

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.

Ok, E. A. (2007). *Real analysis with economic applications*. Princeton: Princeton University Press.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies 68*, 205–229.

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 (2004). Consumption and risk sharing over the life cycle. *Journal of Monetary Economics 51*, 609–633.

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.

Optimal Government Policies in Models with Heterogeneous Agents

Radim Boháček and Michal Kejak
CERGE-EI

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 (see Figure A.1 below for illustration). For these agents, i.e. for those with $(z,k) \in [Z \times \underline{k}, \underline{k}(z))$ and the next period savings h equal to $\underline{k}$, the Euler equation is satisfied in the form of inequality (allocations in the next period are denoted with a superscript plus sign)

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

where

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

implying $h = \underline{k}$.

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

$$\widetilde{\mathcal{F}}(h;\pi,\pi_x) \equiv \begin{cases} \mathcal{F}(h;\pi,\pi_x) & \text{for } (z,k) \in Z \times [\underline{k}(z), \overline{k}(\overline{z})], \\ h & \text{for } (z,k) \in Z \times [\underline{k}, \underline{k}(z)), \end{cases}$$

and thus the operator equation in the form $\widetilde{\mathcal{F}}(\tilde{h};\pi,\pi_x) = 0$ 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

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

where $\lambda(z', \underline{k})$ is the mass of agents with the next period capital $\underline{k}$ and the next-period

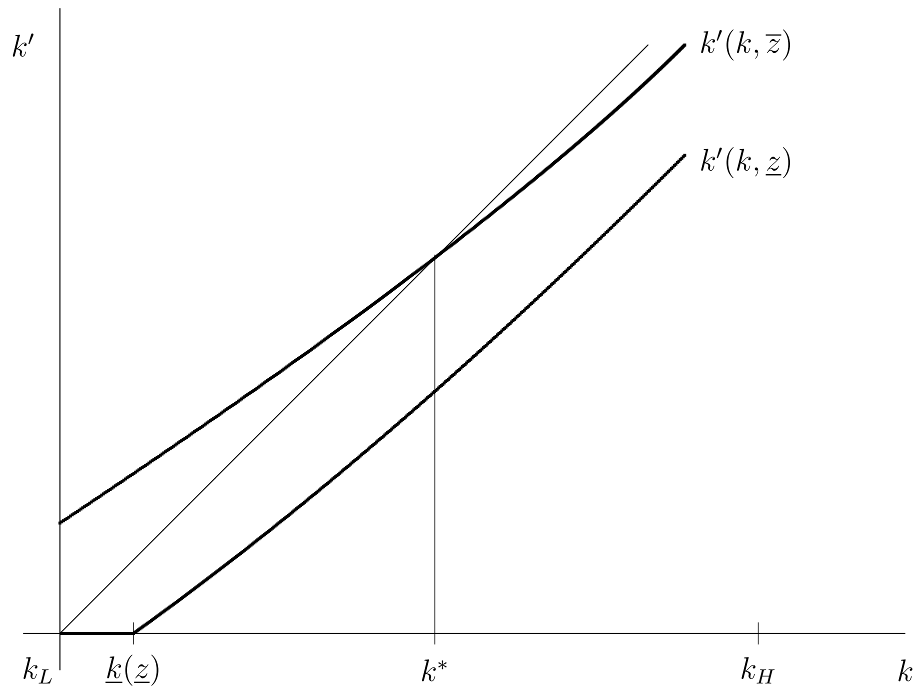

Figure A.1: Policy functions for the next period capital stock. An example with two productivity shocks $\overline{z} > \underline{z}$. There is an exogenous lower bound k_L and an endogenous upper bound $k^* < k_H$. The stationary distribution has a unique ergodic set $E = [k_L, k^*]$. Agents with shock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

shock z'. Let us note such amended distribution function by $\widetilde{\lambda}$.

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

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

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

$$\int_{\underline{k}}^{\overline{k}(\overline{z})} \widetilde{\lambda}(z, k) dk = \int_{\underline{k}}^{\overline{k}(\overline{z})} \lambda(z, k) dk,$$

for any $z \in Z$ and the distribution functions $\widetilde{\lambda}$ and λ are equivalent. So we can simply consider only the distribution function λ given by $\mathcal{L}$ in (16) and h given by (15).

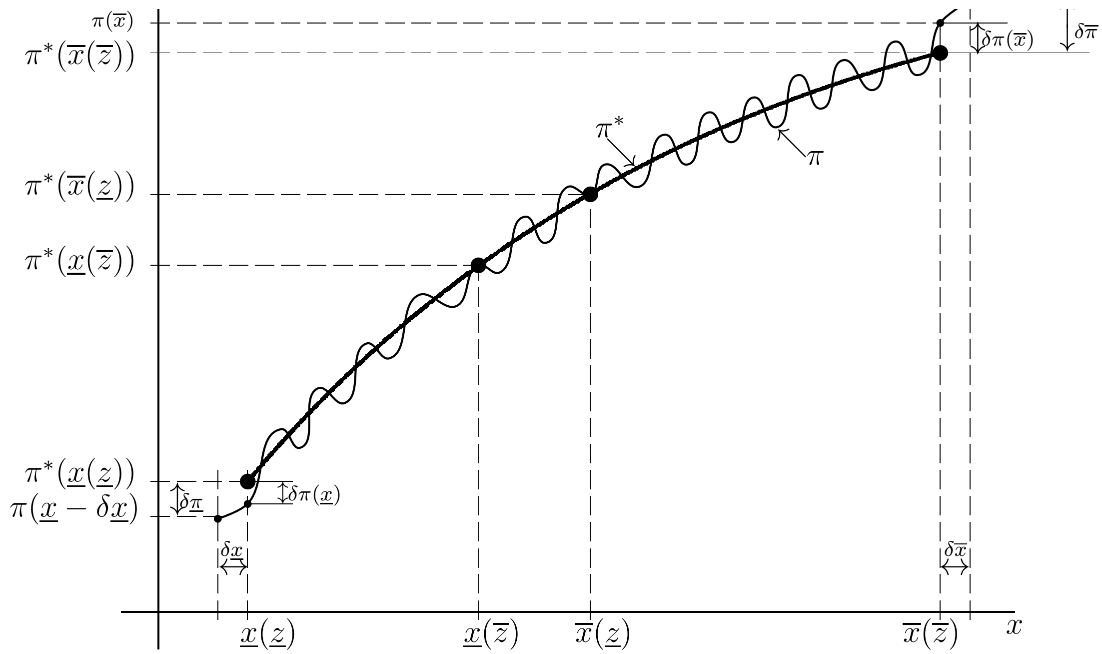

Figure B.1: Variations Around Optimal Tax Schedule

B Appendix: Proofs

B.1 Proof of Theorem 1

First, we derive the first order conditions for the Ramsey problem in (22) for specifying the optimal tax policy function at some period t so we define

$$\begin{aligned} J(\varepsilon) &= \sum_{t=s}^{\infty} \beta^{t-s} \left\{ \int_{\underline{x}_t(\underline{z}_t)-\varepsilon_t\delta\underline{x}_t}^{\overline{x}_t(\underline{z}_t)} \left\{ \mathcal{U}\left[z_t, x_t; \tilde{\pi}_t(x_t), \tilde{\pi}_{tx}(x_t), \tilde{K}_t\right] + \mu_t \mathcal{G}\left[z_t, x_t; \tilde{\pi}_t(x_t), \tilde{\pi}_{tx}(x_t), \tilde{K}_t\right] \right\} dx_t \right. \\ &+ \sum_{z_t \in Z\setminus\{\underline{z},\overline{z}\}} \int_{\underline{x}_t(z_t)}^{\overline{x}_t(z_t)} \left\{ \mathcal{U}\left[z_t, x_t; \tilde{\pi}_t(x_t), \tilde{\pi}_{tx}(x_t), \tilde{K}_t\right] + \mu_t \mathcal{G}\left[z_t, x_t; \tilde{\pi}_t(x_t), \tilde{\pi}_{tx}(x_t), \tilde{K}_t\right] \right\} dx_t \\ &+ \left. \int_{\overline{x}_t(\overline{z}_t)}^{\overline{x}_t(\overline{z}_t)+\varepsilon_t\delta\overline{x}} \left\{ \mathcal{U}\left[z_t, x_t; \tilde{\pi}_t(x_t), \tilde{\pi}_{tx}(x_t), \tilde{K}_t\right] + \mu_t \mathcal{G}\left[z_t, x_t; \tilde{\pi}_t(x_t), \tilde{\pi}_{tx}(x_t), \tilde{K}_t\right] \right\} dx_t \right\}, \end{aligned}$$

where for each t

$$\begin{aligned} \tilde{\pi}_{t+1}(x) &\equiv \pi^*_{t+1}(x) + \varepsilon_{t+1}\delta\pi_{t+1}(x), \\ \tilde{\pi}_{t+1,x}(x) &\equiv \pi^*_{t+1,x}(x) + \varepsilon_{t+1}\delta\pi_{t+1,x}(x), \\ \tilde{K}_{t+1} &\equiv K^*_{t+1} + \varepsilon_{t+1}\delta K_{t+1} \end{aligned}$$

where

$$\mathcal{U}\left[z_t, x_t; \tilde{\pi}_t(x_t), \tilde{\pi}_{tx}(x_t), \tilde{K}_t\right] = u(c(z_t, k(z_t, x_t)))\,\lambda((z_t, k(z_t, x_t))).$$

The dependence of the bounds on the value of shocks $z \in Z$ makes our problem a little harder than the standard calculus of variations 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 B.1 below. Therefore, only the values of the government policy at the boundaries of the maximal interval, $\pi_{t+1}(\underline{x}_{t+1}(\underline{z}))$ and $\pi_{t+1}(\overline{x}_{t+1}(\overline{z}))$, are free while all other interior bounds are fixed. Then the condition for the optimal tax π_{t+1}

$$J_{\varepsilon_{t+1}}(\mathbf{0}) = \lim_{\varepsilon_{t+1}\to 0} \frac{dJ(\varepsilon_{t+1})}{d\varepsilon_{t+1}} = 0$$

leads to the following first-order conditions.

$$\begin{aligned} J_{\varepsilon'}(\mathbf{0}) &= \sum_{z\in Z} \int_{\underline{x}(z)}^{\overline{x}(z)} \left\{ \left[\mathcal{W}_{\pi'}(z,x) + \mu \mathcal{G}_{\pi'}(z,x)\right] \delta\pi'(x) + \left[\mathcal{W}_{\pi'_x}(z,x) + \mu \mathcal{G}_{\pi'_x}(z,x)\right] \delta\pi'_x(x) \right. \\ &+ \left. \left[\mathcal{W}_{K'}(z,x) + \mu \mathcal{G}_{K'}(z,x)\right] \delta K' \right\} dx \\ &- \left[\mathcal{W}(\underline{z},x)|_{\underline{x}(\underline{z})} + \mu \mathcal{G}(\underline{z},x)|_{\underline{x}(\underline{z})}\right](-\delta\underline{x}) + \left[\mathcal{W}(\overline{z},x)|_{\overline{x}(\overline{z})} + \mu \mathcal{G}(\overline{z},x)|_{\overline{x}(\overline{z})}\right] \delta\overline{x}, \end{aligned}$$

where the last line comes from the fact that the lower and upper bounds are free.

Using the following notation

$$\mathcal{W}(z,x) \equiv \mathcal{W}\left[z,x;\tilde{\pi}'^{*}(x),\tilde{\pi}'^{*}_{x}(x),\tilde{K}'^{*}\right], \tag{B.1}$$

$$\begin{aligned} \mathcal{W}\left[z,x;\pi',\pi'_x,K'\right] &= \left\{u\left[c\left(z,k(z,x);\pi',\pi'_x,K'\right)\right]\right. \\ &+ \left.\beta \sum_{z'\in Z} u\left[c\left(z',h(z,k(z,x);\pi',\pi'_x,K'\right)\right] Q(z,z')\right\} \lambda\left[z,k(z,x);\pi',\pi'_x,K'\right]\ k_x(z,x), \end{aligned}$$

and for the side condition,

$$\mathcal{G}(z,x) \equiv \mathcal{G}\left[z,x;\tilde{\pi}'^{*}(x'),\tilde{\pi}x'^{*}(x'),\tilde{K}'^{*}\right],$$

$$\begin{aligned} \mathcal{G}\left[z,x;\pi',\pi'_x,K'\right] &= \sum_{z'\in Z} \left[\pi'\left(x\left(z',h\left(z,x(z,k);\pi',\pi'_x,K'\right)\right)\right) - gy\left(z',h\left(z,x(z,k);\pi',\pi'_x,K'\right);K'\right)\right] \\ &\cdot Q(z,z')\lambda\left[z,k(z,x);\pi',\pi'_x,K'\right]\ k_x(z,x). \end{aligned}$$

Finally, denote

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

whose integration by parts delivers

$$\int_{\underline{x}(z)}^{\overline{x}(z)} \mathbf{L}_{\pi_x}(z,x)\delta\pi'_x(x)dx = \left[\mathbf{L}_{\pi'_x}(z,x)\delta\pi'(x)\right]_{\underline{x}(z)}^{\overline{x}(z)} - \int_{\underline{x}(z)}^{\overline{x}(z)} \frac{d}{dx}\mathbf{L}_{\pi'_x}(z,x)\delta\pi'(x)dx$$

and

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

Thus we can rewrite the formula above in a more compact form as[28]

$$\begin{aligned} J_\varepsilon(0) &= \sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)} \left\{\left[\mathbf{L}_{\pi'}(z,x) - \frac{d}{dx}\mathbf{L}_{\pi'_x}(z,x)\right]\delta\pi'(x) + \mathbf{L}_{K'}(z,x)\delta K'\right\}dx \qquad \text{(B.2)} \\ &+ \left[\mathbf{L}(z,x)\delta\pi'(x)\right]_{\underline{x}(\underline{z})}^{\overline{x}(\overline{z})} - \mathbf{L}(z,x)|_{\underline{x}(\underline{z})}(-\delta\underline{x}) + \mathbf{L}(z,x)|_{\overline{x}(\overline{z})}\delta\overline{x}. \end{aligned}$$

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

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

and

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

This implies that

$$\delta\pi'(\overline{x}) = \delta\overline{\pi}' - \pi'^*_x(\overline{x})\delta\overline{x}, \qquad \text{(B.3)}$$

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

At the equality constrained extreme endpoint $\overline{x}(\overline{z})$ the upper bound for capital, $\overline{k}$, is endogenous and thus implicitly given by the saving function, $\overline{k} = h\left(\overline{z},\overline{k};\pi',\pi'_x,K'\right)$. The total variance differential is

$$\delta\overline{k}\left[h_k\left(\overline{z},\overline{k}\right)-1\right] + h_K\left(\overline{z},\overline{k}\right)\delta K' + h_\pi\left(\overline{z},\overline{k}\right)\delta\overline{\pi}' = 0$$

and thus

$$\delta\overline{k} = \left[1-h_k\left(\overline{z},\overline{k}\right)\right]^{-1}\left[h_K\left(\overline{z},\overline{k}\right)\delta K' + h_\pi\left(\overline{z},\overline{k}\right)\delta\overline{\pi}'\right]. \qquad \text{(B.4)}$$

As $\overline{x} = x\left[\overline{z},\overline{k}\left(\overline{z};\pi',\pi'_x,K'\right);K'\right]$, we determine the variation

$$\delta\overline{x} = x_k\left[\overline{z},\overline{k}\left(\overline{z}\right)\right]\delta\overline{k} + x_K\left[\overline{z},\overline{k}\left(\overline{z}\right)\right]\delta K',$$

where we can further substitute for $\delta\overline{k}$ from (B.4)

$$\delta\overline{x} = \left\{x_k\left(\overline{z},\overline{k}\right)\overline{\omega}_{K'} + x_{K'}\left(\overline{z},\overline{k}\right)\right\}\delta K' + x_k\left(\overline{z},\overline{k}\right)\overline{\omega}_{\pi'}\delta\overline{\pi}', \qquad \text{(B.5)}$$

where

$$\overline{\omega}_{K'} \equiv \frac{\delta\overline{k}}{\delta K'} = \frac{h_{K'}\left(\overline{z},\overline{k}\right)}{1-h_k\left(\overline{z},\overline{k}\right)} \quad \text{and} \quad \overline{\omega}_{\pi'} \equiv \frac{\delta\overline{k}}{\delta\pi'} = \frac{h_{\pi'}\left(\overline{z},\overline{k}\right)}{1-h_k\left(\overline{z},\overline{k}\right)}. \qquad \text{(B.6)}$$

[28] Since we are looking for the optimal tax policy π' the boundaries $\underline{x}(z)$ and $\overline{x}(z)$ are fixed for all $z\in Z$ including $\underline{z}$ and $\overline{z}$, $[\mathcal{U}(z,x)\delta\pi'(x)]_{\underline{x}(z)}^{\overline{x}(z)} = 0$.

We can similarly specify the variation of the lower bound of the policy function,

$$\delta\pi'\left(\underline{x}\right) = -\delta\underline{\pi}' + \pi_x'^*\left(\underline{x}\right)\delta\underline{x}. \tag{B.7}$$

In case when the lower bound is endogenous, we can analogously derive the lower boundary first-order condition in the following way. At the equality constrained endpoint $\underline{x}\left(\underline{z}\right)$ the lower bound for capital, $\underline{k}$, is endogenous and implicitly given by the saving function, $\underline{k}' = h\left(\underline{z}, \underline{k}; \pi', \pi_x', K'\right)$. The total variance differential is

$$\delta\underline{k}'\left[h_k\left(\underline{z}, \underline{k}\right) - 1\right] + h_{K'}\left(\underline{z}, \underline{k}\right)\delta K' + h_{\pi'}\left(\underline{z}, \underline{k}\right)\delta\underline{\pi}' = 0$$

and thus

$$\delta\underline{k}' = \left[1 - h_k\left(\underline{z}, \underline{k}\right)\right]^{-1}\left[h_{K'}\left(\underline{z}, \underline{k}\right)\delta K' + h_{\pi'}\left(\underline{z}, \underline{k}\right)\delta\underline{\pi}'\right]. \tag{B.8}$$

As $\underline{x} = x\left[\underline{z}, \underline{k}\left(\underline{z}; \pi', \pi_x', K'\right); K'\right]$, we determine the variation

$$\delta\underline{x} = x_k\left[\underline{z}, \underline{k}\left(\underline{z}\right)\right]\delta\underline{k} + x_{K'}\left[\underline{z}, \underline{k}\left(\underline{z}\right)\right]\delta K',$$

where we can further substitute for $\delta\underline{k}$ from (B.4)

$$\delta\underline{x} = \left\{x_k\left(\underline{z}, \underline{k}\right)\underline{\omega}_{K'} + x_{K'}\left(\underline{z}, \underline{k}\right)\right\}\delta K' + x_k\left(\underline{z}, \underline{k}\right)\underline{\omega}_{\pi'}\delta\underline{\pi}', \tag{B.9}$$

where

$$\underline{\omega}_{K'} \equiv \frac{\delta\underline{k}}{\delta K'} = \frac{h_{K'}\left(\underline{z}, \underline{k}\right)}{1 - h_k\left(\underline{z}, \underline{k}\right)} \quad \text{and} \quad \underline{\omega}_{\pi'} \equiv \frac{\delta\underline{k}}{\delta\underline{\pi}'} = \frac{h_{\pi'}\left(\underline{z}, \underline{k}\right)}{1 - h_k\left(\underline{z}, \underline{k}\right)}. \tag{B.10}$$

Going back to (B.2) and using (B.3) and (B.7) we obtain

$$\begin{aligned} J_\varepsilon\left(0\right) &= \sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\left\{\left[\mathbf{L}_{\pi'}(z,x) - \frac{d}{dx}\mathbf{L}_{\pi_x'}(z,x)\right]\delta\pi'(x) + \mathbf{L}_{K'}(z,x)\delta K'\right\}dx \\ &+ \mathbf{L}_{\pi_x'}(\underline{z},x)|_{\underline{x}(\underline{z})}\delta\underline{\pi}' - \left[\pi_x'(x)\mathbf{L}_{\pi_x'}(\underline{z},x) - \mathbf{L}(\underline{z},x)\right]_{\underline{x}(\underline{z})}\delta\underline{x} \\ &+ \mathbf{L}_{\pi_x'}(\overline{z},x)|_{\overline{x}(\overline{z})}\delta\overline{\pi}' - \left[\pi_x'(x)\mathbf{L}_{\pi_x'}(\overline{z},x) - \mathbf{L}(\overline{z},x)\right]_{\overline{x}(\overline{z})}\delta\overline{x}. \end{aligned} \tag{B.11}$$

If both the upper and the lower bound are endogenous and therefore equality constrained, $\delta\overline{x}$ and $\delta\underline{x}$, are not independent and we need to use (B.5) to obtain

$$\begin{aligned} J_{\varepsilon'}\left(0\right) &= \sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\left\{\left[\mathbf{L}_{\pi'}(z,x) - \frac{d}{dx}\mathbf{L}_{\pi_x'}(z,x)\right]\delta\pi'(x) + \mathbf{L}_{K'}(z,x)\delta K'\right\}dx \\ &+ \mathbf{L}_{\pi_x'}(\underline{z},x)|_{\underline{x}(\underline{z})}\delta\underline{\pi}' + \left[\mathbf{L}(\underline{z},x) - \pi_x'(x)\mathbf{L}_{\pi_x'}(\underline{z},x)\right]_{\underline{x}(\underline{z})}\left\{\left(\underline{x}_{K'} + \underline{x}_k\underline{\omega}_{K'}\right)\delta K' + \underline{x}_k\underline{\omega}_{\pi'}\delta\underline{\pi}'\right\} \\ &+ \mathbf{L}_{\pi_x'}(\overline{z},x)|_{\overline{x}(\overline{z})}\delta\overline{\pi}' + \left[\mathbf{L}(\overline{z},x) - \pi_x'(x)\mathbf{L}_{\pi_x'}(\overline{z},x)\right]_{\overline{x}(\overline{z})}\left\{\left(\overline{x}_{K'} + \overline{x}_k\overline{\omega}_{K'}\right)\delta K' + \overline{x}_k\overline{\omega}_{\pi'}\delta\overline{\pi}'\right\}. \end{aligned} \tag{B.12}$$

For the variation of the aggregate capital $\delta K'$ we use equation (20)

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

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

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

The variation of the aggregate capital is then

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

where

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

Finally, from (B.12) we get

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

Now, in order to get first-order conditions we assign $\delta J=0$. Since the first two terms have to be zero for any $\delta\pi'(x)$ for all x, and the other two terms have to be zero for any $\delta\underline{\pi}'$ and $\delta\overline{\pi}'$, respectively, we have

$$\sum_{z\in Z}\left[\mathbf{L}_{\pi'}(z,x)+\Psi\mathcal{K}_{\pi'}(z,x)-\frac{d}{dx}\left(\mathbf{L}_{\pi'_x}(z,x)+\Psi\mathcal{K}_{\pi'_x}(z,x)\right)\right]=0,\tag{B.13}$$

$$\left[\mathbf{L}(\underline{z},x)+\Psi\mathcal{K}(\underline{z},x)\right]_{\underline{x}(\underline{z})}+\left[\frac{1}{\underline{x}_k\underline{\omega}_{\pi'}}-\pi'_x(x)\right]\left[\mathbf{L}_{\pi'_x}\left(\underline{z},x\right)+\Psi\mathcal{K}_{\pi'_x}\left(\underline{z},x\right)\right]|_{\underline{x}(\underline{z})}=0,$$

$$\left[\mathbf{L}_{\pi'_x}(\overline{z},x)+\Psi\mathcal{K}_{\pi'_x}(\overline{z},x)\right]|_{\overline{x}(\overline{z})}+\left[\frac{1}{\overline{x}_k\overline{\omega}_{\pi'}}-\pi'_x(x)\right]\left[\mathbf{L}_{\pi'_x}(\overline{z},x)+\Psi\mathcal{K}_{\pi'_x}(\overline{z},x)\right]_{\overline{x}(\overline{z})}=0,$$

where

$$\Psi \equiv \Psi_{K'} \left\{ \sum_{z\in Z} \int_{\underline{x}(z)}^{\overline{x}(z)} \mathbf{L}_{K'}(z,x)dx + \left[\mathbf{L}(\underline{z},x) - \pi'_x(x)\mathbf{L}_{\pi'_x}(\underline{z},x)\right]_{\underline{x}(\underline{z})} \left(\underline{x}_{K'} + \underline{x}_k \underline{\omega}_{K'}\right) \right.$$
$$\left. + \left[\mathbf{L}(\overline{z},x) - \pi'_x(x)\mathbf{L}_{\pi'_x}(\overline{z},x)\right]_{\overline{x}(\overline{z})} \left(\overline{x}_{K'} + \overline{x}_k \overline{\omega}_{K'}\right) \right\}$$

and

$$\left[\frac{\delta x}{\delta \pi'}\right]^{-1} = \frac{1}{\overline{x}_k \overline{\omega}_{K'}}.$$

In case that the lower bound, $\underline{k}\left(\underline{z}\right)$ is exogenous, we obtain

$$J_{\varepsilon'}(0) = \sum_{z\in Z} \int_{\underline{x}(z)}^{\overline{x}(z)} \left\{ \left[\mathbf{L}_{\pi'}(z,x) - \frac{d}{dx}\mathbf{L}_{\pi'_x}(z,x)\right] \delta\pi'(x) + \mathbf{L}_{K'}(z,x)\delta K' \right\} dx \tag{B.14}$$
$$+\mathbf{L}_{\pi'_x}(\underline{z},x)|_{\underline{x}(\underline{z})}\delta\underline{\pi}' + \left[\mathbf{L}(\underline{z},x) - \pi'_x(x)\mathbf{L}_{\pi'_x}(\underline{z},x)\right]_{\underline{x}(\underline{z})} \underline{x}_{K'}\delta K'$$
$$+\mathbf{L}_{\pi'_x}(\overline{z},x)|_{\overline{x}(\overline{z})}\delta\overline{\pi}' + \left[\mathbf{L}(\overline{z},x) - \pi'_x(x)\mathbf{L}_{\pi'_x}(\overline{z},x)\right]_{\overline{x}(\overline{z})} \left\{\left(\overline{x}_{K'} + \overline{x}_k\overline{\omega}_{K'}\right)\delta K' + \overline{x}_k\overline{\omega}_{\pi'}\delta\overline{\pi}'\right\}.$$

The first-order conditions are

$$\sum_{z\in Z} \left[\mathbf{L}_{\pi'}(z,x) + \Psi\mathcal{K}_{\pi'}(z,x) - \frac{d}{dx}\left(\mathbf{L}_{\pi'_x}(z,x) + \Psi\mathcal{K}_{\pi'_x}(z,x)\right)\right] = 0, \tag{B.15}$$
$$\left[\mathbf{L}_{\pi'_x}(\underline{z},x) + \Psi\mathcal{K}_{\pi'_x}(\underline{z},x)\right]|_{\underline{x}(\underline{z})} = 0,$$
$$\left[\mathbf{L}_{\pi'_x}(\overline{z},x) + \Psi\mathcal{K}_{\pi'_x}(\overline{z},x)\right]|_{\overline{x}(\overline{z})} + \left[\frac{1}{\overline{x}_k\overline{\omega}_{\pi'}} - \pi'_x(x)\right]\left[\mathbf{L}_{\pi'_x}(\overline{z},x) + \Psi\mathcal{K}_{\pi'_x}(\overline{z},x)\right]_{\overline{x}(\overline{z})} = 0,$$

where

$$\Psi \equiv \Psi_{K'} \left\{ \sum_{z\in Z} \int_{\underline{x}(z)}^{\overline{x}(z)} \mathbf{L}_{K'}(z,x)dx + \left[\mathbf{L}(\underline{z},x) - \pi'_x(x)\mathbf{L}_{\pi'_x}(\underline{z},x)\right]_{\underline{x}(\underline{z})} \underline{x}_{K'} \right.$$
$$\left. + \left[\mathbf{L}(\overline{z},x) - \pi'_x(x)\mathbf{L}_{\pi'_x}(\overline{z},x)\right]_{\overline{x}(\overline{z})} \left(\overline{x}_{K'} + \overline{x}_k\overline{\omega}_{K'}\right) \right\}.$$

Since we are interested in the limiting tax policy function $\pi \equiv \lim_{t+1\to\infty} \pi_{t+1}$, we need to impose that the derived first-order conditions in (B.15) are at a steady state related to π, i.e. $K = K'$. Recall that as the modified Lagrange function is defined by $\widetilde{\mathbf{L}}(z,x) \equiv \mathbf{L}(z,x) + \Psi\mathcal{K}(z,x)$, then clearly the derived first-order conditions are the same as in (25)-(29). For the case with the endogenous lower bound the first-order conditions are given by (B.13). Q.E.D.

B.1.1 Definition of Terms in Theorem 1

Recall that according to (B.1)

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

and

$$\begin{aligned}\mathcal{W}\left[z,x;\pi',\pi_x',K'\right] &= \left\{u\left[c\left(z,k(z,x);\pi',\pi_x',K'\right)\right]+\beta\sum_{z'\in Z}u\left[c\left(z',h(z,k(z,x);\pi',\pi_x',K'\right)\right]Q(z,z')\right\} \\ &\quad \cdot\lambda\left[z,k(z,x);\pi',\pi_x',K'\right]\ k_x(z,x).\end{aligned}$$

The limit

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

with

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

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

To simplify notation we will omit the obvious arguments. Thus,

$$\mathcal{W}_{\pi'}=\left\{u_c\left(c\right)c_{\pi'}+\beta\sum_{z'\in Z}Q(z,z')\left[u_c\left(c'\right)c_{\pi'}'-\varphi_k h_{\pi'}\right]\right\}\lambda k_x, \tag{B.16}$$

where $c_{\pi'}=-h_{\pi'}$, $c_{\pi'}'=\left[1+r(K')\right]h_{\pi'}$. The Fréchet derivative of the savings policy $h_{\pi'}$ is given by Lemmas 1 and 2, respectively. We have used the fact that $\delta\pi'(x')=\delta\pi'\left(x\left[z',h(z,k(z,x))\right]\right)=\delta\pi'\left(\varphi(x)\right)=\varphi_k(x)h_{\pi'}\delta\pi'(x)$, where $\varphi_k(x)=\pi_x'\left(x\left[z',h(z,k(z,x))\right]\right)x_k\left[z',h(z,k(z,x))\right]=\pi_x'(x')x_k'$.

Similarly,

$$\mathcal{W}_{\pi_x'}=\left\{u_c\left(c\right)c_{\pi_x'}+\beta\sum_{z'\in Z}Q(z,z')\left[u_c\left(c'\right)c_{\pi_x'}'-\varphi_k h_{\pi_x'}\right]\right\}\lambda k_x,$$

where $c_{\pi_x'}=-h_{\pi_x'}$, $c_{\pi_x'}'=\left[1+r(K')\right]h_{\pi_x'}$, and the Fréchet derivative $h_{\pi_x'}$. Finally, for $h_{K'}$

$$\mathcal{W}_{K'}=\left\{u_c\left(c\right)c_{K'}+\beta\sum_{z'\in Z}Q(z,z')\left[u_c\left(c'\right)c_{K'}'-\varphi_k h_{K'}\right]\right\}\lambda k_x,$$

where

$$\begin{aligned}c_{K'} &= -h_{K'}\\ r_{K'} &= F_{11}\left(K',\overline{L}\right),\\ w_{K'} &= F_{21}\left(K',\overline{L}\right),\\ c_{K'} &= r_{K'}(K')h+w_{K'}(K')z'+\left[1+r(K')\right]h_{K'}.\end{aligned}$$

For the equation

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

we obtain,

$$\mathcal{W}_{\pi'_x\pi'_x}(z,x) = \left\{u_c c_{\pi'_x\pi'_x} + u_{cc}\left(c_{\pi'_x}\right)^2 + \beta\sum_{z'\in Z} Q(z,z')\left[u'_c c'_{\pi'_x\pi'_x} + u'_{cc}\left(c'_{\pi'_x}\right)^2 - \varphi_{kk}\left(h_{\pi'_x}\right)^2 - \varphi_k h_{\pi'_x\pi'_x}\right]\right\}\lambda k_x,$$

$$\mathcal{W}_{\pi'_x\pi'}(z,x) = \left\{u_c c_{\pi'_x\pi'} + u'_{cc}c'_{\pi'_x}c'_{\pi'} + \beta\sum_{z'\in Z} Q(z,z')\left[u'_c c'_{\pi'_x\pi'} + u'_{cc}c'_{\pi'_x}c'_{\pi'} - \varphi_{kk}h_{\pi'_x}h_{\pi'} - \varphi_k h_{\pi'_x\pi'}\right]\right\}\lambda k_x,$$

$$\mathcal{W}_{\pi'_x x}(z,x) = \left\{u_c c_{\pi'_x x} + u_{cc}c_{\pi'_x}c_x + \beta\sum_{z'\in Z} Q(z,z')\left[u'_c c'_{\pi'_x x} + u'_{cc}c'_{\pi'_x}c'_x - \varphi_{kk}h_{\pi'_x}h_k k_x - \varphi_k h_{\pi'_x x}\right]\right\}\lambda k_x$$
$$+\left\{u_c\left(c\right)c_{\pi'_x} + \beta\sum_{z'\in Z} Q(z,z')\left[u_c\left(c'\right)c'_{\pi'_x} - \pi'_x(x')x'_k h_{\pi'_x}\right]\right\}\left(\lambda_x k_x - \lambda k_{xx}\right),$$

with $c_{\pi'_x\pi'} = -h_{\pi'_x\pi'}$, $c_{\pi'_x\pi'_x} = -h_{\pi'_x\pi'_x}$, $c'_{\pi'_x\pi'_x} = (1+r(K'))h_{\pi'_x\pi'_x}$, $c'_{\pi'_x\pi'} = (1+r(K'))h_{\pi'_x\pi'}$,$c'_x = -\pi'_x$, and $\varphi_{kk}(x) = \pi'_x\left(x\left[z', h(z,k(z,x))\right]\right)x_k\left[z', h(z,k(z,x))\right] = \pi'_{xx}(x')(x'_k)^2 + \pi'_x(x')x'_{kk}$.

According to (21),

$$\begin{aligned}\mathcal{G}[z,x;\pi',\pi'_x,K'] &= \sum_{z'\in Z}\left[\pi'\left(x\left(z', h\left(z, x(z,k);\pi',\pi'_x,K'\right)\right)\right) - gy\left(z', h\left(z, x(z,k);\pi',\pi'_x,K'\right);K'\right)\right] \\ &\quad \cdot Q(z,z')\lambda(z,k(z,x))k_x(z,x),\end{aligned}$$

and therefore,

$$\begin{aligned}\mathcal{G}_{\pi'}(z,x) &= \sum_{z'\in Z} Q(z,z')\left\{\varphi_k h_{\pi'} - gy'_{\pi'}\right\}\lambda k_x, \\ \mathcal{G}_{\pi'_x}(z,x) &= \sum_{z'\in Z} Q(z,z')\left\{\varphi_k h_{\pi'_x} - gy'_{\pi'_x}\right\}\lambda k_x, \qquad \text{(B.17)} \\ \mathcal{G}_{K'}(z,x) &= \sum_{z'\in Z} Q(z,z')\left\{\varphi_k h_{K'} - gy'_{K'}\right\}\lambda k_x,\end{aligned}$$

$y'_{\pi'} = (1+r(K'))h_{\pi'}$, $y'_{\pi'_x} = (1+r(K'))h_{\pi'_x}$ and $y'_{K'} = r_{K'}(K')h + w_{K'}(K')z' + (1+r(K'))h_{K'}$. Further,

$$\begin{aligned}\frac{d}{dx'}\mathcal{G}_{\pi'_x}(z,x) &= \mathcal{G}_{\pi'_x\pi'_x}(z',x')\pi'_{xx}(x) + \mathcal{G}_{\pi'_x\pi'}(z',x')\pi'_x(x') + \mathcal{G}_{\pi'_x x}(z',x'), \\ \mathcal{G}_{\pi'_x\pi'_x}(z,x) &= \sum_{z'\in Z} Q(z,z')\left\{\varphi_k h_{\pi'_x} - gy'_{\pi'_x}\right\}\lambda k_x, \\ \mathcal{G}_{\pi'_x\pi'}(z,x) &= \sum_{z'\in Z} Q(z,z')\left\{\varphi_{kk}h_{\pi'_x} + \varphi_k h_{\pi'_x\pi'} - gy'_{\pi'_x\pi'}\right\}\lambda k_x, \\ \mathcal{G}_{\pi'_x x}(z,x) &= \sum_{z'\in Z} Q(z,z')\left\{\left[\varphi_{kx}h_{\pi'_x} + \varphi_k h_{\pi'_x k}k_x - gy'_{\pi'_x x}\right]k_x + \left\{\varphi_k h_{K'} - gy'_{K'}\right\}k_{xx}\right\}\lambda,\end{aligned}$$

where $y'_{\pi'_x\pi'_x} = (1+r(K'))h_{\pi'_x\pi'_x}$ and $y'_{\pi'_x\pi'} = (1+r(K'))h_{\pi'_x\pi'}$, $y'_{\pi'_x x} = (1+r(K'))h_{\pi'_x k}k_x$, and $\varphi_{kx}(x) = \left(\pi'_{xx}(x')\left(x'_k\right)^2 + \pi'_x(x')x'_{kk}\right) h_k k_x$.

For the aggregate capital stock according to (20),

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

and thus

$$\begin{aligned} \mathcal{K}_{\pi'}(z,x) &= h_{\pi'}\lambda k_x, \\ \mathcal{K}_{\pi'_x}(z,x) &= h_{\pi'_x}\lambda k_x, \\ \mathcal{K}_{K'}(z,x) &= h_{K'}\lambda k_x, \end{aligned} \tag{B.18}$$

and

$$\begin{aligned} \frac{d}{dx}\mathcal{K}_{\pi'_x}(z,x) &= \mathcal{K}_{\pi'_x\pi'_x}(z,x)\pi'_{xx}(x) + \mathcal{K}_{\pi'_x\pi'}(z,x)\pi'_x(x) + \mathcal{K}_{\pi'_x x}(z,x), \\ \mathcal{K}_{\pi'_x\pi'_x}(z,x) &= h_{\pi'_x\pi'_x}\lambda k_x, \\ \mathcal{K}_{\pi'_x\pi'}(z,x) &= h_{\pi'_x\pi'}\lambda k_x, \\ \mathcal{K}_{\pi'_x x}(z,x) &= h_{\pi'_x}\lambda k_{xx}. \end{aligned}$$

B.2 Proof of Theorem 2

First, we need to prove that the second-order sufficient Legendre condition for a maximum in the optimal tax schedule problem has the form $\sum_{z\in Z} \boldsymbol{L}_{\pi_x\pi_x}(z,x) < 0$ for all $x \in \left[\underline{x}\left(\underline{z}\right), \overline{x}\left(\overline{z}\right)\right]$. See Luenberger (1969), Ok (2007) or Gelfand and Fomin (2000) as standard references.

Applying the second-order Taylor expansion of J,

$$J(\varepsilon) = J(0) + J_\varepsilon(0)\,\varepsilon + J_{\varepsilon\varepsilon}(0)\,\varepsilon^2 + o\left(\varepsilon^2\right),$$

we obtain

$$\begin{aligned} J_{\varepsilon\varepsilon}(0) = {} & \frac{1}{2}\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)} \left\{\boldsymbol{L}_{\pi\pi}(z,x)\left[\delta\pi(x)\right]^2 + 2\boldsymbol{L}_{\pi\pi_x}(z,x)\,\delta\pi(x)\delta\pi_x(x) + \boldsymbol{L}_{\pi_x\pi_x}(z,x)\left[\delta\pi_x(x)\right]^2\right. \\ & \left. +2\boldsymbol{L}_{K\pi}(z,x)\,\delta K\delta\pi(x) + 2\boldsymbol{L}_{K\pi_x}\delta K\delta\pi_x(x) + \boldsymbol{L}_{KK}(z,x)\left(\delta K\right)^2\right\} dx, \end{aligned} \tag{B.19}$$

where

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

The term $\boldsymbol{L}_{K\pi}$ can be written as

$$\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)} \boldsymbol{L}_{K\pi}(z,x)\,\delta K\delta\pi(x)dx = \delta K\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)} \boldsymbol{L}_{K\pi}(z,x)\,\delta\pi(x)dx, \tag{B.20}$$

the term $\boldsymbol{L}_{K\pi_x}$ as

$$\begin{aligned}\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\boldsymbol{L}_{K\pi_x}(z,x)\,\delta K\delta\pi_x(x)dx &= \delta K\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\boldsymbol{L}_{K\pi_x}(z,x)\,\delta\pi_x(x)dx \qquad \text{(B.21)}\\ &= -\delta K\left\{\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\frac{d}{dx}\boldsymbol{L}_{K\pi_x}(z,x)\,\delta\pi(x)dx\right\},\end{aligned}$$

and similarly, the term $\boldsymbol{L}_{KK}$ is,

$$\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\boldsymbol{L}_{KK}(z,x)\,(\delta K)^2\,dx = (\delta K)^2\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\boldsymbol{L}_{KK}(z,x)\,dx.$$

Adding (B.20) and (B.21), we obtain

$$\begin{aligned}&\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\boldsymbol{L}_{K\pi}(z,x)\,\delta K\delta\pi(x)dx + \sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\boldsymbol{L}_{K\pi_x}(z,x)\,\delta K\delta\pi_x(x)dx\\ =\;& \delta K\left\{\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\left[\boldsymbol{L}_{K\pi}(z,x)-\frac{d}{dx}\boldsymbol{L}_{K\pi_x}(z,x)\right]\delta\pi(x)dx\right\}.\end{aligned}$$

Finally,

$$\begin{aligned}\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}2\boldsymbol{L}_{\pi\pi_x}(z,x)\,\delta\pi(x)\delta\pi_x(x) &= \sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\boldsymbol{L}_{\pi\pi_x}(z,x)\,\frac{d}{dx}\left[\delta\pi(x)\right]^2\\ &= -\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\frac{d}{dx}\boldsymbol{L}_{\pi\pi_x}(z,x)\left[\delta\pi(x)\right]^2.\end{aligned}$$

Substituting the variation of the aggregate capital from equation (29) for δK,

$$\delta K = \Psi_K\left\{\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\left[\mathcal{K}_{\pi}(z,x)-\frac{d}{dx}\mathcal{K}_{\pi_x}(z,x)\right]\delta\pi(x)dx\right\}, \qquad \text{(B.22)}$$

the Taylor expansion (B.19) can be expressed as

$$\begin{aligned}J_{\varepsilon\varepsilon}(0) &= \sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\left\{P(z,x)\left[\delta\pi_x(x)\right]^2+Q(z,x)\left[\delta\pi(x)\right]^2\right\}dx \qquad \text{(B.23)}\\ &\quad+\left(\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}A(z,x)\,\delta\pi(x)dx\right)\left(\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}B(z,x)\,\delta\pi(x)dx\right)\\ &\quad+\Phi\left(\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}A(z,x)\,\delta\pi(x)dx\right)^2,\end{aligned}$$

where

$$\begin{aligned}
P(z,x) &\equiv \frac{1}{2}\boldsymbol{L}_{\pi_x\pi_x}(z,x),\\
Q(z,x) &\equiv \frac{1}{2}\left[\boldsymbol{L}_{\pi\pi}(z,x)-\frac{d}{dx}\boldsymbol{L}_{\pi\pi_x}(z,x)\right],\\
A(z,x) &\equiv \Psi_K\left[\mathcal{K}_{\pi}(z,x)-\frac{d}{dx}\mathcal{K}_{\pi_x}(z,x)\right],\\
B(z,x) &\equiv \boldsymbol{L}_{K\pi}(z,x)-\frac{d}{dx}\boldsymbol{L}_{K\pi_x}(z,x),\\
\Phi &\equiv \sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\boldsymbol{L}_{KK}(z,x)\,dx.
\end{aligned}$$

If π is a maximum then for all perturbations $\delta\pi$ the quantity $J_{\varepsilon\varepsilon}(0)$ must be negative. If we consider a family of perturbations $\delta\pi_\varepsilon$ parameterized by a small $\varepsilon>0$, then $\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}P(z,x)\left[\delta\pi_x(x)\right]^2dx$ is not bounded as $\varepsilon\to 0$ due to $\delta\pi_x$ which is of order $1/\varepsilon$. On the other hand, all other terms in (B.23) containing $\delta\pi$ are bounded and, therefore, dominated by $\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}P(z,x)\left[\delta\pi_x(x)\right]^2dx$. This implies that the necessary condition for a maximum is the so called Legendre condition

$$\sum_{z\in Z}\boldsymbol{L}_{\pi_x\pi_x}(z,x)<0 \quad \text{for all } x\in\left[\underline{x}(\underline{z}),\overline{x}(\overline{z})\right].$$

Next we need to prove that this condition is satisfied for our Calculus of Variations Ramsey Problem given in Definition 3. We show that the second-order sufficient condition is satisfied for any shape of the policy schedule π. The definition of $\boldsymbol{L}=\mathcal{W}+\mu\mathcal{G}$ gives us $\boldsymbol{L}_{\pi_x\pi_x}=\mathcal{W}_{\pi_x\pi_x}+\mu\mathcal{G}_{\pi_x\pi_x}$. The expression for $\mathcal{W}$ in (23) leads to

$$\mathcal{W}_{\pi_x}=\left[-u_c h_\pi+\beta\sum_{z'\in Z}\left[u_c'\left(1+r\left(K'\right)\right)-\pi_x'x_k'\right]h_\pi Q\left(z,z'\right)\right]\lambda k_x,$$

which leads to

$$\begin{aligned}
\mathcal{W}_{\pi_x'\pi_x'}(z,x) = \; &\{u_c c_{\pi_x'\pi_x'}+u_{cc}\left(c_{\pi_x'}\right)^2\\
&+\beta\sum_{z'\in Z}Q\left(z,z'\right)\left[u_c'c_{\pi_x'\pi_x'}'+u_{cc}'\left(c_{\pi_x'}'\right)^2-\varphi_{kk}\left(h_{\pi_x'}\right)^2-\varphi_k h_{\pi_x'\pi_x'}\right]\}\lambda k_x.
\end{aligned}$$

Assuming the same specification as the optimal total income tax problem with $c_{\pi_x}=\left[1+r\left(K\right)\right]h_{\pi_x}$ and $c_{\pi_x\pi_x}=\left[1+r\left(K\right)\right]h_{\pi_x\pi_x}$ and $k_x=1/x_k>0$. Since $\mathcal{G}=\left[\pi(x)-gy\right]\lambda k_x$, we obtain

$$\mathcal{G}_{\pi_x}=\sum_{z'\in Z}Q\left(z,z'\right)\left\{\varphi_k h_{\pi'}-gy_{\pi'}'\right\}\lambda k_x$$

and

$$\mathcal{G}_{\pi_x\pi_x}=\sum_{z'\in Z}Q\left(z,z'\right)\left\{\varphi_{kk}h_{\pi_x'}+\varphi_k h_{\pi_x'\pi_x'}-gy_{\pi_x'\pi_x'}'\right\}\lambda k_x,$$

where $y_{\pi_x} = [1 + r(K)]\, h_{\pi_x}$ and $y_{\pi_x\pi_x} = [1 + r(K)]\, h_{\pi_x\pi_x}$.

So we have

$$\begin{aligned} \boldsymbol{L}_{\pi_x\pi_x} &= \{ \left[u_{cc}(c)\,(c_{\pi_x})^2 + u_c(c)\, c_{\pi_x\pi_x} - \mu g y_{\pi_x\pi_x} \right] \lambda \\ &\quad + \left[u_c(c)\, c_{\pi_x} - \mu g y_{\pi_x} \right] \frac{\delta\lambda}{\delta\pi_x} + \left[u(c) + \mu\,(\pi(x) - gy) \right] \frac{\delta^2\lambda}{\delta\pi_x^2} \} k_x. \end{aligned} \tag{B.24}$$

When we consider the higher order terms $c_{\pi_x\pi_x}$, $y_{\pi_x\pi_x}$ and etc negligible, the above equation reduces to the following condition

$$\boldsymbol{L}_{\pi_x\pi_x} = \left\{ u_{cc}(c)\,(c_{\pi_x})^2\, \lambda + [u_c(c) - \mu g]\, c_{\pi_x} \frac{\delta\lambda}{\delta\pi_x} \right\} k_x, \tag{B.25}$$

where the first term is clearly negative due to $u_{cc}(c) < 0$. The sign of the second term follows from the definition of the Frećhet[29] derivative $h_{\pi_x} = \delta h / \delta\pi_x$, where the policy function for the next-period capital depends implicitly on the tax schedule function and its derivative in the Euler equation operator $\mathcal{F}$. $\frac{\delta\lambda}{\delta\pi_x} \equiv (\lambda_{\pi_x} + \lambda_k h_{\pi_x})$ We consider what is the effect on the next-period capital function h when π_x shifts marginally. If this marginal shift of π_x lowers savings, $h_{\pi_x} < 0$, the change in the related segment of the stationary distribution has the same sign, $\lambda_{\pi_x} < 0$. Vice versa, an increase in savings, $h_{\pi_x} > 0$, implies $\lambda_{\pi_x} > 0$. Because from (B.16) the effect on consumption $c_{\pi_x} = -h_{\pi_x}$, the second term in (B.25), $c_{\pi_x}\frac{\delta\lambda}{\delta\pi_x}$, is always negative. Therefore, $\boldsymbol{L}_{\pi_x\pi_x}(z, k; \pi) < 0$ for each $z \in Z$ and $k \in \left[\underline{k}(\underline{z}), \overline{k}(\overline{z}) \right]$, thus $\sum_{z \in Z} \boldsymbol{L}_{\pi_x\pi_x}(z, k) < 0$ for each $k \in \left[\underline{k}(\underline{z}), \overline{k}(\overline{z}) \right]$.

Second, we need to prove that there are no conjugate points on the interval $[\underline{x}(\underline{z}), \overline{x}(\overline{z}))$. Following Liberzon (2011) and considering only the second-order terms, equation (B.23) can be written as

$$\int_a^b \left\{ R(x)\,[\delta\pi(x)]^2 + S(x)\,[\delta\pi_x(x)]^2 \right\} dx,$$

where

$$\begin{aligned} R(x) &\equiv \sum_{z \in Z} Q(z, x) = \frac{1}{2} \sum_{z \in Z} \left[\boldsymbol{L}_{\pi\pi}(z, x) - \frac{d}{dx} \boldsymbol{L}_{\pi\pi_x}(z, x) \right], \\ S(x) &\equiv \sum_{z \in Z} P(z, x) = \frac{1}{2} \sum_{z \in Z} \boldsymbol{L}_{\pi_x\pi_x}(z, x), \end{aligned}$$

with $a \equiv \underline{x}(\underline{z})$ and $b \equiv \overline{x}(\overline{z})$.

For every differentiable function $w = w(x)$ we know that

$$0 = \left[w\delta\pi^2 \right]_a^b = \int_a^b \frac{d}{dx}\left(w\delta\pi^2 \right) dx = \int_a^b \left\{ w_x \delta\pi^2 + 2w\delta\pi\delta\pi_x \right\} dx.$$

[29] For this definition we use the Gateaux derivative which is the same as the Fréchet derivative if it is continuous.

As $\delta\pi(a) = \delta\pi(b) = 0$,

$$\int_a^b \left\{ R(x) \left[\delta\pi(x)\right]^2 + S(x) \left[\delta\pi_x(x)\right]^2 \right\} dx = \int_a^b \left\{ \left[R(x) + w_x(x)\right] \left[\delta\pi(x)\right]^2 \right.$$
$$\left. +2w(x)\delta\pi(x)\delta\pi_x(x) + S(x) \left[\delta\pi_x(x)\right]^2 \right\} dx.$$

Finding a w to make the integrand on the right-hand side into a perfect square means that

$$S\left(R + w_x\right) = w^2, \tag{B.26}$$

which is a quadratic differential equation for the unknown function w

$$\int_a^b S(x) \left[\frac{w(x)}{S(x)} \delta\pi(x) + \delta\pi_x(x) \right]^2 dx$$

with $S(x) < 0$. Excluding the conjugate points is equivalent to ensure the existence of a solution to equation (B.26) on the whole interval $[a, b]$. The first step is to transform the quadratic differential equation (B.26) into a linear second-order differential equation using a substitution

$$w(x) = -\frac{S(x)v_x(x)}{v(x)}, \tag{B.27}$$

where v is a new unknown twice differentiable function. In this way, equation (B.26) transforms into

$$S\left(R - \frac{\frac{d}{dx}\left(Sv_x\right)v - S\left(v_x\right)^2}{v^2}\right) = \frac{S^2\left(v_x\right)^2}{v^2}.$$

Multiplying both sides by v (nonzero) and dividing by S (negative) yields a Jacobi equation

$$Rv = \frac{d}{dx}\left(Sv_x\right). \tag{B.28}$$

Now we need to prove that any solution v to the Jacobi equation does not vanish anywhere on $[a, b]$ since then the desired solution w to the original Ricatti equation (B.26) is given by equation (B.27). Under the condition that R is close to zero in comparison to S (results will be qualitatively the same if R is bounded and small relative to S), equation (B.28) simplifies to $d/dx\left(Sv_x\right) = 0$, implying that $Sv' = A$ where A is a constant, and

$$v(x) = \int_a^x \frac{A}{S(t)} dt.$$

As function v is arbitrary, we can assume initial conditions $v(a) = 0$ and $v_x(a) = 1$. Using the condition $v_x(a) = 1$, we have $S(a) = A$ and

$$v(x) = \int_a^x \frac{S(a)}{S(t)} dt.$$

Now we need to analyze $S(t)$ for $t \in [a, b]$ and show that it is bounded from below and above, i.e.

$$0 > -\underline{S} \geq -S\left(t\right) \geq -\overline{S} > -\infty. \tag{B.29}$$

From (B.25)

$$\begin{aligned} S(x) &= \frac{1}{2}\sum_{z\in Z}\left\{u_{cc}\left(c\left(z,k(x)\right)\right)\left(c_{\pi_x}\left(z,k(x)\right)\right)^2\lambda\left(z,k(x)\right)\right. \\ &\left. +2u_c\left(c\left(z,k(x)\right)\right)c_{\pi'}\left(z,k(x)\right)\lambda_{\pi_x}\left(z,k(x)\right)\right\}. \end{aligned} \tag{B.30}$$

Consider first the terms with consumption. Using the Regularity Condition, the lowest value of consumption is when a household is borrowing constrained at $k = 0$ with total income $\underline{c} = (1-\varepsilon)\, w\underline{z}$ for any $0 < \varepsilon < 1$. Clearly, consumption cannot be larger than the maximum income equal to $\overline{c} = w\overline{z} + r\overline{k}$. Thus, $0 < \underline{c} \leq c \leq \overline{c} < \infty$ and $-\infty < u_{cc}\left(\underline{c}\right) \leq u_{cc}\left(c\right) \leq u_{cc}\left(\overline{c}\right) < 0$. λ is equal to zero at the endogenous upper bound, $\overline{k}$. For the same reason λ_{π_x} is different from zero. For $z = \underline{z}$ and $k \in \left[0, \underline{k}\left(\underline{z}\right)\right]$, $k' = 0$ and $h_{\pi_x}\left(\underline{z}, k\right) = -c_{\pi_x}\left(\underline{z}, k\right) = 0$ but for $z = \overline{z}$, $c_{\pi_x}\left(\overline{z}, k\right)$ is nonzero for any k. Thus, S given by (B.30) is bounded as in (B.29) and

$$v(x) = \int_a^x \frac{S\left(a\right)}{S\left(t\right)}dt \geq \int_a^x \frac{S\left(a\right)}{\overline{S}}dt = \frac{S\left(a\right)}{\overline{S}}\left(x-a\right) > 0 \text{ for all } x > a.$$

Therefore, there are no conjugate points to a on the interval $[a, b]$ and the first-order conditions for the optimal income tax schedule given by the Euler-Lagrange condition are also the sufficiency conditions and the obtained solution is a maximum. Q.E.D.

B.3 Proof of Lemma 1

The Euler equation operator with the variation $\tilde{\pi} \equiv \pi + \varepsilon\delta\pi$ is

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

and

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

Using abbreviated notation $x = x(z,k)$, $c = c(z,k)$, $h = h(z,k)$, $h' = h(z',h)$, $x' = x(z',h)$, $y' = y(z',h)$, $c' = c(z',h)$, and $R' = R\left(z',h\right)$,

$$\mathcal{F}_\pi\left(z,k\right) = u_{cc}\left(c\right)c_\pi - \beta\sum_{z'}\left\{u_{cc}\left(c'\right)c'_\pi R' + u_c(c')R'_\pi\right\}Q(z,z') = 0,$$

where

$$\begin{aligned} c &= y - \tilde{\pi}(x) + k - h, \\ y &= rk + wz, \\ c' &= y' - \tilde{\pi}(x') + h - h', \\ R' &= 1 + y'_k - \pi_x(x')\, x'_k. \end{aligned}$$

Terms for $\mathcal{F}_\pi$ above and in equation (30) are

$$\begin{aligned} c_\pi &= -1 - h_\pi, \\ c'_\pi &= [1 + r - h'_k - \pi_x(x')\, x'_k]\, h_\pi - h'_\pi - 1, \\ R'_\pi &= rh_{k\pi} - \pi_{xx}(x')\,(x'_k)^2\, h_\pi - \pi_x(x')\, x'_{kk} h_\pi. \end{aligned}$$

For

$$\mathcal{F}_{\pi_x}(z,k) = u_{cc}(c)\, c_{\pi_x} - \beta \sum_{z'} \left\{ u_{cc}(c')\, c'_{\pi_x} R' + u_c(c')\, R'_{\pi_x} \right\} Q(z,z') = 0,$$

we use terms

$$\begin{aligned} c_{\pi_x} &= -h_{\pi_x}, \\ c'_{\pi_x} &= [1 + r - h'_k - \pi_x(x')\, x'_k]\, h_{\pi_x} - h'_{\pi_x}, \\ R'_{\pi_x} &= rh_{k\pi_x} - x'_k - \pi_{xx}(x')\,(x'_k)^2 h_{\pi_x} - \pi_x(x')\, x'_{kk} h_{\pi_x}. \end{aligned}$$

For

$$\mathcal{F}_K(z,k) = u_{cc}(c) c_K - \beta \sum_{z'} \{ u_{cc}(c')\, c'_K R' + u_c(c')\, R'_K \} Q(z,z') = 0,$$

the terms are, as well as for equation (30) in the case of $\mathcal{F}_K$,

$$\begin{aligned} c_K &= r_K k + w_K z - \pi_x(x) x_K - h_K, \\ c'_K &= r_K h + w_K z' - \pi_x(x')\, [x'_K + x'_k h_K] - h'_K, \\ R'_K &= r_K h_k + rh_{kK} - \pi_{xx}(x')\, x'_k\, [x'_K + x'_k h_K] - \pi_x(x')\, [x'_{kK} + x'_{kk} h_K]\,. \end{aligned}$$

For

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

we use, as well in equation (30) in the case of $\mathcal{F}_{\pi_x\pi_x}$, the terms

$$
\begin{aligned}
c_{\pi_x\pi_x} &= -h_{\pi_x\pi_x}, \\
c'_{\pi_x\pi_x} &= \left[1+r-\left(h'_{k\pi_x}+h'_{kk}h_{\pi_x}\right)\right]h_{\pi_x}+\left(1+r-h'_k\right)h_{\pi_x\pi_x}-\left(h'_{\pi_x k}h_{\pi_x}+h'_{\pi_x\pi_x}\right) \\
&- \pi_{xx}\left(x'\right)\left(x'_k h_{\pi_x}\right)^2-\pi_x\left(x'\right)\left(x'_{kk}\left(h_{\pi_x}\right)^2+x'_k h_{\pi_x\pi_x}\right), \\
R'_{\pi_x\pi_x} &= rh_{k\pi_x\pi_x}-2x'_{kk}h_{\pi_x}-\pi_{xxx}\left(x'\right)\left(x'_k\right)^3\left(h_{\pi_x}\right)^2 \\
&- \pi_{xx}\left(x'\right)x'_k\left[3x'_{kk}\left(h_{\pi_x}\right)^2+x'_k h_{\pi_x\pi_x}\right]-\pi_x\left(x'\right)\left[x'_{kkk}\left(h_{\pi_x}\right)^2+x'_{kk}h_{\pi_x\pi_x}\right].
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\mathcal{F}_{\pi_x\pi}\left(z,k\right) &= u_{ccc}\left(c\right)c_{\pi_x}c_\pi+u_{cc}\left(c\right)c_{\pi_x\pi}-\beta\sum_{z'}\left\{\left[u_{ccc}\left(c'\right)c'_{\pi_x}c'_\pi+u_{cc}\left(c'\right)c'_{\pi_x\pi}\right]R'\right. \\
&+ \left.u_{cc}\left(c'\right)\left[c'_{\pi_x}R'_\pi+c'_\pi R'_{\pi_x}\right]+u_c\left(c'\right)R'_{\pi_x\pi}\right\}Q(z,z')=0.
\end{aligned}
$$

Terms for $\mathcal{F}_{\pi_x\pi}$ above and in equation (30) are

$$
\begin{aligned}
c_{\pi_x\pi} &= -h_{\pi_x\pi}, \\
c'_{\pi_x\pi} &= -\left(h'_{k\pi}+h'_{kk}h_\pi\right)h_{\pi_x}+\left(1+r-h'_k\right)h_{\pi_x\pi}-\left(h'_{\pi_x k}h_\pi+h'_{\pi_x\pi}\right) \\
&- \left[\pi_{xx}\left(x'\right)\left(x'_k\right)^2+x'_{kk}\right]h_{\pi_x}h_\pi-\pi_x\left(x'\right)x'_k h_{\pi_x\pi}, \\
R'_{\pi_x\pi} &= rh_{k\pi_x\pi}-x'_{kk}h_\pi-\pi_{xxx}\left(x'\right)\left(x'_k\right)^3 h_{\pi_x}h_\pi \\
&- \pi_{xx}\left(x'\right)x'_k\left[2x'_{kk}h_{\pi_x}h_\pi+x'_k h_{\pi_x\pi}+x'_{kk}h_{\pi_x}h_\pi\right] \\
&- \pi_x\left(x'\right)\left[x'_{kkk}h_{\pi_x}h_\pi+x'_{kk}h_{\pi_x\pi}\right].
\end{aligned}
$$

Q.E.D.

B.4 Proof of Lemma 2

The stationary distribution operator is

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

and so

$$
\mathcal{L}_\pi\left(z',k'\right)\delta\pi=\lim_{\varepsilon\to 0}\frac{d\mathcal{L}\left[z',k';K,\tilde{\pi}(x),\pi_x(x)\right]}{d\varepsilon},
$$

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

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

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

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

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

with $h^{-1}_{\pi_x}(z,k') = h_{\pi_x}\left(z, h^{-1}(z,k')\right) / h_k\left(z, h^{-1}(z,k')\right)$, and

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

with $h^{-1}_K(z,k') = h_K\left(z, h^{-1}(z,k')\right) / h_k\left(z, h^{-1}(z,k')\right)$.

Further, we can derive the following total F-derivative of the stationary distribution operator

$$\begin{aligned} \mathcal{L}_{\pi_x\pi_x}(z',k') &= \lambda_{\pi_{xx}}(z',k') - \sum_z \left\{\lambda_{\pi_x\pi_x}\left(z, h^{-1}(z,k')\right) + \lambda_{\pi_x k}\left(z, h^{-1}(z,k')\right) h^{-1}_{\pi_x}(z,k')\right. \\ &+ \left[\lambda_{\pi_x k}\left(z, h^{-1}(z,k')\right) + \lambda_{kk}\left(z, h^{-1}(z,k')\right)\right] \left[h^{-1}_{\pi_x}(z,k')\right]^2 \\ &+ \left.\lambda_k\left(z, h^{-1}(z,k')\right) h^{-1}_{\pi_x\pi_x}(z,k')\right\} Q(z,z') = 0 \end{aligned}$$

where $h^{-1}_{\pi_x\pi_x}(z,k') = h_{\pi_x\pi_x}\left(z, h^{-1}(z,k')\right) / h_{kk}\left(z, h^{-1}(z,k')\right)$.

Finally,

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

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

Q.E.D.

B.5 Proof of Proposition 1

Using the definition for $\widetilde{\mathbf{L}}$ in equation (24), we get

$$\widetilde{\mathbf{L}}\left(\underline{z}, \underline{x}(\underline{z})\right) = \left\{u\left(c\left(\underline{z}, \underline{k}\right)\right) + \mu\left[\pi(\underline{x}(\underline{z})) - gy\left(\underline{z}, \underline{k}\right)\right] + \Psi\underline{k}\right\} \lambda\left(\underline{z}, \underline{k}\right) = 0,$$

where we have used (21)-(23),for $\mathcal{W}$, $\mathcal{G}$, and $\mathcal{K}$, respectively, and also the fact that at the lower endogenous limit on capital, $\underline{k}$, $\lambda\left(\underline{z}, \underline{k}\right) = 0$. Thus the first-order boundary condition (26) becomes

$$\left(\pi_x(\underline{x}(\underline{z})) - \frac{k_x(\underline{z}, \underline{x}(\underline{z}))}{\omega_\pi(\underline{z}, \underline{x}(\underline{z}))}\right) \widetilde{\mathbf{L}}_{\pi_x}(\underline{z}, \underline{x}(\underline{z})) = 0.$$

If we assume that $\pi_x(\underline{x}(\underline{z})) - \frac{k_x(\underline{z}, \underline{x}(\underline{z}))}{\omega_\pi(\underline{z}, \underline{x}(\underline{z}))} \neq 0$, then the condition above is satisfied if the second term $\widetilde{\mathbf{L}}_{\pi_x}(\underline{z}, \underline{x}(\underline{z}))$ is equal to zero. By the inspection of $\mathcal{W}_{\pi_x}$ we see that since $\lambda\left(\underline{z}, \underline{k}\right) = 0$ then again the first term is zero and thus

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

where we again used the fact that in general $\lambda_{\pi_x}(\underline{z}, \underline{k})$, $\lambda_k(\underline{z}, \underline{k})$, $h_{\pi_x}(\underline{z}, \underline{k})$ and $k_x(\underline{z}, \underline{x}(\underline{z}))$ are non-zero. This implies the result of the Proposition.

Looking at the first-order boundary condition for the exogenously given lower bound we see that it leads to the exactly same condition as before

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

and thus he result of the Proposition follows. Q.E.D.

B.6 Proof of Proposition 2

We see that terms $\widetilde{\mathbf{L}}$ and $\widetilde{\mathbf{L}}_{\pi_x}$ appear in the boundary first-order condition in (26). Using the definition of $\widetilde{\mathbf{L}}$ in (24), we get

$$\widetilde{\mathbf{L}}(\overline{z}, \overline{x}(\overline{z})) = \left\{u\left(c\left(\overline{z}, \overline{k}\right)\right) + \mu\left[\pi(\overline{x}(\overline{z})) - gy\left(\overline{z}, \overline{k}\right)\right] + \Psi\overline{k}\right\}\lambda\left(\overline{z}, \overline{k}\right) = 0,$$

where we have used (21)-(23) for $\mathcal{W}$, $\mathcal{G}$, and $\mathcal{K}$, respectively, and also the fact that at the upper endogenous limit on capital, $\overline{k}$, $\lambda\left(\overline{z}, \overline{k}\right) = 0$. Thus the first-order boundary condition (26) becomes

$$\left(\pi_x(\overline{x}(\overline{z})) - \frac{k_x(\overline{z}, \overline{x}(\overline{z}))}{\omega_\pi(\overline{z}, \overline{x}(\overline{z}))}\right)\widetilde{\mathbf{L}}_{\pi_x}(\overline{z}, \overline{x}(\overline{z})) = 0.$$

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

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

where we again used the fact that in general $\lambda_{\pi_x}\left(\overline{z}, \overline{k}\right)$, $\lambda_k\left(\overline{z}, \overline{k}\right)$, $h_{\pi_x}\left(\overline{z}, \overline{k}\right)$ and $k_x(\overline{z}, \underline{x}(\overline{z}))$ are non-zero. This implies the result of the Proposition. Q.E.D.

B.7 Proof of Theorem 3

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: $\Psi = \{x \in B : 0 \leq x \leq \psi(k, z)$ 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 $u : \mathbb{R}_+ \to \mathbb{R}$ 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 $\bar{k} = \max\{\bar{k}(z_1), \ldots, \bar{k}(z_J)\}$; and, (iv) given the tax-schedule function the right-hand side of the Euler equation is strictly positive

$$\beta \sum_{z'} u_c(\psi(h(k, z), z') - h(h(k, z), z'))\psi_1(h(k, z), z')\, Q(z, z') > 0,$$

where

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

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 $\lim_{c\to 0} u_c(c) = \infty$.

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

$$\begin{aligned} u_c(\psi(k,z) - (Th)(k,z)) &= \beta \sum_{z'} u_c(\psi((Th)(k,z),z') - h[(Th)(k,z),z']) \\ &\quad \cdot [(1 + r - \tau_y(y((Th)(k,z),z'))r]\, Q(z,z'). \end{aligned}$$

Then it is easy to prove that T is well defined, continuous and that $T : F \to F$. 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 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 (25);
2. the Euler equation (15) capturing the individual optimal behavior h;
3. five operator equations (30)-(31) for F-derivatives of h based on the Euler equation, h_π, h_{π_x}, h_K, $h_{\pi_x\pi_x}$, and $h_{\pi_x\pi}$;
4. the operator equation for distribution function, λ, in (16); and
5. five operator equations (32)-(33) for F-derivatives of λ based on the operator equation for distribution function: λ_π, λ_{π_x}, λ_K, $\lambda_{\pi_x\pi_x}$, and $\lambda_{\pi_x\pi}$.

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 possibil-

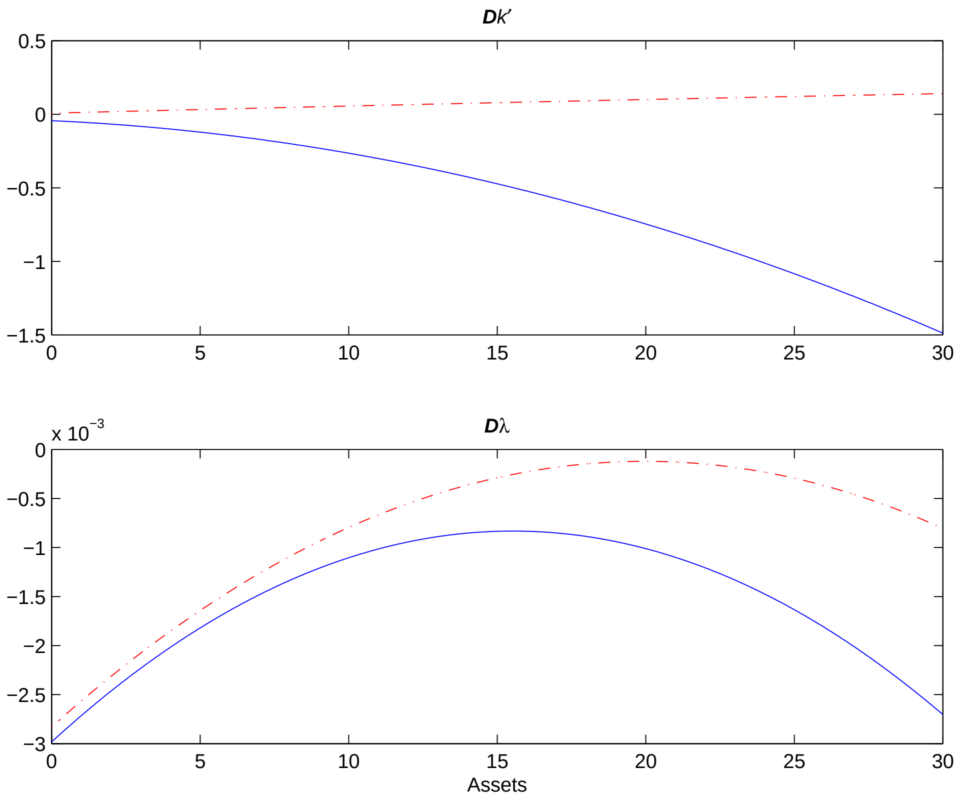

Figure C.1: Savings and distribution sensitivity functions for highest (-) and lowest (-.-) productivity shocks.

ities how to define the projections.[30] 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 (15), (16), (25), (30)-(31), and (32)-(33), there are thirteen unknown classes of functions $\{\pi, h, h_\pi, h_{\pi_x}, h_K, h_{\pi_x\pi_x}, h_{\pi_x\pi}, \lambda, \lambda_\pi, \lambda_{\pi_x}, \lambda_K, \lambda_{\pi_x\pi_x}, \lambda_{\pi_x\pi}\}$. Since we assume that the shocks are discrete, $z \in Z = \{z_1, z_2, \ldots, z_J\}$ and $J > 1$, we define the following family of policy and distribution functions, and their derivatives $\{h^i(k), \lambda^i(k), h^i_\pi(k), h^i_{\pi_x}(k), h^i_K(k), h^i_{\pi_x\pi_x}(k), h^i_{\pi_x\pi}(k), \lambda^i_\pi(k), \lambda^i_{\pi_x}(k), \lambda^i_K(k), \lambda^i_{\pi_x\pi_x}(k), \lambda^i_{\pi_x\pi}(k)\}_{i=1}^J$, for each shock value $z_1, z_2, \ldots, z_J$. 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 $\widehat{\overline{k}}$, greater than the endogenous upper bound on the stationary distribu-

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

tion. Let the interval of approximation be $[\underline{k}, \widehat{\overline{k}}]$ and the degrees of approximation for $\{h^i(k), h^i_\pi(k), h^i_{\pi_x}(k), h^i_K(k), h^i_{\pi_x\pi_x}(k), h^i_{\pi_x\pi}(k), \lambda^i(k), \lambda^i_\pi(k), \lambda^i_{\pi_x}(k), \lambda^i_K(k), \lambda^i_{\pi_x\pi_x}(k), \lambda^i_{\pi_x\pi}(k)\}$ be $M, M_\pi, M_{\pi_x}, M_K, M_{\pi_x\pi_x}, M_{\pi_x\pi}, N, N_\pi, N_{\pi_x}, N_K, N_{\pi_x\pi_x}, N_{\pi_x\pi}, P \geq 2$, respectively.[31]

Thus, we obtain

$$\widehat{h}^i_m(k; a^i_m) \equiv \sum_{j=1}^{M_m} a^i_{m,j}\phi_j(k), \qquad \widehat{\lambda}^i_m(k; b^i_m) \equiv \sum_{j=1}^{N_m} b^i_{m,j}\phi_j(k),$$

with $i \in \{1, 2, ..., J\}$ and $m \in \{\emptyset, \pi, \pi_x, K, \pi_x\pi_x, \pi_x\pi\}$, and $\widehat{\pi}(x; c) \equiv \sum_{j=1}^{P} c_j\phi_j(x)$, 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 (15), (16), (25), (30)-(31), and (32)-(33). Substituting the above approximations for the unknown functions,

$$R^{\widetilde{\mathbf{L}}}(x; \mathbf{p}) = \sum_{z_j} \left[\widetilde{\mathbf{L}}_\pi \left(\widehat{\mathbf{h}}, \widehat{\mathbf{\Lambda}}, \widehat{\pi} \right) + \frac{d}{dx} \widetilde{\mathbf{L}}_{\pi_x} \left(\widehat{\mathbf{h}}, \widehat{\mathbf{\Lambda}}, \widehat{\pi} \right) \right], \tag{C.1}$$

$$R^{\mathcal{F}^i}_m(k; \mathbf{p}) = \mathcal{F}^i_m(\widehat{\mathbf{h}}, \widehat{\pi}), \tag{C.2}$$

$$R^{\mathcal{L}^i}_m(k; \mathbf{p}) = \mathcal{L}^i_m(\widehat{\mathbf{h}}, \widehat{\mathbf{\Lambda}}, \widehat{\pi}), \tag{C.3}$$

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

$$\begin{aligned} \mathbf{p} &\equiv (\mathbf{a}, \mathbf{a}_\pi, \mathbf{a}_{\pi_x}, \mathbf{a}_K, \mathbf{a}_{\pi_x\pi_x}, \mathbf{a}_{\pi_x\pi}, \mathbf{b}, \mathbf{b}_\pi, \mathbf{b}_{\pi_x}, \mathbf{b}_K, \mathbf{b}_{\pi_x\pi_x}, \mathbf{b}_{\pi_x\pi}, \mathbf{c}), \\ \mathbf{a}_m &\equiv (a^1_m, a^2_m, ..., a^J_m), \\ \mathbf{b}_m &\equiv (b^1_m, b^2_m, ..., b^J_m), \end{aligned}$$

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

$$\begin{aligned} \widehat{\mathbf{h}} &\equiv (\widehat{h}, \widehat{h}_\pi, \widehat{h}_{\pi_x}, \widehat{h}_K, \widehat{h}_{\pi_x\pi_x}, \widehat{h}_{\pi_x\pi}), \\ \widehat{\mathbf{h}}_m &\equiv (\widehat{h}^1_m, ..., \widehat{h}^J_m), \\ \widehat{\mathbf{\Lambda}} &\equiv (\widehat{\lambda}, \widehat{\lambda}_\pi, \widehat{\lambda}_{\pi_x}, \widehat{\lambda}_K, \widehat{\lambda}_{\pi_x\pi_x}, \widehat{\lambda}_{\pi_x\pi}), \\ \widehat{\lambda}_m &\equiv (\widehat{\lambda}^1_m, ..., \widehat{\lambda}^J_m), \end{aligned}$$

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

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

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

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

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

$$\min_{\mathbf{p}\in\mathbb{R}^S} \sum_{\check{k}} \left\{ \sum_{i=1}^{J} \sum_{m} \left([R_m^{\mathcal{F}^i}(k;\mathbf{p})]^2 + [R_m^{\mathcal{L}^i}(k;\mathbf{p})]^2 \right) + [R^{\widetilde{\mathbf{L}}}(x(k);\mathbf{p})]^2 \right\}, \tag{C.4}$$

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. $\max\{M, M_\pi, M_{\pi_x}, M_K, M_{\pi_x\pi_x}, M_{\pi_x\pi}, N, N_\pi, N_{\pi_x}, N_K, N_{\pi_x\pi_x}, N_{\pi_x\pi}, P\}$.

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, where 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}.

Figure C.1 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.[32]

D Example: Optimal Tax on Total Income

D.1 Definitions of Terms in Theorem 1

The terms from Theorem 1 for Example are those in Appendix B.1.1 together with $y(z, k(z, x; K)) = x, k_x(z, x; K) = 1/r(K), k_K(z, x; K) = -\{w_K(K) z + r(K) + [x - w(K) z] r_K(K)\}/[r(K)]^2$, and $k_{xK}(z, x; K) = -r_K(K)/[r(K)]^2$.

D.2 Detailed Results

Individual welfare and distributional gains in the steady states are shown in Figure D.1 for agents with the lowest and the highest labor productivity shocks. The steady state gains are obtained from comparing per-period consumption equivalents needed equalize

[32]Our numerical solution is only very close to zero due to approximation errors.

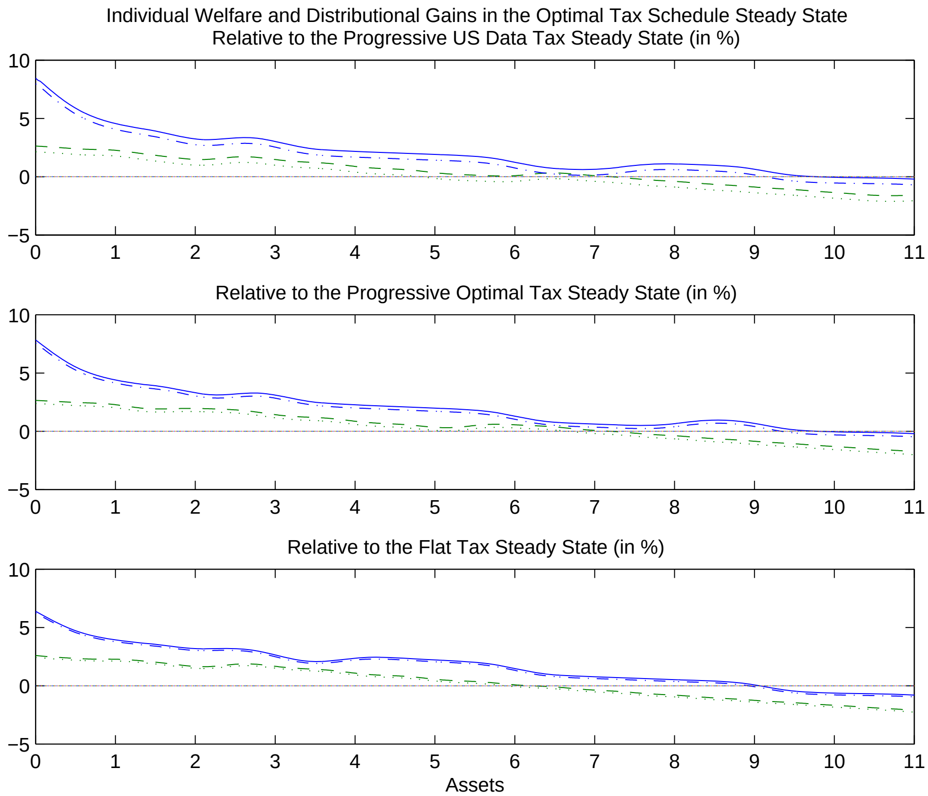

Figure D.1: Individual welfare and distributional gains in the optimal tax schedule steady state. Welfare gains: $z_L(-)$, $z_H(--)$. Distributional gains: $z_L(-\cdot)$, $z_H(\cdot\cdot)$.

the value functions of the progressive tax (with US data and optimal progressivity) and the flat tax steady states to the value function from the optimal tax steady state.

This simple steady state comparison shows that individual welfare gains are decreasing functions of wealth. The asset-poor agents have both welfare and distributional gains while the very rich have losses relative to the steady state with the progressive or flat tax schedules. Agents with a low realization of labor income have larger gains that agents with high labor income. There are two forces present: the tax rate (lower for rich agents in the flat tax steady state) and general equilibrium effects. The large welfare gains for poor agents are mostly due to the higher wage in the optimal steady state. The welfare losses for the asset-rich agents are due to the lower interest rate and their higher tax rate in the optimal tax steady state.

Table D.1 shows the distribution of resources by quintiles of the wealth distribution for each steady state. The top panel shows the average level of capital held by each quintile in the stationary distribution. It illustrates the incentives provided by the optimal limiting tax schedule on savings. The poorest quintile owns on average 85% more assets than in the optimal progressive steady state and more than two times as much as in the flat tax steady state. On the other hand, the top quintile in the optimal steady state owns less assets than in the other economies. Overall, the bottom three quintiles accumulate much more savings.

Compared to the other economies, the average total income is much higher in the

Steady State Distribution of Resources					
	Average Level in Each Quintile				
Tax Schedule	1st	2nd	3rd	4th	5th
			Assets		
Progressive US Data	1.583	3.668	5.518	7.298	11.355
Progressive Optimal	1.726	3.791	5.554	7.600	10.428
Flat Tax	1.336	3.299	5.269	7.893	13.328
Optimal	2.933	5.240	6.660	7.996	9.863
			Total Income		
Progressive US Data	1.256	1.358	1.428	1.506	1.636
Progressive Optimal	1.259	1.361	1.433	1.512	1.658
Flat Tax	1.261	1.360	1.430	1.519	1.700
Optimal	1.322	1.410	1.467	1.518	1.601
			Consumption		
Progressive US Data	0.938	1.007	1.053	1.103	1.185
Progressive Optimal	0.932	1.006	1.056	1.109	1.208
Flat Tax	0.916	0.995	1.048	1.114	1.248
Optimal	0.936	1.022	1.074	1.119	1.187
			Tax Contribution		
Progressive US Data	0.313	0.349	0.375	0.404	0.455
Progressive Optimal	0.321	0.354	0.378	0.404	0.454
Flat Tax	0.338	0.364	0.383	0.406	0.455
Optimal	0.375	0.384	0.393	0.403	0.424

Table D.1: Steady State Distribution of Resources: Average Levels in Quintiles.

bottom quintile of the optimal steady state and is lower in the top quintile. This is due to higher accumulated assets and higher wage in equilibrium. The ratio of total income between the top and the bottom quintile is only 1.21 while in the other economies it is more than 1.3. Despite the incentives by the optimal limiting tax function to save more, agents in the bottom quintile do not consume on average less than in the other economies. This is because the distribution is more concentrated at higher levels of wealth and wages are higher. On the other hand, the consumption is lower in the top quintile. 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 88% of the total average consumption. Similar relationship holds for the distribution of income. The distribution of the tax contributions shows that the optimal tax is more equitable than in the other steady states. The ratio of tax contribution by the top to the bottom quintile is 1.13 while it is 1.45 in the US progressive steady state. Relative to the mean, the bottom quintile pays only 5% less and the top quintile only 7% more than the middle quintile.

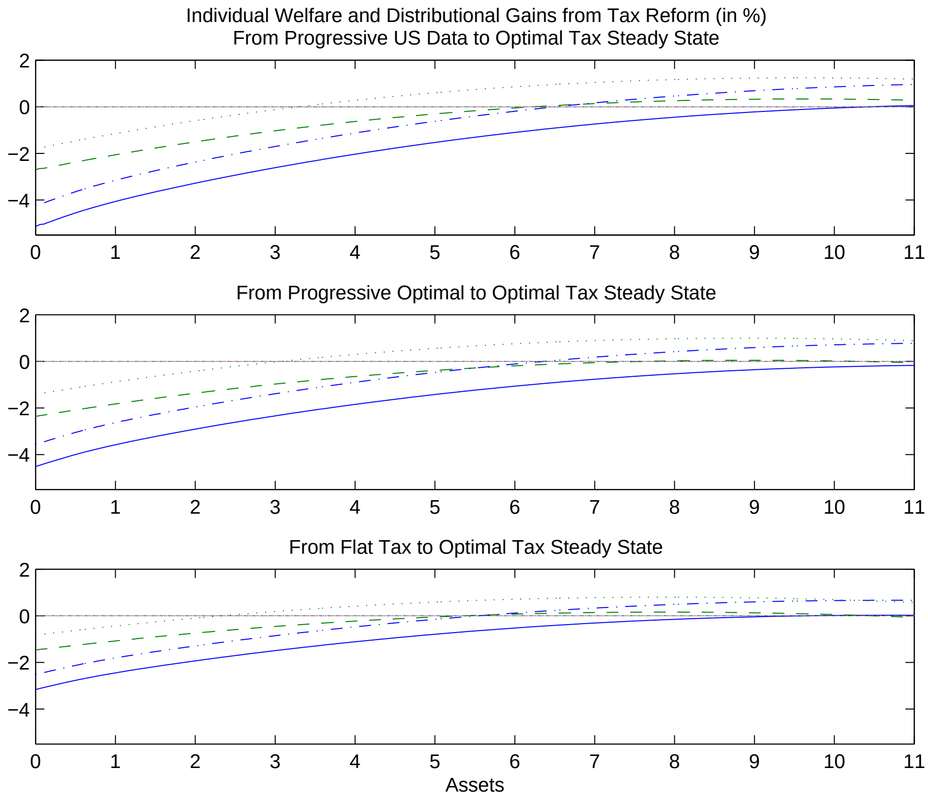

Figure D.2: Individual welfare and distributional gains: Transition from the progressive and the flat tax steady state to the optimal tax schedule steady state. Welfare gains: $z_L(-)$, $z_H(--)$. Distributional gains: $z_L(-\cdot)$, $z_H(\cdot\cdot)$.

D.3 Transition to the Optimal Tax Schedule Steady State

Finally, Figure D.2 shows the efficiency and distributional individual gains from transition presented in Table 3.[33] In contrast to the simple steady state comparison, the Figure shows that the cost of transition falls heavily on the asset-poor agents. The welfare cost of transition is up to 5 percent for the poor agents with a low productivity shock in the initial period of transition. Only the very rich agents are better off from the reform after taking the full cost of transition into account. Note that the distributional costs are smaller but still negative on average. These graphs illustrate the low political support for adopting the limiting optimal tax policy without providing compensatory policies to a large fraction of the agents. These compensation schemes could come from accumulating a debt later repaid from the efficiency gains in the terminal steady state.

[33]These 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 Domeij and Heathcote (2004).

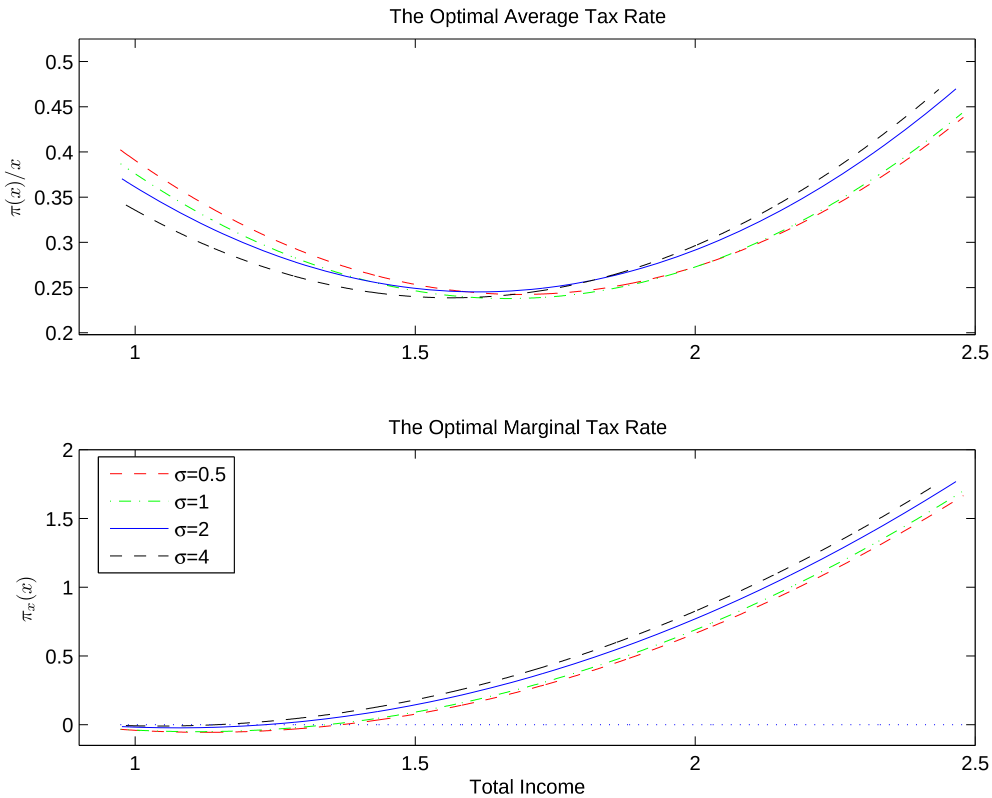

Figure E.1: Sensitivity analysis for different values of σ. The optimal tax schedule and the optimal marginal tax rate. Benchmark $\sigma = 2$.

E Sensitivity Analysis

In this Appendix we analyze the optimal limiting tax for different values of intertemporal elasticity of substitution. Table E.1 shows the steady state and transition allocations for four different values of $\sigma \in \{0.5, 1, 2, 4\}$, where $\sigma = 2$ is the benchmark economy. All other parameters, including $\beta = 0.976$, are kept the same as in the benchmark economy.

Figure E.1 displays the optimal tax schedules and optimal marginal tax schedules for different values of σ. The U-shape of the optimal limiting tax schedule is preserved in all simulated economies. Similar to Imrohoroglu (1998), with higher risk aversion, the agents have their own incentives to accumulate more capital as a buffer stock against the idiosyncratic risk and the optimal tax schedules are less steep at low incomes but steeper at higher incomes. The optimal marginal tax functions are also U-shaped and negative at low levels of income.

In Table E.1, we observe that capital accumulation increases with higher values of σ. The pure steady state welfare gains from the optimal tax reform increase substantially in σ. Similar relationship holds for the efficiency and distributional gains. When transitions from the flat or the progressive steady states to the optimal steady states are taken into account, efficiency and welfare gains become more negative for higher values of σ. Political support for the optimal tax reform is highest for $\sigma = 1/2$ but still remains at about one-third of the population. As in the benchmark economy, the unanticipated reform brings too high short-term welfare costs. The more capital agents need to accumulate on the path to the new steady state, the worse are the welfare outcomes from the transition. Again, the reform is not an optimal as we do not allow the limiting optimal tax policy to change during the transition (especially at initial periods) and we do not allow for debt (that could be financed from the resulting steady-state efficiency gains).

These simulations show that the methodology developed in the paper delivers consistent results in terms of the shape of the optimal tax schedule for the usual values of the intertemporal elasticity of substitution. All other graphs and results similar to those analyzed for the benchmark economy in Appendix D are available upon request.

Sensitivity Analysis σ: Steady State and Transition Results

	$\sigma = 0.5$				$\sigma = 1$				$\sigma = 2$				$\sigma = 4$			
	Progressive		Flat	Opt.	Progressive		Flat	Opt.	Progressive		Flat	Opt.	Progressive		Flat	Opt.
Average	US	Opt.			US	Opt.			US	Opt.			US	Opt.		
τ	0.151	0.097	0.268	—	0.151	0.095	0.268	—	0.151	0.094	0.268	—	0.151	0.092	0.269	—
Capital	5.73	5.90	6.18	6.50	5.74	5.91	6.20	6.51	5.76	5.95	6.25	6.56	5.83	6.05	6.35	6.69
Output	1.86	1.92	1.95	1.98	1.90	1.92	1.95	1.98	1.89	1.92	1.95	1.99	1.91	1.93	1.97	2.00
Capital-output ratio	3.02	3.08	3.17	3.28	3.03	3.09	3.18	3.28	3.04	3.10	3.20	3.30	3.06	3.13	3.23	3.34
Consumption	1.06	1.06	1.06	1.07	1.06	1.06	1.06	1.07	1.06	1.06	1.07	1.07	1.06	1.06	1.07	1.07
Median wealth	5.61	5.86	5.19	6.60	5.51	5.79	5.19	6.63	5.36	5.71	5.27	6.64	5.36	5.63	5.28	6.75
Median/mean wealth ratio	0.98	0.99	0.84	1.02	0.96	0.98	0.84	1.02	0.93	0.96	0.84	1.01	0.92	0.93	0.83	1.01
Interest rate (%)	3.90	3.68	3.34	2.98	3.90	3.67	3.32	2.97	3.86	3.62	3.26	2.92	3.76	3.50	3.15	2.78
Wage	1.19	1.21	1.23	1.25	1.19	1.21	1.23	1.25	1.19	1.21	1.23	1.25	1.20	1.22	1.24	1.26
Gini wealth	0.33	0.32	0.39	0.23	0.33	0.31	0.40	0.22	0.33	0.31	0.39	0.22	0.31	0.30	0.39	0.20
Constrained agents (%)	0.12	0.05	0.47	0.01	0.22	0.13	0.45	0.03	0.26	0.19	0.38	0.08	0.20	0.18	0.27	0.01
Gains from the Optimal Tax Schedule (in %)																
Steady State Gains																
Welfare Gain	0.65	0.53	0.75	—	0.68	0.55	0.79	—	1.02	0.88	1.14	—	1.32	1.17	1.57	—
Efficiency Gain	0.37	0.31	0.35	—	0.38	0.32	0.36	—	0.55	0.48	0.31	—	0.67	0.60	0.45	—
Distributional Gain	0.28	0.22	0.40	—	0.30	0.23	0.43	—	0.47	0.40	0.83	—	0.65	0.57	1.12	—
Transition Gains																
Welfare Gain	-0.66	-0.53	-0.42	—	-0.72	-0.59	-0.48	—	-1.08	-0.91	-0.67	—	-1.15	-1.02	-0.86	—
Efficiency Gain	-0.62	-0.49	-0.40	—	-0.67	-0.55	-0.47	—	-0.99	-0.84	-0.64	—	-1.01	-0.92	-0.79	—
Distributional Gain	-0.04	-0.04	-0.02	—	-0.05	-0.04	-0.01	—	-0.09	-0.07	-0.03	—	-0.14	-0.10	-0.07	—
Median Welfare Gain	-0.41	-0.34	-0.27	—	-0.45	-0.36	-0.29	—	-0.73	-0.62	-0.33	—	-0.82	-0.74	-0.45	—
Political Support	25.8	26.5	34.5	—	23.1	24.8	30.2	—	19.6	20.3	22.1	—	18.3	18.7	20.6	—

Table E.1: Steady State and Transition Results for Different Values of σ.