# Optimal Government Policies in Models with Heterogeneous Agents

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

February 12, 2018

**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 Tomáš 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 heterogeneous agents. We formulate the optimal government policy problem as a calculus of variations problem where the government maximizes an objective functional subject to a system of operator constraints: 1) the first order 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 an 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 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. There are two restrictions we 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 must belong 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.

We formulate the government problem as a modified Golden Rule. That is, we solve for an optimal limiting 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 that attains a constrained optimum by directly manipulating the savings decision of each agent. They derive a functional first-order necessary condition with an added pecuniary externality arising from general equilibrium effects and use the variational approach for its characterization. Compared with Davila et al. (2012), the contribution of our paper is in the formulation of the Euler-Lagrange equations and the joint consideration of general equilibrium and distributional effects of the optimal limiting tax policy function.

We illustrate this methodology on a Ramsey problem, solving for the optimal limiting tax schedule on total income that maximizes average welfare in a steady state of a standard neoclassical, dynamic general equilibrium model with heterogeneous agents and incomplete markets calibrated as in Davila et al. (2012). For this calibration with a realistic wage and wealth inequality, we compare steady state allocations corresponding to the optimal limiting tax schedule with those related to a progressive tax schedule approximated by Heathcote et al. (2016) and to a flat-tax reform.

The optimal limiting average tax schedule is a U-shaped function. The marginal tax rates are also U-shaped, balancing a trade-off between the general equilibrium and the distributional effect. The former effect arises from providing incentives to accumulate a higher stock of aggregate capital that increases productivity of labor and, therefore, the income of poor agents (the average price effect in Davila et al. (2012)). The latter effect redistributes resources across agents in the stationary equilibrium. The marginal tax rates at low incomes induce agents to save more in order to escape relative poverty and secure better insurance against idiosyncratic risk, while the high tax rates on wealthy agents provide resources for short-run redistribution.

Without agent-specific lump-sum transfers the optimal limiting tax schedule cannot attain the constrained optimum in Davila et al. (2012). The optimal limiting tax function only slightly increases the aggregate capital stock but significantly reduces inequality. We compute transitions to the optimal limiting tax steady state and find that only when a high initial capital stock can be consumed during a transition, the tax reform improves welfare of the majority of the population. Following Farhi and Werning (2007), Farhi et al. (2012), and Krueger and Ludwig (2018), we analyze the effects of the optimal limiting tax schedule when the government puts different weights on current redistribution compared to long-term equilibrium effects. Finally, we perform several sensitivity tests and discuss important differences and effects from an alternative parameterization based on Aiyagari (1994). The U-shape of the optimal limiting average tax schedule is obtained in most of the simulated economies.

Our work contributes to recent advances in the literature on dynamic optimal taxation. Throughout the paper we compare our methodology and results to Davila et al. (2012) and show the equivalence of our Euler-Lagrange equation to the first-order condition of their constrained efficiency problem. Several papers have also built on this seminar work: Park (2017) adds an important dimension by introducing human capital. The first-order condition with respect to human capital investment also has an extra term whose sign is opposite to that of savings. Because the planner can improve welfare only by altering equilibrium prices, qualitative results of Davila et al. (2012) do not change. However, endogenous human capital decreases inequality relative to the economy with exogenous labor. Park (2014) derives positive capital income taxes from the pecuniary externality of the aggregate capital stock in a limited commitment economy. Evans (2017) studies the role of capital taxation in a model with uninsurable investment risk. The optimal capital tax rate balances the pecuniary effect of increased savings with redistribution to agents whose investment failed. Krueger and Ludwig (2018) analyze capital taxes in a Ramsey economy where the pecuniary externality cancels out the precautionary savings effect.

Influential studies in the literature analyze the steady state implications of a flat-tax or a capital income tax reform (Lucas (1990), Ventura (1999), and Conesa et al. (2009)), or restrict the tax schedule to a specific functional form: Heathcote et al. (2016) and Conesa and Krueger (2006) compute gains from the optimal *progressivity* of the income tax code. Heathcote et al. (2017) allow the degree of tax progressivity to vary with age. Bakis et al. (2015) apply the same parametric form to compute the optimal tax policy for a dynastic economy. Useful insights have been obtained by imposing restrictions on information available to the government in Golosov et al. (2003), Golosov et al. (2011), Kapicka (2013), or Heathcote and Tsujiyama (2017). Our paper shows that narrowing the analysis to monotone functions may be rather restrictive and that the shape of the optimal tax schedule is sensitive to parameterization and the resulting stationary distribution (see Mirrlees (1971), Saez (2001), Mankiw et al. (2009), or Diamond and Saez (2011)). Compared with the Mirrleesian literature, it is the tax schedule that attains a steady state where the endogenous distribution is optimal with respect to average welfare.

Finally, our paper adds to the new literature on quantitative methods. In a partial equilibrium framework, Golosov, Tsyvinski, and Werquin (2014) also use variational approach to 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. Perturbation methods have been recently used to analyze the optimal government responses to aggregate shocks. Bhandari et al. (2017a) study public debt in an economy where taxes and transfers are chosen optimally subject to heterogeneous agents' borrowing constraints and the distribution of debt ownership. In Bhandari et al. (2017b), the Ramsey planner optimally sets nominal interest rates, transfers and proportional labor taxes in response to aggregate shocks in a New Keynesian model where agents are heterogeneous with respect to co-movements of aggregate variables and measures of inequality.

In our example, we apply our methodology to an optimal limiting tax schedule on total income from labor and capital. There are several reasons why we choose this setup. First, the tax on 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 total income is the complexity of the optimization problem. We discuss how our methodology can be extended to address important issues with respect to endogenous labor supply, taxation of capital income, borrowing constraints, government debt, or more detailed life-cycle features.

The paper is organized as follows. The following section defines the stationary Ramsey problem in a competitive equilibrium. Section 3 formulates the limiting Ramsey problem in the calculus of variations. The necessary and sufficient conditions in terms of a generalized Euler-Lagrange equations and Legendre condition are developed in Section 4. Sections 5 and 6 present an example with the optimal limiting income tax schedule. Section 7 concludes. Appendices contain proofs, additional results, and a sensitivity analysis.

## 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

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

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 $g$ of total output net of depreciation, not returned back to the agents.[1] The tax schedule is applied to a broadly defined taxable activity of each agent, $x_t \in \mathbb{R}_+$. We assume $x_t = x(z_t, k_t)$ where $x : Z \times B \to \mathbb{R}_+$ and $x_z, x_k > 0$. 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 $y_t - \pi_t(x_t)$.[2] An 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 $\{\pi_t\}_{t=0}^{\infty}$ 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 $\lambda_0$, 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

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

[1] Specifying government expenditures net of depreciation simplifies the derivation of analytical properties that can be related to agents' incomes. In numerical simulations we use the usual formulation of expenditures as a fraction of total output. Our methodology equally applies to the case when government finances any level of expenditures $\{G_t\}_{t=0}^{\infty}$ and the corresponding revenue-neutral reforms.

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

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,

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

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 $\lambda_0$.

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 $\lambda_0$ 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.[3]

**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 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

The limiting optimal policy $\pi$ needs to take into account its effects on equilibrium prices and agents' decisions.[4] Define the value function of each agent as $v : Z \times B \to \mathbb{R}$ and the savings function as $h : Z \times B \to B$. Given $\pi$ and equilibrium prices $r(K)$ and $w(K)$, an

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

[4]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.

agent $(z,k)$ solves the following dynamic programming problem

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

subject to a budget constraint

$$c(z,k) + h(z,k) \leq y(z,k) + k - \pi(x(z,k)),$$

with a taxable activity $x(z,k)$, total income $y(z,k) = r(K)k + w(K)z$, and a borrowing constraint, $h(z,k) \geq \underline{k}$.

**Definition 2 (Recursive Competitive Equilibrium)** *For a given share of government expenditures $g$ and the government policy $\pi$ 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,*

$$\lambda'(z', B') = \sum_z \int_{\{(z,k)\in Z\times B:\, h(z,k)\in B'\}} Q(z,z')\, \lambda(z,k)\, dk, \qquad \textit{for all } (z', B') \in \mathcal{Z}\times\mathcal{B}; \tag{8}$$

4. *the aggregation conditions hold, $K = \sum_z \int k\lambda(z,k)dk$, and $L = \sum_z \int z\lambda(z,k)dk$;*
5. *the government budget constraint, $g(F(K,L) - \delta K) = \sum_z \int \pi(x(z,k))\lambda(z,k)dk$.*

In the recursive formulation, the optimal limiting government policy maximizes

$$W(\lambda) = \max_{\pi} \int \sum_z u(c(z,k))dk + \beta W(\lambda'), \tag{9}$$

subject to allocations satisfying the conditions in Definition 2.[5]

The steady state of the economy corresponding to the limiting optimal government policy $\pi$ is characterized by a time-invariant distribution $\lambda$. That the optimal limiting government policy $\pi$ 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 $\pi$ 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 $k \in [\underline{k}, \underline{k}(z)]$.*

[5] Note that if the allocations satisfy the definition of the recursive competitive equilibrium then they are also feasible.

A similar condition is required for the existence of a unique stationary recursive equilibrium in all models with heterogeneous agents (see Stokey et al. (1989)). It implies that the savings function does not display pathological features (for example, that wealthy agents save less than poor agents) so that the stationary distribution has a unique ergodic set. We want to make explicit here that this assumption is completely innocuous.[6]

# 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.[7] The calculus of variations is much more suitable for solving a problem with complicated 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 $\pi$, 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 $\lambda : Z \times B \to [0, 1]$ and on the savings function $h$. We assume that these functions are square integrable functions on some closed domain[8]: $h, \lambda \in L^2(Z \times B)$ 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 a 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 $\pi$ and its derivative $\pi_x$.

[6]The Regularity Condition guarantees that for all $z \in Z$, the government policy function $\pi$ 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 $k = h^{-1}(z, h)$. The Regularity Condition is used only in the law of motion for the distribution $\lambda$ in the operator for the distribution function in equation (11). 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 a government policy function.

[7]Mirrlees (1976), Davila et al. (2012) or Golosov et al. (2014) use the same approach.

[8]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.

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

$$\mathcal{F}(h) \equiv u_c(c) - \beta \sum_{z'} u_c(c') \left[1 + y'_k - \pi_x(x')x'_k\right] Q(z, z'), \tag{10}$$

where in the next period $c' = y(z', h(z,k)) - \pi(x(z', h(z,k))) + h(z,k) - h(z', h(z,k))$, $y' = r(K')h(z,k) + w(K')z'$, and $x' = x(z', h(z,k))$. The term $[1 + y'_k - \pi_x(x')x'_k]$ is the after-tax marginal return to capital when $\pi_x(x')$ is the next-period marginal government policy. Finally, $y'_k = y_k(z', h(z,k)) = r(K')$ is the marginal effect of individual savings on total income and $x'_k = x_k(z', h(z,k))$ is the marginal effect of individual savings on the taxable activity $x'$ next period.[10] 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 (8) is

$$\mathcal{L}(\lambda, \lambda', h) \equiv \lambda'(z', k') - \sum_{z} Q(z, z') \frac{\lambda\left[z, h^{-1}(z, k')\right]}{\frac{d}{dk} h\left(z, h^{-1}(z, k')\right)}, \tag{11}$$

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

A variational approach to the Ramsey problem is based on a sequential formulation of the government optimization problem. For descriptive purposes we present the recursive formulation.[11] Only two adjacent periods from the infinite time series are relevant for the government maximization problem. For a given tax policy $\pi'$ and a savings function $k''$ in the next period, and for a given distribution $\lambda$, the government chooses $\pi$ to maximize

$$\sum_{z} \int u\big[k(1 + r(K)) + zw(K) - \pi(x) - k'\big]\, \lambda(z,k) dk$$
$$+\beta \sum_{z'} \int u\big[k'(1 + r(K')) + z'w(K') - \pi'(x') - k''\big]\, \lambda'(z',k') dk',$$

with the updating operator for the next-period distribution

$$\lambda'(z', k') = \sum_{z} Q(z, z') \frac{\lambda\left[z, h^{-1}(z, k')\right]}{\frac{d}{dk} h\left(z, h^{-1}(z, k')\right)}.$$

The term $h^{-1}(z, k')$ denotes the value at $k'$ of the inverse of $h$ for each $z \in Z$. We use a change of variable $k' = h(z,k)$ to express $dk' = \frac{d}{dk} h(z,k) dk$. Merging the sums over $z$ and

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

[10] The Euler equation for an individual agent is standard. When the taxable activity equals total income, $\mathcal{F}(h) \equiv u_c(c) - \beta \sum_{z'} u_c(c') \left[1 + r(K') - \pi_x(x')r(K')\right] Q(z, z')$.

[11] See Appendix for derivation and a similar discussion in Davila et al. (2012).

the integration with respect to $k$, the optimal limiting tax function $\pi$ maximizes

$$\sum_z \int \Big\{ u[k(1+r(K)) + zw(K) - \pi(x) - h(z,k)] \\ + \beta \sum_{z'} Q(z,z') u\big[h(z,k)(1+r(K')) + z'w(K') - \pi'(x') - k''\big] \Big\} \lambda(z,k)dk,$$

where the aggregate capital stock in the next period equals

$$K' = \sum_{z'} \int k' \sum_z Q(z,z') \frac{\lambda\left[z, h^{-1}(z,k')\right]}{\frac{d}{dk} h\left(z, h^{-1}(z,k')\right)} dk' = \sum_z \int h(z,k)\lambda(z,k)dk.$$

The Ramsey problem in the calculus of variations is then

$$\max_\pi \sum_z \int_{\underline{k}(z)}^{\overline{k}(z)} \Big[ \mathcal{W}(z,k;\pi,\pi_x) + \beta \sum_{z'} Q(z,z') \mathcal{W}'(z,z',k;\pi,\pi_x) \Big] \lambda(z,k)dk,$$

where

$$\begin{aligned} \mathcal{W}(z,k;\pi,\pi_x) &\equiv u[k(1+r(K)) + zw(K) - \pi(x(z,k)) - h(z,k)], \\ \mathcal{W}'(z,z',k;\pi,\pi_x) &\equiv u[h(z,k)(1+r(K')) + z'w(K') - \pi'(x(z',h(z,k))) - k'']. \end{aligned}$$

We make explicit the dependence of operators $\mathcal{W}$ and $\mathcal{W}'$ on the tax function $\pi$ and also on its derivative $\pi_x$. Solving for the optimal tax function, we need to move from the coordinates $k$ to the taxable activity $x = x(z,k)$.[12] Therefore,

$$\max_\pi \sum_z \int_{\underline{x}(z)}^{\overline{x}(z)} \Big[ \mathcal{W}(z,x;\pi,\pi_x) + \beta \sum_{z'} Q(z,z') \mathcal{W}'(z,z',x;\pi,\pi_x) \Big] d\lambda(x), \tag{12}$$

where

$$\begin{aligned} \mathcal{W}(z,x;\pi,\pi_x) &\equiv u[k(z,x)(1+r(K)) + zw(K) - \pi(x(z,x)) - h(z,k(z,x))], \\ \mathcal{W}'(z,z',x;\pi,\pi_x) &\equiv u[h(z,k(z,x))(1+r(K')) + z'w(K') - \pi'(x(z',h(z,k(z,x)))) - k''], \\ d\lambda(x) &\equiv \lambda(z,k(z,x))k_x(z,x). \end{aligned}$$

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 $\underline{x}(z) = \underline{x}(z,\underline{k})$ depends on $z$, on the exogenously given lower bound on capital $\underline{k}$, and on the equilibrium prices. Similar arguments apply to the upper bound $\overline{x}(z) = \overline{x}(z,\overline{k})$.[13]

[12] The taxable activity $x$ is now the independent variable and $k = k(z,x)$ is its function. We want to stress again the dependence of the operator $\mathcal{W}$ and $\mathcal{W}'$ on the aggregate capital stock $K$ and $K'$ through general equilibrium effects, although it is not written as one of its arguments.

[13] 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, $[\underline{x}(z), \overline{x}(z)] \subset [\underline{x}(\underline{z}), \overline{x}(\overline{z})]$.

The aggregate capital stock in the next period is defined as

$$K' \equiv \sum_z \int_{\underline{x}(z)}^{\overline{x}(z)} h(z, k(z, x))\, d\lambda(x), \tag{13}$$

and the side conditions for the government budget constraint is

$$\sum_z \int_{\underline{x}(z)}^{\overline{x}(z)} \mathcal{G}[z, x; \pi, \pi_x]\, d\lambda(x) = 0, \quad \text{where } \mathcal{G}[z, x; \pi, \pi_x] \equiv \pi(x) - gy(z, k(z, x)). \tag{14}$$

**Definition 3 (Calculus of Variations Ramsey Problem)** *The Ramsey Problem in the calculus of variations is a generalized isoperimetric maximization problem (12), subject to the government budget constraint (14), with the individual policy function h given implicitly by the operator Euler equation $\mathcal{F}(h) = 0$, the law of motion for the distribution function, $\lambda$, given implicitly by the operator equation $\mathcal{L}(\lambda, \lambda', h) = 0$, the aggregate capital stock (13), the endogenous bounds of taxable activity, $\underline{x}(z)$ and $\overline{x}(z)$, 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 bounds $\overline{k}(z)$ 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, $\pi$ and $\pi_x$. 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)).[14]

**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*

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

[14] 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 $\mathcal{N}_u = \lim_{\varepsilon \to 0} \mathcal{N}(u + \varepsilon \delta h)/\varepsilon$ when the obtained derivative is continuous.

*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

$$\mathbf{L}(z,x) = \begin{cases} 0 & \text{for } x \in [\underline{x}(\underline{z}), \underline{x}(z)), \\ \mathcal{W}(z,x) + \mu\mathcal{G}(z,x) & \text{for } x \in [\underline{x}(z), \overline{x}(z)], \\ 0 & \text{for } x \in (\overline{x}(z), \overline{x}(\overline{z})]. \end{cases} \tag{15}$$

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

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

$$\tilde{\mathbf{L}}(z,x) = \begin{cases} 0 & \text{for } x \in [\underline{x}(\underline{z}), \underline{x}(z)), \\ \left[\mathbf{L}(z,x) + \beta \sum_{z'} Q(z,z')\mathbf{L}'(z,z',x)\right] \lambda(z,x) k_x(z,x) & \text{for } x \in [\underline{x}(z), \overline{x}(z)], \\ 0 & \text{for } x \in (\overline{x}(z), \overline{x}(\overline{z})], \end{cases}$$

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

1. *the Euler-Lagrange equation,*

$$\sum_z \left( \tilde{\mathbf{L}}_\pi(z,x) - \frac{d}{dx} \tilde{\mathbf{L}}_{\pi_x}(z,x) \right) = 0; \tag{16}$$

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

$$\left[ \tilde{\mathbf{L}}(\underline{z},x) - \left( \pi_x(x) - \frac{k_x(\underline{z},x)}{\omega_\pi(\underline{z},x)} \right) \tilde{\mathbf{L}}_{\pi_x}(\underline{z},x) \right]_{x=\underline{x}(\underline{z})} = 0; \tag{17}$$

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

$$\left[ \tilde{\mathbf{L}}(\overline{z},x) - \left( \pi_x(x) - \frac{k_x(\overline{z},x)}{\omega_\pi(\overline{z},x)} \right) \tilde{\mathbf{L}}_{\pi_x}(\overline{z},x) \right]_{x=\overline{x}(\overline{z})} = 0; \tag{18}$$

4. *and the condition on the Lagrange multiplier, $\mu$, at which (14) is satisfied.*

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

The total variation with respect to the tax schedule is equal to the sum of the total variation of utilities and the total variation of the budget constraint weighted by the shadow price $\mu$ in two consecutive periods. The total variation with respect to the optimal limiting tax policy schedule, i.e. $\pi = \lim_{t\to\infty} \pi_t$, is equal to zero. Denoting $\Delta \equiv \frac{\delta}{\delta\pi} - \frac{d}{dx}\frac{\delta}{\delta\pi_x}$ as the operator for the total variation, the Euler-Lagrange equation is simply $\sum_z \Delta\tilde{\mathbf{L}} = 0$.

The Lagrange-Euler conditions in Theorem 1 contain the tradeoff between the tax level, $\pi$, and its curvature captured by the marginal tax rate, $\pi_x$. At the optimum, the marginal effect of a change in the level of the tax schedule $\pi$ on 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, $\pi_x$.[15]

**Lemma 1** *The Euler-Lagrange equation (16) can be written as*

$$\begin{aligned}
&\sum_z \Bigg\{\Big[-u_c(c) + \beta\sum_{z'} Q(z,z')u_c(c')[1+r-\pi_x(x')x'_k]\Big]h^*_\pi \qquad (19)\\
&-u_c(c)+\mu\\
&+\beta\sum_{z'} Q(z,z')\Big(\int \psi(z,z',x)\big[u_c(c')-\mu g\big]h^*_\pi\lambda^*(z,k(z,x))k_x(z,x)dx + \mu\xi(z,z',x)h^*_\pi\Big)\\
&-\frac{d}{dx}\Big[\beta\sum_{z'} Q(z,z')\Big(\int \psi(z,z',x)\big[u_c(c')-\mu g\big]h^*_{\pi_x}\lambda^*(z,k(z,x))k_x(z,x)dx + \mu\xi(z,z',x)h^*_{\pi_x}\Big)\Big]\\
&-\varepsilon_x^{\lambda^* k_x}(z,x)\Bigg\}\lambda^*(z,k(z,x))k_x(z,x)dx \le 0,
\end{aligned}$$

*where*

$$\begin{aligned}
\psi(z,z',x) &\equiv h^*(z,k(z,x))r_K(K^*) + z'w_K(K^*),\\
\xi(z,z',x) &\equiv \pi_x(x(z',h^*(z,k(z,x))))x_k(z',h^*(z,k(z,x))) - gr(K^*),\\
\varepsilon_x^{\lambda^* k_x}(z,x) &\equiv \frac{\frac{d}{dx}\left(\lambda^*(z,k(z,x))k_x(z,x)\right)}{\lambda^*(z,k(z,x))k_x(z,x)} = \varepsilon_x^{\lambda^*}(z,x) + \varepsilon_x^{k_x}(z,x),
\end{aligned}$$

**Proof** The Euler-Lagrange equation results from a substitution of terms defined in Appendix B into Theorem 1.

The government constructs its optimal limiting tax schedule by balancing the *distributional* effect on the *individual* savings function, $h^*$, and the *general equilibrium* effect on the *aggregate* capital stock. The first line in the Euler-Lagrange equation (19) represents an agent's intertemporal first-order condition that takes into account the tax schedule and is zero except for the borrowing-constrained agents.[16] In the second line the optimal limiting tax schedule directly affects the disposable income and, therefore, the marginal utility

[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 $\tilde{\mathbf{L}}_\pi(z,x) = 0$.

[16] The Euler-Lagrange equation is written with inequality as it includes agents at the lower bound. If written with equality, the transversality condition at the lower bound applies to the case of borrowing constrained agents.

of consumption in the current period, $-u_c(c)$, counterbalanced by an opposite effect of the Lagrange multiplier, $\mu$, on the government budget constraint. The indirect effects of agents' savings, $h^*_\pi$, are aggregated to changes in the next-period aggregate capital stock weighted by marginal utility, $\beta \sum_{z'} Q(z,z') \int \psi u_c(c') h^*_\pi \lambda^* k_x dx$. In a similar way the shape of the tax schedule influences the equilibrium through the effect of the marginal tax rate $\pi_x$ on individual savings function, $h^*_{\pi_x}$, expressed by $\beta \sum_{z'} Q(z,z') \int \psi u_c(c') h^*_{\pi_x} \lambda^* k_x dx$. Note that both aggregate effects arise through the sensitivity functions $h^*_\pi$ and $h^*_{\pi_x}$. Finally, the last line of the Euler-Lagrange equation is the semi-elasticity of the transformed distribution function with respect to the taxed activity, $\varepsilon_x^{\lambda^* k_x}$. It can be decomposed into a sum of the semi-elasticity of the distribution function and the semi-elasticity of the related capital $k_x$ with respect to $x$.

The pecuniary externality effect operates through changes in equilibrium prices and can be written as in Davila et al. (2012),

$$\psi(z,z',x) = r_K(K^*)K^* \left( \frac{h^*(z,k(z,x)))}{K^*} - \frac{z'}{L} \right).$$

For a labor intensive income the term in the brackets is negative and $\psi > 0$ (since $r_K < 0$).

Compared to Davila et al. (2012), the government budget constraint causes additional pecuniary externality effects. First, there are aggregate capital effects on the government budget balance through the change of the tax rate and the marginal tax rate, $-g\beta \sum_{z'} Q(z,z') \int \psi u_c(c') h^*_\pi \lambda^* k_x dx$, and $-g\beta \sum_{z'} Q(z,z') \int \psi u_c(c') h^*_{\pi_x} \lambda^* k_x dx$. Second, individual savings decisions impact the government budget constraint through the change in the tax schedule, $\xi(z,z',x) = [\pi_x(x')x_k(x') - gr(K^*)]h^*_\pi$. The first term captures the change in tax contributions and the second the change in government spending due to the change in the agent's capital income. The total effect is $\beta \sum_z \sum_{z'} Q(z,z') \mu \xi(z,z',x) h^*_\pi \lambda^*(z,k(z,x)) k_x(z,x) dx$, weighted by $\lambda^*(z,k(z,x))k_x(z,x)dx$ across all agents with $(z,x)$. A similar decomposition applies to marginal tax changes $h^*_{\pi_x}$.

Finally, to relate our results to Davila et al. (2012), we formulate the Euler-Lagrange equation for their constrained efficiency problem in which the social planner manipulates savings of each agent. As each agent receives back a lump-sum transfer equal to changes in investment, there is neither the government budget constraint (Lagrange multiplier $\mu$) nor the direct tax effect on disposable income, $-u_c(c)$. Since the tax is not a function of $x$, the terms with $\pi(x)$ as well as $d/dx$ will disappear from the equation. As the social planner chooses savings directly, $h^*_\pi = 1$. The first-order condition is then

$$\begin{aligned} -u_c(c) &+ \beta \sum_{z'} Q(z,z') u_c(c')[1+r] \\ &+ \beta \sum_z \sum_{z'} Q(z,z') \int \psi(z,z',x) u_c(c') \lambda^*(z,k(z,x)) k_x(z,x) dx \leq 0, \end{aligned}$$

identical to the first-order condition in Davila et al. (2012) with the last term equal to the extra term $\Delta$.[17] In our formulation, the function $\psi$ is weighted by agents' marginal utilities and the whole term also enters an individual agent's Euler equation as a number.

**Theorem 2 (Second Order Sufficient Conditions)** *A tax schedule $\pi$ satisfying the first-order conditions in Theorem 1 attains a strict maximum if and only if (i) the Lagrange function defined in equation (15) satisfies the second-order Legendre condition*

$$\sum_{z\in Z} \tilde{\mathbf{L}}_{\pi_x\pi_x}(z,x) < 0 \quad \textit{for all } x \in [\underline{x}(\underline{z}), \overline{x}(\overline{z})]\,,$$

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

**Proof** For the proof see the Appendix.

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. In the following Sections we illustrate these effects numerically.

## 4.1 The Effects of Government Policy on the Equilibrium

If we knew how agents' saving policies $h$ and *simultaneously* how the distribution $\lambda$ 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 conditions for equilibrium prices.

The next Lemma derives the effects of the government policy function $\pi$ on the operator Euler equation by specifying four unknown "sensitivity" functions $h_\pi : Z \times B \longrightarrow \mathbb{R}$, $h_{\pi_x} : Z \times B \longrightarrow \mathbb{R}$, $h_{\pi_x\pi} : Z \times B \longrightarrow \mathbb{R}$, and $h_{\pi_x\pi_x} : Z \times B \longrightarrow \mathbb{R}$. Denote the next-period after-tax marginal return to capital as $R' = 1 + y'_K - \pi_x(x')x'_K$.

**Lemma 2 (The Effects of $\pi$ and $\pi_x$ on the Euler Equation)** *The total F-derivatives of the operator Euler equation $\mathcal{F}$ with respect to the government policy function $\pi$*

[17] Note that the summation over current shocks $z$ disappears as the planner chooses allocations contingent on each shock. The Euler-Lagrange equation for a flat-tax schedule is in the Appendix.

*and to its derivative* $\pi_x$ *are*

$$\mathcal{F}_i = u_{cc}(c)c_i - \beta \sum_{z'} Q(z,z')\big\{u_{cc}(c')c'_i R' + u_c(c')R'_i\big\} = 0, \tag{20}$$

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

$$\begin{aligned} \mathcal{F}_{ij} \;&=\; u_{ccc}(c)c_i c_j + u_{cc}(c)c_{ij} - \\ \beta \sum_{z'} Q(z,z')\Big\{&\big[u_{ccc}(c')c'_i c'_j + u_{cc}(c')c'_{ij}\big]R' + u_{cc}(c')\big[c'_i R'_j + c'_j R'_i\big] + u'(c')R'_{ij}\Big\} = 0. \end{aligned} \tag{21}$$

*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.

We obtain five functional equations (10), and (20)-(21) in unknown functions $h$, $h_\pi$, $h_{\pi_x}$, $h_{\pi_x\pi_x}$, and $h_{\pi_x\pi}$, respectively. Finally, by adding the first-order conditions from Theorem 1, and the functional equation for the distribution function, $\lambda$, the problem of finding the optimal government policy $\pi$ is a system of seven functional equations in seven unknown functions with two side conditions and one condition on the Lagrange multiplier.

# 5 An Example: The Optimal Income Tax Schedule

In this section we demonstrate our method by finding the optimal limiting government policy $\pi$ defined as a tax on total income from capital (net of depreciation) and labor. The taxable activity is

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

and the individual budget constraint is

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

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. In the abbreviated notation, the Euler equation for a $(z,k)$-agent's optimal savings function $h$ 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 $h' = h(z', h(z,k))$. 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.[18] In order to characterize the admissible tax functions and to prove the Schauder Theorem for economies with distortions, we follow the notation in Stokey et al. (1989). For each agent $(z,k) \in B \times Z$, denote the after-tax gross income as $\varphi(z,k) \equiv x(z,k) - \pi(x(z,k)) + k$, and rewrite the Euler equation as

$$u'(\varphi(z,k) - h(z,k)) = \beta \sum_{z'} u'(\varphi(z', h(z,k)) - h(z', h(z,k)))\varphi_k(z', h(z,k))Q(z,z'),$$

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

**Theorem 3** *For a given tax schedule $\pi : \mathbb{R}_+ \to \mathbb{R}$, if for each $(z,k) \in B \times Z$, $\varphi_k(z,k) > 0$, and $\varphi$ is quasi-concave, then the solution to each agent's maximization problem and the stationary recursive competitive equilibrium exist.*

**Proof** See the Appendix.

The following corollary characterizes the set of admissible tax functions.

**Corollary 1 (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 $\Upsilon$,*

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

*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 $\varphi_k(z,k) > 0$ and that $\varphi$ 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.[19] Application of Theorem 1 and Theorem 2 then implies necessary and sufficient conditions for the unique maximum attained by the tax schedule.

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

[19] For the targeted equilibrium interest rate in the progressive-tax steady state $r = 0.04$, the upper bound on the marginal tax rate is equal to 26. 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.

Table 1: Parameters of the Benchmark Economy

| | | | | |
|---|---|---|---|---|
| $\beta = 0.9309$ | $\sigma = 2.0$ | $\alpha = 0.36$ | $\delta = 0.08$ | $g$=0.189 |
| Earnings Process (Davila et al. (2012)): | | | | |
| $z \in Z = \{1.000, 5.290, 46.550\}$ | | $Q(z, z') = \begin{bmatrix} 0.992 & 0.008 & 0.000 \\ 0.009 & 0.980 & 0.011 \\ 0.000 & 0.083 & 0.917 \end{bmatrix}$ | | |

# 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 progressive tax schedule in the U.S. economy and from the usual flat-tax reform. In order to evaluate welfare implications, we conduct a transition analysis.

We use the same calibration as Davila et al. (2012). The three-state, first order Markov process of the uninsurable idiosyncratic shock to labor productivity is based on Diaz-Jimenez et al. (2003) with large and asymmetric wage risk necessary for a realistic dispersion of both earnings and wealth. We set the discount factor so that in the progressive-tax steady state the capital-output ratio equals 3 and the equilibrium interest rate is 4 percent. Other parameters in Table 1 are standard, $\alpha = 0.36$, $\delta = 0.08$, and the risk aversion parameter $\sigma = 2$ (intertemporal elasticity of substitution of 1/2). The stationary distribution of productivity shocks is $\{0.498, 0.443, 0.059\}$.

We model the progressive tax schedule as Heathcote, Storesletten, and Violante (2016) by specifying a tax on total income as a function

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

where $\tau$ is a parameter of the rate of progressivity and $\chi$ 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.[20] For the U.S. economy, Heathcote et al. (2016) estimate from the 2000-2006 PSID data that $\tau = 0.161$. In each period the government is required to collect tax revenues equal to 18.9% of the total output. As depreciation of capital is deducted from taxable income, in the flat-tax rate steady state $\tau = g/(1 - \delta K/Y)$.

The optimal limiting tax policy is a solution to the system of functional equations defined in Theorem 1, the functional equation for the stationary distribution $\lambda$, the two side conditions, and the Lagrange multiplier condition. We solve this functional equations problem by the least squares projection method described in Appendix C. Definitions of functional equations are in Appendix D, together with additional results and a sensitivity analysis for different parameters of risk aversion.[21] In Appendix E we also compute an

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

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

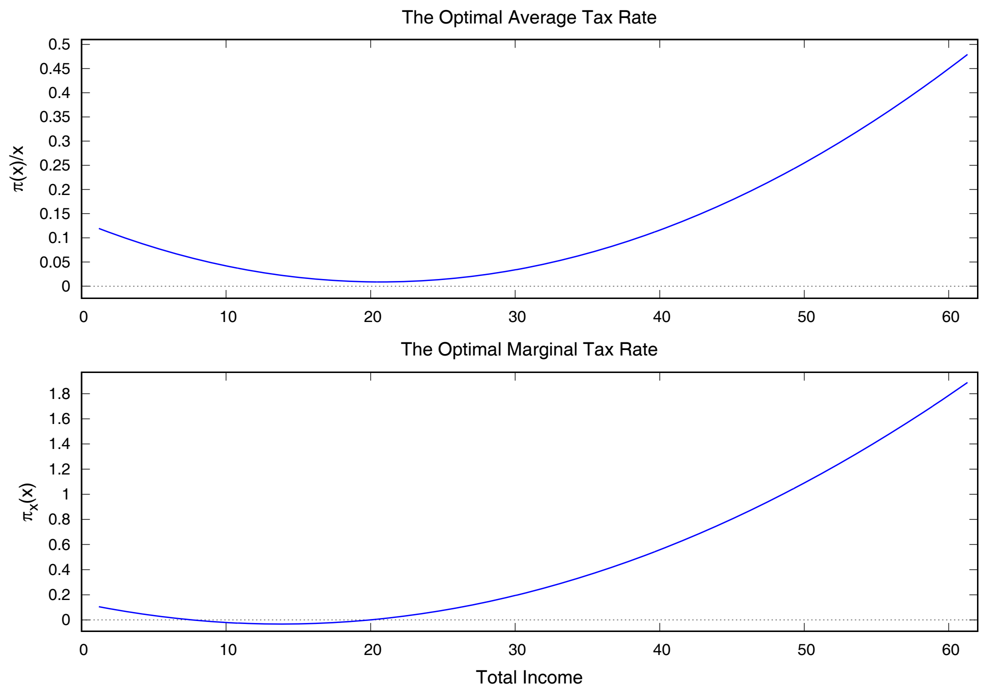


Figure 1: The optimal average tax rate and the optimal marginal tax rate.

alternative calibration for a low wealth dispersion economy based on Aiyagari (1994).

## 6.1 Steady State Results

The optimal limiting average tax schedule is U-shaped. In Figure 1, the average tax rate at the lowest total income is 14%, decreases to a minimum of 2%, and rises to 48% at the highest level of total income. The marginal tax rates are also U-shaped: close to zero for low incomes, negative at medium incomes, and positive at higher income levels.[22]

Table 2 shows steady state outcomes of the progressive, flat, and optimal limiting tax schedules. Under the progressive tax, the persistent earnings process generates a large fraction of poor agents and a substantial right tail of the wealth distribution with a Gini coefficient of wealth above 0.8, exactly as in Davila et al. (2012). The flat-tax reform increases the aggregate capital stock by 31% and output by 10%, while only slightly increasing inequality. The optimal limiting income tax increases the aggregate capital stock by only 3 percent. The coefficient of variation of wealth falls substantially and the Gini coefficient of wealth inequality becomes 0.698. While in the progressive- and flat-tax steady

problems. For a detailed explanation of applying projection methods to stationary equilibria in economies with a continuum of heterogeneous agents see Bohacek and Kejak (2002).

[22]The optimal limiting tax schedule easily satisfies the admissibility condition in Corollary 1. Similar marginal tax rates are found in dynamic models of Kapicka (2013) or Golosov et al. (2011) and in static models of Diamond (1998) and Saez (2001). 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.

Table 2: Steady State Results

| | Progressive | Flat | Optimal |
|---|---|---|---|
| $\tau$ | 0.161 | 0.265 | — |
| Aggregate assets | 3.000 | 3.934 | 3.093 |
| Output | 1.000 | 1.102 | 1.011 |
| Capital-output ratio | 3.000 | 3.568 | 3.059 |
| Interest rate (%) | 4.000 | 2.092 | 3.771 |
| Coeff. of variation of wealth | 2.125 | 2.313 | 1.501 |
| Gini wealth | 0.812 | 0.838 | 0.698 |
| Perc. of wealth of the top 5% | 0.433 | 0.475 | 0.276 |

states the top five percent of richest agents hold more than 40 percent of all assets, the optimal limiting tax schedule lowers this share to 27.6 percent.

There are two important effects the optimal limiting tax schedule takes into account. The first is the general equilibrium price effect derived in Davila et al. (2012). Poor agents in the benchmark economy have labor intensive incomes and an increased aggregate capital stock improves their welfare through higher equilibrium wages: the extra term defined in Davila et al. (2012) is positive in all our simulations, suggesting an overaccumulation of capital in a competitive equilibrium relative to the constrained efficient allocation by a social planner.

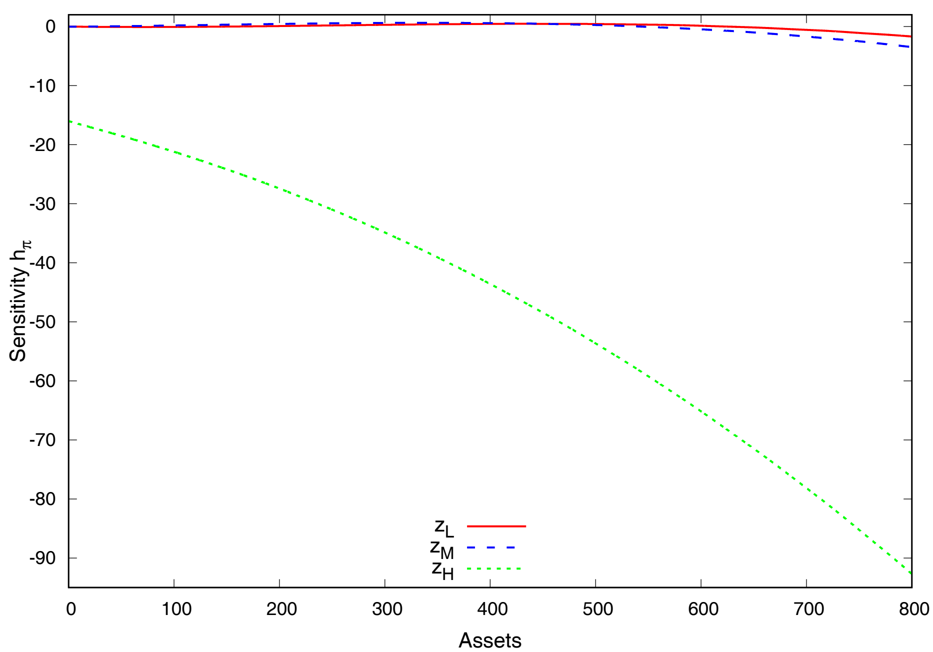


Figure 2: Sensitivity of savings to changes in the tax schedule $h_\pi$.

The second, distributional effect is related to agents' insurance and mobility within the stationary distribution. The large and persistent dispersion of earnings makes redistribution important. While in Davila et al. (2012) the wealthy agents are only induced to save

Table 3: Steady State Distribution of Assets and Tax Contributions

| | Assets | | | Tax Contributions | | |
|---|---|---|---|---|---|---|
| | Progress. | Flat | Optimal | Progress. | Flat | Optimal |
| 1st Quintile | 0.1 | 0.1 | 0.2 | -3.4 | 5.8 | 2.6 |
| 2nd Quintile | 0.2 | 0.2 | 0.3 | -3.4 | 5.9 | 2.9 |
| 3rd Quintile | 1.7 | 1.5 | 8.1 | 4.0 | 14.4 | 3.3 |
| 4th Quintile | 9.9 | 7.0 | 18.8 | 9.6 | 16.4 | 4.3 |
| 5th Quintile | 88.3 | 91.9 | 72.7 | 89.9 | 57.4 | 89.5 |
| Top 10% | 66.4 | 71.6 | 48.5 | 81.7 | 54.3 | 87.0 |
| Top 5% | 43.3 | 47.5 | 27.6 | 64.5 | 41.1 | 76.3 |
| Top 1% | 12.1 | 13.5 | 6.6 | 14.5 | 8.8 | 17.4 |

Note: Each entry is the percentage share of assets owned or taxes paid by each group.

more, under the optimal limiting tax schedule they are taxed in order to provide transfers to agents with high marginal utility. Figure 2 shows the sensitivity functions of savings to changes in the optimal tax schedule, $h_\pi$, for three labor productivity shocks $z_L < z_M < z_H$. The response of savings by agents with the low and medium productivity shocks is negligible across the ergodic set of the wealth distribution as all perturbations of the tax schedule are consumed by the agents. Only agents with the highest labor productivity alter their savings: a perturbation of the convex increasing tax schedule reduces savings, $h_\pi(k, \underline{z}) < 0$.

The substantial right tail of the income distribution contributes a large fraction of the total tax revenues. Table 3 shows that the optimal tax schedule puts most of the tax burden on the top decile of the wealth distribution. The top decile pays almost 90 percent and the top five percent pays more than 3/4 of the total tax revenues, respectively, much more than in the progressive-tax steady state. For a comparison, the flat-tax reform dramatically increases the aggregate capital stock without taking into account the distribution of agents. Under the flat tax, the share of the tax burden of the top decile is only 54.3% of all tax contributions compared to above eighty percent in the other steady states.

In general, the low and initially decreasing marginal tax rates of the limiting optimal tax policy insure poor agents and improve efficiency by lowering distortions in the economy. The marginal tax rates motivate the savings of poor agents towards the desired aggregate capital level while the increasing rates on high incomes deliver revenues for redistribution. The optimal limiting tax schedule lowers the share of assets held by the top quintile to 72.7%; the top decile owns less than one half of total assets. This redistribution is reflected in the low wealth inequality in the optimal limiting tax steady state.

## 6.2 Welfare Gains from the Optimal Limiting Tax Schedule

The top part of Table 4 shows that the optimal tax schedule delivers large average welfare gains in the steady state, 1.7% and 6.1% with respect to the progressive and the flat

Table 4: Welfare Gains from the Optimal Tax Schedule (in %)

| | Progressive | Flat |
|---|---|---|
| Steady State | 1.735 | 6.129 |
| Transition | | |
| Average Welfare | -0.767 | 14.884 |
| Aggregate Component | -0.580 | 5.670 |
| Distributional Component | -0.188 | 8.720 |
| Political Support | 46.471 | 94.129 |

tax steady state, respectively.[23] Because a pure steady state comparison involves different stationary distributions and ignores transition costs, the rest of Table 4 shows welfare gains in terms of expected present discounted values from an unanticipated reform in which the optimal limiting tax schedule is imposed on the initial progressive- or flat-tax steady state.[24] We adopt the approach of Domeij and Heathcote (2004) and decompose the average welfare gain into an aggregate and a distributional component. The former denotes a hypothetical expected present value of per-period consumption if a household consumes each period the same fraction of aggregate consumption as in the pre-reform steady state. The latter is the difference between the average welfare gain and the aggregate component.[25]

Imposing the optimal limiting tax schedule on the progressive-tax steady state leads to short-run costs in both distributional and aggregate components. The progressive tax steady state provides better short-time insurance to the poorest agents as well as requires more savings. On the other hand, in the flat-tax steady state the aggregate capital stock is very high and can be deaccumulated during the transition. This additional consumption as well as increased insurance delivers a 14.8% welfare gain per period and a majority political support. Detailed results in Appendix D show that all but the high productivity agents support the tax reform.

Obviously, we do not compute an optimal transition process that also adjust the shape of the optimal tax policy in each period of the transition. Also, efficiency gains in the terminal steady state are large and compensation schemes could be designed to alleviate welfare losses from the transition in order to obtain political support for the reform.[26]

[23]Steady state average welfare is defined as the expected discounted value of being born into a stationary equilibrium, expressed in consumption units per period.

[24]The average welfare gain from the optimal tax reform is defined as a constant percentage increase in consumption of each household in the pre-reform steady state that delivers the same expected utility as when the optimal limiting tax schedule is implemented. We guess a sufficiently large number of convergence periods and iterate on paths of equilibrium prices and tax levels to clear markets and the government budget constraint in each period of the transition.

[25]In a representative agent economy, the distributional component is zero. If the aggregate component is positive, the reform is Pareto-improving if it leaves the distribution of consumption unchanged. In Domeij and Heathcote (2004), the distributional component is negative and outweighs the aggregate component as the capital income tax reform shifts taxation to labor income.

[26]See Gottardi et al. (2011) for an analysis of issuing debt against subsequent efficiency gains.

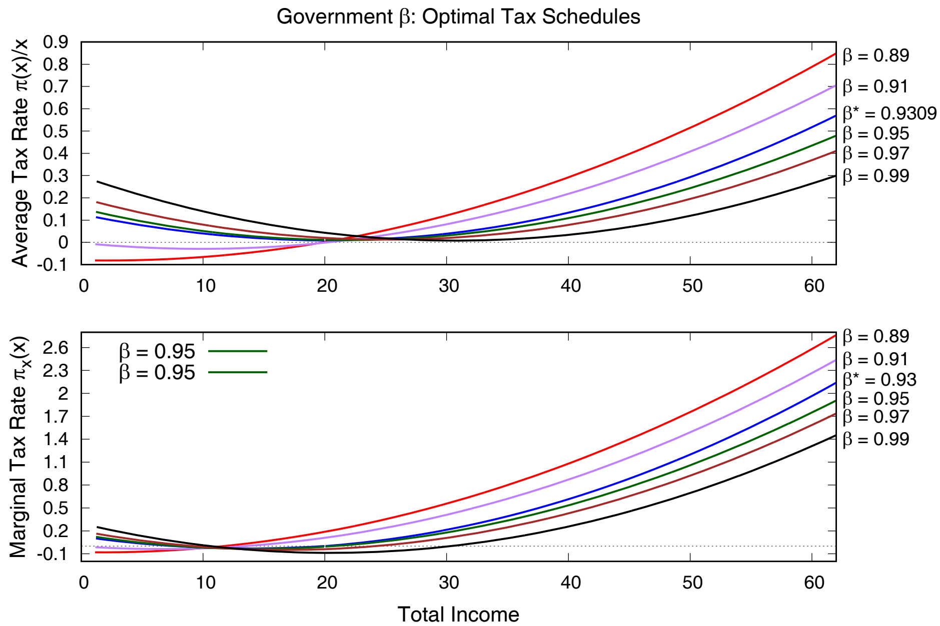


Figure 3: Optimal average and marginal tax rates for different government discount factors.

## 6.3 Different Welfare Criteria

We now relate our results to recent findings in the private information literature that society should be more patient than individual households.[27] In dynamic models, the social discount factor affects the way how government (social planner) balances efficiency and equality between the current and future generations (or periods). Indeed, the high aggregate capital stock in the constrained optimum in Davila et al. (2012) might arise in an economy whose social planner is more patient than agents. To study this issue, we set the government's discount factor in the calculus of variations Ramsey problem in equation (12) to $\hat{\beta}$ that differs from the benchmark discount factor $\beta$ used by agents in the Euler equation (10).

Figure 3 shows that when the social discount factor is lower than that of agents, $\hat{\beta} < \beta$, the optimal tax schedule puts more weight on current redistribution and less weight on the long-term general equilibrium effect from the aggregate capital stock accumulation. The optimal tax schedule for a very low $\hat{\beta} = 0.89$ is actually increasing and convex. On the other hand, a very high social discount factor leads to a more convex U-shaped average tax

[27] This issue has been studied in an asymmetric information framework as a tool to overcome the immiseration result of Atkeson and Lucas (1992). Phelan (2006) achieves this by assigning equal weights on all future generations, while Farhi and Werning (2007) place a positive and vanishing Pareto weight on expected welfare of future generations. Farhi and Werning (2010) analyze a dynamic Mirrleesian model with productivity shocks and find that a progressive estate tax implements efficient allocations by providing the necessary mean reversion and insurance across generations. Farhi et al. (2012) study efficient non-linear taxation of labor and capital in a dynamic Mirrleesian model without commitment. In order to lower the capital stock and, therefore, the gains from a deviation, the marginal tax on capital income is progressive.

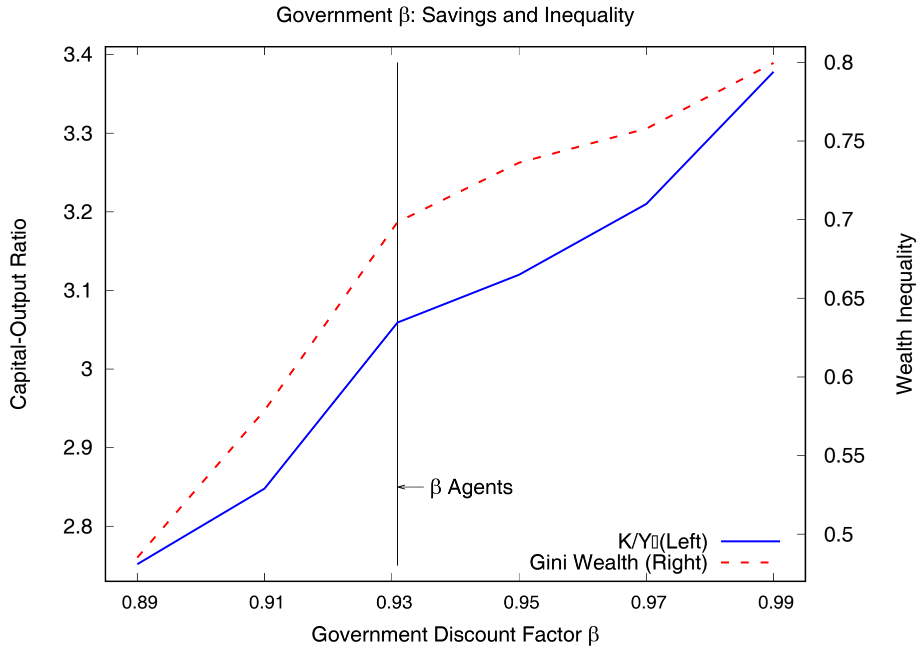


Figure 4: Capital-output ratio (left axis) and wealth inequality (right axis) for different government discount factors.

function needed for saving incentives (see Bakis et al. (2015) for similar results in a dynastic economy). Figure 4 shows that both the capital-output ratio and wealth inequality increase in $\hat{\beta}$. When $\hat{\beta} = 0.99$, the capital-output ratio rises to 3.38 and Gini coefficient of wealth to 0.80. On the other hand, a very low discount rate $\hat{\beta} = 0.89$ decreases wealth inequality to 0.485. In Appendix D we report all steady state and transition results. For $\hat{\beta} < \beta$, the optimal tax reform of the progressive-tax steady state is supported by the majority of the population. This is because for low $\hat{\beta}$s the government prefers a lower aggregate capital stock, so agents can deaccumulate and consume their capital stock during a transition, and vice versa for high social discount factors. For the same reason, the optimal tax reform of the flat-tax economy makes the majority of agents better off.

The shape of the optimal tax schedule is necessarily sensitive to the government's objective function. Heathcote and Tsujiyama (2017) characterize the mapping between a taste for redistribution in a class of Pareto weight functions and the progressivity parameter in Heathcote et al. (2016). When the taste for redistribution increases, the marginal tax schedule moves from being upward sloping to becoming a U-shaped function. In their simulations, the welfare gains from an optimal progressive tax reform are very small. Appendix F discusses these extensions and other results in detail.

## 6.4 Sensitivity Analysis

Finally, we analyze the optimal limiting tax schedule in the benchmark calibration for different values of risk aversion $\sigma$. The U-shape of the optimal limiting average tax schedule holds in all simulated economies. An increasing risk aversion puts more weight on the distributional effect relative to the general equilibrium effect and the average tax schedules become more progressive with higher $\sigma$. The Gini coefficient of wealth falls to 0.59 when $\sigma = 4$. In the flat-tax economy, a higher risk aversion induces agents to accumulate more capital as a buffer stock against idiosyncratic risk (see Imrohoroglu (1998)), rising transitional gains as a consequence. Because general equilibrium effects are more important at low risk aversion, the negative slope of the average tax rates is steeper at low incomes.

## 6.5 A Low Wealth Dispersion Economy

Mirrlees (1971), Mankiw et al. (2009), or Diamond and Saez (2011) note that the shape of the marginal tax schedule is sensitive to the distribution of skill or income. In Appendix E, we evaluate our example in Section 6 for a more traditional calibration based on Aiyagari (1994). The three-state, first order Markov process of the idiosyncratic shock to labor productivity is estimated from the household annual labor income process in the PSID data by a first-order autoregression with a persistence parameter 0.6 and a volatility 0.2.

This calibration of the earnings process leads to a steady state with an unrealistic, low dispersion of wealth. The optimal limiting average tax schedule is also U-shaped with an important upward shift: the average tax rate at the lowest total income is 37%, decreases to a minimum of 24%, and rises to 48% at the highest level of total income. The marginal tax rates are close to zero for low incomes and rise at higher income levels to provide resources for redistribution and insurance. In this low wealth dispersion economy, the earnings process is not very persistent, wealth inequality is low and poor agents can escape poverty easily. This allows the optimal tax schedule to focus more on the general equilibrium price effect, increasing the aggregate capital stock by 18.7%, to a higher level than in the flat-tax steady state. The savings decisions of all types of agents are very sensitive to perturbations of the optimal limiting tax schedule. The degree of persistence in the idiosyncratic labor productivity process has similar effects to those found in Evans (2017). In that paper, the optimal capital tax rate balances the general equilibrium price effect and redistribution. When the productivity shocks are iid, the former effect dominates and the planner subsidizes savings decisions as in Davila et al. (2012). When persistence of the productivity shocks increases, the latter effect is more important and the long-run capital tax is positive. Golosov et al. (2016) and Mankiw et al. (2009) also find U-shaped marginal rates originating from a large fraction of agents with low productivity.[28]

# 7 Conclusion

In this paper, we provide a solution method for 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. In our example, we find the optimal limiting tax schedule on total income in a stationary Ramsey problem. As in Diamond (1998), Saez (2001), or Golosov et al. (2011), the optimal limiting average tax schedule is U-shaped. The shape of the optimal limiting tax schedule balances the efficiency-equality trade-off between general equilibrium price effects and redistribution. In the calibrated economy with high wage and wealth dispersion, the distributional effects dominate and the high income agents contribute a large fraction of the total tax revenues. Initially low and decreasing marginal tax rates improve efficiency by lowering distortions in the economy. When a government discounts future more than agents or if the latter are more risk averse, the short-term distributional effect is more important than the long-term, general equilibrium effect from changes in the aggregate capital stock. Compared with Davila et al. (2012), the aggregate capital stock increases only slightly but the optimal limiting tax function significantly reduces inequality. In other words, without agent-specific lump-sum transfers the optimal tax policy would not subsidize rich agents' savings at the cost of not providing insurance to poor agents with high marginal utility. Our results confirm the conjecture in Davila et al. (2012) that simple uniform policies cannot achieve the constrained optimum.

This paper is a first step for analyzing more realistic models with important policy implications. Our example shows that analyzing linear tax functions or restricting functional forms to progressive taxation might miss a large part of efficiency and equality effects. The most important extension of the model would introduce the extensive labor margin. Labor decisions are more relevant for poor agents while savings decisions are more important for wealthy agents. In a simplified two-period model, Davila et al. (2012) find that the pecuniary externality might induce the agents to supply inefficiently low effort. Park (2017) analyzes the constrained optimum with endogenous human capital, motivated by Huggett et al. (2011) who find that the determinants of endogenous human capital dispersion are more important than idiosyncratic shocks as a source of lifetime earnings dispersion. Because the planner in Park (2017) can only alter equilibrium prices, the qualitative results

[28] Heathcote and Tsujiyama (2017) the shape depends on how much fiscal pressure the government faces. When fiscal pressure is low, the marginal tax schedule is monotone increasing. As fiscal pressure increases, the optimal schedule becomes first flatter and then U-shaped.

of Davila et al. (2012) do not change. However, numerical simulations lead to a lower capital-labor ratio and a reduced inequality relative to the economy with exogenous labor. While the optimal limiting tax schedule might motivate poor agents to increase labor effort, the welfare of poor households could be better improved by reducing the risky part of their income through a lower wage. Therefore, the redistribution channel arising from the endogenous labor income dispersion could further reduce inequality and lower the optimal capital-labor ratio. It is also not clear whether the optimal tax schedule would still display increasing marginal rates as they reduce the returns to labor and the returns to human capital investment.

In our future research we also plan to study the optimal limiting tax schedule with life-cycle income profiles as in Golosov et al. (2011) and Farhi and Werning (2012). 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. Further restrictions on available policy tools imply additional aggregate and distributional effects. One can also analyze different calibrations with heterogeneous preferences, opportunities to participate in labor or asset markets. A major task would be an extension of our methodology to the private information literature. Finally, we plan to make our methodology applicable to time-varying optimal policies for heterogeneous response of agents to aggregate shocks.

# References


Ahn, S., G. Kaplan, B. Moll, T. Winberry, and C. Wolf (2017). When inequality matters for macro and macro matters for inequality. *NBER Macroeconomic Annual 32*.

Aiyagari, R. S. (1994). Uninsured idiosyncratic risk and aggregate savings. *Quarterly Journal of Economics 109*(3), 659–684.

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

Atkeson, A. and R. E. Lucas, Jr. (1992). On efficient distribution with private information. *Review of Economic Studies 59*, 427–453.

Bakis, O., B. Kaymak, and M. Poschke (2015). Transitional dynamics and the optimal progressivity of income redistribution. *Review of Economic Dynamics 18*(3), 679–693.

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2017a). Fiscal policy and debt management with incomplete markets. *Quarterly Journal of Economics 132*(2), 617–663.

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2017b). Inequality, business cycles and fiscal-monetary policy. *Working Paper*.

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

Conesa, J. C., S. Kitao, and D. Krueger (2009). Taxing capital? Not a bad idea after all! *American Economic Review 99*(1), 25–48.

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

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 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.

Evans, D. (2017). Optimal taxation with persistent idiosyncratic investment risk. *Working Paper*.

Farhi, E., C. Sleet, I. Werning, and S. Yeltekin (2012). Non-linear capital taxation without commitment. *Review of Economic Studies 80*, 1–25.

Farhi, E. and I. Werning (2007). Inequality and social discounting. *Journal of Political Economy 115*, 365–402.

Farhi, E. and I. Werning (2010). Progressive estate taxation. *Quarterly Journal of Eco-*

*nomics 125*(2), 635–673.

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., M. Troshkin, and A. Tsyvinski (2016). Redistribution and social insurance. *American Economic Review 106*, 359–386.

Golosov, M., A. Tsyvinski, and M. Troshkin (2011). Optimal taxation: Merging micro and macro approaches. *Journal of Money, Credit and Banking 43*(5), 147–174.

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. (2005). Fiscal policy with heterogeneous agents and incomplete markets. *Review of Economic Studies 72*, 161–188.

Heathcote, J., K. Storesletten, and G. Violante (2016). Optimal tax progressivity: An analytical framework. *Working Paper*.

Heathcote, J., K. Storesletten, and G. Violante (2017). Optimal progressivity with age-dependent taxation. *Working Paper*.

Heathcote, J. and H. Tsujiyama (2017). Optimal income taxation: Mirrlees meets ramsey. *Working Paper*.

Huggett, M., G. Ventura, and A. Yaron (2011). Sources of lifetime inequality. *American Economic Review 101*(7), 2923–2954.

Imrohoroglu, S. (1998). A quantitative analysis of capital income taxation. *International Economic Review 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.

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.

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

Krueger, D. and A. Ludwig (2018). Precautionary savings and pecuniary externalities: Optimal taxes on capital in the olg model with idiosyncratic income risk. *Working Paper*.

Liberzon, D. (2011). *Calculus of variations and optimal control theory: a concise intro-*

*duction*. Princeton: Princeton University Press.

Lucas, Jr., R. E. (1990). Supply-side economics: An analytical review. *Oxford Economic Papers* (42), 293–316.

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.

Park, Y. (2014). Optimal taxation in a limited commitment economy. *Review of Economic Studies 81*(2), 884–918.

Park, Y. (2017). Constrained efficiency in a human capital model. *Working paper*.

Phelan, C. (2006). Opportunity and social mobility. *Review of Economic Studies 73*, 487–504.

Reiter, M. (2009). Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control 33*(3), 649–665.

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.

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

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

implying $h = \underline{k}$.

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

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

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

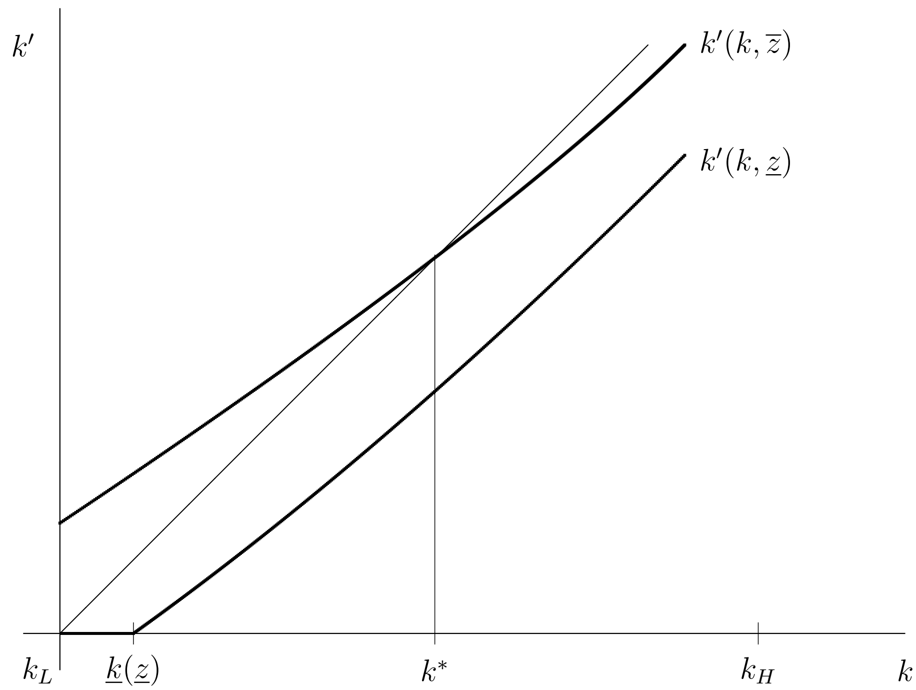


Figure A.1: Policy functions for the next period capital stock. An example with two productivity shocks $\overline{z} > \underline{z}$. There is an exogenous lower bound $k_L$ and an endogenous upper bound $k^* < k_H$. The stationary distribution has a unique ergodic set $E = [k_L, k^*]$. Agents with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

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

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

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

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 $\lambda$ are equivalent. So we can simply consider only the distribution function $\lambda$ given by $\mathcal{L}$ in (11) and $h$ given by (10).

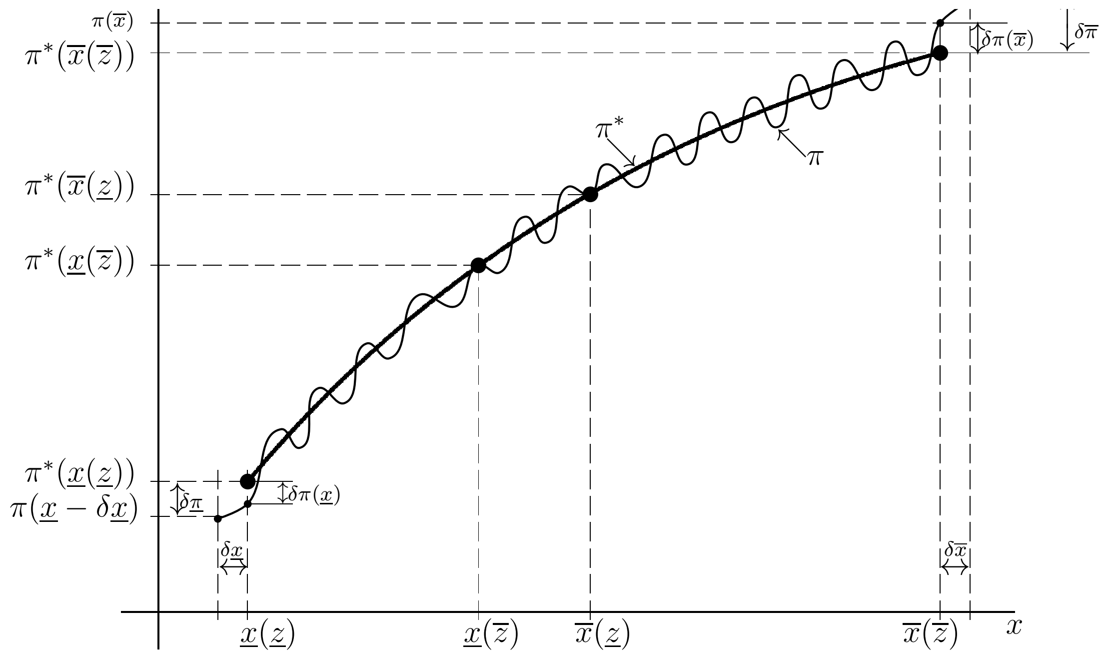


Figure B.1: Variations Around Optimal Tax Schedule

# B Appendix: Proofs

## B.1 Proof of Theorem 1

First, we derive the first order conditions for the Ramsey problem in (12), specifying the optimal tax policy function at some period $t$. Define

$$
\begin{aligned}
J(\varepsilon) \quad = \quad & \int_{\underline{x}(\underline{z})-\varepsilon\delta\underline{x}}^{\overline{x}(\underline{z})} \left\{ \mathbf{L}\left[\underline{z}, x; \tilde{\pi}(x), \tilde{\pi}_x(x)\right] + \beta \sum_{z'} Q\left(\underline{z}, z'\right) \mathbf{L}'\left[\underline{z}, z', x; \tilde{\pi}(x), \tilde{\pi}_x(x)\right] \right\} d\lambda(x) \\
+ \quad & \sum_{z \in Z \backslash \{\underline{z}, \overline{z}\}} \int_{\underline{x}(z)}^{\overline{x}(z)} \left\{ \mathbf{L}\left[z, x; \tilde{\pi}(x), \tilde{\pi}_x(x)\right] + \beta \sum_{z'} Q\left(z, z'\right) \mathbf{L}'\left[z, z', x; \tilde{\pi}(x), \tilde{\pi}_x(x)\right] \right\} d\lambda(x) \\
+ \quad & \int_{\overline{x}(\overline{z})}^{\overline{x}(\overline{z})+\varepsilon\delta\overline{x}} \left\{ \mathbf{L}\left[\overline{z}, x; \tilde{\pi}(x), \tilde{\pi}_x(x)\right] + \beta \sum_{z'} Q\left(\overline{z}, z'\right) \mathbf{L}'\left[\overline{z}, z', x; \tilde{\pi}(x), \tilde{\pi}_x(x)\right] \right\} d\lambda(x),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{L}\left[z, x; \tilde{\pi}(x), \tilde{\pi}_x(x)\right] \quad &\equiv \quad \mathcal{W}\left[z, x; \tilde{\pi}(x), \tilde{\pi}_x(x)\right] + \mu \mathcal{G}\left[z, x; \tilde{\pi}(x), \tilde{\pi}_x(x)\right], \\
\mathbf{L}'\left[z, z', x; \tilde{\pi}(x), \tilde{\pi}_x(x)\right] \quad &\equiv \quad \mathcal{W}'\left[z, z', x; \tilde{\pi}(x), \tilde{\pi}_x(x)\right] + \mu' \mathcal{G}'\left[z, z', x; \tilde{\pi}(x), \tilde{\pi}_x(x)\right], \\
d\lambda(x) \quad &= \quad \lambda(z, k(z, x)) k_x(z, x) dx,
\end{aligned}
$$

with

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

The current period and next period utilities are

$$\begin{aligned}\mathcal{W}\left[z,x;\tilde{\pi}(x),\tilde{\pi}_x(x)\right] &= u\left(c\left(z,k(z,x);\tilde{\pi}(x),\tilde{\pi}_x(x)\right)\right),\\ \mathcal{W}'[z,z',x;\tilde{\pi}(x),\tilde{\pi}_x(x)] &\equiv u'(c'(z',h(z,k(z,x);\tilde{\pi}(x),\tilde{\pi}_x(x));\tilde{\pi}(x),\tilde{\pi}_x(x))),\end{aligned}$$

with consumption in the current period

$$c\left(z,k(z,x);\tilde{\pi}(x),\tilde{\pi}_x(x)\right)=k[1+r(K)]+zw(K)-\tilde{\pi}(x)-h(z,k(z,x);\tilde{\pi}(x),\tilde{\pi}_x(x)),$$

and in the next period

$$\begin{aligned}c'(z',h(z,k(z,x);\tilde{\pi}(x),\tilde{\pi}_x(x));\tilde{\pi}(x),\tilde{\pi}_x(x)) = {}& h(z,k(z,x);\tilde{\pi}(x),\tilde{\pi}_x(x))[1+r(K')]+z'w(K')\\ &-\pi(x(z',h(z,k;\tilde{\pi}(x),\tilde{\pi}_x(x))))-k''.\end{aligned}$$

These terms will be maximized with the side conditions (isoperimetric tasks) of balanced budget in all periods. The two relevant budget constraints are

$$\sum_z\int_{\underline{x}(z)}^{\overline{x}(z)}\mathcal{G}[z,x;\tilde{\pi}(x),\tilde{\pi}_x(x)]\,d\lambda(x)=\sum_z\int_{\underline{x}(z)}^{\overline{x}(z)}\left[\tilde{\pi}(x)-gy(z,k(z,x)))\right]d\lambda(x)=0,$$

and in the next period,

$$\sum_{z'\in Z}\int_{\underline{k}'}^{\overline{k}'}[\pi(x(z',k'))-gy(z',k')]\lambda'[z',k']dk'=0,$$

which is in coordinates $x$,

$$\begin{aligned}&\sum_z\sum_{z'}Q(z,z')\int_{\underline{x}}^{\overline{x}}\mathcal{G}'[z,z',x;\tilde{\pi}(x),\tilde{\pi}_x(x)]d\lambda(x)=\\ &\qquad\sum_z\sum_{z'}Q(z,z')\int_{\underline{x}}^{\overline{x}}[\tilde{\pi}(x(z',h(z,k(z,x))))-gy(z',h(z,k(z,x)))]d\lambda(x)=0.\end{aligned}$$

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

$$J_\varepsilon\left(\mathbf{0}\right)=\lim_{\varepsilon\to 0}\frac{dJ\left(\varepsilon\right)}{d\varepsilon}=0$$

leads to the following first-order conditions

$$J_\varepsilon(\mathbf{0})=\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\left\{\tilde{\mathbf{L}}_\pi(z,x)\delta\pi(x)+\tilde{\mathbf{L}}_{\pi_x}(z,x)\delta\pi_x(x)\right\}dx-\left[\tilde{\mathbf{L}}_\pi(\underline{z},x)|_{\underline{x}(\underline{z})}\right](-\delta\underline{x})+\left[\tilde{\mathbf{L}}_\pi(\overline{z},x)|_{\overline{x}(\overline{z})}\right]\delta\overline{x},$$

where for $\iota \in \{\pi(x), \pi_x(x)\}$ we use a simplified notation $\tilde{\mathbf{L}}_\iota(z,x) = \tilde{\mathbf{L}}_\iota\left[z, x; \tilde{\pi}^*(x), \tilde{\pi}^*_x(x)\right]$, and

$$\tilde{\mathbf{L}}_\iota(z,x) \equiv \Big[\mathbf{L}_\iota(z,x) + \beta \sum_{z'} Q(z,z')\mathbf{L}'_\iota\left(z,z',x\right)\Big] \lambda\left(z, k(z,x)\right) k_x(z,x).$$

Further, the integration by parts delivers

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

Thus we can rewrite the formula above in a more compact form as[29]

$$\begin{aligned} J_\varepsilon\left(0\right) &= \sum_{z\in Z} \int_{\underline{x}(z)}^{\overline{x}(z)} \left[\tilde{\mathbf{L}}_\pi(z,x) - \frac{d}{dx}\tilde{\mathbf{L}}_{\pi_x}(z,x)\right] \delta\pi(x)dx \\ &+ \Big[\tilde{\mathbf{L}}_{\pi_x}(z,x)\delta\pi(x)\Big]_{\underline{x}(\underline{z})}^{\overline{x}(\overline{z})} - \tilde{\mathbf{L}}(z,x)|_{\underline{x}(\underline{z})}\left(-\delta\underline{x}\right) + \tilde{\mathbf{L}}(z,x)|_{\overline{x}(\overline{z})}\delta\overline{x}. \end{aligned} \tag{B.1}$$

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

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

and

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

This implies that

$$\delta\pi(\overline{x}) = \delta\overline{\pi} - \pi^*_x(\overline{x})\delta\overline{x}, \tag{B.2}$$

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, $\overline{k} = h\left(\overline{z}, \overline{k}; \pi(x), \pi_x(x)\right)$. The total variance differential is

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

and thus

$$\delta\overline{k} = [1 - h_k(\overline{z},\overline{k})]^{-1} h_\pi(\overline{z},\overline{k})\delta\overline{\pi}. \tag{B.3}$$

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

$$\delta\overline{x} = x_k(\overline{z}, \overline{k}(\overline{z}))\delta\overline{k},$$

where we can further substitute for $\delta\overline{k}$ from (B.3)

$$\delta\overline{x} = x_k(\overline{z},\overline{k})\overline{\omega}_\pi\delta\overline{\pi}, \tag{B.4}$$

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

where

$$\overline{\omega}_\pi \equiv \frac{\delta\overline{k}}{\delta\pi} = \frac{h_\pi(\overline{z},\overline{k})}{1-h_k\left(\overline{z},\overline{k}\right)}. \tag{B.5}$$

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

$$\delta\pi(\underline{x}) = -\delta\underline{\pi} + \pi_x^*(\underline{x})\delta\underline{x}. \tag{B.6}$$

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

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

and thus

$$\delta\underline{k}' = [1-h_k(\underline{z},\underline{k})]^{-1}h_\pi(\underline{z},\underline{k})\delta\underline{\pi}. \tag{B.7}$$

As $\underline{x} = x(\underline{z},\underline{k}(\underline{z};\pi(x),\pi_x(x)))$, we determine the variation

$$\delta\underline{x} = x_k[\underline{z},\underline{k}(\underline{z})]\delta\underline{k},$$

where we substitute for $\delta\underline{k}$ from (B.7)

$$\delta\underline{x} = x_k(\underline{z},\underline{k})\underline{\omega}_\pi\delta\underline{\pi}, \tag{B.8}$$

where

$$\underline{\omega}_\pi \equiv \frac{\delta\underline{k}}{\delta\pi} = \frac{h_\pi(\underline{z},\underline{k})}{1-h_k(\underline{z},\underline{k})}. \tag{B.9}$$

Going back to (B.1) and using (B.2) and (B.6) we obtain

$$\begin{aligned} J_\varepsilon(0) &= \sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\left\{\left[\tilde{\mathbf{L}}_\pi(z,x) - \frac{d}{dx}\tilde{\mathbf{L}}_{\pi_x}(z,x)\right]\delta\pi(x)\right\}dx \\ &+ \tilde{\mathbf{L}}_{\pi_x}(\underline{z},x)|_{\underline{x}(\underline{z})}\delta\underline{\pi} - \left[\pi_x(x)\tilde{\mathbf{L}}_{\pi_x}(\underline{z},x) - \tilde{\mathbf{L}}(\underline{z},x)\right]_{\underline{x}(\underline{z})}\delta\underline{x} \\ &+ \tilde{\mathbf{L}}_{\pi_x}(\overline{z},x)|_{\overline{x}(\overline{z})}\delta\overline{\pi} - \left[\pi_x(x)\tilde{\mathbf{L}}_{\pi_x}(\overline{z},x) - \tilde{\mathbf{L}}(\overline{z},x)\right]_{\overline{x}(\overline{z})}\delta\overline{x}. \end{aligned} \tag{B.10}$$

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.8) to obtain

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

Finally, from (B.11) we get

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

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

$$\begin{aligned}
\sum_{z\in Z}\left[\tilde{\mathbf{L}}_{\pi}(z,x)-\frac{d}{dx}\tilde{\mathbf{L}}_{\pi_x}(z,x)\right] &= 0, \qquad\qquad (B.12)\\
\tilde{\mathbf{L}}(\underline{z},x)|_{\underline{x}(\underline{z})}+\left[\frac{1}{\underline{x}_k\underline{\omega}_{\pi}}-\pi_x(x)\right]\tilde{\mathbf{L}}_{\pi_x}\left(\underline{z},x\right)|_{\underline{x}(\underline{z})} &= 0,\\
\tilde{\mathbf{L}}_{\pi_x}(\overline{z},x)|_{\overline{x}(\overline{z})}+\left[\frac{1}{\overline{x}_k\overline{\omega}_{\pi}}-\pi_x(x)\right]\tilde{\mathbf{L}}_{\pi_x}(\overline{z},x)|_{\overline{x}(\overline{z})} &= 0,
\end{aligned}$$

and

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

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

$$\begin{aligned}
J_{\varepsilon}(0) &= \sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)}\left[\tilde{\mathbf{L}}_{\pi}(z,x)-\frac{d}{dx}\tilde{\mathbf{L}}_{\pi_x}(z,x)\right]\delta\pi(x)dx \qquad\qquad (B.13)\\
&+\tilde{\mathbf{L}}_{\pi_x}(\underline{z},x)|_{\underline{x}(\underline{z})}\delta\underline{\pi}\\
&+\tilde{\mathbf{L}}_{\pi_x}(\overline{z},x)|_{\overline{x}(\overline{z})}\delta\overline{\pi}+\left[\tilde{\mathbf{L}}(\overline{z},x)-\pi_x(x)\tilde{\mathbf{L}}_{\pi_x}(\overline{z},x)\right]_{\overline{x}(\overline{z})}\overline{x}_k\overline{\omega}_{\pi}\delta\overline{\pi}.
\end{aligned}$$

The first-order conditions are

$$\begin{aligned}
\sum_{z\in Z}\left[\tilde{\mathbf{L}}_{\pi}(z,x)-\frac{d}{dx}\tilde{\mathbf{L}}_{\pi_x}(z,x)\right] &= 0, \qquad\qquad (B.14)\\
\tilde{\mathbf{L}}_{\pi_x}(\underline{z},x)|_{\underline{x}(\underline{z})} &= 0,\\
\tilde{\mathbf{L}}_{\pi_x}(\overline{z},x)|_{\overline{x}(\overline{z})}+\left[\frac{1}{\overline{x}_k\overline{\omega}_{\pi}}-\pi_x(x)\right]\tilde{\mathbf{L}}_{\pi_x}(\overline{z},x)|_{\overline{x}(\overline{z})} &= 0.
\end{aligned}$$

Since we are interested in the limiting tax policy function $\pi\equiv\lim_{t\to\infty}\pi_t$, we need to impose that the derived first-order conditions in (B.14) are at a steady state related to $\pi$, i.e. $K = K'$. Then clearly the derived first-order conditions are the same as in Theorem 1. For the case with the endogenous lower bound the first-order conditions are given by (B.12). Q.E.D.

### B.1.1 Definition of Terms in Theorem 1

For the definition of terms in Theorem 1 recall that

$$\tilde{\mathbf{L}}_\iota(z,x) = \Big[\mathbf{L}_\iota(z,x) + \beta \sum_{z'} Q(z,z')\mathbf{L}'_\iota(z,z',x)\Big]\lambda\left(z,k(z,x)\right)k_x(z,x),$$

where

$$\begin{aligned} \mathbf{L}_\iota(z,x) &= \mathcal{W}_\iota(z,x) + \mu\mathcal{G}_\iota(z,x), \\ \mathbf{L}'_\iota(z,z',x) &= \mathcal{W}'_\iota(z,z',x) + \mu'\mathcal{G}'_\iota(z,z',x), \end{aligned}$$

where $\iota \in \{\pi(x), \pi_x(x)\}$ and $\mu$ and $\mu'$ be Lagrangian multipliers for side conditions.

We can express the Euler-Lagrange equation evaluated at the limiting $\pi_t \to \pi^*$ at the steady state as (omitting $(z,x)$ and $(z,z',x)$)

$$\begin{aligned} \sum_z \Bigg\{ & \Big(\mathcal{W}_\pi + \mu\mathcal{G}_\pi + \beta\sum_{z'} Q(z,z')\Big(\mathcal{W}'_\pi + \mu'\mathcal{G}'_\pi\Big)\Big)\lambda k_x \; - \\ & \frac{d}{dx}\Bigg[\Big(\mathcal{W}_{\pi_x} + \mu\mathcal{G}_{\pi_x} + \beta\sum_{z'} Q(z,z')\Big(\mathcal{W}'_{\pi_x} + \mu'\mathcal{G}'_{\pi_x}\Big)\Big)\lambda k_x\Bigg]\Bigg\} = 0, \end{aligned}$$

which is again in the more compact formulation,

$$\sum_z \Bigg\{\Big(\mathbf{L}_\pi + \beta\sum_{z'} Q(z,z')\mathbf{L}'_\pi\Big)\lambda k_x - \frac{d}{dx}\Bigg[\Big(\mathbf{L}_{\pi_x} + \beta\sum_{z'} Q(z,z')\mathbf{L}'_{\pi_x}\Big)\lambda k_x\Bigg]\Bigg\} = 0,$$

with

$$\begin{aligned} & \frac{d}{dx}\Bigg[\Big(\mathbf{L}_{\pi_x} + \beta\sum_{z'} Q(z,z')\mathbf{L}'_{\pi_x}\Big)\lambda k_x\Bigg] \\ = \; & \Big(\frac{d}{dx}\mathbf{L}_{\pi_x} + \beta\sum_{z'} Q\left(z,z'\right)\frac{d}{dx}\mathbf{L}'_{\pi_x}\Big)\lambda k_x + \Big(\mathbf{L}_{\pi_x} + \beta\sum_{z'} Q\left(z,z'\right)\mathbf{L}'_{\pi_x}\Big)\left(\lambda_x k_x + \lambda k_{xx}\right). \end{aligned}$$

To determine individual terms, recall that in the full notation

$$\begin{aligned} \mathcal{W}\left[z,x;\tilde{\pi}(x),\tilde{\pi}_x(x)\right] &= u\left(c\left(z,k(z,x);\tilde{\pi}(x),\tilde{\pi}_x(x)\right)\right), \\ \mathcal{W}'[z,z',x;\tilde{\pi}(x),\tilde{\pi}_x(x)] &\equiv u'(c'(z',h(z,k(z,x);\tilde{\pi}(x),\tilde{\pi}_x(x));\tilde{\pi}(x),\tilde{\pi}_x(x))). \end{aligned}$$

with

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

where $\tilde{\pi}(x) \equiv \pi^*(x) + \varepsilon\delta\pi(x)$, and $\tilde{\pi}_x(x) \equiv \pi_x^*(x) + \varepsilon\delta\pi_x(x)$.

To simplify notation we will omit the obvious arguments. Thus at the limit

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

for $\iota \in \{\pi(x), \pi_x(x)\}$,

$$
\begin{aligned}
\mathcal{W}_\iota &= u_c c_\iota, \\
\mathcal{W}'_\iota &= u_c\left[1 + r(K') - \tilde{\pi}_x(x')x'_k\right] h_\iota + \int u_c \psi(z, z', x) h_\iota, \\
c_\pi &= -1 - h_\pi, \\
c_{\pi_x} &= -h_{\pi_x}, \\
\psi(z, z', x) &= h r_k(K') + z' w_k(K').
\end{aligned}
$$

The sensitivity function of the savings policy $h_\pi$ is given by Lemma 2. Further,

$$
\begin{aligned}
\mathcal{G}_\pi &= 1, \\
\mathcal{G}_{\pi_x} &= 0, \\
\mathcal{G}'_\iota &= -\int g\psi(z, z', x)h_i + \xi(z, z', x)h_i, \\
\xi(z, z', x) &= \pi_x(x')x'_k - gr.
\end{aligned}
$$

For the equation

$$
\begin{aligned}
\frac{d}{dx}\mathcal{W}_{\pi_x}(z,x) &= \mathcal{W}_{\pi_x\pi_x}(z,x)\tilde{\pi}_{xx}(x) + \mathcal{W}_{\pi_x\pi}(z,x)\tilde{\pi}_x(x) + \mathcal{W}_{\pi_x x}(z,x), \\
\frac{d}{dx}\mathcal{W}'_{\pi_x}(z,z',x) &= \mathcal{W}'_{\pi_x\pi_x}(z,z',x)\tilde{\pi}_{xx}(x) + \mathcal{W}'_{\pi_x\pi}(z,z',x)\tilde{\pi}_x(x) + \mathcal{W}'_{\pi_x x}(z,z',x),
\end{aligned}
$$

we obtain, for $\iota \in \{\pi(x), \pi_x(x)\}$,

$$
\begin{aligned}
\mathcal{W}_{\pi_x\iota}(z,x) &= u_c c_{\pi_x\iota} + u_{cc} c_{\pi_x} c_\iota, \\
\mathcal{W}_{\pi_x x}(z,x) &= u_c c_{\pi_x x} + u_{cc} c_{\pi_x} c_x,
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{W}'_{\pi_x\iota}(z,z',x) &= [u_{cc}c'_\iota R' + u_c R'_\iota]h_{\pi_x} + u_c R h_{\pi_x\iota} + \int \left\{[u_{cc}c'_\iota\psi + u_c\psi_\iota]h_{\pi_x} + u_c\psi h_{\pi_x\iota}\right\}, \\
\mathcal{W}'_{\pi_x x}(z,z',x) &= [u_{cc}c'_x R' + u_c R'_x]h_{\pi_x} + u_c R h_{\pi_x k}k_x + u'_c\psi h_{\pi_x},
\end{aligned}
$$

with

$$
\begin{aligned}
c_{\pi_x\iota} &= -h_{\pi_x\iota}, \\
c_{\pi_x x} &= -h_{\pi_x k}k_x \\
c_x &= [1 + r(K) - h_k]\,k_x - \tilde{\pi}_x(x), \\
c'_\iota &= h_\iota R + \psi K_\iota, \\
c'_x &= [1 + r(K') - \tilde{\pi}_x(x')x'_k]\,h_k k_x. \\
R &= 1 + r(K') - \tilde{\pi}_x(x')x'_k, \\
R_\iota &= r_K(K')K'_\iota - \left[\tilde{\pi}_{xx}(x')(x'^2_k) + \tilde{\pi}_x(x')x'_{kk}\right](h_\iota)^2, \\
R_x &= -\left[\tilde{\pi}_{xx}(x')\left(x'^2_k\right) + \tilde{\pi}_x(x')x'_{kk}\right]h_k k_x, \\
K'_\iota &= \sum_z \int h_\iota \lambda dk, \\
\psi_\iota &= h_\iota r(K') + [h r_{KK}(K') + z' w_{KK}(K')]\,K'_\iota.
\end{aligned}
$$

According to (14),

$$
\begin{aligned}
\frac{d}{dx}\mathcal{G}_{\pi_x}(z,x) &= 0, \\
\frac{d}{dx}\mathcal{G}'_{\pi_x}(z,z',x) &= \mathcal{G}'_{\pi_x\pi_x}(z,z',x)\tilde{\pi}_{xx}(x) + \mathcal{G}'_{\pi_x\pi}(z,z',x)\tilde{\pi}_x(x) + \mathcal{G}'_{\pi_x x}(z,z',x), \\
\mathcal{G}'_{\pi_x\iota}(z,z',x) &= -\int g\left[\psi_\iota h_{\pi_x} - \psi h_{\pi_x\iota}\right] + \xi_\iota h_{\pi_x} + \xi h_{\pi_x\iota}, \\
\mathcal{G}'_{\pi_x x}(z,z',x) &= -g\psi h_{\pi_x} + \xi_x h_{\pi_x} + \xi h_{\pi_x k}k_x,
\end{aligned}
$$

where

$$
\begin{aligned}
\xi_\iota &= \left[\tilde{\pi}_{xx}(x')(x'_k)^2 + \tilde{\pi}_x(x')x'_{kk}\right]h_\iota - g r(K')K'_\iota, \\
\xi_x &= \left[\tilde{\pi}_{xx}(x')(x'_k)^2 + \tilde{\pi}_x(x')x'_{kk}\right]h_k k_x.
\end{aligned}
$$

## B.2 The Euler-Lagrange Equation for a Flat-Tax Schedule

If the Euler-Lagrange equation (19) were constrained to be a flat tax rate $\tau$. Without $\pi(x)$ there is no need for the calculus of variation and we can take just simple derivatives with respect to $\tau$. Since the tax is not a function of $x$ the term $d/dx$ will disappear from the equation. The boundary transversality conditions for $\pi$ do not exist and the Euler-Lagrange equation holds at inequality,

$$\sum_z \Big\{ \Big[ -u_c(c) + \beta \sum_{z'} Q(z,z') u_c(c')[1+r-\tau x'_k] \Big] h^*_\pi \tag{B.15}$$
$$-u_c(c) + \mu$$
$$+\beta \sum_{z'} Q(z,z') \Big( \int \psi(z,z',x) \big[u_c(c') - \mu g\big] h^*_\tau \lambda^*(z,k(z,x)) k_x(z,x) dx$$
$$+\mu \xi(z,z',x) h^*_\tau - u_c(c') x' \Big) \Big\} \lambda^*(z,k(z,x)) k_x(z,x) dx \leq 0,$$

with $\xi(z,z',x) = \tau x_k(z', h^*(z,k(z,x))) - gr(K^*)$. The return on the next period capital in the Euler-Lagrange equation is $[1+r-\tau x'_k]$.

## B.3 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 $\sum_{z\in Z} \mathbf{L}_{\pi_x \pi_x}(z,x) < 0$ for all $x \in [\underline{x}(\underline{z}), \overline{x}(\overline{z})]$ . 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

$$J_{\varepsilon\varepsilon}(0) = \frac{1}{2} \sum_{z\in Z} \int_{\underline{x}(z)}^{\overline{x}(z)} \left\{ \mathbf{L}_{\pi\pi}(z,x)[\delta\pi(x)]^2 + 2\mathbf{L}_{\pi\pi_x}(z,x)\delta\pi(x)\delta\pi_x(x) + \mathbf{L}_{\pi_x\pi_x}(z,x)[\delta\pi_x(x)]^2 \right\} dx,$$

where

$$\mathbf{L}(z,x) \equiv \Big[ \mathcal{L}(z,x) + \beta \sum_{z'} Q(z,z') \mathcal{L}'(z,z',x) \Big] \lambda(z,k(z,x)) k_x(z,x).$$

The second term is

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

The Taylor expansion can be expressed as

$$J_{\varepsilon\varepsilon}(0) = \sum_{z\in Z} \int_{\underline{x}(z)}^{\overline{x}(z)} \left\{ P(z,x) [\delta\pi_x(x)]^2 + Q(z,x) [\delta\pi(x)]^2 \right\} dx, \tag{B.16}$$

where

$$P(z,x) \equiv \frac{1}{2}\mathbf{L}_{\pi_x\pi_x}(z,x),$$
$$Q(z,x) \equiv \frac{1}{2}\left[\mathbf{L}_{\pi\pi}(z,x) - \frac{d}{dx}\mathbf{L}_{\pi\pi_x}(z,x)\right].$$

If $\pi$ 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 $\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)} P(z,x)\,[\delta\pi_x(x)]^2\,dx$ 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.16) containing $\delta\pi$ are bounded and, therefore, dominated by $\sum_{z\in Z}\int_{\underline{x}(z)}^{\overline{x}(z)} P(z,x)\,[\delta\pi_x(x)]^2\,dx$. This implies that the necessary condition for a maximum is the so called Legendre condition

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

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 $\pi$. The definition of $\mathbf{L}=\mathcal{W}+\mu\mathcal{G}$ gives us $\mathbf{L}_{\pi_x\pi_x} = \mathcal{W}_{\pi_x\pi_x} + \mu\mathcal{G}_{\pi_x\pi_x}$. The expression for $\mathcal{W}$ in (12) leads to

$$\mathbf{L}_{\pi_x} = \Big(\beta\sum_{z'} Q(z,z')\,[u_c'(c') - \mu g]\,\psi(z,z',x)K'_{\pi_x} + \mu\,[\pi_x(x')x'_k - gr(K')]\,h_{\pi_x}\Big)\lambda k_x,$$

and

$$\begin{aligned}\mathbf{L}_{\pi_x\pi_x} \;=\; & \Big\{\beta\sum_{z'} Q(z,z')\{u_{cc}(c')c'_{\pi_x}\psi K'_{\pi_x} + u_c'(c')[\psi_{\pi_x}K'_{\pi_x} + \psi K'_{\pi_x}] \\ & +\mu\{[(\pi_{xx}(x')(x'_k)^2 + \pi_x(x')x'_{kk})h_{\pi_x} - gr(K')K'_{\pi_x}]h_{\pi_x} \\ & +[\pi_x(x')x'_k - gr(K')]h_{\pi_x\pi_x} - g[\psi_{\pi_x}K'_{\pi_x} + \psi K'_{\pi_x\pi_x}]\}\}\Big\}\lambda(z,k(z,x))k_x(z,x).\end{aligned}$$

When we consider the second and higher order terms negligible, the above equation reduces to the following condition

$$\begin{aligned}\mathbf{L}_{\pi_x\pi_x} = \beta\sum_{z'} Q(z,z')\,\big\{& u_{cc}(c')\psi(z,z',x)K'_{\pi_x}c'_{\pi_x} + \\ & \mu\left[(\pi_{xx}(x')(x'_k)^2 + \pi_x(x')x'_{kk})h_{\pi_x} - gr(K')K'_{\pi_x}\right]h_{\pi_x}\big\}\,\lambda k_x,\end{aligned} \tag{B.17}$$

where the first term is clearly negative due to $u'_{cc}(c) < 0$. For persistent idiosyncratic shocks at the upper bound $\psi(\overline{z},\overline{z},\overline{x}(\overline{z})) < 0$ and $c'_{\pi_x}(\overline{z},\overline{x}(\overline{z})) > 0$ (note that $c_{\pi_x} = -h_{\pi_x}$). Using $K'_{\pi_x} < 0$, the first term is negative.

The sign of the second term follows from the sensitivity function with respect to the tax schedule derivative $h_{\pi_x}$ at $(\overline{z},\overline{x}(\overline{z}))$ from the functional equation given by the Frechet derivative of the Euler equation operator $\mathcal{F}_\pi$. We consider what is the effect on the next-period capital function $h$ when $\pi_x$ shifts marginally. The marginal shift of $\pi_x$ lowers savings,

$h_{\pi_x}(\overline{z}, \overline{x}(\overline{z})) < 0$. The first term inside the square brackets on the second line is negative. As $K'_{\pi_x} < 0$ the second term is positive and much smaller than the first term due to $r(K') \ll 1$. Since $\mu < 0$ and $h_{\pi_x} < 0$ the whole term on the second line is negative and the whole expression $\mathbf{L}_{\pi_x\pi_x} < 0$. Considering the other boundary point at the lower bound, where $\psi(\underline{z}, \underline{z}, \underline{x}(\underline{z})) > 0$ and $c'_{\pi_x} < 0$, the first term in (B.17) is again negative. Since $x'_{kk}$ is close to zero (in the case of total income $x'_{kk} = 0$) the first term on the second line is negative even when $\pi_x < 0$, and $\mathbf{L}_{\pi_x\pi_x} < 0$. Therefore, $\mathbf{L}_{\pi_x\pi_x}(z,k) < 0$ for each $z \in Z$ and $k \in [\underline{k}(\underline{z}), \overline{k}(\overline{z})]$, thus $\sum_{z\in Z} \mathbf{L}_{\pi_x\pi_x}(z,k) < 0$ for each $k \in [\underline{k}(\underline{z}), \overline{k}(\overline{z})]$.

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.16) can be written as

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

where

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

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

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

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

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

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

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

$$S\left(R + w_x\right) = w^2, \tag{B.18}$$

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.18) on the whole interval $[a, b]$. The first step is to transform the quadratic differential equation (B.18) into a linear second-order differential equation using a substitution

$$w(x) = -\frac{S(x)v_x(x)}{v(x)}, \tag{B.19}$$

where $v$ is a new unknown twice differentiable function. In this way, equation (B.18) 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

$$Rv = \frac{d}{dx}\left(Sv_x\right). \tag{B.20}$$

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.18) is given by equation (B.19). 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.20) 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 $v\left(a\right) = 0$ and $v_x(a) = 1$. Using the condition $v_x\left(a\right) = 1$, 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.

$$0 > -\underline{S} \geq -S(t) \geq -\overline{S} > -\infty. \tag{B.21}$$

From (B.17)

$$\begin{aligned} S(x) = \frac{1}{2}\beta\sum_{z'} Q(z, z')\left\{u_{cc}(c')\psi(z, z', x)K'_{\pi_x}c'_{\pi_x} + \right. \\ \left. \mu\left[(\pi_{xx}(x')(x'_k)^2 + \pi_x(x')x'_{kk})h_{\pi_x} - gr(K')K'_{\pi_x}\right]h_{\pi_x}\right\}\lambda k_x. \end{aligned} \tag{B.22}$$

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 $\overline{c} = w(K)\overline{z} + r(K)\overline{k}$. Thus, $0 < \underline{c} \leq c \leq \overline{c} < \infty$ and $-\infty < u_{cc}(\underline{c}) \leq u_{cc}(c) \leq u_{cc}(\overline{c}) < 0$. All other terms in the formula are clearly bounded. Thus, $S$ given by (B.22) is bounded as in (B.21) and

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

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 equation are also the sufficiency conditions and the obtained solution is a maximum. Q.E.D.

## B.4 Proof of Lemma 2

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

$$\mathcal{F}(h)(z,k;\tilde{\pi}(x),\tilde{\pi}_x(x)) \equiv u_c(c(z,k;\tilde{\pi}(x),\tilde{\pi}_x(x))) -$$
$$\beta\sum_{z'} u_c(c(z',h(z,k;\tilde{\pi}(x),\tilde{\pi}_x(x));\tilde{\pi}(x),\tilde{\pi}_x(x)))R(z',h(z,k;\tilde{\pi}(x),\tilde{\pi}_x(x));\tilde{\pi}(x),\tilde{\pi}_x(x))Q(z,z'),$$

and

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

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 $R' = R\left(z',h\right)$,

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

where

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

Terms for $\mathcal{F}_\pi$ are

$$\begin{aligned}
c_\pi &= -1 - h_\pi,\\
c'_\pi &= \left[1 + r(K')\right]h_\pi + \left[r_k(K')h + w_k(K')z'\right]K'_\pi,\\
R'_\pi &= r_k(K')K'_\pi - \Pi_k(x')h_\pi,
\end{aligned}$$

with

$$\begin{aligned}
\Pi(x') &\equiv \tilde{\pi}_x(x')x'_k\\
\Pi_k(x') &= \tilde{\pi}_{xx}(x')(x'_k)^2 + \tilde{\pi}_x(x')x'_{kk}.
\end{aligned}$$

For

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

the terms are

$$
\begin{aligned}
c_{\pi_x} &= -h_{\pi_x}, \\
c'_{\pi_x} &= [1 + r(K') - \Pi(x')]\, h_{\pi_x} + [r_k(K')h + w_k(K')z']\, K'_{\pi_x}, \\
R'_{\pi_x} &= r_k(K')K'_{\pi_x} - \Pi_k(x')h_{\pi_x}.
\end{aligned}
$$

For

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

we use terms

$$
\begin{aligned}
c_{\pi_x\pi_x} &= -h_{\pi_x\pi_x}, \\
c'_{\pi_x\pi_x} &= [1 + r(K') - \Pi(x')]\, h_{\pi_x\pi_x} + [r_k(K')h + w_k(K')z']\, K'_{\pi_x\pi_x} \\
&\quad + \left[r_{KK}(K')K'_{\pi_x} - \Pi_k(x')h_{\pi_x}\right] h_{\pi_x} \\
&\quad + \left[r_k(K')h_{\pi_x} + \left(r_{KK}(K')h + w_{KK}(K')z'\right) K'_{\pi_x}\right] K'_{\pi_x}, \\
R'_{\pi_x\pi_x} &= r_{KK}(K')K'_{\pi_x} + r_k(K')K'_{\pi_x\pi_x} - \Pi_{kk}(x')(h_{\pi_x})^2 - \Pi_k(x')h_{\pi_x\pi_x}.
\end{aligned}
$$

Finally,

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

with

$$
\begin{aligned}
c_{\pi_x\pi} &= -h_{\pi_x\pi}, \\
c'_{\pi_x\pi} &= [1 + r(K') - \Pi(x')]\, h_{\pi_x\pi} + [r_k(K')h + w_k(K')z']\, K'_{\pi_x\pi} \\
&\quad + [r_{KK}(K')K'_{\pi} - \Pi_k(x')h_{\pi}]\, h_{\pi_x} \\
&\quad + [r_k(K')h_{\pi} + (r_{KK}(K)h + w_{KK}(K')z')\, K'_{\pi}]\, K'_{\pi_x} \\
R'_{\pi_x\pi} &= r_{KK}(K')K'_{\pi} + r_k(K')K'_{\pi_x\pi} - \Pi_{kk}(x')h_{\pi}h_{\pi_x} - \Pi_k(x')h_{\pi_x\pi},
\end{aligned}
$$

and

$$
\Pi_{kk}(x') = \tilde{\pi}_{xxx}(x')(x'_k)^3 + 3\tilde{\pi}_{xx}(x')x'_k x'_{kk} + \tilde{\pi}_x(x')x'_{kkk}.
$$

Q.E.D.

## B.5 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: $\Psi = \{x \in B : 0 \leq x \leq \varphi(k,z) \text{ for all } (k,z) \in B \times Z\}$. If the function $\varphi$ is increasing and quasi-concave, then the set $\Psi$ 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 $\bar{k}(z) > 0, \varphi(k,z) - k$ 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

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

where

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

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 $\varphi$ is increasing in $k$, i.e. $\varphi_k > 0$.

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 $0 \leq h(k,z) \leq \varphi(k,z)$, all $(k,z) \in B \times Z$, and $h$ and $\varphi - h$ 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 $F$,

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

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, $\pi$, is a solution of the following system of operator equations:

1. FOC for $\pi$ given by the Euler-Lagrange equation in (16);
2. the Euler equation (10) capturing the individual optimal behavior $h$;
3. five operator equations (20)-(21) for F-derivatives of $h$ based on the Euler equation, $h_\pi$, $h_{\pi_x}$, $h_K$, $h_{\pi_x\pi_x}$, and $h_{\pi_x\pi}$; and
4. the operator equation for distribution function, $\lambda$, in (11).

In order to solve the problem numerically, we first approximate all the unknown functions by combinations of polynomials from a polynomial base. Approximated solutions are specified by unknown parameters transforming the original infinitely dimensional problem into a finite dimensional one. After substituting the approximated functions into the original operator equations we construct the *residual equations*. Ideally, the residual functions should be uniformly equal to zero. In practical situations, however, this is not achievable and we limit the problem to a finite number of conditions, the so called *projections*, whose satisfaction guarantees a reasonably good approximation. There are many possibilities how to define the projections.[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 (10), (11), (16), and (20)-(21), there are seven unknown classes of functions $\{\pi, h, h_\pi, h_{\pi_x}, h_{\pi_x\pi_x}, h_{\pi_x\pi}, \lambda\}$. 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 $\{h^i(k), \lambda^i(k), h^i_\pi(k), h^i_{\pi_x}(k), h^i_{\pi_x\pi_x}(k), h^i_{\pi_x\pi}(k)\}_{i=1}^J$, 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 $\lambda^i$ 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 distribution. Let the interval of approximation be $[\underline{k}, \widehat{\overline{k}}]$ and the degrees of approximation for $\{h^i(k), h^i_\pi(k), h^i_{\pi_x}(k), h^i_{\pi_x\pi_x}(k), h^i_{\pi_x\pi}(k), \lambda^i(k), \pi(x)\}$ be $M, M_\pi, M_{\pi_x}, M_{\pi_x\pi_x}, M_{\pi_x\pi}, N, P \geq 2$, respectively.[31]

---

[30]For an excellent survey and description of these methods see Chapter 11 in Judd (1998). Reiter (2009) uses perturbations to solve for aggregate dynamics around a steady-state equilibrium. Evans (2017) uses the perturbation theory with a truncated Taylor expansion to study the role of capital taxation in an economy with uninsurable investment risk. Golosov, Tsyvinski, and Werquin (2014) use variational approach for their analysis of optimal tax systems.

[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 $\xi(k) = 2(k - \underline{k})/(\overline{k} - \underline{k}) - 1$.

Thus, we obtain

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

with $i \in \{1, 2, ..., J\}$ and $m \in \{\emptyset, \pi, \pi_x, \pi_x\pi_x, \pi_x\pi\}$, and $\widehat{\pi}(x; c) \equiv \sum_{j=1}^{P} c_j\phi_j(x)$, for any $k \in [\underline{k}, \overline{k}]$ and $x \in [r(K)\underline{k} + w(K)\underline{z}, r(K)\overline{k} + w(K)\overline{z}]$, 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 (10), (11), (16), and (20)-(21). Substituting the above approximations for the unknown functions,

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

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

$$R^{\mathcal{L}^i}(k; \mathbf{p}) = \mathcal{L}^i(\widehat{\mathbf{h}}, \widehat{\lambda}), \quad \text{(C.3)}$$

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

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

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

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

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,

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

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

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

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

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, $\lambda$, the related sensitivity functions $h_\pi$ and the optimal government policy function, $\pi$, 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 $h_\pi$ of the order $10^{-2}$.

# 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 $y(z, k(z, x)) = x$, and $k_x(z, x) = 1/r(K)$.

## D.2 Extra Term for Pecuniary Externality of Aggregate Capital

In our economy, the extra term in the constrained optimal allocation in Davila, Hong, Krusell, and Rios-Rull (2012) corresponds to

$$\Delta \equiv \beta \sum_{z} \int \sum_{z'} u_c(h(z,k), z') \left[ h(k,z) F_{KK}(K) + z' F_{LK}(K) \right] Q(z,z') \, \lambda(k,z) dk.$$

The term $F_{KK}(K) < 0$ and $F_{LK}(K) > 0$. It is thus the sum of next period changes in total income weighted by marginal utility of each agent. In a representative agent economy $\Delta = 0$.

The sign of $\Delta$ depends on the persistence properties of the individual labor productivity shocks. The fraction of poor (in particular borrowing constrained) agents with high marginal utility is the main determinant of the extra term. It is positive if a sufficient number of poor agents have labor-intensive income and would benefit from more aggregate capital stock. If the labor productivity shock is iid, then agents would use savings to insure against idiosyncratic risk. Then the capital income would be relatively high and the extra term negative.

Finally, the extra term is an aggregate effect which is the same for all agents and more distortionary for agents with low marginal utility. This implies that changes in either direction will be mostly implemented by the rich agents.

The Euler-Lagrange equation for the first-order condition of the constrained efficiency problem is

$$\begin{aligned} -u_c(c) &+ \beta \sum_{z'} Q(z,z') u_c(c')[1 + r] \\ &+ \beta \sum_{z} \sum_{z'} Q(z,z') \int \psi(z,z',x) u_c(c') \lambda^*(z, k(z,x)) k_x(z,x) dx \le 0, \end{aligned}$$

where the last term equal to the extra term $\Delta$.

## D.3 Additional Distributional Effects

Table D.1: Steady State Distribution of Income and Consumption

| | Consumption | | | Income | | |
|---|---|---|---|---|---|---|
| | Progress. | Flat | Optimal | Progress. | Flat | Optimal |
| 1st Quintile | 5.97 | 6.40 | 5.39 | 9.07 | 6.34 | 7.53 |
| 2nd Quintile | 6.04 | 6.80 | 6.15 | 9.20 | 6.86 | 7.96 |
| 3rd Quintile | 10.52 | 16.40 | 11.88 | 14.19 | 16.66 | 15.59 |
| 4th Quintile | 17.68 | 18.60 | 18.42 | 20.92 | 20.91 | 24.09 |
| 5th Quintile | 65.76 | 65.00 | 63.56 | 55.69 | 62.44 | 51.16 |
| Top 10% | 54.25 | 54.34 | 51.76 | 38.19 | 44.83 | 34.53 |
| Top 5% | 40.26 | 41.12 | 39.04 | 23.61 | 28.29 | 21.12 |
| Top 1% | 8.88 | 8.82 | 8.10 | 6.14 | 7.52 | 4.84 |

Note: Each entry is the percentage share by each group.

Table D.1 shows additional results on distribution of resources by quintiles and top groups of agents in the stationary equilibrium for each economy. Table 3 in the main paper shows that in the benchmark, high wealth dispersion economy, the optimal limiting tax schedule reduces taxation of the persistently poor agents. As income is more equally distributed with no major differences across the three steady states, consumption by agents in the third and fourth quintiles increases. Agents in the top quintile still consume more than a half of the aggregate consumption.

## D.4 Transition to the Optimal Tax Schedule Steady State

Figure D.1 shows welfare gains by individual wealth and productivity shock, $(k, z)$, where $z$ is either low, medium, or high as in Table 1. These individual welfare gains aggregate to the average welfare gain from transition in Table 4.

The welfare gains and losses are much smaller on the transition from the progressive-tax steady state than from the flat-tax steady state. Compared to the progressive tax schedule, the optimal limiting tax schedule reduces insurance to poor agents and brings their welfare losses. In the progressive-tax steady state, agents whose total income falls into the trough of the U-shape optimal limiting tax schedule benefit from the reform.

In the transition from the flat-tax steady state, agents with the high idiosyncratic shock lose from the reform while the other two types of agents benefit. The welfare gains of agents with the low and medium productivity shocks are positive for two reasons. First, their insurance is improved. Second, the aggregate capital stock in the optimal tax steady state is much lower than in the flat tax steady state. This allows agents to consume individual savings during the initial periods of transition. The bottom panel shows large welfare gains for agents with medium and low productivity shocks. Only the high-skill agents suffer losses as they pay much higher taxes than in the pre-reform steady state.

## D.5 Different Welfare Criteria

Table D.2 exhibits detailed results for the benchmark economy illustrated by Figures 3 and 4 in the text. Each column shows a steady state in which the government discounts

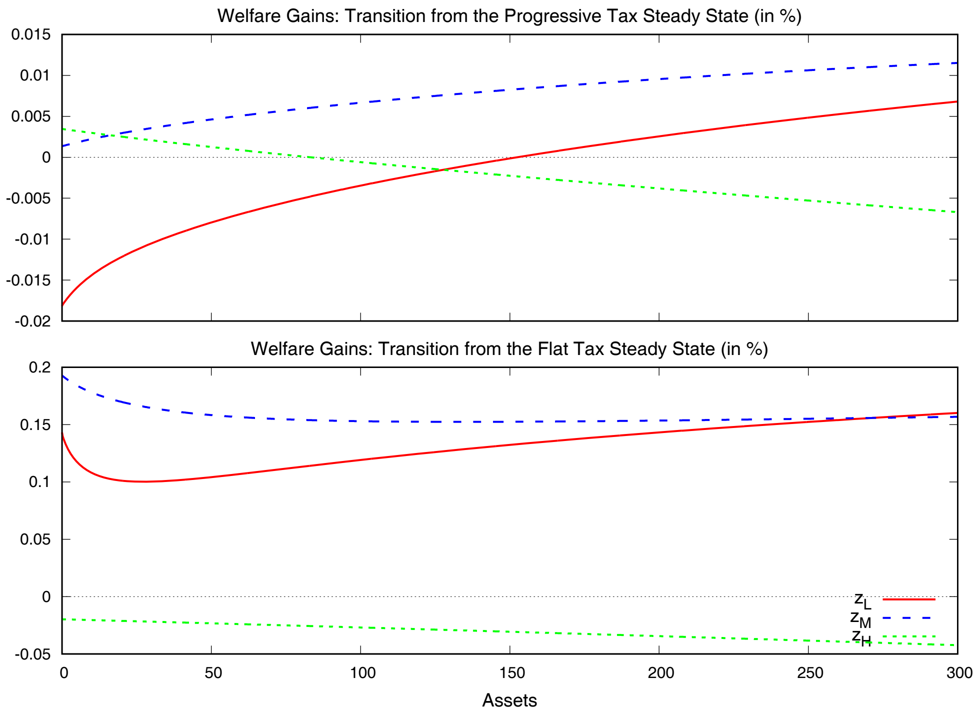


Figure D.1: Welfare gains from transition.

future by a different $\hat{\beta}$. In all steady states agents discount future by $\beta = 0.9303$ as in the benchmark economy.

As social discount factor grows, capital accumulation increases, together with wealth inequality. Because inequality and lower insurance hurts more the poor agents with high marginal utility, the steady state average welfare falls. When $\hat{\beta} = 0.89$ the Gini coefficient of inequality falls to 0.485 with the top five percent of the wealth distribution owning only 13% of the capital stock. For $\hat{\beta} < \beta$, the optimal tax reform of the progressive tax steady state is supported by the majority of the population. This is because agents can consume the capital stock during the transition. For the same reason, the optimal tax reform of the flat tax economy makes the majority of agents always better off. For low social discount factors the average welfare gains are very large. When $\hat{\beta} \leq \beta$ in the flat-tax economy, only agents with the high labor productivity are hurt by the large drop in equilibrium wages. Gains and losses from the optimal tax schedule are calculated using the agents' discount factor.

## D.6 Sensitivity Analysis for Different Values of $\sigma$

We analyze the optimal limiting tax for four different values of $\sigma \in \{0.5, 1, 2, 4\}$, where $\sigma = 2$ is the benchmark value and $\theta = 1/\sigma$ is the intertemporal elasticity of substitution. The discount factor is adjusted for each $\sigma$ so that in the progressive tax steady state the capital-output ratio is 3 and the equilibrium interest rate is 4 percent. All other parameters are kept the same as in the benchmark economy.

Figure D.2 shows the optimal tax schedules and optimal marginal tax schedules for different values of $\sigma$. For the highest degree of risk aversion the average tax rate is negative

Table D.2: Government Discount Factor: Steady State Allocations

| | Optimal Tax Schedule Steady State with $\hat{\beta}$ | | | | | |
|---|---|---|---|---|---|---|
| | 0.89 | 0.91 | 0.9309* | 0.95 | 0.97 | 0.99 |
| Steady State | | | | | | |
| Aggregate assets | 2.621 | 2.766 | 3.093 | 3.165 | 3.322 | 3.611 |
| Output | 0.953 | 0.971 | 1.011 | 1.015 | 1.033 | 1.069 |
| Capital-output ratio | 2.752 | 2.848 | 3.059 | 3.118 | 3.216 | 3.378 |
| Interest rate (%) | 5.085 | 4.648 | 3.771 | 3.596 | 3.342 | 2.661 |
| Coeff. of variation of wealth | 0.848 | 1.076 | 1.501 | 1.534 | 1.680 | 1.907 |
| Gini wealth | 0.485 | 0.579 | 0.698 | 0.736 | 0.758 | 0.799 |
| Perc. of wealth of the top 5% | 0.131 | 0.192 | 0.286 | 0.294 | 0.326 | 0.378 |
| Steady State Average Welfare | 5.506 | 5.418 | 5.277 | 5.269 | 5.205 | 5.099 |
| Transition Gains from the Progressive Tax Steady State (in %) | | | | | | |
| Average Welfare | 9.890 | 6.267 | -0.767 | -1.397 | -4.327 | -9.785 |
| Aggregate Component | 2.140 | 1.120 | -0.580 | -0.800 | -1.850 | -4.000 |
| Distributional Componen | 7.588 | 5.090 | -0.188 | -0.602 | -2.524 | -6.026 |
| Political Support | 68.573 | 57.161 | 46.471 | 40.201 | 22.641 | 10.546 |
| Transition Gains from the Flat Tax Steady State (in %) | | | | | | |
| Average Welfare | 26.956 | 22.892 | 14.884 | 14.163 | 10.809 | 4.551 |
| Aggregate Component | 9.340 | 8.170 | 5.670 | 5.420 | 4.260 | 1.900 |
| Distributional Component | 16.111 | 13.610 | 8.720 | 8.293 | 6.281 | 2.602 |
| Political Support | 94.129 | 94.129 | 94.129 | 63.934 | 58.746 | 54.130 |

Note: Agents discount factor $\beta = 0.9309$ in all steady states.

around the median wealth. The optimal marginal tax functions are also U-shaped and negative at low levels of income.

For the benchmark economy calibrated according to Davila, Hong, Krusell, and Rios-Rull (2012) with a high wealth dispersion in the steady state, in Table D.3 we observe that capital accumulation increases with higher values of $\sigma$ in the flat-tax steady states but remains almost constant in the optimal-tax steady states. Wealth inequality falls in all steady states, but most dramatically in steady states with the optimal tax schedules. The optimal tax schedule delivers the highest steady state average welfare level in economies where agents are most risk-averse. The flat tax reform increases the aggregate capital stock so much that its deaccumulation allows for a sufficient consumption stream during the transition process: in the $\sigma$=4 economy the average welfare gain is almost 20 percent.

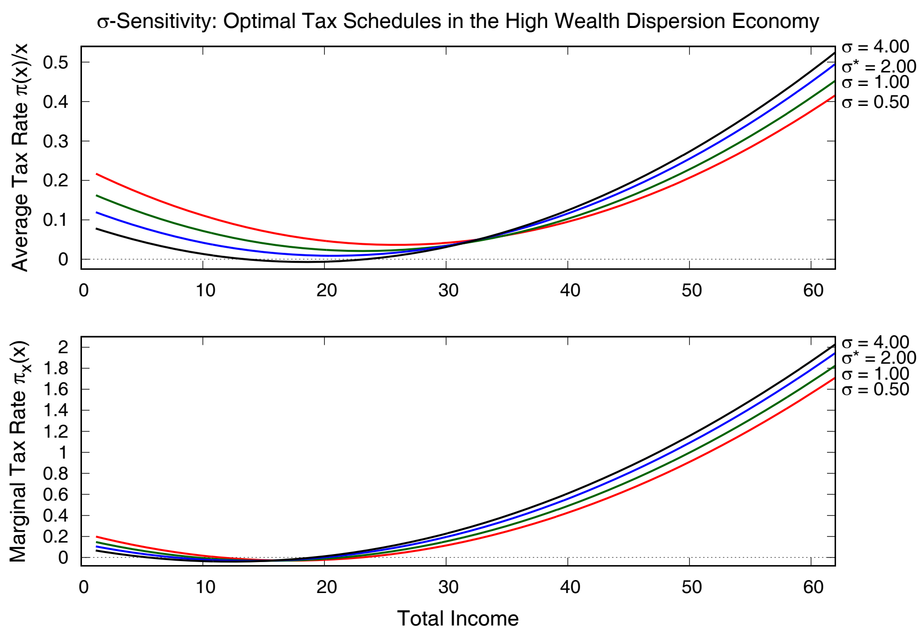


Figure D.2: Optimal average and marginal tax rates for different values of the risk aversion coefficient.

Table D.3: Sensitivity Analysis $\sigma$: Steady State and Transition

| | $\sigma = 0.5$ ($\beta$=0.966) | | | $\sigma = 1$ ($\beta$=0.956) | | | $\sigma = 2$ ($\beta$=0.9309) | | | $\sigma = 4$ ($\beta$=0.869) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Average | Prog. | Flat | Opt. | Prog. | Flat | Opt. | Prog. | Flat | Opt. | Prog. | Flat | Opt. |
| $\tau$ | 0.161 | 0.255 | — | 0.161 | 0.258 | — | 0.161 | 0.265 | — | 0.161 | 0.273 | — |
| Aggregate assets | 3.000 | 3.365 | 3.178 | 3.000 | 3.563 | 3.171 | 3.000 | 3.934 | 3.093 | 3.000 | 4.420 | 3.076 |
| Output | 1.000 | 1.042 | 1.021 | 1.000 | 1.064 | 1.020 | 1.000 | 1.102 | 1.011 | 1.000 | 1.149 | 1.010 |
| Capital-output ratio | 3.000 | 3.229 | 3.113 | 3.000 | 3.349 | 3.108 | 3.000 | 3.568 | 3.059 | 3.000 | 3.847 | 3.049 |
| Interest rate (%) | 4.000 | 3.153 | 3.569 | 4.000 | 2.759 | 3.583 | 4.000 | 2.092 | 3.771 | 4.000 | 1.361 | 3.810 |
| Coeff. of variation of wealth | 2.280 | 2.511 | 1.455 | 2.175 | 2.464 | 1.553 | 2.125 | 2.313 | 1.501 | 1.998 | 2.124 | 1.318 |
| Gini wealth | 0.847 | 0.871 | 0.733 | 0.834 | 0.867 | 0.725 | 0.812 | 0.838 | 0.698 | 0.776 | 0.797 | 0.596 |
| Perc. of wealth of the top 5% | 0.443 | 0.515 | 0.292 | 0.437 | 0.504 | 0.283 | 0.433 | 0.475 | 0.276 | 0.413 | 0.440 | 0.259 |
| Steady State Average Welfare | 5.543 | 5.409 | 5.462 | 5.352 | 5.152 | 5.341 | 5.188 | 4.975 | 5.277 | 5.254 | 5.136 | 5.329 |
| Transition Gains (in %) | | | | | | | | | | | | |
| Average Welfare | -3.152 | 2.956 | — | -2.151 | 8.362 | — | -0.767 | 14.884 | — | -0.682 | 18.988 | — |
| Aggregate Component | -0.640 | 0.590 | — | -0.770 | 1.880 | — | -0.580 | 5.670 | — | -0.528 | 11.790 | — |
| Distributional Component | -2.528 | 2.352 | — | -1.392 | 6.362 | — | -0.188 | 8.720 | — | -0.154 | 6.439 | — |
| Political Support | 17.446 | 52.568 | — | 25.793 | 56.794 | — | 46.471 | 94.129 | — | 48.402 | 94.129 | — |

Note: The benchmark high wealth dispersion economy corresponds to $\sigma = 2$.

# E Low Wealth Dispersion Economy

In this Appendix, we also evaluate our example in Section 6 for the more traditional calibration based on Aiyagari (1994). The three-state, first order Markov process of the uninsurable idiosyncratic shock to labor productivity is estimated from the household annual labor income process in the PSID data by a first-order autoregression with a persistence parameter 0.6 and a volatility 0.2. The discount factor is set at $\beta = 0.975$ in order to arrive at the capital-output ratio equal to 3.0 in the progressive-tax steady state (the equilibrium interest rate is 4 percent).

Table E.1: Parameters of the Low Wealth Dispersion Economy

| $\beta = 0.9750$ | $\sigma = 2.0$ | $\alpha = 0.36$ | $\delta = 0.08$ | $g = 0.189$ |
|---|---|---|---|---|
| Earnings Process (Aiyagari (1994)): | | | | |
| $z \in Z = \{0.780, 1.000, 1.270\}$ | | $Q(z, z') = \begin{bmatrix} 0.660 & 0.270 & 0.070 \\ 0.280 & 0.440 & 0.280 \\ 0.070 & 0.270 & 0.660 \end{bmatrix}$ | | |

Other parameters in Table E.1 are the same as in the benchmark economy in Section 6. This calibration of the earnings process leads to a steady state with an unrealistic, low dispersion of wealth. In the progressive-tax steady state, the Gini coefficient of wealth is only 0.304.

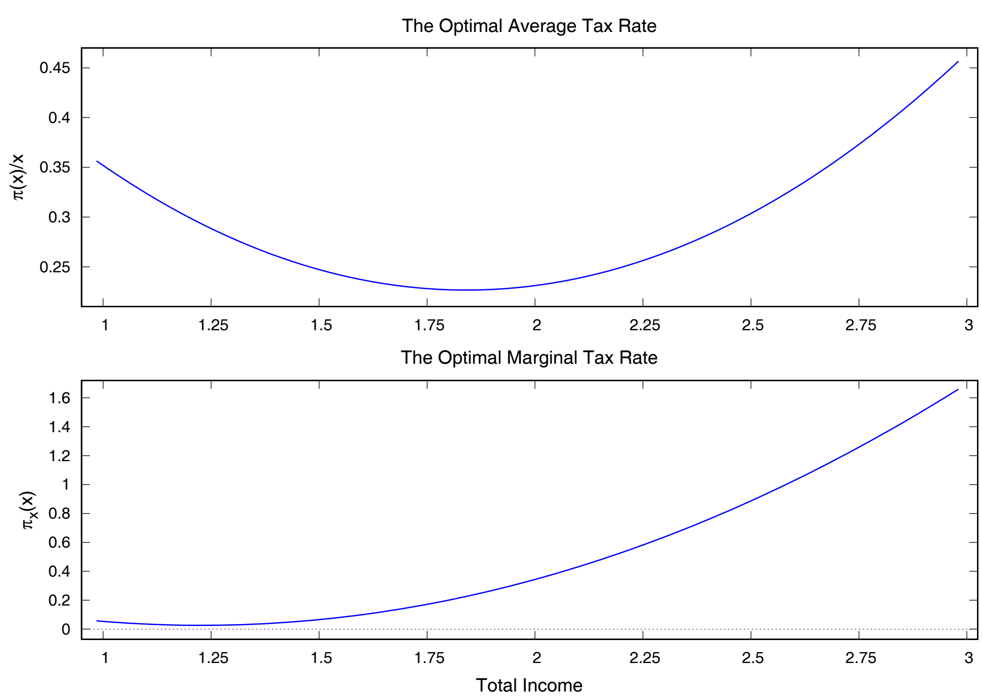


Figure E.1: The optimal average tax rate and the optimal marginal tax rate in the low wealth dispersion economy.

In the low wealth dispersion economy, the optimal limiting average tax schedule is also U-shaped with an important upward shift. In Figure E.1, the average tax rate at the lowest

Table E.2: Steady State Results: Low Wealth Dispersion Economy

| | Progressive | Flat | Optimal |
|---|---|---|---|
| $\tau$ | 0.161 | 0.253 | — |
| Aggregate assets | 3.000 | 3.273 | 3.561 |
| Output | 1.000 | 1.032 | 1.064 |
| Capital-output ratio | 3.000 | 3.172 | 3.348 |
| Interest rate (%) | 4.000 | 3.352 | 2.751 |
| Coeff. of variation of wealth | 0.516 | 0.673 | 0.498 |
| Gini wealth | 0.304 | 0.389 | 0.290 |
| Perc. of wealth of the top 5% | 0.113 | 0.148 | 0.103 |

total income is 37%, decreases to a minimum of 24%, and rises to 48% at the highest level of total income. The marginal tax rates are close to zero for low incomes and rise at higher income levels.[32]

Table E.2 shows steady state outcomes of the progressive, flat, and optimal limiting tax schedules. In this Aiyagari's low wealth dispersion economy, the earnings process is not very persistent, wealth inequality is low and poor agents can escape poverty easily. This allows the optimal limiting tax schedule to focus more on the general equilibrium price effect, increasing the aggregate capital stock by 18.7%, to a higher level than in the flat-tax steady state. The Gini coefficient for wealth declines from 0.304 to 0.29. The top five percent of richest agents hold only around 10 percent of all assets. As before, we find that the extra term of Davila et al. (2012) is positive in all simulations. This implies an overaccumulation of capital in a competitive equilibrium relative to the constrained efficient allocation attained by a social planner.

Table E.3: Steady State Distribution of Assets and Tax Contributions

| | Assets | | | Tax Contributions | | |
|---|---|---|---|---|---|---|
| | Progress. | Flat | Optimal | Progress. | Flat | Optimal |
| 1st Quintile | 6.08 | 4.28 | 6.29 | 12.46 | 14.51 | 18.76 |
| 2nd Quintile | 13.28 | 10.58 | 13.70 | 16.13 | 17.32 | 19.04 |
| 3rd Quintile | 19.13 | 16.82 | 19.58 | 19.48 | 19.71 | 19.41 |
| 4th Quintile | 25.40 | 25.05 | 25.56 | 23.69 | 22.68 | 20.46 |
| 5th Quintile | 36.10 | 43.27 | 34.88 | 28.24 | 25.78 | 22.33 |
| Top 10% | 20.27 | 25.68 | 19.21 | 14.89 | 13.54 | 11.50 |
| Top 5% | 11.13 | 14.80 | 10.33 | 7.77 | 7.08 | 5.89 |
| Top 1% | 2.54 | 3.77 | 2.34 | 1.66 | 1.54 | 1.23 |

Note: Each entry is the percentage share by each group in the low wealth dispersion economy.

The average tax rates are convex and U-shaped in income, providing incentives for agents to concentrate around the desired aggregate capital level: The low or negative marginal tax rates at low income levels motivate savings while the increasing rates on high incomes deliver tax revenues for redistribution. These distributional results are illustrated by shares of assets owned and taxes paid by different percentiles of agents in Table E.3. In the Aiyagari economy with reduced wealth inequality and low shock persistence, the

[32] Again, the optimal limiting tax schedule satisfies the admissibility condition from Corollary 1.

optimal tax schedule equalizes tax burden across agents (quintile shares between 18.8% and 22.3%). The low wealth dispersion leads to only minor differences in asset distribution in the top quintile of agents.

Table E.4: Steady State Distribution of Consumption and Income

| | Consumption | | | Income | | |
|---|---|---|---|---|---|---|
| | Progress. | Flat | Optimal | Progress. | Flat | Optimal |
| 1st Quintile | 14.79 | 14.51 | 14.92 | 17.69 | 17.13 | 17.04 |
| 2nd Quintile | 17.43 | 17.35 | 17.32 | 19.07 | 18.68 | 18.89 |
| 3rd Quintile | 19.75 | 19.70 | 19.69 | 19.90 | 19.67 | 20.03 |
| 4th Quintile | 22.53 | 22.67 | 22.72 | 20.88 | 20.93 | 21.12 |
| 5th Quintile | 25.49 | 25.77 | 25.35 | 22.46 | 23.58 | 22.93 |
| Top 10% | 13.23 | 13.56 | 13.02 | 11.56 | 12.4 | 11.81 |
| Top 5% | 6.82 | 7.08 | 6.65 | 5.92 | 6.48 | 6.06 |
| Top 1% | 1.43 | 1.55 | 1.37 | 1.24 | 1.42 | 1.27 |

Note: Each entry is the percentage share by each group in the low wealth dispersion economy.

Otherwise, the dispersion of the stationary distribution is very low and there are no significant differences in income or consumption shares in Table E.4.

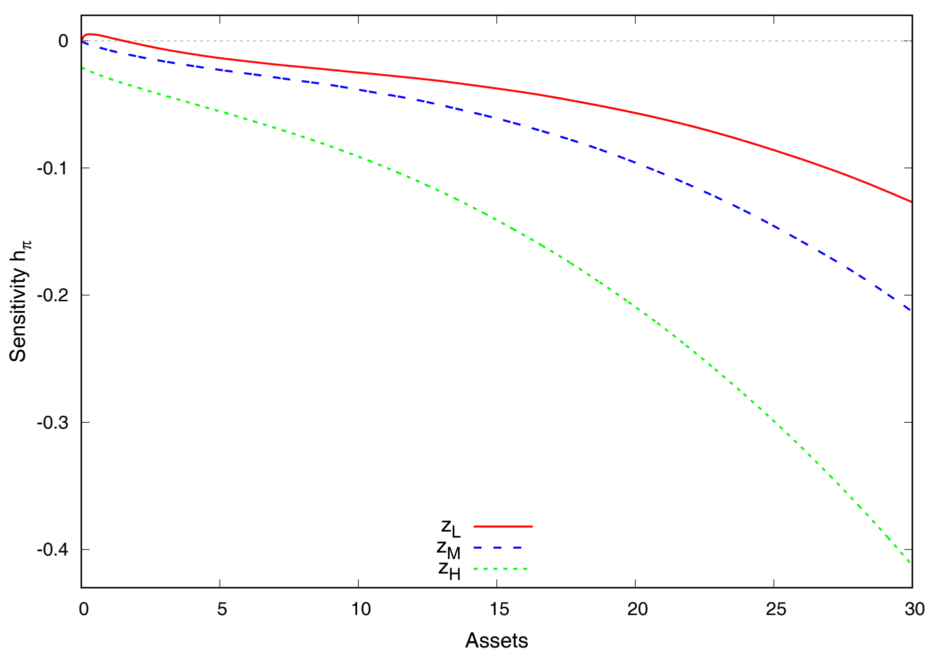


Figure E.2: Sensitivity of savings to changes in the tax schedule $h_\pi$ in the low wealth dispersion economy.

The different general equilibrium and redistribution effects of the Aiyagari (1994) calibration can be seen in Figure E.2. The sensitivity functions of savings to changes in the optimal tax schedule, $h_\pi$, for three labor productivity shocks $z_L < z_M < z_H$, show that in this calibration the general equilibrium effects dominate and a small perturbation of the optimal tax schedule induces large responses from all types of agents (compared to the benchmark calibration). Importantly, the shape of the tax schedule increases savings of the poor agents with the low labor productivity shock, $h_\pi(\underline{k}, \underline{z}) > 0$. Naturally, at higher wealth levels, a perturbation of the convex increasing tax schedule reduces savings.

## E.1 Gains from the Optimal Limiting Tax Schedule

Table E.5 shows that the optimal tax schedule increases steady state average welfare much less than in the benchmark calibration. In the low wealth dispersion economy, steady state average welfare in the optimal tax steady state is 0.9% and 0.1% higher than in the progressive and flat-tax steady state, respectively.

Table E.5: Welfare Gains (in %): Low Wealth Dispersion Economy

| | Progressive | Flat |
|---|---|---|
| Steady State | 0.908 | 0.131 |
| Transition | | |
| Average Welfare | -1.718 | -1.040 |
| Aggregate Component | -1.330 | -0.730 |
| Distributional Component | -0.393 | -0.312 |
| Political Support | 15.494 | 18.397 |

The rest of Table E.5 shows welfare gains in terms of expected present discounted values from an unanticipated reform in which the optimal limiting tax schedule is imposed on the initial progressive- or flat-tax steady state. The costs incurred during the transition are substantial.

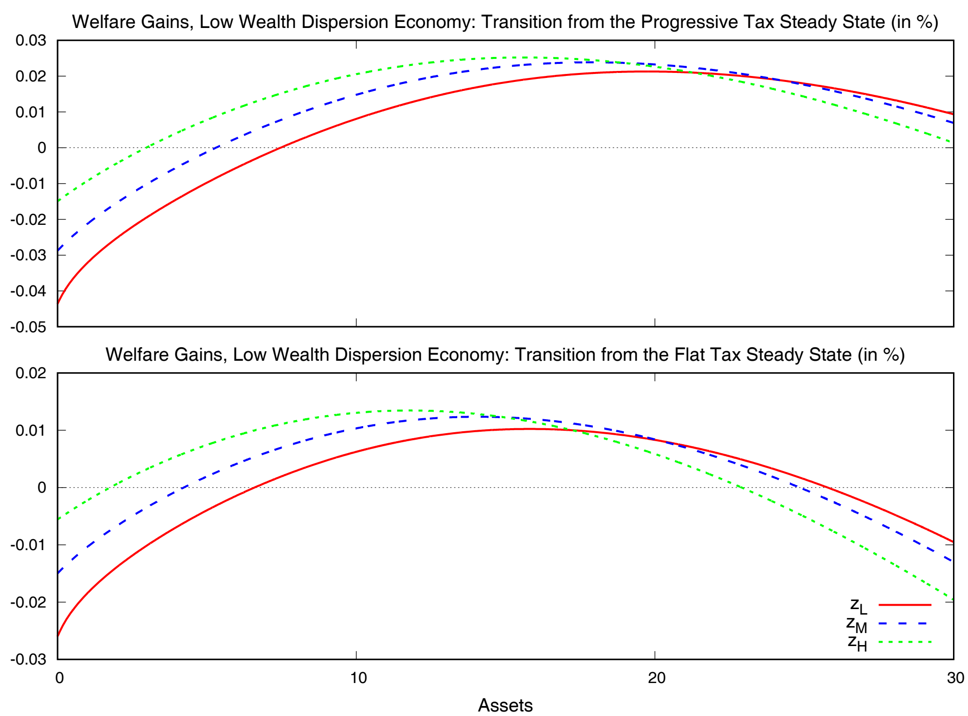


Figure E.3: Welfare gains from transition in the low wealth dispersion economy.

In the Domeij and Heathcote (2004) decomposition, the aggregate component is negative due to lower consumption during the transition as agents must accumulate a higher stock of capital. Only a small fraction of the population would be better off from the optimal tax reform. Figure E.3 shows that the cost of transition falls heavily on the asset-poor agents. In both transitions from the progressive steady state (top panel) and from the flat tax steady state (bottom panel), the gains and losses have an inverted U-shape across wealth. In the latter case, welfare losses can be observed for the poorest and richest agents who face higher tax rates.

The costs accrue mostly during the initial periods of the transition during which agents in both steady states must accumulate additional capital. Compared to the benchmark calibration, the unanticipated optimal tax reform is welfare improving for only a small fraction of the population around the median of the wealth distribution.

## E.2 Different Welfare Criteria

In the low wealth dispersion economy, aggregate general equilibrium effects are stronger and distributional effects weaker due to extremely low wealth inequality. Figure E.4 shows the optimal limiting average tax rates and the optimal marginal tax rates for different values of social discount factors $\hat{\beta}$. As in the high wealth dispersion economy, low social discount factor puts more weight on insurance against idiosyncratic risk and less weight on the long-term average price effect from the aggregate capital stock accumulation. Again, a very high social discount factor uses the U-shaped average tax schedule for long-term saving incentives.

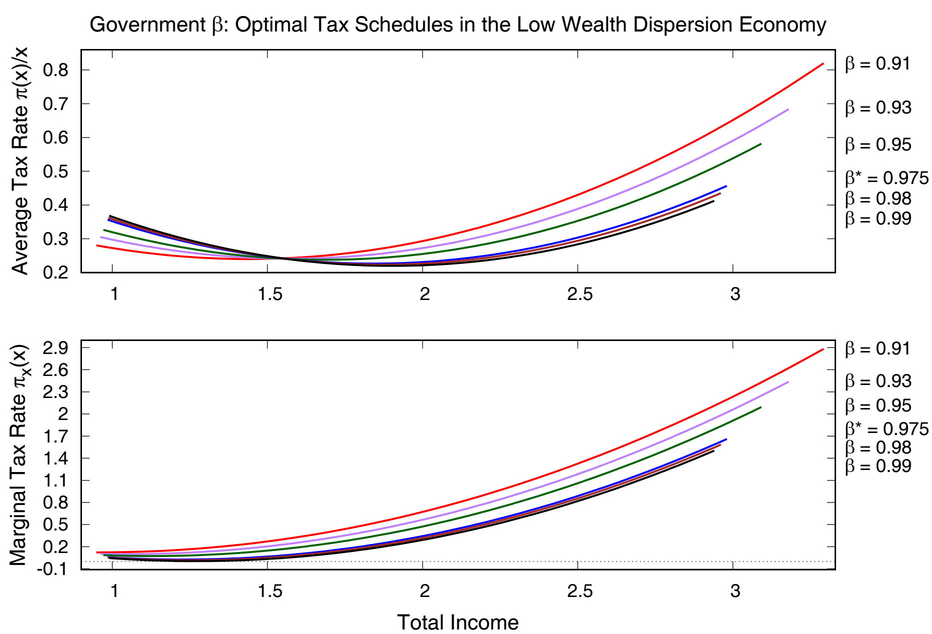


Figure E.4: Optimal average and marginal tax rates for different government discount factors in the low wealth dispersion economy.

Figure E.5 documents a very close relationship between the capital-output ratio and wealth inequality. Both increase as social discounting $\hat{\beta}$ rises relative to that of agents.

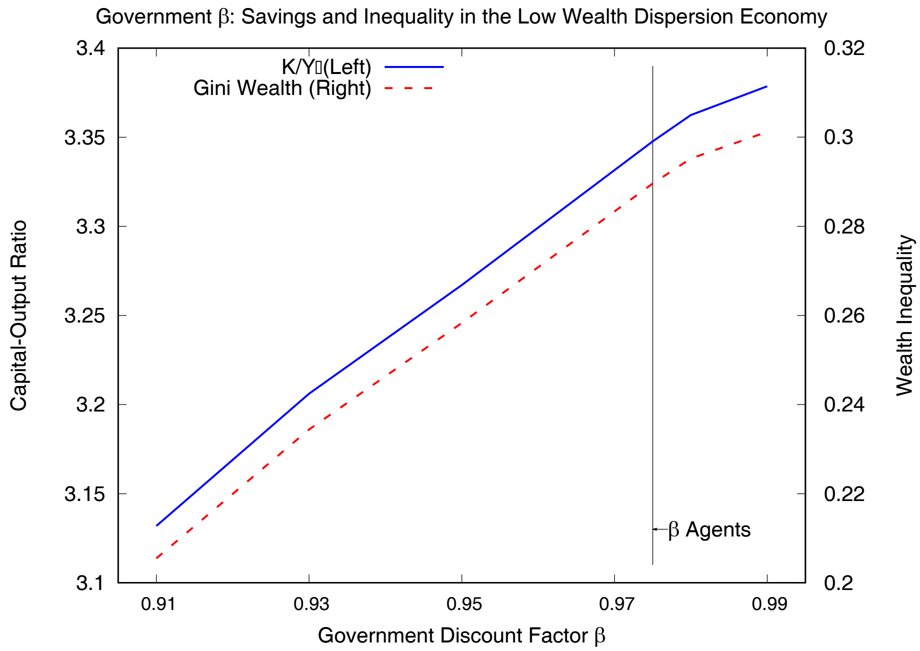


Figure E.5: Capital-output ratio (left axis) and wealth inequality (right axis) for different government discount factors in the low wealth dispersion economy.

For different levels of the government discount factor, Table E.6 reveals capital accumulation effects are stronger and distributional effects weaker due to extremely low wealth inequality. Without dissaving, there is no political support for the tax reform. Steady state average welfare levels are similar across all cases.

The distributional effect dominates at high income levels, where it is optimal to increase the marginal tax rates to raise as much tax revenue as possible. The average tax rate is convex in income, as the richest households with a low marginal utility provide most of the tax revenues. With high risk aversion or low social discount factor, the taste for redistribution also dominates at low income levels and the marginal tax schedule increases on the whole income distribution. When general equilibrium effects are more important for low risk aversion or high discount factor, the marginal tax rate schedule is declining at low income levels. First, poor agents save little relative to rich agents so the efficiency losses are small. Second, high marginal rates at the bottom of the distribution apply to a larger tax base and, therefore, their reduction lowers the average marginal rates.

Table E.6: Government Discount Factor: Low Wealth Dispersion Economy

| | Optimal Tax Schedule Steady State with $\hat{\beta}$ | | | | | |
|---|---|---|---|---|---|---|
| | 0.91 | 0.93 | 0.95 | 0.975* | 0.98 | 0.99 |
| Steady State | | | | | | |
| Aggregate assets | 3.209 | 3.328 | 3.427 | 3.561 | 3.585 | 3.612 |
| Output | 1.025 | 1.038 | 1.049 | 1.064 | 1.066 | 1.069 |
| Capital-output ratio | 3.132 | 3.206 | 3.267 | 3.348 | 3.362 | 3.378 |
| Interest rate (%) | 3.493 | 3.236 | 3.028 | 2.751 | 2.715 | 2.661 |
| Coeff. of variation of wealth | 0.357 | 0.407 | 0.445 | 0.498 | 0.509 | 0.519 |
| Gini wealth | 0.205 | 0.234 | 0.258 | 0.290 | 0.295 | 0.301 |
| Perc. of wealth of the top 5% | 0.084 | 0.090 | 0.095 | 0.103 | 0.105 | 0.107 |
| Steady State Average Welfare | 1.082 | 1.084 | 1.086 | 1.088 | 1.088 | 1.088 |
| Transition Gains from the Progressive Tax Steady State (in %) | | | | | | |
| Average Welfare | -0.610 | -0.958 | -1.264 | -1.718 | -1.806 | -1.905 |
| Aggregate Component | -0.350 | -0.670 | -0.950 | -1.330 | -1.410 | -1.490 |
| Distributional Component | -0.261 | -0.290 | -0.317 | -0.393 | -0.402 | -0.421 |
| Political Support | 23.792 | 19.496 | 18.788 | 15.494 | 10.991 | 6.133 |
| Transition Gains from the Flat Tax Steady State (in %) | | | | | | |
| Average Welfare | 0.078 | -0.260 | -0.576 | -1.040 | -1.133 | -1.235 |
| Aggregate Component | 0.090 | -0.030 | -0.320 | -0.730 | -0.810 | -0.890 |
| Distributional Component | -0.012 | -0.230 | -0.257 | -0.312 | -0.326 | -0.348 |
| Political Support | 49.667 | 34.741 | 26.839 | 18.397 | 9.022 | 7.042 |

Note: Agents discount factor $\beta = 0.975$ in all steady states.

## E.3 Sensitivity Analysis for Different Values of $\sigma$

Finally, we analyze the optimal limiting tax in the low wealth dispersion economy for four different values of $\sigma \in \{0.5, 1, 2, 4\}$, where $\sigma = 2$ is the benchmark value and $\theta = 1/\sigma$ is the intertemporal elasticity of substitution. Discount factor is adjusted for each $\sigma$ so that in the progressive tax steady state the capital-output ratio is 3 and the equilibrium interest rate is 4 percent. All other parameters are kept the same as in the benchmark economy.

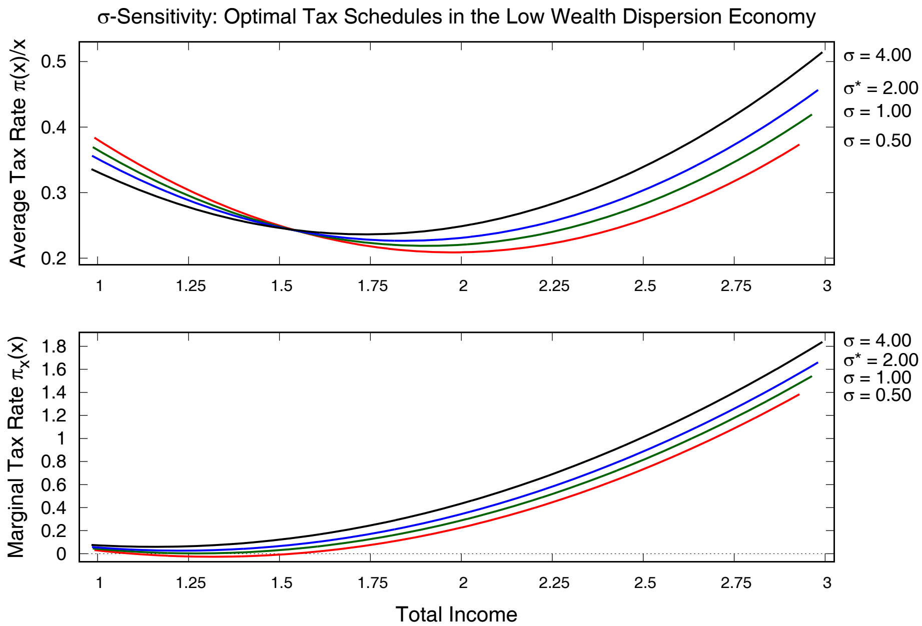


Figure E.6: Optimal average and marginal tax rates in the low wealth dispersion economy for different values of the risk aversion coefficient.

Figure E.6 displays the optimal average and marginal tax rates for different values of $\sigma$ in the low wealth dispersion economy. The U-shape of the optimal limiting average tax schedule is preserved in all simulations. 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 increasing for high levels of risk aversion.

Table E.7 shows the steady state allocations and transition gains of the optimal tax reform. In general, the Aiyagari's parameterization with low wealth dispersion is not very sensitive to risk aversion. The optimal tax schedule increases the aggregate capital stock more than the flat tax as the average price effect dominates the distributional effect. As a result, inequality slightly increases with risk aversion. Steady state average welfare levels are similar across all steady states. When transitions from the flat or the progressive steady states to the optimal steady states are taken into account, average welfare gains are negative for all values of risk aversion.

In general, as the government discounts future more than agents or if the latter are more risk averse, the short-term distributional effect is more important than the long-term, general equilibrium effect from changes in the aggregate capital stock.

Table E.7: Sensitivity Analysis $\sigma$: Steady State and Transition Results in the Low Wealth Dispersion Economy

| | $\sigma = 0.5$ ($\beta$=0.9753) | | | $\sigma = 1$ ($\beta$=0.9752) | | | $\sigma = 2$ ($\beta$=0.975) | | | $\sigma = 4$ ($\beta$=0.9743) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Average | Prog. | Flat | Opt. | Prog. | Flat | Opt. | Prog. | Flat | Opt. | Prog. | Flat | Opt. |
| $\tau$ | 0.161 | 0.253 | — | 0.161 | 0.253 | — | 0.161 | 0.253 | — | 0.161 | 0.254 | — |
| Aggregate assets | 3.000 | 3.261 | 3.642 | 3.000 | 3.265 | 3.594 | 3.000 | 3.273 | 3.561 | 3.000 | 3.289 | 3.512 |
| Output | 1.000 | 1.030 | 1.072 | 1.000 | 1.031 | 1.067 | 1.000 | 1.032 | 1.064 | 1.000 | 1.034 | 1.058 |
| Capital-output ratio | 3.000 | 3.164 | 3.396 | 3.000 | 3.167 | 3.368 | 3.000 | 3.172 | 3.348 | 3.000 | 3.182 | 3.319 |
| Interest rate (%) | 4.000 | 3.382 | 2.603 | 4.000 | 3.375 | 2.694 | 4.000 | 3.352 | 2.751 | 4.000 | 3.334 | 2.858 |
| Coeff. of variation of wealth | 0.369 | 0.694 | 0.381 | 0.452 | 0.687 | 0.476 | 0.516 | 0.673 | 0.498 | 0.556 | 0.646 | 0.493 |
| Gini wealth | 0.214 | 0.384 | 0.261 | 0.263 | 0.389 | 0.276 | 0.304 | 0.389 | 0.290 | 0.325 | 0.394 | 0.296 |
| Perc. of wealth of the top 5% | 0.089 | 0.154 | 0.095 | 0.100 | 0.152 | 0.099 | 0.113 | 0.148 | 0.103 | 0.119 | 0.142 | 0.104 |
| Steady State Average Welfare | 1.079 | 1.087 | 1.088 | 1.079 | 1.087 | 1.090 | 1.078 | 1.087 | 1.088 | 1.078 | 1.085 | 1.085 |
| Transition Gains (in %) | | | | | | | | | | | | |
| Average Welfare | -1.111 | -0.715 | — | -1.500 | -0.993 | — | -1.718 | -1.040 | — | -1.971 | -0.890 | — |
| Aggregate Component | -1.220 | -0.670 | — | -1.420 | -0.860 | — | -1.330 | -0.730 | — | -1.200 | -0.440 | — |
| Distributional Component | 0.110 | -0.045 | — | -0.081 | -0.134 | — | -0.393 | -0.312 | — | -0.780 | -0.452 | — |
| Political Support | 17.207 | 15.211 | — | 16.843 | 16.859 | — | 15.494 | 18.397 | — | 13.841 | 19.116 | — |

Note: The benchmark low wealth dispersion economy corresponds to $\sigma = 2$.

# F Extensions

Labor decisions are more relevant for the poor agents while savings decisions are more important for the wealthy agents. Park (2017) analyzes the Davila et al. (2012) economy with endogenous human capital accumulation. The first order condition with respect to human capital investment also has an extra term whose sign is opposite to that of savings. Because the planner can improve welfare only by changing equilibrium prices, qualitative results of Davila et al. (2012) do not change by introducing endogenous human capital. However, introducing human capital decreases inequality relative to the economy with exogenous labor.

In our paper, we assume that in each steady state the government must collect a fraction $g = 0.189$ of the total output. In Heathcote and Tsujiyama (2017) the shape of the optimal labor tax schedule is sensitive to the amount of fiscal pressure the government faces to raise revenue. When the fiscal pressure is low, the optimal marginal tax schedule is increasing as in the U.S. economy. As fiscal pressure grows, the marginal tax rates increase and the tax schedule becomes U-shaped, as in Saez (2001). This means that in economies with a higher share of government consumption the optimal limiting tax schedule might be more U-shaped than we found in this paper.

Another possible extension would be an optimal limiting tax schedule with life-cycle income profiles as in Golosov, Tsyvinski, and Troshkin (2011) and Farhi and Werning (2012). Heathcote, Storesletten, and Violante (2017) find that welfare gains from allowing the degree of tax progressivity to vary with age are very small. The optimal tax system is U-shaped in age due to the life-cycle productivity and insurance effects.

The exogenous lower bound on savings $\underline{k}$ has important implications on savings decisions of poor agents. In Heathcote (2005), the magnitude of the response of aggregate variables to tax changes depends on the fraction of wealth-poor and thus potentially borrowing-constrained agents. A combination of distortionary taxation and capital market imperfections leads to quantitatively important departures from Ricardian equivalence: after a tax cut, the average percentage increase in consumption is almost twice as large as the increase in investment. Intergenerational redistribution of the tax burden is the least important source of non-neutrality.

The variational approach used in our paper is well suited for analyzing time-varying optimal policies that take into account heterogeneous response of agents to aggregate shocks. Bhandari, Evans, Golosov, and Sargent (2017b) study how the Ramsey planner optimally sets nominal interest rates, transfers and proportional labor taxes in response to aggregate shocks in a New Keynesian model where agents are heterogeneous with respect to co-movements of aggregate variables and measures of inequality. During a recession, productivity shocks change the distribution of labor earnings or markup shocks. Optimal fiscal and monetary tools redistribute resources towards agents who are adversely affected. They find that this kind of redistributive hedging is quantitatively very important. Ahn, Kaplan, Moll, Winberry, and Wolf (2017) show how inequality matters for the dynamics of macroeconomic aggregates over the business cycle.

The shape of the optimal tax schedule is necessarily sensitive to the government's objective function. Heathcote and Tsujiyama (2017) characterize the mapping between the taste for redistribution in a class of Pareto weight functions and the progressivity parameter in Heathcote, Storesletten, and Violante (2016). When the taste for redistribution increases, the optimal marginal tax schedule moves from being upward sloping to becoming U-shaped. For the utilitarian welfare function with equal Pareto weights, the average welfare gain from an optimal progressive tax reform are very small. It would be straightforward to implement this parameterization in our model.

In Heathcote and Tsujiyama (2017), the planner maximizes a social welfare function

$$\int W(\alpha;\theta)U(\alpha)d\lambda(\alpha),$$

using Pareto weights $W(\alpha;\theta)$ that might vary with individual productivity $\alpha$. The parameter $\theta$ controls the extent to which the planner puts weight on low relative to high productivity agents. The weight function represents the taste for redistribution,

$$W(\alpha;\theta) = \frac{\exp(\theta\alpha)}{\int \exp(\theta\alpha)d\lambda(\alpha)}.$$

A positive weight $\theta$ puts more weight on the less productive agents, and vice versa for a negative $\theta$. The case of $\theta = 0$ corresponds to our utilitarian welfare function. The Rawlsian case has $\theta \to \infty$ as the planner maximizes the minimum level of welfare in the economy.