Pareto-Improving Optimal Capital and Labor Taxes*

Katharina Greulich†  Sarolta Laczó‡  Albert Marcet§

July 2015

Abstract

We study Pareto-optimal fiscal policy in a model with agents who are heterogeneous in their labor productivity and wealth. In our model long-run capital taxes are zero. We focus on Pareto-improving policies and we find that a gradual reform is crucial in achieving a Pareto improvement: labor taxes should be cut and capital taxes should remain high for a very long time before reaching zero. Therefore, the long-run optimal tax mix is the opposite of the short- and medium-run one. This policy redistributes wealth in favor of workers so that all agents benefit, and it favors quick capital growth after the reform. The labor tax cut is financed by deficits which lead to a positive level of government debt in the long run, reversing the standard prediction that the government accumulates savings in models with optimal capital taxes. The welfare benefits from the tax reform are relatively large and they can be shifted entirely to capitalists or workers by varying the length of the transition. We address a number of technical issues such as sufficiency of Lagrangian solutions in a Ramsey problem, relation of Pareto-improving allocations with welfare functions, asymptotic behavior, and solution algorithms.

JEL classification: E62, H21

Keywords: fiscal policy, Pareto-improving tax reform, redistribution

*We wish to thank Jess Benhabib, Jordi Caballé, Begoña Dominguez, Joan M. Esteban, Michael Golosov, Andreu Mas-Colell, Michael Reiter, Sevi Rodríguez, Raffaele Rossi, Kjetil Storesletten, Jaume Ventura, Iván Werning, Philippe Weil, Fabrizio Zilibotti, and seminar participants in various institutions for useful comments and suggestions. Michael Reiter provided the implementation of Broyden’s algorithm used in this paper. Greulich acknowledges support from the National Centre of Competence in Research “Financial Valuation and Risk Management” (NCCR FINRISK) and the Research Priority Program on Finance and Financial Markets of the University of Zurich. Laczó acknowledges funding from the JAE-Doc grant co-financed by the European Social Fund. Marcet acknowledges funding from AGAUR, Plan Nacional (Spanish Ministry of Science), the Axá Foundation, the Excellence Program of Banco de España, and the European Research Council under the EU 7th Framework Programme (FP/2007-2013), Grant Agreement n. 324048 - APMPAL. The views expressed in this paper are not those of the Bank of England or Swiss Re.

†Swiss Re and Institut d’Anàlisi Econòmica (IAE-CSIC).
‡Swiss Re and Institut d’Anàlisi Econòmica (IAE-CSIC), Threadneedle Street, EC2R 8NH, London, United Kingdom. Email: sarolta.laczo@gmail.com.
§Institut d’Anàlisi Econòmica (IAE-CSIC), ICREA, UAB, MOVE, Barcelona GSE & CEPR, Campus UAB, 08193 Bellaterra, Barcelona, Spain. Email: albert.marcet@iae.csic.es.
1 Introduction

A large literature on dynamic taxation concludes that long-run capital taxes should be zero. This result, which originally goes back to Chamley (1986) and Judd (1985), is resilient to many modifications of the basic model. We denote this result as $\tau_k^\infty = 0$.\(^1\)

This is a controversial policy recommendation as it implies increasing labor taxes. It seems that this will hurt less wealthy taxpayers and, given the highly skewed distribution of wealth, a large part of the population. But it has been shown that the optimal policy with heterogeneous agents also involves $\tau_k^\infty = 0$, even if the government only considers allocations that improve the welfare of consumers with very little wealth (see, for example, Judd, 1985, and Atkeson, Chari, and Kehoe, 1999).\(^2\) This suggests that there is no equity-efficiency trade-off: everybody gains from lowering capital taxes. Our aim is to study carefully Ramsey-Pareto-optimal (RPO) allocations when all agents gain relative to a status quo.

We consider a model in which agents are heterogenous in their labor productivity and wealth. The government can only levy proportional labor and capital taxes, lump-sum transfers are not available. To insure smooth transitions we also introduce an upper bound on capital taxes below 100 percent. We focus on Pareto-improving tax reforms and study the entire path of optimal taxes.\(^3\) We find that there is in fact an equity-efficiency trade-off. Pareto-improving allocations are achieved only if labor taxes are low and capital taxes are high for a very long time. In other words, if all agents should benefit from lowering capital taxes, the limit $\tau_k^\infty = 0$ has to be reached very slowly.

More generally, our paper speaks to the issue of how to implement economic reforms. Economists often promote reforms which improve aggregate efficiency, but these reforms may come at the cost of a welfare decrease for many agents. This may or may not be considered ‘unfair,’ but it certainly acts as an obstacle for the actual implementation of such reforms. Considering Pareto improvements addresses the potential fairness issue, and it facilitates the implementation of such reforms, given that they become more widely accepted. This is in line

---

\(^1\)A very incomplete summary of a large literature is that in the few cases where $\tau_k^\infty \neq 0$, capital taxes are often small or even negative. Recently (Straub and Werning, 2014) have shown that $\tau_k^\infty = 0$ does not obtain nearly as generally as had been thought. In fact, $\tau_k^\infty > 0$ in many previously considered models, while capital and consumption go to zero. We consider a simple modification and show that $\tau_k^\infty = 0$ in our model. This allows us to focus on issues of optimal redistribution during the transition.

\(^2\)Aiyagari (1995) shows that, due to capital overaccumulation, $\tau_k^\infty > 0$ with heterogeneous agents and incomplete markets. We do not focus on this implication of heterogeneous agents for two reasons. First, the result is tenuous: one obtains $\tau_k^\infty < 0$ depending on the form of income shocks (see Chamley, 2001) or with endogenous labor supply (see Marcet, Obiols-Homs, and Weil, 2007). Second, the Aiyagari result holds under ‘the veil of ignorance’ welfare function, and it may not hold for other welfare functions.

\(^3\)An early paper studying the transition of optimal taxes is Jones, Manuelli, and Rossi (1993).
with the literature on gradualism of political reforms, which has been at the center of some policy debates. In our case, the slow transition to $\tau^k_\infty = 0$ means that a gradual reform is needed for all consumers’ welfare to improve. Therefore, high capital taxes that are observed are not necessarily a failure of a political system or a result of frequent voting, as has been suggested. They could be a sign of perfectly functioning institutions.

We first show analytically that $\tau^k_\infty = 0$ obtains in our model, but capital taxes may be high and at the upper bound for many periods. We need a careful proof of this result, because the analysis of Straub and Werning (2014) puts into question the validity of many available $\tau^k_\infty = 0$ results. We show analytically that introducing a constraint that consumption is bounded away from zero gives conditions to ensure that $\tau^k_\infty = 0$ in our model. To find the length of the transition period and the effects of the tax reform on allocations and welfare, we turn to numerical methods.

To demonstrate the effects of heterogeneity in isolation we first study a model with completely inelastic labor supply. This is the starkest case to demonstrate the efficiency-equity trade-off: in a homogeneous-agent world capital taxes should be zero in all periods. But with heterogeneous agents capital taxes equal to zero in all periods is not Pareto improving. Instead a long period of high capital taxes (between 13 and 26 years for our calibration) is needed in order to raise more tax revenues from the capitalists and less from the workers so as to ensure that all consumers gain from the tax reform. Therefore, even though the planner has access to non-distortive labor taxes, she has to resort to distortive capital taxation to achieve a Pareto improvement. The redistribution comes at a cost, as there are significant welfare losses relative to the case where lump-sum transfers are available.

If labor supply is elastic we also find that redistributive concerns cause the transition to be very long: capital taxes should be high for 11 to 26 years before they are set to zero. In addition, labor taxes should be lower than at the status quo during the transition, thus promoting growth in the early periods. Therefore, optimal Pareto-improving factor taxation depends very much on heterogeneity.

Our results are complementary to some papers which establish that a large part of the population would suffer a large utility loss if capital taxes were suddenly abolished, see Correia (1999), Correia (2010), Domeij and Heathcote (2004), Conesa and Krueger (2006), Flodén (2009) and Garcia-Milà, Marcet, and Ventura (2010). In contrast, Lucas (1990) showed that the welfare of a representative agent would increase for this same tax reform. Our results give
the following interpretation to this literature: it is crucial to follow the optimal transition in order to achieve a Pareto improvement under heterogeneity, while in the case where agents are homogeneous the optimal transition may be less important.

In our main model government debt is positive in the long run, while the government often accumulates savings under homogeneous agents. This is because the government initially runs a deficit to finance the initial drop in labor taxes. This shows how a positive level of government debt can be a by-product of an optimal reform.

The results are robust to various parameter changes and to the introduction of progressive taxation. We also investigate numerically issues of time consistency. We find that if the reform can only be overturned by Pareto improving new policies, the tax reform is time consistent. Therefore, the requirement of consensus to change previous policies builds in time consistency. Armenter (2004) finds a related result analytically in a simpler model.

To our knowledge, Ramsey-Pareto-optimal (RPO) allocations in models of factor taxation have not been solved for before. We had to resolve a number of technical issues. Our approach is to summarize equilibrium conditions in such a way that the decisions of all agents but one are summarized by the ratio of marginal utilities, denoted $\lambda$. Then the computational cost of this model is, essentially, the same as of a homogeneous-agents model. Another issue is that agents’ relative Pareto weight in the welfare function does not coincide with the ratio of marginal utilities $\lambda$, instead this ratio has to be chosen optimally. A further difficulty is that the set of competitive equilibria is potentially not convex. That this can be a problem in models of Ramsey taxation is well known but often ignored. It is an important issue in our paper, because we want to trace out a large part of the Ramsey Pareto frontier, so if the duality gap is non-empty we are likely to miss some RPO allocations using a Lagrangian. To deal with this problem we show a sufficient condition, which can be checked numerically, insuring that the duality gap is empty. We also show how to find allocations that are not Pareto optimal but are on the frontier of competitive equilibria, by analyzing a planner’s problem where some welfare weights are negative. In one case we study the set of possible equilibria is non-standard in that its frontier has an increasing part.

Most papers on optimal policy with heterogeneous agents focus on the case where all agents receive the same Pareto weight. Many papers interpret Pareto weights as representing agents’ political power. We treat Pareto weights as another multiplier to be solved for.

---

$^5$ Flodén (2009) states that, for a model similar to ours, “[w]hen more than one optimized household is considered [...] solving the problem is computationally challenging” (page 288).

$^6$ An exception is Bassetto (2014) who states in section 3.1 that “we show how heterogeneity may lead to situations in which the first-order conditions are not sufficient even in the simplest case.”
the requirement of Pareto-improvement. We find that equal weights are not necessarily related to Pareto improvements, thus the focus on equal weighting is not to be interpreted as giving rise to ‘equitable’ reforms, see section 4.3.2.

Recent literature on optimal policy in models with wealth heterogeneity includes Niepelt (2004) and Bassetto (2014), who study how taxes affect taxpayers of different wealth in stochastic models without capital. Bassetto and Benhabib (2006) establish a median voter theorem in a model with production, Gorman aggregation, and wealth heterogeneity, and characterize some properties of the taxes chosen by the median voter. Bhandari, Evans, Golosov, and Sargent (2013) study transfers along the business cycle under incomplete markets. Werning (2007) studies redistribution with progressive taxation. Conesa, Kitao, and Krueger (2009) study a large overlapping-generations (OLG) model, and they also find a role for capital taxes. Flodén (2009) considers a model with many labor productivity-wealth types and capital/labor taxation, like the present paper, but with Gorman-aggregable preferences. He shows how to analyze many different feasible policies by studying policies that cater to a certain agent who has measure zero, but his approach does not characterize all RPO allocations.\footnote{See Appendix D for a discussion of how this claim relates to the work of Flodén (2009).}

The rest of the paper is organized as follows. In Section 2 we lay out our baseline model and discuss further the motivation for our assumptions. Section 3 presents some properties of the model obtained analytically, including a proof that capital taxes are zero at the steady state, about the form of the transition, and a sufficient condition for the Lagrangian to deliver all RPO solutions. Our numerical results are in Section 4. Section 5 concludes. Appendices contain some algebraic details, a description of our computational approach, a sensitivity analysis, as well as details on the relation of our solution method to other approaches in the literature.

2 The model

We consider an economy with two heterogeneous consumers, discrete time, capital accumulation, endogenous labor supply and no uncertainty. Our emphasis differs from the bulk of the Ramsey factor taxation literature in four aspects:

i) We study the whole path of taxes.

ii) We preclude agent-specific redistributive lump-sum transfers. This assumption seems reasonable given the focus on proportional taxation in this literature. Furthermore,
most tax codes and even constitutions stipulate that all individuals are equal in front of the law, preventing individual-specific lump sum transfers.

iii) We search for Ramsey Pareto-optimal allocations and we focus on Pareto improving allocations relative to a status quo.

iv) We oppose that consumption be bounded away from zero.

v) We impose an upper bound on the capital tax rate in each period.

Chamley (1986) and Atkeson, Chari, and Kehoe (1999) assume an upper bound of 100 percent for capital taxes in all periods. Many other papers in the optimal taxation literature assume a bound only in the initial period. Optimal policies under these constraints imply that capital taxes should be very high in the first few periods, much higher than current actual capital taxes which, by all measures, are already high. The initial tax hike recommended by these models could have devastating effects on investment in the real world if there is partial credibility of government policy, or if agents form their expectations by learning from past experience. To avoid this initial tax hike we use a capital tax ceiling lower than 100 percent. In particular, in all of our computational exercises we fix this ceiling to the status-quo capital tax rate. Alternatively, this bound can be interpreted as a value that avoids massive capital flight in an open economy with partial mobility of capital.

Next we present our model formally. We refer to Garcia-Milà, Marcet, and Ventura (2010) for some details on how to characterize competitive equilibria. For details on formulating Ramsey equilibria and the primal approach in general, see Chari and Kehoe (1999) or Ljungqvist and Sargent (2012).

2.1 The environment

There are two types of consumers $j = 1, 2$ with utility $\sum_{t=0}^{\infty} \beta^t [u(c_{j,t}) + v(l_{j,t})]$ each, where $c_{j,t}$ is consumption and $l_{j,t}$ is labor of consumer $j$ in period $t$. This is for simplicity, our analysis is easily extended to many consumers and non-separable utility. We assume $u_c > 0$, $v_l < 0$, and the usual Inada and concavity conditions. Agents differ in their initial wealth $k_{j,-1}$ and their labor productivity $\phi_j$. Agent $j$ obtains income in period $t$ from renting his/her capital at the rental price $r_t$ and from selling his/her labor for a wage $w_t \phi_j$. Agents pay taxes

---

8Lucas (1990) offered a similar reasoning to motivate his study of a tax reform that abolishes capital taxes immediately. Of course, one could introduce credibility and learning concerns explicitly. The time-consistency literature deals, in a way, with issues of credibility. An analysis of capital taxes under learning can be found in Giannitsarou (2006). In this paper we keep the more common assumptions of rational expectations and full commitment.
at rate $\tau_t^l$ on labor income and $\tau_t^k$ on capital income net of depreciation allowances. Therefore, the period-$t$ budget constraint of consumer $j$ is given by

$$c_{j,t} + k_{j,t} = w_t \phi_j l_{j,t} (1 - \tau_t^l) + k_{j,t-1} \left[1 + (r_t - \delta)(1 - \tau_t^k)\right], \text{ for } j = 1, 2. \tag{1}$$

Consumers and firms take the whole sequence of prices and tax rates as given.

Firms maximize profits and have a production function $F(k_{t-1}, e_t)$, where $k$ is total capital and $e$ is total efficiency units of labor. $F$ is concave and increasing in both arguments, differentiable, has constant returns to scale, and $F_k (k, e) \to 0$ as $k \to \infty$, where a subindex denotes the partial derivative with respect to the corresponding variable.

The government chooses capital and labor taxes, has to spend $g$ in every period, saves in capital, and has initial capital $k^g_{t-1}$. The government can hold debt. In that case $k^g_t < 0$. Ponzi schemes for consumers and the government are ruled out.

We normalize the mass of each group to $\frac{1}{2}$. Market clearing conditions for all $t$ are

$$\frac{1}{2} \sum_{j=1}^{2} \phi_j l_{j,t} = e_t, \tag{2}$$

$$k_t = k^g_t + \frac{1}{2} \sum_{j=1}^{2} k_{j,t},$$

$$\frac{1}{2} \sum_{j=1}^{2} c_{j,t} + g + k_t - (1 - \delta) k_{t-1} = F(k_{t-1}, e_t). \tag{3}$$

### 2.2 Conditions of competitive equilibria

Our competitive-equilibrium (CE) concept is standard: consumers and firms take prices and taxes as given and maximize their utility and profits, respectively. Markets clear and the budget constraint of the government is satisfied.

Consumers’ first-order conditions (FOCs) with respect to consumption and labor yield

$$u'(c_{j,t}) = \beta u'(c_{j,t+1}) \left(1 + (r_{t+1} - \delta) (1 - \tau_{t+1}^k)\right), \forall t, \tag{4}$$

$$-\frac{v'(l_{j,t})}{u'(c_{j,t})} = w_t (1 - \tau_t^l) \phi_j, \forall t, \tag{5}$$

i.e., the Euler equation and the consumption-labor optimality condition, respectively, for all $j$. For most of the paper we assume the current utility function

$$u(c) = \frac{c^{1-\sigma_c}}{1-\sigma_c} \text{ and } v(l) = -\omega \frac{l^{1+\sigma_l}}{1+\sigma_l}, \tag{6}$$
where $\omega$ is the relative utility weight of hours, $\sigma_c > 0$ is the coefficient of relative risk aversion, and $\sigma_l > 0$ is the inverse of the (constant) Frisch elasticity of labor supply. In this case the above FOCs imply
\[
\frac{c_{2,t}}{c_{1,t}} = \lambda \quad \text{and} \quad \frac{l_{2,t}}{l_{1,t}} = \lambda^{-\frac{\sigma_c}{\sigma_l}} \left( \frac{\phi_2}{\phi_1} \right)^{\frac{1}{\sigma_l}}, \quad \forall t,
\] (7)
for some $\lambda$ constant through time.\(^9\)

Firms behave in a competitive fashion, hence factor prices equal marginal products, i.e.,
\[
r_t = F_k (k_{t-1}, e_t) \quad \text{and} \quad w_t = F_e (k_{t-1}, e_t).
\]
Using these conditions we can eliminate factor prices from the characterization of competitive equilibria.

Using equation (4) the budget constraints of consumer $j$ for all $t = 0, 1, \ldots$ can be summarized in the present-value budget constraint
\[
\sum_{t=0}^{\infty} \beta^t \left( u'(c_{j,t}) c_{j,t} - w_t \phi_j l_{j,t} (1 - \tau_t) \right) = k_{j,-1} \left( 1 + (r_0 - \delta) (1 - \tau_0^k) \right), \quad \text{for } j = 1, 2. \quad (8)
\]
Then, using (5) and rearranging, for consumer 1 we have
\[
\sum_{t=0}^{\infty} \beta^t \left( u'(c_{1,t}) c_{1,t} + v'(l_{1,t}) l_{1,t} \right) = u'(c_{1,0}) k_{1,-1} \left( 1 + (r_0 - \delta) (1 - \tau_0^k) \right). \quad (9)
\]
Using (4), (5), and (7), we can write the present-value budget constraint of consumer 2 as
\[
\sum_{t=0}^{\infty} \beta^t \left( u'(c_{1,t}) \lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'(l_{1,t}) f(\lambda, l_{1,t}) \right) = u'(c_{1,0}) k_{2,-1} \left( 1 + (r_0 - \delta) (1 - \tau_0^k) \right), \quad (10)
\]
where
\[
f(\lambda, l_{1,t}) \equiv \lambda^{-\frac{\sigma_c}{\sigma_l}} \left( \frac{\phi_2}{\phi_1} \right)^{\frac{1}{\sigma_l}} l_{1,t}, \quad (11)
\]
which gives $l_{2,t}$ for each possible value of the endogenous variables $\lambda$ and $l_{1,t}$ from (7).\(^{10}\)

Formally, let the set of CE allocations be
\[
\mathcal{S}^{CE} \equiv \{ \text{sequences } \{(c_{j,t}, l_{j,t})_{j=1,2}, k_t\}_{t=0}^{\infty} \text{ which are a CE} \}
\]
for given initial conditions on capital. Elements of $\mathcal{S}^{CE}$ are characterized by (3), (7), (9), and (10). It is clear that one can substitute out the sequences $\{c_{1,t}^2, l_{1,t}^2\}_{t=0}^{\infty}$ for a given value of $\lambda$.

---

\(^9\)Note that labor supply depends not only on labor productivity but also on the distribution of consumption/wealth through $\lambda$. Under Gorman aggregation this would not be the case.

\(^{10}\)Walras’ law guarantees that the budget constraint of the government is implied by the above equations plus feasibility, hence it can be ignored.
Therefore, the necessary and sufficient conditions for sequences \( \{c^1_t, l^1_t, k_t\}_{t=0}^{\infty} \) and a constant \( \lambda \) to be a CE allocation are feasibility (3) and the implementability constraints (9) and (10). Note that the ratio \( \lambda \) is so far unknown. It has to be determined in equilibrium, and it has to be consistent with all equilibrium conditions.

Given a set of CE allocations, taxes are found from (4) and (5), and individual capital is backed out from the budget constraint period by period.

### 2.3 The policy problem

A Ramsey Pareto-optimal (RPO) allocation is an element of \( S^{CE} \) such that the utility of one or more agents cannot be improved within the set \( S^{CE} \). We now show how to formulate this equilibrium as the solution to a planner’s problem.

A standard argument shows that RPO allocations can be found by solving a problem where a planner maximizes the utility of, say, consumer 1, subject to the constraint that the utility of consumer 2 has a minimum value of \( U_2 \), i.e.,

\[
\sum_{t=0}^{\infty} \beta^t [u(c_{2,t}) + v(l_{2,t})] \geq U_2, \tag{12}
\]

where \( U_2 \) is restricted to belong to the set of utilities that can be attained for agent 2 within \( S^{CE} \). Varying the value of the minimum utility \( U_2 \) along all possible utilities for consumer 2 in \( S^{CE} \), we can trace out the whole set of RPO allocations.

We assume the planner faces a tax limit, denoted \( \tilde{\tau} \), and impose \( \tau^k_t \leq \tilde{\tau} \) for all \( t = 0, 1, \ldots \), for a constant \( \tilde{\tau} \) exogenously given. Combining this limit with the Euler equation of consumer 1, it is easy to see that the tax limit is satisfied in equilibrium if and only if

\[
u'(c_{1,t}) \geq \beta u'(c_{1,t+1}) (1 + (r_{t+1} - \delta) (1 - \tilde{\tau})) \quad \forall t > 0, \tag{13}\]

\[
\tau^k_0 \leq \tilde{\tau}. \tag{14}\]

The first equation ensures that the actual capital tax \( \tau^k_t, t = 1, 2, \ldots \), implied by (4) satisfies the limit, and it allows us to use the primal approach, where taxes at \( t = 1, 2, \ldots \) do not appear explicitly in the government’s problem.

Finally, we introduce a constraint with respect to admissible consumptions on the planner’s problem. Straub and Werning (2014) show that in some models the optimal policy entails \( c^*_j,t \to 0 \) as \( t \to \infty \) and \( \tau^k_\infty \neq 0 \). Since we want to study the transition for Pareto-improving policies when capital taxes go to zero, we abstract from this type of solutions. For this reason we assume that the government finds it inadmissible to immiserize all consumers.
in the future and we add the constraint \( c_{j,t} \geq \tilde{c}_j, \forall j, \forall t \), for some \( \tilde{c}_j > 0 \). Given (7) these constraints can be substituted, without loss of generality, by

\[
c_{1,t} \geq \tilde{c}, \forall t.
\]

As is standard in the Ramsey taxation literature, we assume that the government has full credibility, i.e., it fully commits to the announced policies, and both the government and the agents have rational expectations.

Collecting all the above, all RPO allocations can be found by solving

\[
\begin{align*}
\max_{\tau, \lambda, \{c_{1,t}, k_{1,t}, l_{1,t}\}} & \sum_{t=0}^{\infty} \beta^t [u(c_{1,t}) + v(l_{1,t})] \\
\text{s.t.} & \sum_{t=0}^{\infty} \beta^t [u(\lambda c_{1,t}) + v(f(\lambda, l_{1,t}))] \geq U_2,
\end{align*}
\]

and subject to feasibility (3), implementability (9)-(10), tax limits (13) and (14), and consumption limits (15) for a given level of utility \( U_2 \). We have used (7) and (11) to substitute for \( c_2 \) and \( l_2 \) to obtain (16). \( U_2 \) has to satisfy the requirements discussed above to guarantee that the feasible set is not empty.

Notice that a special feature of this problem is that the constant \( \lambda \) appears as an argument in the optimization problem. Formulating the maximization problem as a function of \( \lambda \) simplifies finding a RPO allocation in a heterogeneous-agents model relative to previous approaches that kept the series of all agents as arguments of the maximization problem.\(^{11}\) Using our approach the number of variables to solve for is essentially the same as in a homogeneous-agents model, see Appendix B for details.

We concentrate our attention on those RPO allocations which are also Pareto-improving with respect to a benchmark feasible allocation, the status quo. We call these ‘POPI’ allocations. Let the utilities attained by agent \( j \) at the status quo be \( U_{1j}^{SQ} \).\(^{12}\) POPI allocations can be found by considering only minimum utility values \( U_2 \) such that \( U_2 \geq U_2^{SQ} \) and by making sure that the planner’s objective at the maximum satisfies

\[
\sum_{t=0}^{\infty} \beta^t [u(c_{1,t}^*) + v(l_{1,t}^*)] \geq U_1^{SQ},
\]

where * denotes the optimized value of each variable for a given \( U_2 \).

\(^{11}\)This was the case for example in Atkeson, Chari, and Kehoe (1999) and Bassetto (2014). We give more details about their approach in Appendix D.

\(^{12}\)The status-quo utilities depend on \( k_{1,-1} \) and \( k_{2,-1} \) in general. We leave this dependence implicit.
2.4 Optimality conditions

To solve the above planner’s problem, we find the first-order conditions of an appropriate Lagrangian. Let $\psi$ be the Lagrange multiplier of the minimum utility constraint (16), let $\Delta_1$ and $\Delta_2$ be the multipliers of implementability constraints (9) and (10), respectively, and $\mu_t$, $\gamma_t$, and $\xi_t$ be the multipliers of the feasibility constraint (3), the tax limit (13), and the consumption limit (15), respectively, at time $t$. The Lagrangian for the government’s problem is

$$
\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \left\{ u(c_{1,t}) + v(l_{1,t}) + \psi [u(\lambda c_{1,t}) + v(f(\lambda, l_{1,t}))]ight.
$$

$$
+ \xi_t(c_{1,t} - \bar{c}) + \Delta_1 [u'(c_{1,t}) c_{1,t} + v'(l_{1,t}) l_{1,t}]
$$

$$
+ \Delta_2 \left[ u'(c_{1,t}) \lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'(l_{1,t}) f(\lambda, l_{1,t}) \right]
$$

$$
+ \gamma_t [u'(c_{1,t}) - \beta u'(c_{1,t+1}) (1 + (r_{t+1} - \delta)(1 - \bar{\tau}))]
$$

$$
+ \mu_t \left[ F(k_{t-1}, e_t) + (1 - \delta)k_{t-1} - k_t - \frac{1 + \lambda}{2} c_{1,t} - g \right] \right\} - \psi U_2 - W,
$$

where $W = u'(c_{1,0}) (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}) (1 + (r_0 - \delta)(1 - \tau_{0}^k))$. Further, $\xi_t$, $\gamma_t$, $\mu_t \geq 0$, $\forall t$, and $\psi \geq 0$, with complementary slackness conditions. We know that the resource constraint binds in each period, hence, $\mu_t > 0$, $\forall t$.

The first line of this Lagrangian has the usual interpretation: finding a Pareto-efficient allocation amounts to maximizing a welfare function where the planner weights linearly the utility of the two consumers. The weight of consumer 1 is normalized to one and the weight of consumer 2 is the Lagrange multiplier of the minimum utility constraint. This weight $\psi$ is not chosen arbitrarily in our setup, it has to be such that the minimum utility constraint is satisfied. The next two lines in (18) correspond to the minimum consumption and the equilibrium deficits of consumers. The fourth line ensures that $\tau_t^k \leq \bar{\tau}$ for all $t > 0$. The last line is the feasibility constraint. The term $W$ collects the period-0 terms in the budget constraints of the consumers.

As is often the case in optimal-taxation models, the feasible set of sequences for the planner is non-convex. This means that we need to be careful about necessity and sufficiency of FOCs. We address these issues in detail in Section 3.2.

The tax limit is a forward-looking constraint, therefore standard dynamic programming does not apply. Using a promised-utility approach would be complicated, because of the appearance of a vector of state variables (marginal utilities of consumption for all agents) that has to be bounded to stay in the set of feasible marginal utilities, and, since there is also
a natural state variable \((k)\), characterizing this set would be quite difficult. The Lagrangian approach of Marcet and Marimon (2011) is easier to use under these circumstances. Appendix A shows the recursive Lagrangian and the FOCs with respect to consumption, labor, and capital. In the rest of this section we comment on features of the remaining FOCs which differ from other papers on dynamic taxation.

The multipliers have to satisfy complementary slackness conditions. For \(\psi\), the multiplier of (16), we have that

\[
either \psi > 0 \text{ and } \sum_{t=0}^{\infty} \beta^t [u(c_{2,t}) + v(l_{2,t})] = U_2, \\
or \psi = 0 \text{ and } \sum_{t=0}^{\infty} \beta^t [u(c_{2,t}) + v(l_{2,t})] \geq U_2.
\]

In other words, the minimum utility constraint may or may not be binding. In the first case, the Lagrangian amounts to maximizing the weighted sum of utilities of consumers 1 and 2 with weights 1 and \(\psi\), respectively. If the minimum utility constraint is not binding, the planner gives zero weight to consumer 2. The latter case would only occur in models without frictions if the planner would be willing to give a very low utility to consumer 2. We show in section 4.2 a case where even if the lower bound \(U_2\) is the status-quo utility, \(\psi = 0\). This is because even if \(\psi = 0\), consumer 2 has to consume due to the fact that the allocations are determined in equilibrium, which implies that his budget constraint has to be satisfied, assuring him some revenue for any tax policies.

In standard models with heterogeneous agents, the ratio of marginal utilities equals the relative Pareto weight, i.e., \(\lambda^{-\frac{1}{\sigma_c}} = \psi\). Key to our approach is the fact that this equality does not hold and that the relative consumption of consumers (\(\lambda\)) has to be chosen optimally. Hence we set the derivative of \(L\) with respect to \(\lambda\) equal to zero to obtain the optimality condition

\[
\sum_{t=0}^{\infty} \beta^t \left\{ \psi [u'(\lambda c_{1,t}) c_{1,t} + u' (f (\lambda, l_{1,t})) f_\lambda (\lambda, l_{1,t})] \\
+ \Delta_2 \left( u' (c_{1,t}) c_{1,t} + \frac{\phi_2}{\phi_1} u'(l_{1,t}) f_\lambda (\lambda, l_{1,t}) \right) \\
- \gamma_{l-1} u' (c_{1,t}) F_{ke} (k_{t-1}, e_t) \frac{\phi_2}{2} f_\lambda (\lambda, l_{1,t}) (1 - \tilde{r}) - \frac{\mu_t}{2} (c_{1,t} - F_e (k_{t-1}, e_t) \phi_2 f_\lambda (\lambda, l_{1,t})) \right\} \\
- u'(c_{1,0}) (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}) F_{ke} (k_{-1}, e_0) \frac{\phi_2}{2} f_\lambda (\lambda, l_{1,0}) (1 - \tau^k) = 0.
\]

The fact that \(\lambda\) is a choice for the government reflects the fact that the government can vary consumers’ utility by varying the total tax burden of labor and capital in discounted present
value.

For $\gamma_t$ for each $t$ we have that

$$
either \quad \gamma_t > 0 \quad \text{and} \quad u'(c_{1,t}) = \beta u'(c_{1,t+1}) (1 + (r_{t+1} - \delta) (1 - \tilde{\tau})) ,$$

or $\gamma_t = 0$ and $u'(c_{1,t}) \geq \beta u'(c_{1,t+1}) (1 + (r_{t+1} - \delta) (1 - \tilde{\tau})).$

Below we use these conditions to characterize the path of capital taxes.

It turns out that the $\Delta_j$’s may be positive or negative, since the corresponding present-value budget constraints have to be satisfied as equality. This becomes clear by looking at the following interpretation. With two consumers the marginal utility cost of distortive taxation is

$$\frac{\partial c}{\partial \tau_0} = u'(c_{1,0}) (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}) (r_0 - \delta).$$

Hence,

$$\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1} \geq 0,$$

with strict inequality as long as any taxes are raised after the initial period. This allows for one of the $\Delta_j$’s to be negative, which will indeed be the case whenever the constraints on redistribution imposed by the CE conditions and the Pareto-improvement requirement are sufficiently severe. To see this, consider a slightly modified model in which the social planner is allowed to redistribute initial wealth between consumers by means of lump-sum transfers $T_j, \ j = 1, 2$, such that $T_1 = -T_2$. All this modification does to the Lagrangian is that it changes the implementability constraints, in particular, the term $u'(c_{1,0}) (\Delta_1 - \Delta_2) T_1$ is added to $\mathbf{W}$. Now the derivative of the Lagrangian with respect to the lump-sum transfer between consumers is $\frac{\partial c}{\partial T_1} = u'(c_{1,0}) (\Delta_1 - \Delta_2)$. For any given $T_1$, and in particular for $T_1 = 0$ as in our baseline model, this expression is a measure of the marginal utility cost of the transfer not being optimal. If the planner were free to choose $T_1$ optimally, we would have $\Delta_1 = \Delta_2 > 0$. If the planner would like to redistribute more towards consumer 2, then $\Delta_1 - \Delta_2 > 0$, and vice versa. If the transfer is much too low (high), the derivative will be large in absolute value and $\Delta_2 (\Delta_1)$ will be negative. In sum, while the weighted sum of the multipliers on the present-value budget constraints is related to the cost of distortive taxation, their difference indicates the cost of not being able to redistribute using lump-sum transfers. Hence, these multipliers capture in a simple way the two forces which drive the solution of our model away from the first best: the absence of lump-sum taxes and of agent-specific lump-sum transfers.

For the government’s problem to be well defined, we should ensure that the set of feasible equilibria is non-empty. This is guaranteed, for example, if the status-quo allocation is feasible in the planner’s problem (i.e., if $\tilde{\tau}$ is larger or equal to the status-quo capital tax) and $U_2$ is close to the status-quo utility.
3 Characterization of equilibria

Here we describe some analytical results.

3.1 Qualitative behavior of capital taxes

To the best of our knowledge no available result guarantees that $\tau^\infty_k = 0$ in our model. This is so even if we took for granted that multipliers have a steady state.\(^{13}\) Furthermore, as emphasized in Lansing (1999), Straub and Werning (2014), and Reinhorn (2014) (and in a 2008 version of the present paper) previous results proving that $\tau^\infty_k = 0$ assuming that Lagrange multipliers have a finite steady state were flawed, since Lagrange multipliers should not be constrained to be bounded. Therefore, we take for granted the existence of a steady state in quantities but not in multipliers. This is the right way to proceed, because real variables have natural bounds and existence of a finite steady state can be expected. On the other hand, if arbitrary bounds are imposed on the multipliers, the Lagrangian is not guaranteed to give a maximum, and a steady state in allocations could be compatible with multipliers which go to infinity or to zero.

**Proposition 1.** Assume $\bar{c} > 0$, $F(k, 0) = F(0, e) = 0$, and $0 < \bar{\tau} < 1$. Further, assume that allocations have a limiting steady state, denoted $(c, k, l, e)$. Then $\tau^k_t \to 0$ as $t \to \infty$. Take $N < \infty$ such that $\tau^k_{N+1} < \bar{\tau}$, such an $N$ is guaranteed to exist. If in addition

$$c_{1,t} > \bar{c}, 0.1cm \forall t > N,$$  \hspace{1cm} (20)

then $\tau^k_t = 0$ for all $t \geq N + 2$.

**Proof.** It would be trivial to show that the proposed features of the solution satisfy the FOCs of the planner’s problem. But this approach would be questionable, because the feasible set is non-convex and FOCs are not sufficient for a maximum.

We proceed in two steps. First, we show that $\tau^k_t \to 0$. Second, we show that capital taxes go from the limit to zero in two periods.

\(^{13}\)Chamley (1986) and Atkeson, Chari, and Kehoe (1999) (ACK) also prove that the tax limit cannot be binding forever and that the transition takes two periods but the results in those papers are not directly applicable to our model. First because they do not consider a tax limit and heterogeneity at the same time, and, more importantly, their proof is for the case where $\bar{\tau} = 100\%$. ACK discard a tax limit binding forever by showing that if $\bar{\tau} = 100\%$ feasibility would be violated. The same line of argument does not apply here, because we assume that $\bar{\tau}$ is equal to the status-quo capital tax rate, and the economy could stay at the status quo forever.
We first find some relations that hold asymptotically among multipliers. Evaluating the planner’s FOC for capital (see Appendix A) at steady state for large $t$ we have that

$$\mu_t = A\mu_{t+1} - \gamma_t\beta c_1^{-\sigma_c} F_{kk}(k,e)(1 - \bar{\tau})$$

(21)

for a constant

$$A \equiv \beta(1 - \delta + F_k(k,e)).$$

(22)

Collecting the terms in the FOC for labor for $t > 0$ (see Appendix A) at steady state which do not depend on the multipliers $\gamma$ or $\mu$, asymptotically we have

$$B + C c_1^{-\sigma_c} \gamma_{t-1} = \mu_t,$$

(23)

where

$$B = \frac{\omega l_1}{F_e(k,e)} \left[ 1 + \frac{\phi_2 f_1(\lambda, l_1)}{\phi_1} + \left( \Delta_1 + \Delta_2 \phi_2 f_1(\lambda, l_1) \right) (1 + \sigma_t) \right],$$

$$C = \frac{F_{ke}(k,e)}{F_e(k,e)} (1 - \bar{\tau}).$$

Notice that the terms $A$, $B$, and $C$ do not depend on the multipliers $\gamma$ or $\mu$. Note that because of (15) $c_1^{-\sigma_c}$ is finite. Together with assumptions $F(k,0) = F(0,e) = 0$, we have that $k > 0$ and $e > 0$, and this implies that $0 < F_e(k,e) < \infty$. Constant returns to scale and concavity imply that $F_{ke}(k,e)k = -F_{ee}(k,e)e > 0$. All these observations imply that $C > 0$ and finite. Combining (21) and (23) we have that asymptotically

$$\mu_t = D\mu_{t+1} - E$$

(24)

for constants

$$D = A - \frac{\beta F_{kk}(k,e)(1 - \bar{\tau})}{C},$$

$$E = -\beta F_{kk}(k,e)(1 - \bar{\tau}) \frac{B}{C}.$$

The fact that the linear difference equation (24) holds asymptotically implies that either $\mu_t \to \bar{\mu}$ as $t \to \infty$ for a finite $\bar{\mu}$ (this would happen in the case where $D > 1$, and also in the case where $D \leq 1$ if $\mu_t$ converged on a stable saddle path to $\bar{\mu}$) or $|\mu_t| \to \infty$ as $t \to \infty$. The

---

14We use in the proof the idea that we take difference equations that are linear in the multipliers for large $t$. Strictly speaking we should take into account that the coefficients in these difference equations should depend on $t$ as well. For example we should have $A_t$ in the equation below. However, it is clear that because all the equations are linear in the multipliers we can substitute $A_t$ by $A$, and the limiting argument about the behavior of multipliers is equally correct.
latter is impossible, because $\mu$ cannot be negative, and because $\mu$ is the marginal improvement in utility of one more unit of production and, since $c_1 \geq \tilde{c} > 0$, marginal utility is bounded in equilibrium so that $\mu$ has a finite upper bound. Therefore, the only possibility is that $\mu_t$ has a finite steady state, $\mu_t \to \overline{\mu}$. Since $\gamma \geq 0$ and $F_{kk} \leq 0$, we have that (21) implies

$$\overline{\mu}(1 - A) \geq 0. \quad \text{(25)}$$

Now assume, towards a contradiction, that the tax limit is binding in all periods for $t$ large enough. Then the FOCs of the consumer give

$$\beta [1 + (F_k(k, e) - \delta)(1 - \tilde{\tau})] = 1. \quad \text{(26)}$$

Since $\tilde{\tau} < 1$, this implies $A > 1$, and in turn (25) implies $\overline{\mu} \leq 0$. Since multipliers are non-negative, the only possibility is $\overline{\mu} = 0$. But this is a contradiction, because $\mu$ is the marginal improvement in utility and, given non-satiation, it must be that one more unit of production increases utility, hence $\overline{\mu} > 0$. Therefore, the tax limit cannot be binding forever. Equivalently, this shows that $\gamma_t = 0$ infinitely often.

Now, equation (23) implies

$$\gamma_{t-1} = \frac{\mu_t - B}{Cc_1^{\sigma c}}. \quad \text{(27)}$$

Since $\mu$ has a steady state, the right-hand side converges to $\frac{\mu - B}{Cc_1^{\sigma c}}$. The fact that $\gamma_t = 0$ infinitely often implies that $\frac{\mu - B}{Cc_1^{\sigma c}} = 0$. Further, (27) implies $\gamma_t \to 0$. Plugging this limit into the right-hand side of the planner’s FOC for $k$ gives

$$\mu_t - \beta (1 - \delta + F_k(k_t, e_{t+1})) \mu_{t+1} \to 0. \quad \text{(28)}$$

Combining this with the FOCs of the consumer a standard argument gives $\tau^k_t \to 0$.

Now we prove the second part of the proposition. Given $N$, consider the following modification to the baseline model. Assume that instead of the uniform tax limit in all periods, we impose

$$\tau^k_t \leq \overline{\tau}, \; \forall t \neq N + 1,$$

but $\tau^k_{N+1}$ is unconstrained. Let us call this ‘modified model 1’ (MM1). It is clear that the solution to this problem is equal to the solution of the baseline model, because we have just relaxed a tax limit which was not binding in the optimum of the baseline model. Let us keep this fact in store for a while.

Now consider a second modified model, which we dub MM2, where we impose

$$\tau^k_t \leq \overline{\tau}, \; \forall t \leq N,$$
but $\tau^k_t$ is unconstrained for all $t > N$. Let us denote with a $\hat{}$ the solution to MM2. Clearly, the FOCs for this model are the same as for the baseline model except that

$$\hat{\gamma}_t = 0, \quad \forall t \geq N. \quad (29)$$

Notice that $\gamma_t$ is the multiplier associated with the constraint on $\tau^{k+1}_t$, so that $\tau^k_N + 1$ being the first unconstrained tax means $\gamma_N$ is the first multiplier that must be 0.

With the current utility function (6), the FOC with respect to consumption for $t > 0$ (see Appendix A) becomes

$$c_{1,t}^{-\sigma} \left(1 + \psi \lambda^{1-\sigma} + (\Delta_1 + \lambda \Delta_2) (1 - \sigma_c)\right)$$

$$\quad - \sigma_c c_{1,t}^{-\sigma} \left[\gamma_t - \gamma_{t-1} (1 + (F_k (k_{t-1}, e_t) - \delta) (1 - \tilde{r}))\right] + \xi_t = \mu_t \frac{1 + \lambda}{2}. \quad (30)$$

Combining (29) with (30), and since (20) gives $\xi_t = 0$, at the solution to MM2 we have

$$\hat{c}_{1,t}^{-\sigma} \left(1 + \psi \lambda^{1-\sigma} + (\Delta_1 + \lambda \Delta_2) (1 - \sigma_c)\right) = \hat{\mu}_t \frac{1 + \hat{\lambda}}{2}, \quad \forall t \geq N + 1. \quad (31)$$

This last equation does not hold for $t = N$, because $\hat{\gamma}_{N-1} \neq 0$ appears in (30). Plugging (29) into the FOC with respect to capital (see Appendix A) we get

$$\hat{\mu}_t = \beta \hat{\mu}_{t+1} \left(1 - \delta + F_k \left(\hat{k}_t, \hat{e}_{t+1}\right)\right), \quad \forall t \geq N.$$ 

Then, using (31) we have

$$\hat{c}_{1,t}^{-\sigma} = \beta \hat{c}_{1,t+1}^{-\sigma} \left(1 - \delta + F_k \left(\hat{k}_t, \hat{e}_{t+1}\right)\right), \quad \forall t \geq N + 1.$$ 

Comparing this equation with the Euler equation of the consumer, we conclude that

$$\hat{\tau}_t^k = 0, \quad \forall t \geq N + 2. \quad (32)$$

Therefore, the properties for taxes mentioned in the statement of the proposition hold for MM2.

Since the optimal solution to MM2 is also feasible in MM1, even though the latter is more restrictive, because $\tau_t^k$ for $t > N + 1$ are (potentially) constrained, $\hat{\tau}_t^k$ is also the optimal tax in MM1, $\forall t$. This proves that in MM1

$$\tau_t^k = 0, \quad \forall t \geq N + 2.$$ 

Since we have already argued that the solution to MM1 was equal to the solution of the baseline model, this completes the proof.\textsuperscript{15}

\textsuperscript{15}Notice that for the proof to work we do need to consider the two modified models MM1 and MM2. If we tried to compare MM2 to the solution of the baseline model directly, we would not be able to rule out that $\hat{\tau}_{N+1}^k > \tilde{r}$. The solution to MM2 would then be infeasible in the baseline model and could not be compared to it.
It follows that if (20) holds for $N \equiv \arg \min_t \{\tau_{t+1}^k < \bar{\tau}\}$, then the capital tax jumps from the tax limit to zero in two periods, that is,

$$\tau_t^k = \begin{cases} 
\bar{\tau}, & \forall t \leq N, \\
0, & \forall t \geq N + 2.
\end{cases} \quad (33)$$

In practice it is quite likely that if the initial capital stock is lower than the steady-state one, and if we choose a very small $\bar{c}$, then (20) holds and (33) follows. Therefore, for a sufficiently low $\bar{c}$, it is likely that $\min_{t>N} \{c_{1,t}\} > \bar{c}$ and (20) holds.

This argument falls short of providing a formal proof that such a sufficiently low $\bar{c}$ exists for any set of parameter values. But it provides a way of checking that a numerical solution is correct: given a candidate numerical solution computed using (33), one can check that indeed (20) holds, as we do in the simulations we report, and this guarantees that the transition in two periods is indeed the solution. This is useful because using (33) simplifies the computations considerably.

### 3.2 The frontier of the equilibrium set

We now describe how to trace out the frontier of equilibrium utilities and RPO allocations.

As is well known, under proportional taxes the set of CE allocations $S^{CE}$ is not convex. In this case a Lagrangian approach is strictly speaking not guaranteed to give all the RPO allocations. Formally, the duality gap (i.e. the set of optimal allocations that are not a saddle point of the corresponding Lagrangian) could be non-empty. Most of the literature on optimal taxation with homogeneous agents ignores this issue. In most cases this is not a problem: if one is interested in the equilibrium for a given initial debt level, it would be ‘bad luck’ if the equilibrium happened to belong to the duality gap. In any case, a careful researcher would notice if this was the case, because there would generically be several solutions to the FOCs.

But in our application this is a relevant issue, because we are interested in tracing out the entire Ramsey Pareto frontier. If the duality gap is non-empty we are bound ignore some relevant RPO allocations. To be precise, let the feasible set of utilities

$$S^U \equiv \left\{(U_1, U_2) \in \mathbb{R}^2 : U_j = \sum_{t=0}^{\infty} \beta^t \left[ u(c_{j,t}) + v(l_{j,t}) \right] \text{ for some } \{(c_{j,t}, l_{j,t})_{j=1,2}, k_t\} \in S^{CE} \right\},$$

and let $F$ be the boundary (or ‘frontier’) of $S^U$. In the standard case without distortions and a concave utility function, it is well known that $F$ corresponds to the RPO allocations, and it defines $U_1$ as a decreasing and concave function of $U_2$ (or vice versa). In that case an

---

16This is because the multipliers $\Delta_1, \Delta_2$, and $\lambda$, which determine steady-state capital $k$ in principle also depend on $\bar{c}$. 

---
allocation is Pareto optimal if and only if it optimizes a welfare function with fixed weights for consumers. But if $S^U$ is not convex, its frontier may have a non-concave part, and the equilibria with utilities in that non-concave part are not solutions of a Lagrangian with fixed weights. Further, parts of the frontier $\mathcal{F}$ may now be increasing, and in that case $\mathcal{F}$ will not coincide with the set of RPO allocations. Indeed, this is the case in the model of Section 4.2, where labor supply is fixed. For all these reasons we now show a sufficient condition guaranteeing that in our model we can compute all RPO equilibria. We will see that this condition can be checked numerically in our application.

Fix $\psi \in [-\infty, \infty]$. Consider the following modified model (MM3):

$$\max_{\tau_t, \lambda, \{c_{1,t}, l_{1,t}\}} \sum_{t=0}^{\infty} \beta^t \left\{ u(c_{1,t}) + v(l_{1,t}) + \psi [u(\lambda c_{1,t}) + v(f(\lambda, l_{1,t}))] \right\},$$

subject to all CE constraints and the tax and consumption limits. Notice that we allow for negative $\psi$'s and that we consider the case $\psi = \infty$ as a convention to denote the case where consumer 1 receives no weight. In order for MM3 to trace out all RPO allocations and a large part of the frontier $\mathcal{F}$ we need the following assumption. Let $U_j(\psi)$ be the utility of consumer $j = 1, 2$ at the solution to MM3.

**A1.** There is a unique solution to MM3 for all $\psi \geq 0$. Furthermore, $U_2(\cdot)$ is invertible on $[0, \infty]$.

**Proposition 2.** Assume A1.

1. A solution to MM3 for any $\psi \in [0, \infty]$ is Ramsey Pareto optimal.
2. Every Ramsey-Pareto-optimal allocation is also the solution of MM3 for some $\psi \in [0, \infty]$.
3. Given $\psi \in [-\infty, \infty]$, if the solution of MM3 exists, it defines a point on the frontier, i.e., $(U_1(\psi), U_2(\psi)) \in \mathcal{F}$.

**Proof.** Part 1 is obvious. For part 2, consider a pair of utilities $(\overline{U}_1, \overline{U}_2) \in S^U$ that correspond to a RPO allocation. Invertibility in A1 guarantees that there is a $\overline{\psi}$ such that $\overline{U}_2 = U_2(\overline{\psi})$. Consider a finite $\overline{\psi}$. We have

$$\overline{U}_1 + \overline{\psi} \overline{U}_2 \leq U_1(\overline{\psi}) + \overline{\psi} U_2(\overline{\psi}),$$

since the equilibrium that gives rise to $(\overline{U}_1, \overline{U}_2)$ is feasible in MM3, and the right-hand side is the value of the objective function of MM3 at the maximum with $\overline{\psi}$. Since $\overline{U}_2 = U_2(\overline{\psi})$, the above inequality implies $\overline{U}_1 \leq U_1(\overline{\psi})$. But the fact that $(\overline{U}_1, \overline{U}_2)$ is the utility of a RPO allocation implies $\overline{U}_1 \geq U_1(\overline{\psi})$. Therefore, the allocation that gives rise to $(\overline{U}_1, \overline{U}_2)$ attains
the maximum of MM3 with $\psi$. Uniqueness implies that this RPO allocation solves MM3 with $\psi$.

The case $\psi = \infty$ can be treated as $\psi = 0$ when agents 1 and 2 switch places in the objective function.

Part 3 follows from Part 2 when $\psi \geq 0$. For a given $\psi < 0$ we need to find points in $\mathbb{R}^2$ outside $S^U$ which are arbitrarily close to $(U_1(\psi), U_2(\psi))$. For any $\varepsilon > 0$ we have $(U_1(\psi) + \varepsilon, U_2(\psi) - \varepsilon) \notin S^U$, since this point achieves a higher value of the objective function of MM3 than its maximum. Since $(U_1(\psi) + \varepsilon, U_2(\psi) - \varepsilon)$ can be made arbitrarily close to $(U_1(\psi), U_2(\psi))$, this last point is in $\mathcal{F}$.

Part 2 of Proposition 2 implies that we can find all RPO allocations by solving MM3 varying $\psi$ from zero to infinity. Part 3 guarantees that we obtain additional points on the frontier $\mathcal{F}$ from $(U_1(\psi), U_2(\psi))$ for negative $\psi$. These points are not Pareto optimal, since both consumers’ utilities could be increased along the frontier. More points on the frontier can be found if the consumers switch places in the objective function of MM3, that is, if $\psi$ multiplies the utility of consumer 1 and we take $\psi \leq 0$. We will see in the model of section 4.2 how these can be used to find an increasing part of the frontier $\mathcal{F}$ which is not Pareto optimal.\footnote{Notice that if we had a standard model without distortions and $u(0) = -\infty$, then MM3 with $\psi < 0$ would not have a solution and, in that case, it would not define a point on the frontier of utilities.}

Since the feasible set is non-convex, A1 may not hold. But it can be checked numerically if it does hold in a given application. For uniqueness we search for more solutions to the FOCs, as is done in scores of papers where the maximum is found by searching for all critical points. For invertibility of $U_2(\cdot)$ we record all utilities for a fine grid of $\psi$’s and check that $U_2(\psi)$ is increasing and continuous (see our discussion about Figure 3 for a check of invertibility in the model of Section 4.3). Both of these checks can only be done approximately, as they rely on numerical approximations, but the two assumptions give us a clear indication of aspects of the solution which need to be checked.

The Ramsey Pareto-optimal and Pareto-improving (POPI) plans can be found with $\psi$ such that $(U_1(\psi), U_2(\psi))$ are larger than the status-quo utilities of consumer 1 and 2, respectively. We do this in practice in the next section.

4 Numerical results

We now present and discuss our numerical results. Details on our computational strategy are in Appendix B. In the next subsection, we discuss how we calibrate the model. Afterwards,
Table 1: Parameter values of the baseline economy

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.96</td>
</tr>
<tr>
<td>$\sigma_c$</td>
<td>1</td>
</tr>
<tr>
<td>$\sigma_l$</td>
<td>3</td>
</tr>
<tr>
<td>$\omega$</td>
<td>221.1</td>
</tr>
<tr>
<td>$\phi_c/\phi_w$</td>
<td>1.1</td>
</tr>
<tr>
<td>$k_{c,-1}$</td>
<td>3.831</td>
</tr>
<tr>
<td>$k_{w,-1}$</td>
<td>-0.611</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.394</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.074</td>
</tr>
<tr>
<td>$g$</td>
<td>0.094</td>
</tr>
<tr>
<td>$k_{g,-1}$</td>
<td>-0.315</td>
</tr>
<tr>
<td>$\tau^l$</td>
<td>0.214</td>
</tr>
<tr>
<td>$\tau^k$</td>
<td>0.401</td>
</tr>
</tbody>
</table>

we first analyze the case where labor supply is fixed. Second, in Section 4.3, we study optimal policy in an economy with flexible labor supply. We also contrast the results with those from an extension of our model in which lump-sum transfers are permitted, in order to gain intuition for the forces at work.

4.1 Calibration

We calibrate the model at a yearly frequency. An overview of our parameter choices is provided in Table 1.

We calibrate our parameters using the private sector’s FOCs at steady state, taking as given average effective tax rates and government debt, to match United States data. The macro variables are computed using the data for the period 2001-2010 provided by Trabandt and Uhlig (2012), who collected data from the OECD and other sources.\(^{18}\) We use the effective tax rates they have computed for the status quo, in particular, $\tau^l = 0.214$ and $\tau^k = 0.401$. The average level of government debt was 66.8 percent of GDP. Note that the choice of tax rates and government indebtedness at the status quo matters in several ways. First of all, they influence the status-quo steady-state (and hence initial) capital stock. Secondly, status-quo utilities depend on these variables, and thus restrict the scope for Pareto improvements. Thirdly, we assume that during the reform the capital tax rate can never increase above its initial level, i.e., we set $\bar{\tau} = 0.401$ as at the status quo.

\(^{18}\)https://sites.google.com/site/mathiastrabandt/home/downloads/LafferNberDataMatlabCode.zip
We set some preference parameters a priori. The utility function is as stated in Section 2. We set the annual discount factor to the most-commonly-used value, i.e., $\beta = 0.96$. We choose $\sigma_c = 1$ in keeping with a large part of the literature on taxation. The choice of $\sigma_l = 3$ is for the case of an elastic supply of labor, to be discussed in Section 4.3, which prevents hours from greatly differing across consumers with different wealth.\footnote{See Garcia-Milà, Marcet, and Ventura (2010) for a discussion of the trade-offs in choosing $\sigma_l$.} Note that this implies a lower Frisch elasticity of labor supply than in many applications of real-business-cycles (RBC) model, but is in line with micro estimates.

We assume that the production function is Cobb-Douglas with a capital elasticity of output of $\alpha = 0.394$ to match the labor income share. There is no productivity growth.

Our two types of consumers are heterogeneous with respect to both their labor efficiency $\phi_j$ and their initial wealth $k_{j,-1}$. The relevant aspect of heterogeneity when studying optimal proportional labor and capital income taxation is the agents’ wage-wealth ratio, see e.g. Garcia-Milà, Marcet, and Ventura (2010). We use their calculations from the Panel Study of Income Dynamics (PSID) when splitting the population into two groups: (i) those with above the median wage-wealth ratio, whom we call ‘workers,’ indexed $w$, and (ii) those with below the median wage-wealth ratio, called ‘capitalists,’ indexed $c$. Capitalists are richer relative to their earnings potential, however, both types of consumers work and save. We choose $\phi_c / \phi_w = 1.10$ to match the observed ratio of labor earnings, and $\lambda = c_w / c_c = 0.54$ is the ratio of consumptions of the two groups in the data.

Finally, we find $\omega, \delta, g, k_{t-1}^d$, and the initial wealth of each group in the model, $k_{c,-1}$ and $k_{w,-1}$, consistent with the steady state given the status-quo tax rates, using (i) the consumption-labor FOC of capitalists and that average hours should match the fraction of time worked for the working age population, 0.245,\footnote{Hours to be allocated between work and leisure: 13.64.} (ii) that capital held by the agents and the government has to equal steady-state aggregate capital, (iii) that $g$ over output has to equal government consumption over GDP (iv) that $k_{t-1}^d$ over output has to match the public assets-GDP ratio from the data, and (v-vi) the present-value budget constraint of both groups.

\subsection*{4.2 Results with fixed labor supply}

In our model, the set of POPI plans deviates from the first best for two reasons. One is that, as is standard in models of factor taxation, the need to raise tax revenue discourages the supply of capital and/or labor. The second reason is the lack of non-distortive means of redistribution between types of consumers. Since our paper is mostly about the latter, we
first analyze a case where only the redistributive effect is present. To do so, we assume fixed labor supply. Formally, in this section we take \( v(l) = 1 \) and \( l_{j,t} \leq \bar{l} \). We set hours worked \( \bar{l} = 0.245 \) to match the data, see above. All parameters unrelated to the utility from leisure are as in Table 1.

In a model with homogeneous agents and fixed labor supply the policy-maker would abolish capital taxes immediately, collect all revenues from taxes on labor, and thus implement the first-best allocation. In a model with heterogeneous agents, if the government could stipulate agent-specific lump-sum transfers at time 0 (with \( T_w = -T_c \) as introduced at the end of Section 2) so as to achieve a Pareto improvement, it could still achieve the first best. But in the case of interest where lump-sum redistribution is not possible, deviations from the first-best policy are necessary for distributive reasons.

In Figure 1 we compare the set of POPI plans to the first best. Units in this graph are consumption-equivalent welfare gains.\(^\text{21}\) The dashed black line labeled ‘first-best PI’ represents optimal allocations with \( \tau^k_t = 0 \) for all \( t \) where agent-specific lump-sum transfers are available and which are Pareto improving. The frontier of the set of possible competitive equilibria \( \mathcal{F} \) is depicted as the union of the solid blue and the dot-dashed green lines in Figure 1. This frontier is non-standard as it has an increasing part. The allocations in the increasing part of \( \mathcal{F} \), depicted with a (green) dot-dashed line, are not Pareto optimal, while the POPI allocations coincide with the decreasing part of \( \mathcal{F} \), depicted with a solid (blue) line. Figure 1 shows how the Ramsey Pareto frontier never reaches the welfare of capitalists at the status quo. Mechanically, the decreasing part of \( \mathcal{F} \) is found with \( \psi > 0 \) in MM3, higher \( \psi \) corresponding to points further to the right along the blue line. These points imply a longer period of high capital taxes. When \( \psi \to \infty \) (i.e. the planner cares only about workers) the POPI allocation converges to the point \( w^{max} \) in Figure 1. At that point capital taxes are at the upper bound for 24 periods. The points along the green line (the increasing part of \( \mathcal{F} \)) imply an even longer period of high capital taxes. This lowers the efficiency of the economy so much that it worsens both agents’ stance compared to the point \( w^{max} \). Mechanically, the increasing part of \( \mathcal{F} \) is found according to Proposition 2 by assigning a negative Pareto weight to capitalists.

Clearly, the absence of transfers significantly reduces the scope for Pareto improvements. All POPI plans depicted on the solid line are inferior to the first best. Why? If \( \tau^k_t = 0 \)

\(^{21}\)More precisely, in all the figures reporting results on welfare, the welfare gains for each consumer are measured as the percentage of a permanent increase in status-quo consumption which would give the consumer the same utility as the optimal tax reform. Therefore, the origin of the graph represents status-quo utilities, and the positive orthant contains utilities which correspond to Pareto-improving allocations.
for all $t$ (as in the first best with homogenous agents) and with $T_w = T_c = 0$ the worker would be worse off than at the status quo, as has been shown previously in a number of contributions.\textsuperscript{22} Therefore, all the Pareto-improving first-best allocations involve positive transfers to the worker, i.e., $T_w > 0$. This is because capital taxes at the status quo are disproportionately borne by capitalists, and when capital taxes are abolished, labor taxes have to rise in order for the government to meet its budget constraint. This increase in labor taxes due to an immediate reform has a strong redistributive effect and, from the perspective of the worker, it would overcompensate the welfare gains arising from increased efficiency. The only thing the planner can do to make the abolition of capital taxes palatable for the worker is to keep capital taxes high for a long time (the $N$ periods of Proposition 1) before setting capital taxes to zero from time $N + 2$ onwards. In this way the government raises more tax revenue from capitalists and less from workers. This is why POPI plans are second best even though taxation could be entirely non-distortive, and this would be the Ramsey optimum in a homogeneous-agents model or if lump-sum redistribution were available.\textsuperscript{23}

It is worthwhile to note that the utility loss relative to the first best is small if we only focus on equilibria which leave the worker indifferent and give all the benefits of the reform to the capitalist (i.e., if we focus on points where the frontiers cross the vertical axis of Figure 1). This requires the capital tax to stay at the upper bound for 12 years. But the utility loss becomes larger as we try to give some of the benefits to the worker. The most we can give to the worker is a 1.08 percent improvement (at point $w^{\text{max}}$), which is about one-seventh of the most the worker could gain with lump-sum redistribution. This requires capital taxes to stay at their upper bound for 24 years. There is little to be gained from cutting capital taxes if the worker must enjoy most of the benefits, but capitalists stand much more to gain even if we only look at Pareto-improving reforms. Finally, the optimal policy under ‘the veil of ignorance,’ i.e., when $\psi = 1$, gives welfare gains of 2.87 and 0.48 percent for capitalists and workers, respectively. We discuss this issue further in the next section.

4.3 Main results

We return to our benchmark model featuring an elastic labor supply. In particular, we set $\sigma_l = 3$, which means that the Frisch elasticity of labor supply is $1/3$.


\textsuperscript{23}Notice that in the case of a fixed labor supply the evolution of labor taxes is undetermined. All that matters is that the net present value of labor taxes balances the government’s budget given the optimal path for capital taxes found.
4.3.1 Th welfare frontier and capital taxes

Figure 2 reports the set of POPI plans in terms of welfare gains. Again, we contrast our main model with the case where agent-specific lump-sum transfers $T_w = -T_c$ are available. Note that even with access to transfers the first best is not attained in this case, because distortive capital and/or labor taxes are needed to raise tax revenue to finance government spending.

As with fixed labor supply, the absence of redistributive transfers clearly constitutes an extra constraint on the feasible set, and the welfare gains are smaller for POPI allocations than with lump-sum transfers. However, the limits to redistribution are less severe here than with fixed labor supply. The equilibrium frontier $F$, the solid (blue) line, is now decreasing in the range of Pareto-improving allocations, hence it is now feasible to leave either the worker or the capitalist indifferent relative to the status quo without violating Pareto optimality. In addition, the total welfare loss relative to the case with transfers is now much lower. If we focus, for example, on points that give equal gain to both consumers (the points where each frontier crosses the 45° line), we see that the welfare gain is roughly 1.3 percent for both consumers in the POPI allocation, only slightly below the 1.5 percent to be gained by both consumers with lump-sum redistribution. We conjecture, though, that for sufficiently high $\sigma_l$ and correspondingly close-to-inelastic labor supply, the picture would start resembling Figure 1.

To numerically verify that Assumption A1 (see Section 3.2) holds, Figure 3 shows the welfare gain of workers as a function of their relative Pareto weight. As required, $U_w(\psi)$ appears invertible.

As the distribution of welfare gains varies along the frontier of POPI plans, so do the corresponding capital tax schedule and relative consumption of agents. Qualitatively the properties of capital taxes over time are always the same: capital taxes are at their upper bound for all but the last period of the transition, and then they stay at zero, as we know from (33). Note that we will see in the next subsection that consumption grows after period $N$, thus providing a check that (20) holds so the second part of Proposition 1 is applicable. A typical time path for capital taxes is drawn in Figure 4.

The length of the transition increases as welfare gains are shifted towards the worker. This is illustrated in the first panel of Figure 5 showing the duration of the transition in the vertical axis for each POPI allocation indexed by the welfare gain of the worker on the horizontal axis. We see that the number of periods before capital taxes drop to zero increases from eleven to twenty-six years as we increase the welfare gain of the worker from zero (i.e.,
leaving the worker indifferent with the status quo) to 1.8 percent (which leaves the capitalist indifferent with the status quo). Along with the duration of the transition, the present-value share of capital taxes in government revenues increases from 12.7 to 21.7 percent, as the second panel in Figure 5 reveals. This is the clue to why a longer period of high capital taxes is beneficial for the worker: the worker contributes to the public coffers primarily through labor taxes, which means that his burden in the long run stands to increase through the reform, while the capitalist’s long-run burden decreases. The earlier capital taxes are suppressed, the more revenue has to be raised from labor taxes, and the bigger is the relative tax burden of the worker.

The final panel in Figure 5 depicts $\psi$, the multiplier on the minimum utility constraint (16) (or, equivalently, the relative Pareto weight of the worker in MM3), and $\lambda$, the ratio of the worker’s consumption to the capitalist’s in equilibrium. We put these two variables in the same picture because $\psi = \lambda$ would hold in a first-best situation without distortive taxation or distributive conflict ($\Delta_1 = \Delta_2 = 0$) and if the upper bound on capital taxes never binds ($\gamma_t = 0, \forall t$). In our second-best world, by contrast, as we increase the welfare of the worker, the marginal cost of doing so (as measured by $\psi$) explodes, while his consumption share increases only mildly. In fact, it always remains very close to its value at the status quo, which is 0.54. This shows that it is very difficult to alter the ratio of consumptions $\lambda$ even if the planner cares very differently about the two types of consumers, given that she has access only to proportional taxes.

If optimal lump-sum transfers were possible, the graphs in Figure 5 would look very different. We find that for all RPO allocations capital taxes would be suppressed after 9 years for all $\psi$, and the share of capital taxes would always be 10.3 percent. The multiplier $\psi$ would increase very little with $U_w$, while $\lambda$ would rise much more than without transfers. This is because in this case the redistribution can be achieved with agent-specific lump-sum taxes independently of the fact that the planner lowers quickly capital taxes to achieve aggregate efficiency. The policies and the path of the economy would hardly depend on the distribution of the gains from the reform. Shifting welfare gains and consumption between agents would be much easier.

In Appendix C we show that the main features described here are robust to changes in parameter values. In particular, we consider different measurements for the relevant tax rates and consumption inequality at the status quo. We recalibrate and solve our baseline

\footnote{For comparison, the share of capital taxes in revenues is about 37.1 percent at the status quo.}

\footnote{Note that even with lump-sum transfers, we do not obtain $\psi = \lambda$, which only holds in optimal allocations if there is no distortionary taxation.}
model considering both a lower and a higher value for each of the three data moments. In addition, we recalibrate the heterogeneity parameters \( \phi_j \) and \( k_{j,-1} \) to match the top and bottom quintiles of the wage-wealth distribution (rather than the top and bottom half as in Table 1). The POPI frontier is shown in Figure 6 and the numerical results are in Appendix C. In all these cases the results are similar to the ones for the benchmark calibration.

### 4.3.2 Interpreting \( \psi \)

In the literature on optimal policy with heterogeneous agents it is customary to fix certain agents’ weight \( \psi \) and to provide an interpretation for this choice. Some papers interpret \( \psi \) as probabilistic voting or as the bias of the planner in favor of some agents. Many authors focus on the case \( \psi = 1 \), justified by a moral choice under the ‘veil of ignorance.’

Given our focus on Pareto-improving allocations, the weight \( \psi \) is just the Lagrange multiplier of the promise-keeping constraint (16), and its value is determined in equilibrium. This provides a very different view of the role of \( \psi \). From this point of view, there is no reason that the case \( \psi = 1 \) should reflect an ‘equitable’ reform.

To be precise, we dub as ‘equitable reform’ a RPO solution which implies that both agents gain more or less equally.\(^{26}\) Graphically, equitable reforms are points on the frontiers of Figures 1, 2, and 6 which are near the 45\(^o\) line. Our calculations show that even within our model but for different parameter values, equitable reforms can imply very different values of \( \psi \). In fact, the reform corresponding to \( \psi = 1 \) can be very far from equitable.

For example, Figure 1 shows that with a fixed labor supply \( \psi = 1 \) gives most of the welfare gains to the capitalist. In this economy, the closest to an equitable reform we can get is the point \( w^{\text{max}} \) corresponding to \( \psi = \infty \). This shows that a very large relative Pareto weight may be required in order to achieve an equitable reform.

In the case where labor supply is flexible, the optimal policy for \( \psi = 1 \) does give a similar welfare gain to both agents: 1.10 percent to the worker and 1.35 percent to the capitalist, see Figure 2. Therefore, in this case \( \psi = 1 \) is roughly equitable.

The case where we calibrate our heterogeneity parameters the top and bottom quintiles of the wage-wealth distribution is shown in Figure 6. In that case \( \psi = 1 \) is not even Pareto improving, and, therefore, it is far from equitable. An equitable reform is achieved by setting \( \psi = 0.507 \).

This shows that \( \psi = 1 \) corresponds to an ‘equitable’ reform (or similar Nash-bargaining

---

\(^{26}\)Such a reform could be the outcome of Nash bargaining by agents’ representatives about which reform to implement when both agents’ representatives have a similar bargaining power and the outside option is the status quo.
power of both types of agents) only by chance. The value of $\psi$ that achieves an equitable reform differs strongly depending on the economy and on the ability of the planner to redistribute given the policy variables at her disposal. This can be seen, for example, in the final panel of Figure 5: $\psi$ has to increase a lot in order to achieve a small redistribution, reflecting the difficulties the planner faces in redistributing wealth from one consumer to the other when only proportional capital and labor taxes are available.

However, other aspects of the solution are similar across the different model versions that we have considered if we focus on Pareto-improving allocations. For example, the transition period of high capital taxes is similar in all the models if we focus on the point which gives all the welfare gain to the worker (24, 26, and 26 years, for the benchmark calibration with inelastic and elastic labor supply, and the quintiles calibration, respectively) or to the capitalist (12, 11, and 16 years, respectively).

### 4.3.3 The time path of the economy

The evolution of aggregate capital, labor, consumptions, tax rates, and government deficit are pictured in Figure 7. Different paths in each graph show different policies along the POPI frontier. First, note that qualitatively the paths are very similar. The horizontal shifts in the graphs occur because the more a plan benefits the worker, the longer capital taxes remain at their initial level. The kinks in the paths of labor taxes and government deficit occur precisely in the intermediate period when capital taxes transit from their maximum to zero.

The most surprising observation is, perhaps, that labor taxes should be initially lowered, and they should remain low for a long time. The reason for this behavior is the following: the planner wants to frontload capital taxes for the usual reason described at length in the literature with homogeneous agents,\textsuperscript{27} namely, that early capital taxes imply taxing capital that is inelastically supplied as it is already in place. Therefore, it is optimal to keep capital taxes at the upper limit in the first few periods and then let them go to zero. But with such high capital taxes investors would not invest much. However, the government has another instrument which can be used to boost output and capital accumulation in the early periods. The government can lower labor taxes, inducing an increase in labor supply, causing the return on capital to go up, increasing investment in the initial periods, and achieving a faster convergence to the optimal long-run capital-labor ratio compatible with zero capital taxes.

The upper right panel in Figure 7 shows that aggregate labor supply is very high in the early periods.\textsuperscript{28}

\textsuperscript{27}Section III of Jones, Manuelli, and Rossi (1993) find similar behavior in a model with homogeneous agents, where labor taxes should be very negative and capital taxes very high. In their paper this occurs only in one period.
periods. Note that the accumulation of capital accelerates around the period when capital
taxes become zero, as can be seen by comparing the kink in the graph for labor taxes with the
capital accumulation graph. Therefore, eventually the zero capital tax is the one promoting
growth and helping the economy converge to the new steady state. Absent this backloading of
labor taxes, capital would initially grow only to the extent that the expectation of low capital
taxes in the distant future raises incentives to save early on. In this case capital accumulation
would be much slower, as in the fixed-labor-supply case of Section 4.2. Therefore, low early
labor taxes are an instrument to induce investment in the early periods in the case of elastic
labor supply.

The same pattern can be observed in our model if optimal transfers are allowed. We
have computed that in the case with agent-specific lump-sum transfers, the period of low
labor taxes would be much shorter, 5 to 6 years in particular, matching the lower duration
of the transition to zero capital taxes. However, implementing this policy without lump-
sum transfers would leave the worker worse off than the status quo. Distributive concerns
lengthen the transition up to more than four times, as described in the previous paragraph.

It is interesting that with flexible labor supply the redistributive effect and the effect of
promoting growth go in the same direction: they both induce the planner to set low initial
labor taxes. This explains why with flexible labor supply the POPI frontier is closer to
the frontier with optimal transfers, as shown in Figure 2, than it is with fixed labor supply
(Figure 1). With elastic labor supply the desire to boost investment early on is not in conflict
with the redistribution objective.

A somewhat surprising pattern which emerges from the figures is that the long-run labor
tax rate is higher for a policy that favors the worker more. This may seem paradoxical,
because the worker is interested in low labor taxes. Note, however, that even though the
long-run labor tax rate is higher if the worker is favored, the initial cut is even larger, and
the share of labor taxes in the total present value of government revenues is lower for these
policies, as the second panel of Figure 5 shows. This suggests that the long-run labor tax
rate is high for two reasons. First, when capital taxation is abandoned late, the initial boost
to capital accumulation comes mainly from extremely low initial labor taxes. That is, the
backloading of labor taxes is strongest in these cases. Second, long-run labor supply is lower
the later capital taxes are suppressed, while the gross wage is always the same.\(^{28}\)

Since government expenditures are constant, low initial labor taxes translate into govern-

\(^{28}\)Since the long-run real return on capital is determined by the rates of time preference and depreciation,
and the production function is Cobb-Douglas, the long-run capital-labor ratio and wage are independent of
the policy, as long as capital taxes are zero eventually.
ment deficits. Only as labor taxes rise and output grows the government budget turns into surplus. Once capital taxes are suppressed and revenues fall again the government deficit quickly reaches its long-run value, which can be positive or negative depending on whether during the transition the government accumulated wealth or not. We can see from Figure 7 that most POPI policies imply that the government runs a primary surplus in the long run. This implies that the government is indebted in the long run, because the primary surplus is needed to pay the interest on debt. Therefore, for most POPI tax reforms low taxes in the initial periods generate a positive level of long-run government debt. This feature of the model is quite different from that of Chamley (1986), where the government accumulates savings in the early periods to lower the labor tax bill in the long run. Here, the early drop in labor taxes is financed in part with long-run government debt, showing that one possible reason for government debt is to finance the initial stages of a reform.

4.4 Extensions

Now we explore several variations of the model to consider issues of progressive taxation, political sustainability of equilibrium, and time consistency.

4.4.1 Progressive taxes

Given that we set out to analyze the consequences of distributive concerns for optimal tax policy, it might strike the reader as restrictive to allow proportional factor taxation only. After all, one of the prime instruments of redistribution in the real world is progressive taxation, so it is natural to ask if allowing for a progressive tax code would help solve the issue of redistribution and cause the economy to be closer to the first best. We now allow for non-proportional taxes in a simple way.

We assume that the planner can choose a lump-sum payment \( D \) which is paid in period zero uniformly across all consumers. Following Werning (2007), under complete markets this is equivalent to a fixed deductible from the tax base in each period. A positive \( D \) means progressive taxation.

Introducing this in the model is simple: the only change in all our equations is that we need to add \( u'(c_{1,0})[\Delta_1 + \Delta_2]D \) to the \( W \)-term in (18). We then let the planner maximize over \( D \) additionally.

We find that if we restrict our attention to a non-negative \( D \) (progressive taxation), the optimal choice is to set \( D = 0 \). Therefore, access to progressive taxation does not change any of our conclusions since the government optimally decides not to use progressivity.
The reason for this result is the following. There are two forces at work in the determination of the optimal $D$. On the one hand, distributive concerns would advise the government to choose a positive $D$, since capitalists are richer. But a negative $D$ is equivalent to a lump-sum tax, and it allows to raise revenue in a distortion-free manner. In the standard case of a representative-agent model, where only this second force is present, the first best can be achieved by choosing a negative $D$ big enough (in absolute value) to raise all government revenue ever needed. In our model with heterogeneous agents it turns out that the second force is stronger.

How can a negative $D$ be Pareto improving? The government now redistributes by choosing very negative labor taxes for many periods. In fact, the present value of revenues from labor taxes is not only negative but even bigger in absolute value than the revenue from capital taxes. The transition is 6 and 25 years at the two extremes of the POPI frontier. Welfare gains are larger than in the case with optimal transfers: capitalists can gain maximum 5.0 percent and workers 3.7 percent in welfare-equivalent consumption units.\footnote{Recall that we have calibrated our model according to wage-wealth ratios, because, as shown in Garcia-Milà, Marcet, and Ventura (2010), this is appropriate when only proportional taxes are allowed. In the real world some consumers with a high wage-wealth ratio are rich (say, some young stockbrokers) and some consumers with a low wage-wealth ratio are poor (say, some farmers in economically depressed areas). Hence, the wage-wealth ratio is not sufficient once progressive taxation is considered, instead the total income of the consumer is also relevant. Therefore, a careful study of progressive taxation should introduce total income in the calibration. In that case the optimal scheme described above would unlikely be Pareto improving. This is left for future research. However, the results in this subsection show that progressive taxation may have difficulties in solving the redistribution problem.}

4.4.2 The evolution of wealth and welfare and time consistency

One might conjecture that the welfare of workers and capitalists drift apart over time, with capitalists profiting from the abolition of capital taxes and workers suffering from high labor taxes in the long run. It might seem that such a scenario would render the tax reform politically unsustainable. We now study this issue, first by exploring the evolution of welfare and wealth and then more formally by addressing issues of time consistency.

The time paths of consumers’ welfare and wealth are plotted in Figure 8. Welfare increases along with the accumulation of capital, and, contrary to the conjecture, both consumers’ welfare evolves more or less in lockstep. The reason is that, by the CE conditions (7), both relative consumption and relative leisure are roughly constant over time. Therefore, it is not the case that workers lose dramatically when capital taxes finally drop to zero.

This is an implication of the permanent income hypothesis. Agents’ income net of taxes varies through time, hence consumers will save or dissave in order to smooth consumption.
and hours. The smooth time path of welfare is made possible by a less-smooth path of individual wealth. Since the workers’ main contribution to the public coffers is due in later periods when labor taxes are high, and in the early years of the new policy they benefit from extremely low labor taxes, they accumulate wealth to provide for the higher tax burden later on. The capitalists’ tax burden, by contrast, tends to decrease over time, since initial capital taxes are very high and they are later suppressed. By deferring wealth accumulation until their tax burden drops, capitalists can afford a smoother consumption profile.

The fact that the welfare of both types increases over time in a similar fashion suggests that the solution is, in some informal sense, politically sustainable. We can study if the solution we have found is time consistent more formally by performing some numerical checks. In particular, we study whether the planner would want to reoptimize if the new plan, just as the initial plan, has to be Pareto improving.

We assume that the optimal plan is followed for \( M - 1 \) periods and then in period \( M \) the planner reoptimizes if there is consensus for a new policy, taking \( k_{g,M-1}, k_{w,M-1}, \) and \( k_{c,M-1} \) as given. That is, a reoptimization takes place only if a Pareto-improving allocation can be found relative to the consumers’ continuation utilities at the period of reoptimization.

From our numerical experiments it seems impossible to make one consumer strictly better off without hurting the other. That is, reoptimizing with consensus always leads to the confirmation of the original plan in terms of taxes and allocations, only the Lagrange multipliers change. Let \( \tilde{\psi} \) and \( \tilde{\Delta}_j, j = 1, 2, \) denote the appropriately-chosen, reoptimized values of the time-invariant multipliers. The time-variant multipliers \( \mu_t \) and \( \gamma_t \) are rescaled by a factor \( \frac{1+\tilde{\psi}}{1+\psi} \). Moreover, we have the relationship \( \gamma_{M-1} = \frac{1+\tilde{\psi}}{1+\psi} \left( \tilde{\Delta}_1 k_{1,M-1} + \tilde{\Delta}_2 k_{2,M-1} \right) \). Inspection of the FOCs reveals that the remainder of the original optimal plan satisfies the FOCs of the reoptimization problem. Interestingly, \( \tilde{\psi} \) always turns out to be smaller than \( \psi \). For instance, in the case of \( \psi = 1 \) and reoptimization in period \( M = 5 \), the continuation utilities are respected if \( \tilde{\psi} = 0.63 \). Hence, the influence of the worker on the solution under consensus reform, measured by his relative Pareto weight, has to be lower at the point of reoptimization.

This suggests that in order to sustain the tax reform it is not necessary to write it as part of a constitution that cannot ever be changed. It is enough to require that the constitution can only be changed under wide consensus for the tax reform to be sustainable. This result is reminiscent of the one found by Armenter (2004) analytically in a simpler model.
5 Conclusion

We find that there is an equity-efficiency trade-off in the determination of capital and labor taxes. Capital taxes should be zero in the long run, but this is an optimal Pareto-improving policy only if capital taxes are very high (and labor taxes very low) for a very long time after the reform starts. The government typically accumulates debt in the long run in order to finance the initial cut in labor taxes. Lower initial labor taxes are necessary for two reasons: first, to redistribute wealth in favor of workers to ensure that they also gain from the reform, and, second, to boost investment in the initial periods.

Many of our results are numerical, for a given calibration of heterogeneity according to wage-wealth ratios. The results are robust to variations in parameter values and even to the introduction of progressive taxation. If labor supply is inelastic, it is very costly to make workers enjoy significant benefits from the capital tax cut, while an elastic labor supply makes it possible for the government to ensure that workers enjoy a larger welfare gain. The reason is that with a flexible labor supply an initial cut in labor taxes promotes both efficiency and redistribution at the same time. The solution is time consistent if consensus is required at the time of reoptimization, suggesting that the tax reform is credible if it can only be overturned when all agents agree. We also find that optimal policies that give equal weights to the two types of agents can be very far from equitable and, for some parameters, they can even lead to non-Pareto-improving allocations.

Our analysis suggests that issues of redistribution are crucial in designing optimal policies involving capital and labor taxes, even though the Chamley-Judd result survives with heterogeneous agents. Therefore, much is to be learnt from more research on these issues, both from an empirical and a theoretical point of view. One avenue for research is to study other policy instruments which may be used to compensate the workers for the elimination of capital taxes. For example, promoting certain types of government spending or cuts to other taxes could play this role. More empirical work on the relevant aspects of heterogeneity so that issues of progressivity can be addressed carefully is certainly needed. The transition in our model is very long, therefore partial credibility or absence of rational expectations might render this policy ineffective in practice. Introducing issues of partial credibility, learning about expectations, and political economy would therefore be of interest and might influence the picture on what an optimal policy should do.
References


Appendices

A First-order conditions of the policy-maker’s problem

Using the derivations in Section 2, the Lagrangian of the policy-maker’s problem in recursive form is

\[
\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \left\{ u(c_{1,t}) + v(l_{1,t}) + \psi \left[ u(\lambda c_{1,t}) + v(f(\lambda, l_{1,t})) \right] \right. \\
+ \Delta_1 \left[ u'(c_{1,t}) c_{1,t} + v'(l_{1,t}) l_{1,t} \right] + \Delta_2 \left[ u'(c_{1,t}) \lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'(l_{1,t}) f(\lambda, l_{1,t}) \right] \\
+ \xi_t (c_{1,t} - \bar{c}) + \gamma_t u'(c_{1,t}) - \gamma_{t-1} u'(c_{1,t}) (1 + (r_t - \delta) (1 - \tau)) \\
+ \mu_t \left[ F(k_{t-1}, e_t) + (1 - \delta) k_{t-1} - k_t - \frac{1 + \lambda}{2} c_{1,t} - g \right] - \psi U_2 \\
- u'(c_{1,0}) (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}) (1 + (r_0 - \delta) (1 - \tau^k_0)),
\]

with \( \psi \geq 0, \xi_t \geq 0, \) and \( \gamma_t \geq 0, \forall t, \) with the usual complementary slackness conditions, and \( \gamma_{-1} = 0. \) The FOCs are:

- for consumption at \( t > 0: \)

\[
u'(c_{1,t}) + \psi \lambda u'(\lambda c_{1,t}) + (\Delta_1 + \lambda \Delta_2) [u'(c_{1,t}) + u''(c_{1,t}) c_{1,t}] + \xi_t \\
+ \gamma_t u''(c_{1,t}) - \gamma_{t-1} u''(c_{1,t}) (1 + (r_t - \delta) (1 - \tau)) = \mu_t \frac{1 + \lambda}{2}
\]

- for consumption at \( t = 0: \)

\[
u'(c_{1,0}) + \psi \lambda u'(\lambda c_{1,0}) + (\Delta_1 + \lambda \Delta_2) [u'(c_{1,0}) + u''(c_{1,0}) c_{1,0}] + \xi_0 \\
+ \gamma_0 u''(c_{1,0}) - u''(c_{1,0}) (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}) (1 + (r_0 - \delta) (1 - \tau^k_0)) = \mu_0 \frac{1 + \lambda}{2}
\]

- for labor at \( t > 0, \) noting that \( r_t = F_k(k_{t-1}, e_t) = F_k \left( k_{t-1}, \frac{\phi_1 l_{1,t} + \phi_2 f(\lambda, l_{1,t})}{2} \right): \)

\[
u'(l_{1,t}) + \psi v'(f(\lambda, l_{1,t})) f_t(\lambda, l_{1,t}) \\
+ \Delta_1 [v'(l_{1,t}) + v''(l_{1,t}) l_{1,t}] + \Delta_2 \frac{\phi_2}{\phi_1} [v'(l_{1,t}) f_t(\lambda, l_{1,t}) + v''(l_{1,t}) f(\lambda, l_{1,t})] \\
- \gamma_{t-1} u'(c_{1,t}) F_k e(k_{t-1}, e_t) \frac{1}{2} (\phi_1 + \phi_2 f_t(\lambda, l_{1,t})) (1 - \tau) \\
= -F_k(k_{t-1}, e_t) \frac{1}{2} (\phi_1 + \phi_2 f_t(\lambda, l_{1,t})) \mu_t
\]
• for labor at \( t = 0 \):

\[
v'(l_{1,0}) + \psi v'(f(\lambda, l_{1,0})) f_1(\lambda, l_{1,0}) \\
+ \Delta_1[v'(l_{1,0}) + v''(l_{1,0}) l_{1,0}] + \Delta_2 \frac{\phi_2}{\phi_1} [v'(l_{1,0}) f_1(\lambda, l_{1,0}) + v''(l_{1,0}) f(\lambda, l_{1,0})] \\
- u'(c_{1,0}) (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}) F_{ke}(k_{-1}, e_0) \frac{1}{2} (\phi_1 + \phi_2 f_1(\lambda, l_{1,0})) (1 - \tau_0^k) \\
= -F_e(k_{-1}, e_0) \frac{1}{2} (\phi_1 + \phi_2 f_1(\lambda, l_{1,0})) \mu_0
\]

• for capital at \( t \geq 0 \):

\[
\mu_t + \gamma_t \beta u'(c_{1,t+1}) F_{kk}(k_t, e_{t+1}) (1 - \tilde{\tau}) = \beta \mu_{t+1} (1 - \delta + F_k(k_t, e_{t+1})).
\]

**B Computational strategy: Approximation of the time path**

1. Fix \( T \) as the number of periods after which the steady state is assumed to have been reached. (We use \( T = 150 \).)

2. Propose a \( 3T + 3 \)-dimensional vector \( X = \{k_0, ..., k_{T-1}, l_0, ..., l_{T-1}, \gamma_0, ..., \gamma_{T-1}, \Delta_1, \Delta_2, \lambda\} \). Note that this is not the minimal number of variables we could find solving a fixed point problem. \( 2T + 3 \) would be sufficient. However, convergence is better if the approximation errors are spread over a larger number of variables.

3. With \( k_{-1} \) and \( g \) known, find \( \{c_t, F_{k,t}, F_{l,t}, F_{kl,t}, F_{kk,t}\} \) from the resource constraint and the production function.

4. Calculate \( \{\mu_t\} \) from the FOC for labor.

5. Calculate \( \{\gamma_t\} \) from the FOC for consumption, making use of \( \{\mu_t\} \) and the guess for \( \{\gamma_{t-1}\} \) from the X-vector.

6. Form the \( 3T + 3 \) residual equations to be set to 0:

   - The FOC for capital (Euler equation) has to be satisfied. \( (T \) equations) 
   - The vector \( \{\gamma_t\} \) has to converge, i.e., old and new guesses have to be equal. \( (T \) equations) 
   - Check for each period whether the constraint on \( \tau^k \) is satisfied. If yes, impose \( \gamma_t = 0 \). Otherwise, the constraint on capital taxes has to be satisfied with equality. \( (T \) equations)
The remaining 3 equations come from the present-value budget constraints (PVBC) and the FOC for $\lambda$. The discounted sums in the PVBCs are calculated using the time path of the variables for the first $T$ periods and adding the net present value of staying at the steady state thereafter.

7. Iterate on $X$ to set the residuals to 0. We use a trust-region dogleg algorithm and Broyden’s algorithm, repeatedly when necessary, to solve this $(3T + 3)$-dimensional fixed point problem. We thank Michael Reiter for providing us his implementation of Broyden’s algorithm.

C Sensitivity analysis

To check the sensitivity of our results to the measurement of relevant tax rates and consumption inequality at the status quo, we recalibrate and solve our baseline model considering both a lower and a higher value for each of the three data moments.

In addition, we recalibrate the heterogeneity parameters in our model to the top and bottom quintiles of the wage-wealth distribution in the PSID rather than the top and bottom half as in the main text. By looking for RPO solutions that improve the welfare of the highest and lowest quintiles we find policies that most likely improve the utility of most of the population. In that case $\phi_{c}/\phi_{w} = 1.05$ and $\lambda^{SQ} = 0.31$.

Table 2 summarizes the results by reporting the duration of the transition and the revenue share of capital taxes for the two extreme points of the set of POPI plans. We always find the same qualitative properties of the optimal policy as for the baseline calibration described in Section 4.

D Alternative solution strategies for RPO allocations

The closest paper to ours is Flodén (2009). It is important to clarify the differences, as in our view Flodén’s strategy of solving a model with a so-called ‘optimized’ agent does not find all RPO solutions. In fact, it is not clear that this strategy gives RPO allocations except in a very special case. Here we describe in detail his approach and review his contribution.

There are several ways in which our solution approach differs from Flodén’s. He assumes that agents have a Greenwood-Hercowitz-Huffman (GHH) utility, i.e., the utility of agent $j$ is

$$U_{j,t} = \frac{1}{1 - \mu} \left( c_{j,t} - \frac{\zeta}{1 + 1/\gamma} \frac{1}{j_{j,t}^{1+1/\gamma}} \right)^{1-\mu}.$$
Table 2: Sensitivity analysis

<table>
<thead>
<tr>
<th>Calibration</th>
<th>workers gain as much as possible</th>
<th>capitalists gain as much as possible</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>duration of revenue share</td>
<td>of $\tau^k(%)$</td>
</tr>
<tr>
<td>benchmark</td>
<td>26</td>
<td>21.7</td>
</tr>
<tr>
<td>$\tau_{SQ}^k = 0.3$</td>
<td>35</td>
<td>25.5</td>
</tr>
<tr>
<td>$\tau_{SQ}^k = 0.57$</td>
<td>17</td>
<td>15.4</td>
</tr>
<tr>
<td>$\tau_{SQ}^k = 0.15$</td>
<td>30</td>
<td>36.1</td>
</tr>
<tr>
<td>$\lambda_{SQ} = 0.3$</td>
<td>14</td>
<td>7.2</td>
</tr>
<tr>
<td>$\lambda_{SQ} = 0.5$</td>
<td>25</td>
<td>21.5</td>
</tr>
<tr>
<td>$\lambda_{SQ} = 0.6$</td>
<td>25</td>
<td>21.3</td>
</tr>
<tr>
<td>quintiles</td>
<td>26</td>
<td>25.1</td>
</tr>
</tbody>
</table>

Notes: The column entitled ‘Calibration’ indicates which data moment has been reset to which value. The subscript ‘SQ’ refers to the status quo.

This is a non-separable utility function, unlike ours, but it is immediate to extend our approach to this case. In addition, Flodén (2009) considers a general measure of agents $\tilde{\lambda}(s)$ ($\lambda(s)$ in Flodén, 2009) of agents of type $s$. Our two-types-of-agents setup is a special case of his, therefore this is not an important difference either. Our approach could also be generalized to a general measure of agents.

Flodén writes the planner’s problem as Atkeson, Chari, and Kehoe (1999), ACK hereafter, by keeping consumption of all agents in the equilibrium conditions, instead of summarizing the allocations of other agents using (7) and $\lambda$ as we do. Although this makes computations different, it should not affect the allocations found. We describe this approach in detail below.

A key difference is that Flodén solves a planner’s problem that maximizes the utility of one agent (the ‘optimized’ agent). Then Proposition 5 in his paper claims that all RPO allocations can be traced out by changing the wage and wealth of the optimized agent. By contrast we solve for all individual allocations directly (through the optimal choice of $\lambda$). These differences are important and we examine them carefully below.

We use the notation

$$u_{jc,t} = \left( c_{j,t} - \frac{\zeta}{1 + 1/\gamma} \right)^{-\mu}_{1+1/\gamma}$$

and similarly for $u_{jl,t}$. 

40
Using an ACK Lagrangian

Instead of representing equilibrium conditions with \((7)\) as we do, Flodén follows ACK and keeps equilibrium conditions
\[
\frac{u_{1c,t}}{u_{1c,t+1}} = \frac{u_{ic,t}}{u_{ic,t+1}} \quad \text{and} \quad \frac{u_{1l,t}}{u_{1l,t} \phi_1} = \frac{u_{il,t}}{u_{ic,t} \phi_i},
\]
for all \(i\), as separate constraints in the planner’s problem. Feasibility, firm behavior, and budget constraints are as in the main text of our paper. For simplicity we do not consider consumption limits or tax limits in this appendix.

We focus on the case where \(\tilde{\lambda}\) is a discrete measure with \(I\) types of agents, where \(I\) a finite integer, and agent \(i\) has mass \(\tilde{\lambda}_i\). This is the case of our main text with \(I = 2\) and \(\tilde{\lambda}_1 = \tilde{\lambda}_2 = 1/2\). It also seems to be the case that Flodén is thinking of, since in the computations he looks at a case with 300 agents, each with the same mass. We comment on the case of a continuum of agents at the end.

The Lagrangian to find the RPO allocations using this approach would be
\[
L = \sum_{t=0}^{\infty} \beta^t \left\{ \left( \sum_{i=1}^{I} \psi_i U_{i,t} + \Delta_i \left[ U_{i,t} (1 - \mu) + \frac{u_{1c,t} \xi_1^{1/\gamma+1}}{\gamma+1} \right] \right) 
+ \rho_{it} \left[ u_{1c,t} u_{ic,t+1} - u_{ic,t} u_{1c,t+1} \right] + \xi_{it} \left[ u_{1c,t} u_{il,t} \phi_{1} - u_{1l,t} u_{ic,t} \phi_i \right] \right) 
+ \mu_t \left( \sum_{i=1}^{I} \tilde{\lambda}_i c_{i,t} + g + k_t - (1 - \delta) k_{t-1} - F(k_{t-1}, e_t) \right) \right\} + \sum_{i=1}^{I} \Delta_i W_{i,-1}.
\]

We use Flodén’s notation except that we use \(\psi\) instead of his agent weights \(\omega_i\), we use \(\Delta_i\) for the multipliers of individual implementability constraints instead of \(\lambda_i\), and for the multiplier of the feasibility constraint we use \(\mu_t\) instead of Flodén’s \(-\nu_t\).

We prefer representing CE in the main text using \((7)\) to substitute out agent 2’s consumption and labor because then the planner’s problem can be written as a maximization over \(\tau^k_0, \lambda, \{c^l_t, k_t, \xi^l_t\}_{t=0}^{\infty}\). This reduces enormously the number of variables and multipliers to be computed, and it is much more convenient for computation. More precisely, given the algorithm described in Appendix B, the number of variables to solve for with \(I\) agents would be \((2 + I) \times T + 2 + I\) using the ACK approach, while using our approach the number of variables to compute is only \(3T + 2 + I\). But solving the Lagrangian \((37)\) is equally valid, and it should give the same solution as we find. Hence in this appendix we characterize RPO solutions to \((37)\), as in done in Flodén (2009).
Using a representative agent

Flodén actually uses a modification of the above Lagrangian applying his Proposition 3. This proposition says that CE constraints can be summarized in an implementability constraint of a representative agent (RA) who has productivity $\phi^{RA} \equiv \left( \sum_{i=1}^{I} \tilde{\lambda}_i \phi_i^{1+\gamma} \right)^{\frac{1}{1+\gamma}}$ and initial wealth $\sum_{i=1}^{I} \tilde{\lambda}_i k_{i,-1} = k_{-1} - k_{-1}^g$. This RA consumes $C_t^{RA} = \sum_{i=1}^{I} \tilde{\lambda}_i c_{i,t}$. His Proposition 3 shows that as long as a CE satisfies

$$\sum_{t=0}^{\infty} \beta^t \left[ u_{C^{RA},t} C_t^{RA} + u_{L^{RA},t} L_t^{RA} \right] = W_{-1}^{RA},$$

(38)

there is a heterogeneous-agents equilibrium which is consistent with the tax policy for this RA economy.

Flodén finds equilibria that arise from the FOC of the Lagrangian on page 300 (all page and equation numbers refer to Flodén (2009)). The reader can check that one can go from the above Lagrangian (37) to Flodén’s on page 300 with the following three modifications:

1. Equation (38) is introduced in the planner’s problem as an additional constraint.
2. The competitive equilibrium conditions (36) are written in terms of ratios of individual marginal utilities to the RA’s marginal utilities.
3. Individual consumptions disappear from the feasibility constraint, i.e., $\sum_{i=1}^{I} \tilde{\lambda}_i c_{i,t}$ is replaced by $C_t^{RA}$ in the feasibility constraint.

Let us comment on the validity of these modifications.

Modification 1 is not needed for an equilibrium, because if all individual implementability constraints are satisfied, constraint (38) is guaranteed to hold. Therefore, modification 1 is redundant. All this means is that the multipliers $\lambda_i$ and $\tilde{\lambda}$ (in Flodén’s notation) are not uniquely defined, but the FOCs obtained from introducing modification 1 should give the same allocations as (37).

Modification 2 is also correct, indeed it implies and is implied by (36).

But modification 3 is incorrect. Only if an additional constraint was added restricting

$$\sum_{i=1}^{I} \tilde{\lambda}_i c_{i,t} = C_t^{RA},$$

(39)

one could put only $C_t^{RA}$ in the feasibility constraint. A similar point applies to aggregate labor.
As written, the Lagrangian on page 300 ignores the fact that the aggregate of all individual consumptions and leisure have to satisfy the feasibility constraint. A proper solution would entail incorporating the constraint (39) into the planner’s problem, since it is not implied by any combination of the other constraints imposed. Therefore FOCs (A.6) to (A.14) in Flodén (2009) do not provide a PO allocation.

That the FOCs of the Lagrangian in Flodén (2009) on page 300 do not give the correct solution can be seen in the following way. Let $L^2$ represent the expression in the first two lines of (37). The correct FOC with respect to $c_{i,t}$ from (37) would involve

$$\frac{\partial L^2}{\partial c_{i,t}} = -\mu_t \tilde{\lambda}_i.$$  \hspace{1cm} (40)

Now, since $\frac{\partial L^2}{\partial c_{i,t}}$ is the expression on the left-hand side of equation (A.6) in Flodén (2009) one can see that he is using the FOC

$$\frac{\partial L^2}{\partial c_{i,t}} = 0,$$  \hspace{1cm} (41)

which are not compatible with optimality. Therefore, the FOCs in Flodén (2009) do not give a RPO solution. In particular, his solution does not insure that

$$\frac{\partial L^2}{\partial c_{i,t}} = \frac{\partial L^2}{\partial c_{1,t}} \tilde{\lambda}_i \lambda_1,$$

as should hold in the optimum for all $i = 1, ..., I$. A similar issue is found in the FOCs with respect to individual labor. In other words, the FOCs on page 300 do not relate correctly the marginal conditions of the RPO solution to the Lagrange multiplier of the feasibility constraint and, therefore, the solution is not RPO.

If we considered a measure $\tilde{\lambda}(i)$ with a continuous density $\tilde{\lambda}'$ (where $\tilde{\lambda}$ represents the measure of agents denoted $\lambda$ on page 283), we would have the same problem. Then to find a RPO solution we would maximize $\sum_{t=0}^{\infty} \beta^t \left( \int_{[0,1]} \psi(i)U_{i,t}di \right)$ for some density $\psi$ and incorporating in the feasibility constraint that

$$C^{RA}_t = \int c(i)d\tilde{\lambda}(i),$$

we would find the FOC

$$\frac{\partial L^2}{\partial c_{i,t}} = -\tilde{\lambda}'(i)\mu_t, \forall i \in [0,1].$$  \hspace{1cm} (42)

This is incompatible with (41). The correct solution would imply $\int_J \frac{\partial L^2}{\partial c_{i,t}}d\tilde{\lambda}(s) = -\mu_t \int_J d\tilde{\lambda}(s)$ for any subset of agents $J$, but Flodén’s FOCs give $\int_J \frac{\partial L^2}{\partial c_{i,t}}d\tilde{\lambda}(s) = 0$.

The only case where (41) is correct is when an agent has $\tilde{\lambda}_i = 0$ in the discrete case or $\tilde{\lambda}'(i) = 0$ in the continuous case. In other words, it seems that the case where the FOCs are
valid is when the planner gives full measure in her objective function to agents who have zero measure in the market.

Later on Proposition 5 in Flodén (2009) argues that all RPO solutions can be traced out by maximizing the utility with respect to one ‘optimized’ agent, whose initial state is denoted $\tilde{\sigma}$. The proof of that proposition shows that the FOCs for this modified problem coincide with the FOCs on page 300 which are as (41). But if (as we think) the latter do not give an PPO allocation, then the conclusion of Proposition 5 does not follow. In fact, most RPO solutions involve giving weight to all agents in the objective function of the planner, hence (40) has to hold instead of (41) and, therefore, it is not true that all RPO solutions can be found by selecting an optimized agent even with GHH utility.

In our opinion one can only find all RPO allocations by taking properly into account the utilities of all agents in the economy, as we do in the main text, or as a direct application of (37) would do.

A rationale for Flodén’s solution

Although we do not provide a careful account, we believe that Flodén’s results can be reinterpreted as follows.

Imagine we consider optimizing a weighted sum of utilities of $I'$ agents (where $I'$ is a discrete number) and that these agents have mass zero in the economy. This can either be because $\tilde{\lambda}_i = 0$ for all $i = 1, \ldots, I'$ and $I' < I$ in the discrete case or because we consider only a discrete number of agents in the continuous case. For this RPO allocation the planner’s FOCs are indeed (41). But this is only a very small share of RPO solutions. For any welfare function that gives positive weight to all agents, (41) does not work.

Hence what Flodén does do is to find some fiscal policies which are feasible (in the heterogeneous-agents economy) by searching those that are optimal from the point of view of infinitesimal agents. This is a useful way of exploring the set of feasible policies in an ordered and easy-to-compute fashion, but it does not trace out all RPO equilibria, and indeed it is not guaranteed that the solutions found are even Pareto optimal for a set of agents of positive measure.
Figure 1: The Ramsey Pareto frontier of Pareto-improving equilibria with fixed labor supply

Notes: Welfare is measured as the percentage increase in status-quo consumption that would give the consumers the same utility as the optimal tax reform. The point $\psi = 1$ corresponds to the policy under ‘the veil of ignorance’ and the point $w^{max}$ represents the case where workers’ utility is highest, i.e., $\psi \to \infty$. 
Figure 2: The Ramsey Pareto frontier of Pareto-improving equilibria in the baseline model

![Ramsey Pareto frontier graph]

Notes: Welfare is measured as the percentage increase in status-quo consumption that would give the consumers the same utility as the optimal tax reform. The point $\psi = 1$ corresponds to the policy under ‘the veil of ignorance.’

Figure 3: Workers’ welfare increase as a function of their relative Pareto weight

![Workers’ welfare increase graph]

Notes: Welfare is measured as the percentage increase in status-quo consumption that would give the workers the same utility as the optimal tax reform.
Figure 4: A typical time path for capital taxes
Figure 5: Properties of POPI programs in baseline model

Duration of transition (years)

Share of capital taxes in government revenues

Workers' relative Pareto weight and consumption share (normalized)
Figure 6: The Ramsey Pareto frontier when calibrating to quintiles

Notes: Welfare is measured as the percentage increase in status-quo consumption that would give the consumers the same utility as the optimal tax reform. The point $\psi = 1$ corresponds to the policy under ‘the veil of ignorance.’
Figure 7: The time paths of selected variables for three POPI plans in the baseline model.
Figure 8: Typical time paths for consumers’ welfare and wealth.