Pareto-Improving Optimal Capital and Labor Taxes

Katharina Greulichy
Sarolta Laczó
Albert Marcet

March 2016

Pareto-Improving Optimal Capital and Labor Taxes[*]

Katharina Greulich[†]   Sarolta Laczó[‡]   Albert Marcet[§]

March 2016

Abstract

We study Pareto-optimal fiscal policy in a model with agents who are heterogeneous in their labor productivity and wealth. We show a natural modification of the standard Ramsey problem to guarantee that 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

---

[†]Swiss Re and Institut d'Anàlisi Econòmica (IAE-CSIC).

[‡]University of Surrey and Institut d'Anàlisi Econòmica (IAE-CSIC), School of Economics, AD Building ground floor, Guildford, Surrey, GU2 7XH, United Kingdom. Email: s.laczo@surrey.ac.uk.

[§]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 concluded 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 $\tau_\infty^k = 0$.[1] This policy recommendation is controversial as it implies increasing labor taxes, therefore, it seems to hurt less wealthy taxpayers and a large part of the population. But some papers argued that the optimal policy with heterogeneous agents also involves $\tau_\infty^k = 0$, even if the government only cares about improving the welfare of poor consumers (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.

We consider optimal policy under full commitment in a model in which agents are heterogenous in their labor productivity and wealth. The government can only levy proportional labor and capital taxes, and lump-sum transfers are not available. We also introduce an upper bound on capital taxes to insure smooth transitions and prevens very high initial values. We focus on Pareto-improving tax reforms and study the entire path of optimal taxes.[3]

We first revisit available results on $\tau_\infty^k = 0$ with heterogeneous agents. This is necessary because some recent results by Lansing (1999), Reinhorn (2014) and Straub and Werning (2015) show that $\tau_\infty^k = 0$ actually does not hold as often as had been previously thought. In fact many available results incorrectly assumed that Lagrange multipliers were bounded when in the optimum this is often not the case. In the same vein, and closer to our paper, Bassetto and Benhabib (2006) contradict previous results and they show that $\tau_\infty^k > 0$.

These recent results may give the impression that anything can happen to long-run capital taxes. We introduce a simple and (we contend) reasonable modification of the Ramsey problem: we prevent the government from immiserating future generations. Under this restriction we prove that $\tau_\infty^k = 0$ even when Lagrange multipliers are treated properly, providing a valid proof of this result under heterogeneous agents.[4]

---

[1] A very incomplete summary of a large literature is that in the few cases where $\tau_\infty^k \neq 0$, capital taxes are often small or even negative.

[2] Aiyagari (1995) shows that, due to capital over-accumulation, $\tau_\infty^k > 0$ 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 $\tau_\infty^k < 0$ 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 for all Pareto-improving allocations.

[3] An early paper studying the transition of optimal taxes with homogeneous agents is Jones, Manuelli, and Rossi (1993).

[4] To compare our result with the literature mentioned in the previous paragraph we show how our proof strategy recovers $\tau_\infty^k = 0$ in a version of Bassetto and Benhabib (2006), see Appendix E.

Then we study the transition using numerical simulations. We find that even if $\tau_\infty^k = 0$ for all Ramsey-Pareto-optimal (RPO) allocations, there is in fact an equity-efficiency trade-off. Pareto-improving allocations are achieved only if labor taxes are initially 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 $\tau_\infty^k = 0$ 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 with the literature on gradualism of political reforms, which has been at the center of some policy debates.[5] In our case, the slow transition to $\tau_\infty^k = 0$ means that a gradual reform is needed if all consumers should benefit from this reform. Therefore, high capital taxes that are observed currently in many economies 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.

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, as 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.

We then consider the case of an elastic labor supply. We find again 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. It recommends a very

---

[5]For example, the desirable speed of transition to market economies of formerly planned economies has been extensively discussed both in policy and academic circles. Within this literature, closest to our approach is Lau, Qian, and Roland (2001), who find a gradual reform which improves all consumers' welfare.

long transition where capital taxes are high and labor taxes are low.

Our results are complementary to some papers which establish that a large part of the population would suffer a large utility loss if the optimal transition is ignored and capital taxes are 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. This suggests that following the optimal transition is very important in order to achieve a Pareto improvement under heterogeneity, while the transition might be less important with homogeneous agents.[6]

In our main model, government debt is positive in the long run, while the government often accumulates savings in the optimal policy 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.[7]

To our knowledge, RPO allocations in models of factor taxation with heterogeneous agents have only been solved for in very special cases. We had to resolve a number of technical issues.[8] 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 consumptions, constant under complete markets and homogenous preferences, denoted λ. Then the computational cost of this model is, essentially, the same as of a homogeneous-agents model. This idea can be used with many utility and production functions. Another issue is that agents' relative Pareto weight in the welfare function does not coincide with the ratio of marginal utilities, instead this ratio has to be chosen optimally. A further difficulty is that the set of competitive equilibria is potentially not convex. It is well known, but often ignored, that this can be a problem in models of Ramsey taxation.[9] But it is possibly more of an issue in our paper because we want to trace

---

[6]Chari, Christiano and Kehoe (1994) show that there may be a slight loss in utility if the transition is ignored in a model with very high risk aversion and with homogeneous agents. By comparison our result about how the transition matters obtains with standard levels of risk aversion and it avoids very large utility losses for some agents, as found in the papers mentioned in this paragraph.

[7]A similar result can be found, for a different model, in Armenter (2004).

[8]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).

[9]An exception is Bassetto (2014) who states in section 3.1 that "we show how heterogeneity may lead to

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 the set of competitive equilibria by analyzing a planner's problem with negative welfare weights. We show one case where the frontier of the set of possible equilibria is non-standard as it has an increasing part.

Most papers on heterogeneous agents nowadays focus on the case where all agents receive the same weight in the welfare function of the planner. This is justified as a decision under the 'veil of ignorance'. Other papers interpret welfare weights as representing agents' political power. We treat welfare weights as another multiplier to be solved for, given the requirement of a Pareto improvement. We find that equal weights under the veil of ignorance are not necessarily related to a Pareto improvement, see Section 4.3.2.

Our focus on Pareto improvements is justified because it represents a list of allocations that could be achieved if a tax reform (under full commitment) needs the support of all agents. Therefore, our comment in the previous paragraph implies that the reform under the veil of ignorance would not be implemented if unanimity was needed.

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.[10]

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 proves analytically

---

situations in which the first-order conditions are not sufficient even in the simplest case."

[10]In the Online Appendix we argue that the approach in Flodén (2009) does not find all RPO allocations, although it does provide a simple and systematic way to search over some competitive equilibria.

some properties of the model, including $\tau_\infty^k = 0$ and some properties of the transition. It also provides 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, and a careful comparison with the result of Bassetto and Benhabib (2006). An online appendix gives details on the relation of our solution method to other approaches in the literature, including Atkeson, Chari, and Kehoe, 1999 and Flodén (2009).

2 The model

We consider an economy with 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 the following 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 RPO allocations and we focus on Pareto-improving allocations relative to a status quo.

iv) We impose that consumption is bounded away from zero. That is, we consider the case where the policy-maker is constrained not to immiserate all consumers.

v) We impose an upper bound below 100 percent 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 income 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.[11] Alternatively, this

---

[11]Lucas (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

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

We focus on the case of two types of consumers, $j = 1, 2$, with utility $\sum_{t=0}^{\infty} \beta^t \left[u\left(c_{j,t}\right) + v\left(l_{j,t}\right) \right]$ each, where $c_{j,t}$ is consumption and $l_{j,t}$ is labor of consumer j in period t. This is for simplicity, and it is immediate to extend our analysis 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 ϕ_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 at rate τ_t^l on labor income and τ_t^k on capital income net of a depreciation allowance. Therefore, the period-t budget constraint of consumer j is given by

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

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 strictly concave and increasing in both arguments, differentiable, has constant returns to scale, $F(k, 0) = F(0, e) = 0$, 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_{-1}^g. The government can hold debt, in that case $k_t^g < 0$. Ponzi schemes for consumers and the government are ruled out.

---

in Giannitsarou (2006). In this paper we keep the more common assumptions of rational expectations and full commitment.

We normalize the mass of each group to $1/2$. Market clearing conditions for all t are

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

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

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

2.2 Conditions of competitive equilibria

Our competitive-equilibrium (CE) concept is standard: consumers and firms take sequences of prices and taxes as given and maximize their utility and profits, respectively. Markets clear and the budget constraint of the government is satisfied. We now find a set of necessary and sufficient conditions that equilibrium allocations satisfy.

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

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

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

i.e., the Euler equation and the consumption-labor optimality condition, respectively, for $j = 1, 2$.

For many of our results we assume that the current utility function is isoelastic:

A1. *The two elements of the current utility function take the form*

$$u(c) = \frac{c^{1-\sigma_c}}{1-\sigma_c} \;\; and \;\; v(l) = -\omega\frac{l^{1+\sigma_l}}{1+\sigma_l}, \tag{6}$$

where ω is the relative utility weight of hours, $\sigma_c > 0$ is the (constant) coefficient of relative risk aversion, and $\sigma_l > 0$ is the inverse of the (constant) Frisch elasticity of labor supply.

Assumption A1 simplifies our characterization. It is clear that (4) for $j = 2$ can be replaced by the condition

$$\lambda \equiv \frac{c_{2,t}}{c_{1,t}}, \;\; \forall t, \tag{7}$$

for some constant λ to be determined in equilibrium. Further, (5) for $j = 2$ can then be replaced by

$$l_{2,t} = \mathcal{K}(\lambda)l_{1,t}, \;\; \forall t, \tag{8}$$

where $\mathcal{K}(\lambda) \equiv \lambda^{-\frac{\sigma_c}{\sigma_l}} \left(\frac{\phi_2}{\phi_1} \right)^{\frac{1}{\sigma_l}}$. Note that the function $\mathcal{K}$ depends only on exogenous objects, the functions $u()$ and $v()$ and labour productivities, ϕ_1 and ϕ_2.[12]

Firms behave in a competitive fashion, hence in equilibrum factor prices equal marginal products, i.e.,

$$ r_t = F_k \left(k_{t-1}, e_t \right) \quad \text{and} \quad w_t = F_e \left(k_{t-1}, e_t \right). $$

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 $t = 0, 1, ...$ can be summarized in the present-value budget constraint

$$ \sum_{t=0}^{\infty} \beta^t \frac{u'(c_{j,t})}{u'(c_{j,0})} \left(c_{j,t} - w_t \phi_j l_{j,t} \left(1 - \tau_t^l \right) \right) = k_{j,-1} \left(1 + (r_0 - \delta) \left(1 - \tau_0^k \right) \right), \quad \text{for } j = 1, 2. \quad (9) $$

Then, using (5) and rearranging, for consumer 1 we have

$$ \sum_{t=0}^{\infty} \beta^t \left(u'(c_{1,t}) c_{1,t} + v'(l_{1,t}) l_{1,t} \right) = u'(c_{1,0}) k_{1,-1} \left(1 + (r_0 - \delta) \left(1 - \tau_0^k \right) \right). \quad (10) $$

Using (4), (5), (7), and (8), we can write the present-value budget constraint of consumer 2 as

$$ \sum_{t=0}^{\infty} \beta^t \left(u'(c_{1,t}) \lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'(l_{1,t}) \mathcal{K}(\lambda) l_{1,t} \right) = u'(c_{1,0}) k_{2,-1} \left(1 + (r_0 - \delta) \left(1 - \tau_0^k \right) \right). \quad (11) $$

It is easy to show that a set of necessary and sufficient conditions for an equilibrium allocation is given by feasibility, the marginal conditions for, say, agent $j = 1$, constant ratios of labor and consumption, and the last two equations, the present-value budget constraints. Formally, let the set of CE allocations be

$$ \mathcal{S}^{CE} \equiv \left\{ \text{sequences } \{ (c_{j,t}, l_{j,t})_{j=1,2}, k_t \}_{t=0}^{\infty} \text{ which are a CE} \right\} $$

for given initial conditions on capital. Elements of $\mathcal{S}^{CE}$ are characterized by (3), (7), (8), (10), and (11) for some λ.[13] It is clear that one can substitute out the sequences $\{ c_t^2, l_t^2 \}_{t=0}^{\infty}$ for a given value of λ so that these necessary and sufficient conditions can be expressed in terms of sequences $\{ c_t^1, l_t^1, k_t \}_{t=0}^{\infty}$ and a constant λ. Note that the ratio λ is so far unknown, it has to be determined in equilibrium, 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.

---

[12]Note that labor supply depends also on the distribution of consumption/wealth through λ. Under Gorman aggregation this would not be the case.

[13]As usual the government budget constraint can be ignored due to Walras' law.

2.3 The policy problem

Now we describe in detail the policy problem and some constraints that we introduce.

A Ramsey Pareto-optimal (RPO) allocation is an element of $\mathcal{S}^{CE}$ such that the utility of one or more agents cannot be improved within the set $\mathcal{S}^{CE}$. We now formulate an RPO allocation 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 $\underline{U}_2$, i.e.,

$$\sum_{t=0}^{\infty} \beta^t \left[u\left(c_{2,t}\right) + v\left(l_{2,t}\right) \right] \geq \underline{U}_2, \tag{12}$$

where $\underline{U}_2$ is restricted to belong to the set of utilities that can be attained for agent 2 within $\mathcal{S}^{CE}$. Varying the value of the minimum utility $\underline{U}_2$ along all possible utilities that consumer 2 can attain in $\mathcal{S}^{CE}$, we can trace out the whole set of RPO allocations.

We assume the planner faces a *tax limit*, denoted $\widetilde{\tau}$, and impose $\tau_t^k \leq \widetilde{\tau}$ for all $t = 0, 1, ...$ for a constant $0 < \widetilde{\tau} < 1$ 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

$$u'\left(c_{1,t}\right) \geq \beta u'\left(c_{1,t+1}\right)\left(1 + \left(r_{t+1} - \delta\right)\left(1 - \widetilde{\tau}\right)\right), \quad \forall t > 0, \tag{13}$$

$$\tau_0^k \leq \widetilde{\tau}. \tag{14}$$

The first equation ensures that the actual capital tax τ_t^k, $t = 1, 2, ...$, implied by (4) satisfies the limit, and it allows us to use the primal approach, where taxes at $t = 1, 2, ...$ do not appear explicitly in the government's problem. The motovation for this tax limit has been discussed above.

Finally, we introduce a constraint on the admissible consumptions in the planner's problem. Lansing (1999), Bassetto and Benhabib (2006), Reinhorn (2014), and Straub and Werning (2015) show that in many models the optimal policy entails $c_{j,t} \to 0$ as $t \to \infty$ and $\tau_\infty^k \neq 0$. These papers include models very similar to ours, specially that of Bassetto and Benhabib (2006). Since our aim is 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 immiserate all consumers in the future. That is, we add the constraint $c_{j,t} \geq \widetilde{c}_j$, $\forall j$, $\forall t$, for some $\widetilde{c}_j > 0$. Given (7) these constraints can be substituted, without loss of generality, by the following assumption:

A2.

$$c_{1,t} \geq \widetilde{c}, \ \forall t, \tag{15}$$

for some given $\widetilde{c} > 0$.

That is, the planner is constrained to policies where consumption is uniformly bounded away from zero. The interpretation is that the government cannot credibly commit today to impose policies that immiserate future generations, either because no amount of commitment will bind future governments in that way, or because of moral concerns about how to treat future generations.

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

$$\max_{\tau_0^k, \lambda, \left\{c_t^1, k_t, l_t^1\right\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left[u\left(c_{1,t}\right) + v\left(l_{1,t}\right)\right]$$

$$\text{s.t.} \quad \sum_{t=0}^{\infty} \beta^t \left[u\left(\lambda c_{1,t}\right) + v\left(\mathcal{K}(\lambda)l_{1,t}\right)\right] \geq \underline{U}_2, \tag{16}$$

for a given level of utility $\underline{U}_2$ and subject to feasibility (3), implementability (10) and (11), tax limits (13) and (14), and consumption limits (15). We have used (7) and (8) to substitute for c_2 and l_2 to obtain (16). $\underline{U}_2$ 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 λ appears as an argument in the optimization problem. Formulating the maximization problem as a function of λ 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. Using our approach the number of variables to solve for is essentially the same as in a homogeneous-agents model.[14]

We concentrate our attention on those RPO allocations which are also Pareto improving relative to a benchmark CE allocation, the status quo, where taxes are set as in the past. We call these 'POPI' allocations. Let the utilities attained by agent j at the status quo be U_j^{SQ}.[15] POPI allocations can be found by considering only minimum utility values $\underline{U}_2$ such

---

[14]In the Online Appendix we discuss in detail the formulation of a Lagrangian as used by Atkeson, Chari, and Kehoe (1999). This has been used in many papers, for example Flodén (2009) and Bassetto (2014). The Lagrangian of ACK, of course, gives rise to valid first-order conditions, but a direct application of that approach leads to a much larger computational problem than our approach. An early version of Bassetto (2014) used a similar characterization to ours.

[15]The status-quo utilities depend on $k_{1,-1}$ and $k_{2,-1}$ in general. We leave this dependence implicit.

that $\underline{U}_2 \geq U_2^{SQ}$ and such that the maximum satisfies

$$\sum_{t=0}^{\infty} \beta^t \left[u\left(c_{1,t}^*\right) + v\left(l_{1,t}^*\right) \right] \geq U_1^{SQ}, \tag{17}$$

where $*$ denotes the optimized value of each variable for a given $\underline{U}_2$. Notice that since we restrict the status quo to be a CE, the set of POPI allocations is not empty.

2.4 Optimality conditions

To solve the above planner's problem, we find the first-order conditions of an appropriate Lagrangian. Let ψ be the Lagrange multiplier of the minimum-utility constraint (16), let Δ_1 and Δ_2 be the multipliers of implementability constraints (10) and (11), respectively, and μ_t, γ_t, and ξ_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

$$\begin{aligned}
\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \Bigg\{ & u\left(c_{1,t}\right) + v\left(l_{1,t}\right) + \psi\left[u\left(\lambda c_{1,t}\right) + v\left(\mathcal{K}(\lambda)l_{1,t}\right)\right] \\
& + \xi_t(c_{1,t} - \widetilde{c}) + \Delta_1 \left[u'\left(c_{1,t}\right)c_{1,t} + v'\left(l_{1,t}\right)l_{1,t}\right] \\
& + \Delta_2 \left[u'\left(c_{1,t}\right)\lambda c_{1,t} + \frac{\phi_2}{\phi_1}v'\left(l_{1,t}\right)\mathcal{K}(\lambda)l_{1,t}\right] \\
& + \gamma_t \left[u'\left(c_{1,t}\right) - \beta u'\left(c_{1,t+1}\right)\left(1 + (r_{t+1} - \delta)(1 - \widetilde{\tau})\right)\right] \\
& + \mu_t \left[F\left(k_{t-1}, e_t\right) + (1 - \delta)k_{t-1} - k_t - \frac{1+\lambda}{2}c_{1,t} - g\right] \Bigg\} - \psi\underline{U}_2 - \mathbf{W},
\end{aligned} \tag{18}$$

where $\mathbf{W} = u'\left(c_{1,0}\right)\left(\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}\right)\left(1 + (r_0 - \delta)(1 - \tau_0^k)\right)$. Further, ξ_t, γ_t, $\mu_t \geq 0$, $\forall t$, and $\psi \geq 0$, with complementary slackness conditions.[16]

The first line of this Lagrangian has the usual interpretation: finding a Pareto-efficient allocation amounts to maximizing a welfare function where the planner weighs 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 ψ 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 $\tau_t^k \leq \widetilde{\tau}$ for all $t > 0$. The last line is the feasibility constraint. The term $\mathbf{W}$ collects the period-0 terms in the budget constraints of the consumers.

---

[16]Strictly speaking, in models of taxation it is not impossible that $\mu_t < 0$ if taxes are very negative. We contend that this is not the case in our model. In any case, $\mu_t \geq 0$ can always be guaranteed if the government can choose g_t endogenously with $g_t \geq g$.

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 the FOCs derived from this Lagrangian. 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 ψ, the multiplier of (16), we have that

$$\text{either} \quad \psi > 0 \quad \text{and} \quad \sum_{t=0}^{\infty} \beta^t \left[u\left(c_{2,t}\right) + v\left(l_{2,t}\right) \right] = \underline{U}_2,$$

$$\text{or} \quad \psi = 0 \quad \text{and} \quad \sum_{t=0}^{\infty} \beta^t \left[u\left(c_{2,t}\right) + v\left(l_{2,t}\right) \right] \geq \underline{U}_2.$$

In other words, the minimum-utility constraint may or may not be binding. In the first case, the Lagrangian amounts to maximizing (subject to equilibrium constraints) the weighted sum of utilities of consumers 1 and 2 with weights 1 and ψ, 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 $\underline{U}_2$ 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. Key to our approach is the fact that this does not hold and that the relative consumption of consumers (λ) has to be chosen optimally. Hence we set the

derivative of $\mathcal{L}$ with respect to λ equal to zero to obtain the optimality condition

$$
\begin{aligned}
\sum_{t=0}^{\infty} \beta^t &\Big\{ \psi \left[u' \left(\lambda c_{1,t} \right) c_{1,t} + v' \left(\mathcal{K}(\lambda) l_{1,t} \right) \; \mathcal{K}'(\lambda) l_{1,t} \right] \\
&+ \Delta_2 \left(u' \left(c_{1,t} \right) c_{1,t} + \frac{\phi_2}{\phi_1} v' \left(l_{1,t} \right) \; \mathcal{K}'(\lambda) l_{1,t} \right) \\
&- \gamma_{t-1} u' \left(c_{1,t} \right) F_{ke} \left(k_{t-1}, e_t \right) \frac{\phi_2}{2} \; \mathcal{K}'(\lambda) l_{1,t} \left(1 - \widetilde{\tau} \right) - \frac{\mu_t}{2} \left(c_{1,t} - F_e \left(k_{t-1}, e_t \right) \phi_2 \; \mathcal{K}'(\lambda) l_{1,t} \right) \Big\} \\
&- u' \left(c_{1,0} \right) \left(\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1} \right) F_{ke} \left(k_{-1}, e_0 \right) \frac{\phi_2}{2} \; \mathcal{K}'(\lambda) l_{1,0} \left(1 - \tau_0^k \right) = 0.
\end{aligned}
\tag{19}
$$

The fact that λ 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 γ_t for each t we have that

$$
\begin{aligned}
&\text{either} \quad \gamma_t > 0 \quad \text{and} \quad u' \left(c_{1,t} \right) = \beta u' \left(c_{1,t+1} \right) \left(1 + \left(r_{t+1} - \delta \right) \left(1 - \widetilde{\tau} \right) \right), \\
&\text{or} \quad \gamma_t = 0 \quad \text{and} \quad u' \left(c_{1,t} \right) \geq \beta u' \left(c_{1,t+1} \right) \left(1 + \left(r_{t+1} - \delta \right) \left(1 - \widetilde{\tau} \right) \right).
\end{aligned}
$$

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

It turns out that the Δ_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 $\frac{\partial \mathcal{L}}{\partial \tau_0^k} = u' \left(c_{1,0} \right) \left(\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1} \right) \left(r_0 - \delta \right)$. 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 Δ_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 $T_1 = -T_2$. The planner in this case still needs to raise some revenue with proportional taxes, but the lump-sum transfers allow for non-distortionary redistribution.

All this modification does to the Lagrangian is that it changes the implementability constraints, in particular, the term $u' \left(c_{1,0} \right) \left(\Delta_1 - \Delta_2 \right) T_1$ is added to $\mathbf{W}$. Now the derivative of the Lagrangian with respect to the lump-sum transfer between consumers is $\frac{\partial \mathcal{L}}{\partial T_1} = u' \left(c_{1,0} \right) \left(\Delta_1 - \Delta_2 \right)$. 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

14

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 (1), then $\Delta_1 - \Delta_2 > 0$ ($<$), and vice versa. If the optimal transfer would be large, Δ_2 (Δ_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, by the existence of an equilibrium status-quo allocation where the tax limit $\widetilde{\tau}$ is not violated.

3   Characterization of equilibria

In this section we describe some analytical results.

3.1   Zero capital taxes in the long run

We now prove that $\tau_\infty^k = 0$ and that capital taxes jump from the tax limit to zero in two periods. This result is of independent interest for various reasons.

Most available proofs of $\tau_\infty^k = 0$ assume that Lagrange multipliers of the Ramsey problem have a finite steady state. But Lagrange multipliers should not be constrained to have a bounded steady state if the Lagrangian is to characterize the maximum. Lansing (1999), Bassetto and Benhabib (2006) (BB hereafter), Reinhorn (2014), and Straub and Werning (2015) show many examples where the multipliers of the feasibility constraint, μ_t, go to infinity and $\tau_\infty^k \neq 0$. Therefore, previous proofs of $\tau_\infty^k = 0$ assuming Lagrange multipliers have a finite steady state should be revised. We dub 'LBBRSW problem' the situation where $\tau_\infty^k \neq 0$ because $\mu_t \to \infty$. In line with Straub and Werning (2015), the proof below takes for granted the existence of a steady state in allocations but not in multipliers. This is a reasonable way to proceed, because real variables have natural bounds. Therefore, existence of a finite steady state can be expected for real variables. But a proper proof cannot restrict multipliers to be unbounded.

We show that a small modification of the optimal-taxation problem recovers $\tau_\infty^k = 0$. This is so even if Bassetto and Benhabib (2006) and Straub and Werning (2015) find the opposite for heterogeneous-agents models very similar to ours.[17]   Our modification is **A**2, i.e., that

---

[17]The following is a list of the differences between our model and BB's. In their model (i) the government

consumption is uniformly bounded away from zero.

This result is of interest for methodological reasons, as it shows a natural way to avoid the LBBRSW problem. It is also important for our application because we want to study the role of the transition in order to redistribute wealth, so we want to avoid cases where $\tau_\infty^k > 0$ and hence there is less need to redistribute along the transition. Finally, this result makes our computations even easier, since the correct solution for τ^k in the long run can be introduced upfront.

We use the following further assumptions:

A3. $0 < \widetilde{\tau} < 1$.

A4. *Allocations have a finite limiting steady state in the Ramsey equilibrium, namely* $(c_{1,t}, k_t, e_t) \to (c^{ss}, k^{ss}, e^{ss}) < \infty$ *as* $t \to \infty$.

A5. $\Delta_1 > 0$.

A3 has been discussed around equation (13). **A**4 is also used in Straub and Werning (2015). **A**5 is a natural assumption, essentially guaranteeing that the taxation problem is interesting. It is only violated if reducing distortive taxes hurts both agents. This rules out, for example, a case where the government has 'too much' savings at the beginning of time so that it has to impose negative proportional taxes on agents. Formally, it requires that there is one agent (say agent 1 without loss of generality) whose utility increases if he is given a lump-sum transfer.

The first part of the proposition states that $\tau_\infty^k = 0$ in our model. The second part shows that if $\widetilde{c} > 0$ is small enough for (15) not to bind in the optimum, then τ_t^k jumps from the tax limit to zero in two periods.

Proposition 1. *Assume* ***A***1-***A***5 *and that a Ramsey-Pareto-optimal allocation exists.*[18] *Then*

$$\tau_t^k \to 0 \quad as \quad t \to \infty.$$

chooses an optimal policy from the point of view of the median voter, which is just one specific Pareto-improving allocation, (ii) labor is fixed, (iii) the government can levy lump-sum taxes common on all agents, (iv) the production function is linear, (v) there is no strictly positive lower bound on consumption, i.e., $\widetilde{c} = 0$. Features (i), (ii), and (iii) are not crucial for the difference in the result: (i) amounts to $\psi = 0$ in our case, (ii) amounts to a very large σ_l, and we introduce (iii) in Section 4.4.1. But notice that in the proof below we use strict concavity of the production function and $\widetilde{c} > 0$ in various steps. Therefore, assumptions (iv) and (v) are key to positive taxes in BB. Appendix E gives a detailed proof that if one introduces strict concavity and $\widetilde{c} > 0$ in the model of BB, then $\tau_\infty^k = 0$, i.e., the LBBRSW problem disappears. The proof in that case may also be useful as it is simpler than the one in the main text, due to the inelastic-labor-supply assumption of BB.

[18]That the set of competitive equilibria is non-empty will be guaranteed typically by the fact that we choose parameter values for which a status-quo equilibrium exists. Continuity of the objective function and boundedness of feasible allocations ensure that a maximum of the planner's problem exists.

Furthermore, consider the case where $c^{ss} > \tilde{c}$. Clearly, $\tau_t^k < \tilde{\tau}$ and $c_{1,t} > \tilde{c}$ for t large enough. Let $N < \infty$ be the smallest integer such that

$$\tau_t^k < \tilde{\tau} \ \ and \ \ c_{1,t+1} > \tilde{c}, \ \ \forall t \geq N.$$

Then

$$\tau_t^k = 0, \ \ \forall t \geq N + 1. \tag{20}$$

Proof. It would be trivial to show that a solution with the proposed features satisfies the FOCs of the planner's problem. But this approach would not provide a formal proof, because the feasible set is non-convex and the FOCs are not sufficient for a maximum. As we want to address issues of non-convexity carefully, we need to argue that the properties in the statement of the proposition hold for any allocation that satisfies the FOCs.

All limits in this proof are taken as $t \to \infty$.

To prove $\tau_t^k \to 0$ we recast the planner's problem by eliminating one variable (λ) and one constraint (11). In particular, we can rewrite (11) in a more compact form as

$$\lambda \mathbf{C} + \mathcal{K}(\lambda)\mathbf{L} = \mathbf{K}, \tag{21}$$

where $\mathbf{C} \equiv \sum_{t=0}^{\infty} \beta^t u'(c_{1,t}) c_{1,t}$, $\mathbf{L} \equiv \frac{\phi_2}{\phi_1} \sum_{t=0}^{\infty} \beta^t v'(l_{1,t}) l_{1,t}$, and $\mathbf{K} \equiv u'(c_{1,0}) k_{2,-1} \left(1 + (r_0 - \delta)\left(1 - \tau_0^k\right)\right)$. Note that $\mathbf{C}$, $\mathbf{L}$, and $\mathbf{K}$ depend on the solution but the function $\mathcal{K}(\cdot)$ does not.

Let

$$\lambda = g(\mathbf{C}, \mathbf{L}, \mathbf{K}) \tag{22}$$

define the value of λ which solves (21). Intuitively, g gives the distributional parameter λ that balances out the budget constraint of agent 2 given the allocations of agent 1, $\mathcal{K}(\cdot)$ and $k_{2,-1}$. Such a λ exists for any $\mathbf{C}, \mathbf{L}, \mathbf{K}$ that are defined using equilibrium allocations. We can reformulate the planner's problem substituting out λ in the objective function of the planner and in the production function using $g(\mathbf{C}, \mathbf{L}, \mathbf{K})$. Then the budget constraint of agent 2 is guaranteed to hold, and we can drop equation (11) from the problem. See Appendix B for the reformulated planner's problem, its Lagrangian, and the FOC for labor with the utility function (6).[19]

The Euler equation of the consumer implies

$$1 - \left[\frac{u'(c_t)}{u'(c_{t+1})\beta} - 1\right] \frac{1}{F_k(k_t, e_{t+1}) - \delta} = \tau_{t+1}^k \leq \tilde{\tau} < 1.$$

---

[19]The remaining FOCs are either unchanged compared to our main formulation or are not needed in this proof.

Since consumption is bounded below by $\tilde{c} > 0$, we have $\infty > u'(c^{ss}) > 0$.[20] Then $\frac{u'(c_t)}{u'(c_{t+1})} \to 1$. A familiar argument in growth theory implies that

$$F_k(k^{ss}, e^{ss}) > \delta. \tag{23}$$

Then the equality above implies that capital taxes have a limit, $\tau_t^k \to \tau^{k,ss}$.

Assume, towards a contradiction, that $\tau^{k,ss} > 0$. The fact that

$$\beta[1 + (F_k(k^{ss}, e^{ss}) - \delta)(1 - \tau^{k,ss})] = 1 \tag{24}$$

and (23) give $\beta[1 + F_k(k^{ss}, e^{ss}) - \delta] > 1$, so that we can pick a constant $\mathbf{A}$ such that

$$1 > \mathbf{A} > \frac{1}{\beta[1 + F_k(k^{ss}, e^{ss}) - \delta]}.$$

Obviously,

$$\mathbf{A} > \frac{1}{\beta(1 - \delta + F_k(k_t, e_{t+1}))} \text{ for } t \text{ large enough.} \tag{25}$$

We can write the planner's FOC for capital (see Appendix B) as

$$\mu_t \frac{1}{\beta(1 - \delta + F_k(k_t, e_{t+1}))} + \gamma_t \frac{u'(c_{1,t+1}) F_{kk}(k_t, e_{t+1})(1 - \tilde{\tau})}{1 - \delta + F_k(k_t, e_{t+1})} = \mu_{t+1}. \tag{26}$$

We have $F_{kk}(k, e) \le 0$ by concavity and $\gamma_t \ge 0$, hence the second term on the left-hand side is non-positive. This, together with $\mu_t \ge 0$ and (25), implies that for t large enough

$$\mu_t \mathbf{A} \ge \mu_{t+1}.$$

Since $\mathbf{A} < 1$ and $\mu_t \ge 0$ we have that $\mu_t \to 0$. Plugging $\mu_t \to 0$ in (43) gives

$$\gamma_t \frac{u'(c_{1,t+1}) F_{kk}(k_t, e_{t+1})(1 - \tilde{\tau})}{1 - \delta + F_k(k_t, e_{t+1})} \to 0. \tag{27}$$

Now we show that the term multiplying γ_t in (27) cannot go to zero. First, notice that (24) and $\tilde{\tau} < 1$ implies $F_k(k^{ss}, e^{ss}) < \infty$. We also need $F_{kk}(k^{ss}, e^{ss}) < 0$. Note that even if F is strictly concave $F_{kk}(k^{ss}, 0) = 0$, but it follows from $F(k, 0) = 0$, $k^{ss} < \infty$, $c^{ss} \ge \tilde{c} > 0$ and feasibility that $e^{ss} > 0$. Therefore strict concavity of F gives $F_{kk}(k^{ss}, e^{ss}) < 0$. Then, existence of a finite limit c^{ss} gives

$$\frac{u'(c_{1,t+1}) F_{kk}(k_t, e_{t+1})(1 - \tilde{\tau})}{1 - \delta + F_k(k_t, e_{t+1})} \to u'(c^{ss}) \frac{F_{kk}(k^{ss}, e^{ss})(1 - \tilde{\tau})}{1 - \delta + F_k(k^{ss}, e^{ss})} < 0.$$

---

[20]A key ingredient of the non-zero capital tax results in LBBRSW is that $u'(c_t) \to \infty$ in the optimal solution. We rule out this possibility by the consumption limit (15).

Therefore, (27) implies

$$\gamma_t \to 0. \tag{28}$$

From (36) in Appendix B we have $\mathbf{L}_t \to 0$. Since $g_{\mathbf{L}}$ does not depend on t we have $\mathbf{L}_t g_{\mathbf{L}} \to 0$. Plugging this and the limits $\mu_t, \gamma_t \to 0$ in the FOC (35) in Appendix B we have

$$-\omega \left(l_1^{ss}\right)^{\sigma_l} \left[1 + \psi\mathcal{K}(\lambda)^{\sigma_l+1} + \Delta_1 \left(1 + \sigma_l\right)\right] = 0, \tag{29}$$

which is impossible, since $\psi, \mathcal{K}(\lambda), \Delta_1$, and σ_l are all strictly positive and $e^{ss} > 0$ implies $l_1^{ss} > 0$.

This shows that $\tau_t^k \to \tau^{k,ss} > 0$ leads to a contradiction under our assumptions. Since negative taxes are ruled out this proves that

$$\tau_t^k \to 0. \tag{30}$$

Now we prove the second part of the proposition. For this purpose it is more convenient to consider again the planner's problem without substituting for λ that satisfies (11), as we do in the main text and Appendix A. The Lagrange multipliers in the rest of this proof correspond to that formulation as well, and they do not take the same values as the analog multipliers of the problem of Appendix B used in the first part of this proof. However, we abuse notation and keep the same symbols. The allocations are unchanged, since the two problems give rise to the same optimal allocation.

Given the way N is chosen it is clear that the tax limit is not binding for $t \geq N$ and the consumption limit is not binding for $t \geq N + 1$. Therefore $\gamma_t = \xi_{t+1} = 0$ for all $t \geq N$, the FOC with respect to consumption for $t \geq N$ (see Appendix A) with the utility function (6) gives[21]

$$(c_{1,t})^{-\sigma_c} \left[1 + \psi\lambda^{1-\sigma_c} + (\Delta_1 + \lambda\Delta_2)(1 - \sigma_c)\right] = \mu_t \frac{1+\lambda}{2}, \quad \forall t \geq N+1. \tag{31}$$

Plugging (31) into the FOC with respect to capital and using $\gamma_t = 0$ again, we get

$$(c_{1,t})^{-\sigma_c} = \beta \left(c_{1,t+1}\right)^{-\sigma_c} \left(1 - \delta + F_k\left(k_t, e_{t+1}\right)\right), \quad \forall t \geq N+1.$$

It is clear that this equation is only compatible with the Euler equation of the consumer (4) if (20) holds. □

This argument falls short of providing a formal proof that such a sufficiently low $\widetilde{c} > 0$ exists, guaranteeing that (15) does not bind. At this writing we do not know if this is true

---

[21]This last equation does not hold for $t = N$, because $\gamma_{N-1} \neq 0$ appears in the FOC for consumption at $t = N$.

or not. But the result is sufficient to facilitate considerably finding a numerical solution: it allows one to compute the infinite 'tail' of the sequence imposing (20) until we find N, then compute the transition with a binding (13) for $t < N$, checking that all Lagrange multipliers have the correct sign, and taking a $\widetilde{c}$ sufficiently small. If such an allocation can be found it must be the optimal solution. This occured in all the equilibria we computed.

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 $\mathcal{S}^{CE}$ is not convex. In this case a Lagrangian approach is 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. If the duality gap is non-empty we are bound to ignore some relevant RPO allocations as we trace out the entire Ramsey Pareto frontier. To be precise, let the feasible set of utilities

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

and let $\mathcal{F}$ be the boundary (or 'frontier') of $\mathcal{S}^U$. In the standard case without distortions and a concave utility function, it is well known that $\mathcal{F}$ corresponds to the RPO allocations, and it defines U_1 as a decreasing and concave function of U_2. In that case an allocation is Pareto optimal if and only if it optimizes a welfare function with fixed weights for consumers. But if $\mathcal{S}^U$ is not convex, its frontier may have a non-concave part, and the equilibria with utilities in that non-concave part can not be found by maximizing a welfare function with some 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 below where labor supply is fixed. For all these reasons we now show a sufficient condition guaranteeing that, despite the non-convexities, in our model we can compute all RPO equilibria using a welfare function. We will see how this condition can be checked numerically in our application.

Fix $\psi \in [-\infty, \infty]$, and consider the following modified model (MM):

$$\max_{\tau_0^k, \lambda, \{c_t^1, k_t, l_t^1\}_{t=0}^\infty} \sum_{t=0}^\infty \beta^t \{u(c_{1,t}) + v(l_{1,t}) + \psi[u(\lambda c_{1,t}) + v(\mathcal{K}(\lambda)\, l_{1,t})]\}, \qquad (32)$$

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

A6. *There is a unique solution to MM for all $\psi \geq 0$. Furthermore, $U_2(\cdot)$ is invertible on $[0, \infty]$.*

Proposition 2. *Assume **A6**.*

1. *A solution to MM for any $\psi \in [0, \infty]$ is a Ramsey-Pareto-optimal allocation.*

2. *Every Ramsey-Pareto-optimal allocation is also the solution of MM for some $\psi \in [0, \infty]$.*

3. *Given $\psi \in [-\infty, \infty]$, if the solution of MM exists, it defines a point on the frontier, i.e., $(U_1(\psi), U_2(\psi)) \in \mathcal{F}$.*

Proof. The proof of part 1 is obvious, it is only stated for future reference. For part 2, consider a pair of utilities $(\overline{U}_1, \overline{U}_2) \in \mathcal{S}^U$ that correspond to a RPO allocation. Invertibility in **A6** guarantees that there is a $\bar{\psi}$ such that $\overline{U}_2 = U_2(\bar{\psi})$. Consider a finite $\bar{\psi}$. We have

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

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

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

Let us now consider Part 3. If $\psi \geq 0$ Part 3 follows from Part 2. Consider now a given $\psi < 0$. We can find points in $\mathbb{R}^2$ outside $\mathcal{S}^U$ which are arbitrarily close to $(U_1(\psi), U_2(\psi))$ as follows: for any $\varepsilon > 0$ we have $(U_1(\psi) + \varepsilon, U_2(\psi) - \varepsilon) \notin \mathcal{S}^U$, since this point achieves a higher

value of the objective function of MM 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 on the frontier $\mathcal{F}$. □

Part 2 of Proposition 2 implies that we can find *all* RPO allocations by solving MM varying ψ from zero to infinity. Part 3 guarantees that we may obtain additional points on the frontier $\mathcal{F}$ using a negative ψ, as long as a maximum of MM exists[22] for this $\psi < 0$. These frontier 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 MM, that is, if ψ multiplies the utility of consumer 1 and we take $\psi < 0$. In section 4.2 we use Part 3 to find an increasing part of the frontier $\mathcal{F}$ which is not Pareto optimal.

Since the feasible set is non-convex, **A6** may not hold for some parameterizations. But it can be checked numerically whether it holds in a given application. To verify uniqueness, we search for more solutions to the FOCs, as is done in scores of papers where the maximum is found by searching numerically for all critical points.[23] To verify invertibility of $U_2(\cdot)$, we record all utilities for a fine grid of ψ'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. Nonetheless, 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 ψs such that $(U_1(\psi), U_2(\psi))$ are larger than the status-quo utilities of consumer 1 and 2, respectively.

4   Numerical results

We now present and discuss our numerical results. Details on our computational strategy are in Appendix C. In the next subsection, we discuss how we calibrate the model. Afterwards, we first analyze the case where labor supply is fixed. Then, in Section 4.3, we introduce flexible labor supply. In order to gain intuition for the forces at work and to study progressive taxation, finally we extend the model to lump-sum transfers, equal across agents, and we

---

[22]Notice that if we had a standard model without distortions and $u(0) = -\infty$, then MM with $\psi < 0$ would not have a solution. In that case Part 3 would, of course, not apply and it would not define a point on the frontier of utilities.

[23]For example, almost all the papers in econometrics using maximum likelihood, or all the papers solving dynamic models under rational expectations by approximation of the first-order conditions in non-convex maximization problems, show a "solution" which is only guaranteed to be a maximum if one searches for all possible critical points.

also study issues of time consistency.

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 so that if taxes and initial government debt are matched to the US average effective tax rates and debt/GDP ratio, the status-quo equilibrium matches certain moments in the US economy. The macro variables are taken from the dataset provided by Trabandt and Uhlig (2012), who collected data from the OECD and other sources.[24] We compute averages for the period 2001-2010, including of effective tax rates at the status quo. The average effective tax rates are: $\tau^l = 0.214$ and $\tau^k = 0.401$. Note that the choice of tax rates at the status quo matters in several ways. Firstly, 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 suppose that during the reform the capital tax rate can never increase above its initial level, which is equal to status-quo rate by assumption, i.e., we set $\widetilde{\tau} = 0.401$.

Table 1: Parameter values of the baseline economy

	Parameter	Value
Preference parameters	β	0.96
	σ_c	1
	σ_l	3
	ω	221.1
Heterogeneity parameters	ϕ_w/ϕ_c	0.91
	$k_{c,-1}$	3.831
	$k_{w,-1}$	-0.611
Production parameters	α	0.394
	δ	0.074
Public sector	g	0.094
	k^g_{-1}	-0.315
	τ^l	0.214
	τ^k	0.401

We set some preference parameters a priori. The utility function is as stated in Assumption A1. We set the annual discount factor to the commonly-used value $\beta = 0.96$. We choose $\sigma_c = 1$ in keeping with a large part of the literature on taxation. The choice of $\sigma_l = 3$ is for

---

[24]https://https://sites.google.com/site/mathiastrabandt/home/downloads/LafferNberDataMatlabCode.zip

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.[25] Note that this implies a lower Frisch elasticity of labor supply than in many applications of the 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 ϕ_j and their initial wealth $k_{j,-1}$. Garcia-Milà, Marcet, and Ventura (2010) argue that the relevant aspect of heterogeneity when studying optimal proportional labor and capital income taxation is agents' wage-wealth ratio. 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 $\phi_w/\phi_c = 0.91$ to match the observed ratio of labor earnings, and $\lambda = c_w/c_c = 0.54$ to match the ratio of consumptions of the two groups in the data.

Finally, we find ω, δ, g, k^g_{-1}, and the initial wealth of each group in the model, $k_{c,-1}$ and $k_{w,-1}$, (along with the steady-state values of all endogenous variables) so that at the status-quo steady state, given the status-quo tax rates, the CE conditions hold, and (i) aggregate hours match the fraction of time worked for the working age population, 0.245,[26] (ii) the consumption ratio, λ, is as in the data, (iii) g over output equals government consumption over GDP (iv) that k^g_{-1} over output has to match the public assets-GDP ratio from the data, -66.8 percent of GDP.

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. All parameters unrelated to the utility from leisure are as in Table 1.

---

[25]See Garcia-Milà, Marcet, and Ventura (2010) for a discussion of the trade-offs in choosing σ_l.
[26]Hours to be allocated between work and leisure: 13.64 (Trabandt and Uhlig, 2012).

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 heterogenous agents, and 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) it could implement any Pareto optimal allocation in the first best. But in the case of interest of this paper, where lump-sum redistribution is not possible (hence $T_w = T_c = 0$), 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.[27] The dashed black line labeled 'first-best PI' represents optimal allocations with $\tau_t^k = 0$ for all t where agent-specific lump-sum transfers $T_w = -T_c$ 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 dot-dashed (green) line, are not Pareto optimal, while the POPI allocations coincide with the decreasing part of $\mathcal{F}$, depicted with a solid (blue) line. Using Proposition 2 part 2, the decreasing part of $\mathcal{F}$ is found with $\psi > 0$ in MM, higher ψ 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, ie $\tau_t^k = \widetilde{\tau}$, for 24 periods.

The points along the dot-dashed (green) line (the increasing part of $\mathcal{F}$) imply an even longer period of high capital taxes. These points are found with $\psi < 0$ according to Proposition 2 part 3. The economy is so inefficient along this line that both agents' stance is worse than at the point w^{max}.

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 $\tau_t^k = 0$ for all t (as in the first best with homogenous agents) and $T_w = T_c = 0$, the worker would be worse off than at the status quo, as has been shown previously in a number of contributions.[28] Such a point is not shown in Figure 1, because it corresponds to negative values of the horizontal axis and, therefore, it is outside the picture. All the Pareto-improving first best

---

[27]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.

[28]See Correia (1999), Domeij and Heathcote (2004), Conesa and Krueger (2006), and Garcia-Milà, Marcet, and Ventura (2010).

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 in favor of the capitalist 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 $N + 1$ 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.[29]

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 $w^{\max}$), 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 the capitalist stands much more to gain even if we only look at Pareto-improving reforms.

Finally, note that 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, therefore it achieves a Pareto improvement. This is not the case in all calibrations, see for example Section 4.3.2.

4.3   Main results

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

---

[29]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 The 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 $\mathcal{F}$, 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 σ_l, and correspondingly close-to-inelastic labor supply, the picture would start resembling Figure 1.

To numerically verify that Assumption **A6** (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, therefore MM fully characterizes all RPO solutions.

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 Proposition 1. Note that we will see in the next subsection that consumption grows after period N, thus providing a check that (15) holds, hence the second part of Proposition 1 is applicable for a sufficiently low $\widetilde{c}$. 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.[30] 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 ψ, the multiplier on the minimum-utility constraint (16) (or, equivalently, the relative Pareto weight of the worker in MM), and λ, 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 ψ) 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 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 ψ, and the share of capital taxes would always be 10.3 percent. The multiplier ψ would increase very little with U_w, while λ would rise much more than without transfers.[31] 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 D 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 model

---

[30]For comparison, the share of capital taxes in revenues is about 37.1 percent at the status quo.

[31]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.

considering both a lower and a higher value for each of the three data moments. In all these cases the results are similar to the ones for the benchmark calibration.

4.3.2 Interpreting the welfare weight ψ

In the literature on optimal policy with heterogeneous agents, it is customary to fix certain agents' weight ψ and to provide an interpretation for this choice. Some papers interpret ψ as due to 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 ψ is just the Lagrange multiplier of the promise-keeping constraint (16), and its value is determined in equilibrium. This provides a very different view for the role of ψ. From this point of view, there is no reason that the case $\psi = 1$ should reflect an equitable reform.

To be precise, we dub 'equitable reform' a RPO solution which implies that both agents gain more or less equally.[32] Graphically, equitable reforms are points on the frontiers of Figures 1, 2, and 6 which are near the 45° line. Our calculations show that even within our model but for different parameter values, equitable reforms can imply very different values of ψ. Furthermore, the reform corresponding to $\psi = 1$ can be very far from equitable.

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

In Appendix D we recalibrate the heterogeneity parameters ϕ_j and $k_{j,-1}$ to match the top and bottom quintiles of the wage-wealth distribution. In this case the wage-wealth ratios are much more different across agents than in our benchmark calibration based on the top and bottom half. The POPI frontier is shown in Figure 6. Figure 6 shows that $\psi = 1$ is not even Pareto improving for this calibration 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, to similar Nash bargaining power of both agents) only by chance. The value of ψ that achieves an equitable reform

---

[32]Such a reform could be the outcome of Nash bargaining game by agents' representatives played at $t = 0$ about which reform to implement when both agents' representatives have a similar bargaining power and the outside option is the status quo.

differs strongly depending on the economy and, in particular, 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: ψ 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.

Some 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 lowered initially, and they should remain low for a long time.[33] 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, 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. Note that the accumulation of capital accelerates around the period when capital

---

[33]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.

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.

In summary, in RPO with heterogeneous agents there are two reasons to lower labor taxes: first, to promote investment, second to redistribute wealth so as to achieve a Pareto improvement.

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.[34]

---

[34]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

31

Since government expenditures are constant, low initial labor taxes translate into government 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 $\mathcal{D}$ which is made 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 $\mathcal{D}$ means progressive taxation.

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

We find that if we restrict our attention to a non-negative $\mathcal{D}$ (progressive taxation), the

---

the policy, as long as capital taxes are zero eventually.

optimal choice is to set $\mathcal{D} = 0$. 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 $\mathcal{D}$. On the one hand, distributive concerns would advise the government to choose a positive $\mathcal{D}$, since capitalists are richer. But a negative $\mathcal{D}$ 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 $\mathcal{D}$ 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 $\mathcal{D}$ 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.[35]

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, given that markets are complete, by the CE conditions (7) and (8), both relative consumption and relative leisure are constant over time. Therefore, it is not the case that workers lose dramatically when capital taxes finally drop to zero.

---

[35]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 more careful study of progressive taxation would introduce total income in the calibration. This is left for future research.

This is an implication of the permanent income hypothesis and, therefore, it relies strongly on the presence of complete markets. 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 of consumers 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 relative to the utility promised in period $t = 0$.

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 $\widetilde{\psi}$ and $\widetilde{\Delta}_j$, $j = 1, 2$, denote the appropriately-chosen, reoptimized values of the time-invariant multipliers. The time-variant multipliers μ_t and γ_t are rescaled by a factor $\frac{1+\widetilde{\psi}}{1+\psi}$. Moreover, we have the relationship $\gamma_{M-1} = \frac{1+\psi}{1+\widetilde{\psi}} \left(\widetilde{\Delta}_1 k_{1,M-1} + \widetilde{\Delta}_2 k_{2,M-1} \right)$. Inspection of the FOCs reveals that the remainder of the original optimal plan satisfies the FOCs of the reoptimization problem. Interestingly, $\widetilde{\psi}$ always turns out to be smaller than ψ. For instance, in the case of $\psi = 1$ and reoptimization in period $M = 5$, the continuation utilities are respected if $\widetilde{\psi} = 0.63$. 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.

We first show that the traditional result $\tau_\infty^k = 0$ holds as long as consumption is bounded away from zero. Our assumption simply says that the government cannot commit to immiserating future generations. As a result, we avoid some equilibria analyzed recently where $\tau_\infty^k \neq 0$ in the optimal allocation.

To achieve an optimal Pareto-improving policy, capital taxes should be high (and labor taxes very low) for a very long time after the reform starts. The government typically accumulates debt in order to finance the initial cut in labor taxes, and has a primary budget surplus in the long run to service its debt. Optimal policy calls for lower initial labor taxes during a long transition. Low 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 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 if 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 when the capital income tax rate is zero in the long run. Therefore, much is to be learnt from studying these issues, both from an empirical and a theoretical point of view. One avenue for research is to study other policy instruments which could 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 that need to be introduced in

the model to address issues of progressivity is 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

Aiyagari, S. R. (1995). Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting. *Journal of Political Economy 103*(6), 1158–75.

Armenter, R. (2004). Redistribution, Time-Consistent Fiscal Policy and Representative Democracy. Mimeo, Federal Reserve Bank of New York.

Atkeson, A., V. Chari, and P. Kehoe (1999). Taxing Capital Income: A Bad Idea. *Federal Reserve Bank of Minneapolis Quarterly Review 23*(3), 3–17.

Bassetto, M. (2014). Optimal Fiscal Policy with Heterogeneous Agents. *Quantitative Economics 5*(3), 675–704.

Bassetto, M. and J. Benhabib (2006). Redistribution, Taxes, and the Median Voter. *Review of Economic Dynamics 9*(2), 211–223.

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2013). Taxes, Debts, and Redistributions with Aggregate Shocks. NBER Working Papers 19470, National Bureau of Economic Research.

Chamley, C. (1986). Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. *Econometrica 54*, 607–622.

Chamley, C. (2001). Capital Income Taxation, Wealth Distribution and Borrowing Constraints. *Journal of Public Economics 79*(1), 55–69.

Chari, V. and P. Kehoe (1999). Optimal Fiscal and Monetary Policy. In J. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1, Chapter 26, pp. 1671–1745. Elsevier.

Conesa, J. C., S. Kitao, and D. Krueger (2009). Taxing Capital? Not a Bad Idea after All! *American Economic Review 99*(1), 25–48.

Conesa, J. C. and D. Krueger (2006). On the Optimal Progressivity of the Income Tax Code. *Journal of Monetary Economics 53*(7), 1425–1450.

Correia, I. (2010). Consumption Taxes and Redistribution. *American Economic Review 100*(4), 1673–1694.

Correia, I. H. (1999). On the Efficiency and Equity Trade-off. *Journal of Monetary Economics 44*(3), 581–603.

Domeij, D. and J. Heathcote (2004). On the Distributional Effects of Reducing Capital Taxes. *International Economic Review 45*(2), 523–554.

Flodén, M. (2009). Why Are Capital Income Taxes So High? *Macroeconomic Dynamics 13*(3), 279–304.

Garcia-Milà, T., A. Marcet, and E. Ventura (2010). Supply Side Interventions and Redistribution. *Economic Journal 120*(543), 105–130.

Giannitsarou, C. (2006). Supply-side Reforms and Learning Dynamics. *Journal of Monetary Economics 53*(2), 291–309.

Jones, L. E., R. E. Manuelli, and P. E. Rossi (1993). Optimal Taxation in Models of Endogenous Growth. *Journal of Political Economy 101*(3), 485–517.

Judd, K. (1985). Redistributive Taxation in a Simple Perfect-Foresight Model. *Journal of Public Economics 28*(1), 59–83.

Lansing, K. J. (1999). Optimal Redistributive Capital Taxation in a Neoclassical Growth Model. *Journal of Public Economics 73*(3), 423–453.

Lau, L., Y. Qian, and G. Roland (2001). Reform without Losers: An Interpretation of China's Dual-Track Approach to Transition. *Journal of Political Economy 108*(1), 120–143.

Ljungqvist, L. and T. J. Sargent (2012). *Recursive Macroeconomic Theory*. Third edition, MIT Press, Cambridge, Massachusetts.

Lucas, R. (1990). Supply Side Economics: An Analytical Review. *Oxford Economic Papers 42*, 293–316.

Marcet, A. and R. Marimon (2011). Recursive Contracts. Mimeo.

Marcet, A., F. Obiols-Homs, and P. Weil (2007). Incomplete Markets, Labor Supply and Capital Accumulation. *Journal of Monetary Economics 54*(8), 2621–2635.

Niepelt, D. (2004). Tax Smoothing versus Tax Shifting. *Review of Economic Dynamics 7*(1), 27–51.

Reinhorn, L. J. (2014). On Optimal Redistributive Capital Taxation. Mimeo.

Straub, L. and I. Werning (2015). Positive Long Run Capital Taxation: Chamley-Judd Revisited. Mimeo.

Trabandt, M. and H. Uhlig (2012). How Do Laffer Curves Differ Across Countries? NBER Working Papers 17862, National Bureau of Economic Research.

Werning, I. (2007). Optimal Fiscal Policy with Redistribution. *Quarterly Journal of Economics 122*(2), 925–967.

Appendices

A   Lagrangian and 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

$$
\begin{aligned}
\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \Bigg\{ & u\left(c_{1,t}\right) + v\left(l_{1,t}\right) + \psi\left[u\left(\lambda c_{1,t}\right) + v\left(\mathcal{K}(\lambda)\, l_{1,t}\right)\right] \\
& + \Delta_1\left[u'\left(c_{1,t}\right)c_{1,t} + v'\left(l_{1,t}\right)l_{1,t}\right] + \Delta_2\left[u'\left(c_{1,t}\right)\lambda c_{1,t} + \frac{\phi_2}{\phi_1}v'\left(l_{1,t}\right)\mathcal{K}(\lambda)\, l_{1,t}\right] \\
& + \xi_t\left(c_{1,t} - \widetilde{c}\right) + \gamma_t u'\left(c_{1,t}\right) - \gamma_{t-1}u'\left(c_{1,t}\right)\left(1 + (r_t - \delta)(1 - \widetilde{\tau})\right) \\
& + \mu_t\left[F\left(k_{t-1}, e_t\right) + (1-\delta)k_{t-1} - k_t - \frac{1+\lambda}{2}c_{1,t} - g\right]\Bigg\} - \psi\underline{U}_2 \\
& - u'\left(c_{1,0}\right)\left(\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}\right)\left(1 + (r_0 - \delta)(1 - \tau_0^k)\right),
\end{aligned}
\tag{33}
$$

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 $t > 0$:

$$
\begin{aligned}
& u'\left(c_{1,t}\right) + \psi\lambda u'\left(\lambda c_{1,t}\right) + \left(\Delta_1 + \lambda\Delta_2\right)\left[u'\left(c_{1,t}\right) + u''\left(c_{1,t}\right)c_{1,t}\right] + \xi_t \\
& + \gamma_t u''\left(c_{1,t}\right) - \gamma_{t-1}u''\left(c_{1,t}\right)\left(1 + (r_t - \delta)(1 - \widetilde{\tau})\right) = \mu_t\frac{1+\lambda}{2}
\end{aligned}
$$

- for consumption at $t = 0$:

$$
\begin{aligned}
& u'\left(c_{1,0}\right) + \psi\lambda u'\left(\lambda c_{1,0}\right) + \left(\Delta_1 + \lambda\Delta_2\right)\left[u'\left(c_{1,0}\right) + u''\left(c_{1,0}\right)c_{1,0}\right] + \xi_0 \\
& + \gamma_0 u''\left(c_{1,0}\right) - u''\left(c_{1,0}\right)\left(\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}\right)\left(1 + (r_0 - \delta)\left(1 - \tau_0^k\right)\right) = \mu_0\frac{1+\lambda}{2}
\end{aligned}
$$

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

$$
\begin{aligned}
& v'\left(l_{1,t}\right) + \psi v'\left(\mathcal{K}(\lambda)\, l_{1,t}\right)\mathcal{K}(\lambda) \\
& + \Delta_1\left[v'\left(l_{1,t}\right) + v''\left(l_{1,t}\right)l_{1,t}\right] + \Delta_2\frac{\phi_2}{\phi_1}\left[v'\left(l_{1,t}\right)\mathcal{K}(\lambda) + v''\left(l_{1,t}\right)\mathcal{K}(\lambda)\, l_{1,t}\right] \\
& - \gamma_{t-1}u'\left(c_{1,t}\right)F_{ke}\left(k_{t-1}, e_t\right)\frac{1}{2}\left(\phi_1 + \phi_2\mathcal{K}(\lambda)\right)(1 - \widetilde{\tau}) \\
& = -F_e\left(k_{t-1}, e_t\right)\frac{1}{2}\left(\phi_1 + \phi_2\mathcal{K}(\lambda)\right)\mu_t
\end{aligned}
$$

- for labor at $t = 0$:

$$v'(l_{1,0}) + \psi v'(\mathcal{K}(\lambda) \, l_{1,0}) \, \mathcal{K}(\lambda)$$

$$+ \Delta_1[v'(l_{1,0}) + v''(l_{1,0}) \, l_{1,0}] + \Delta_2 \frac{\phi_2}{\phi_1}[v'(l_{1,0}) \, \mathcal{K}(\lambda) + v''(l_{1,0}) \, \mathcal{K}(\lambda) \, l_{1,0}]$$

$$- u'(c_{1,0}) \, (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}) \, F_{ke}(k_{-1}, e_0) \, \frac{1}{2}(\phi_1 + \phi_2 \mathcal{K}(\lambda)) \, (1 - \tau_0^k)$$

$$= -F_e(k_{-1}, e_0) \, \frac{1}{2}(\phi_1 + \phi_2 \mathcal{K}(\lambda)) \, \mu_0$$

- for capital at $t \geq 0$:

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

B Alternative formulation and FOC for labor

With λ substituted, all RPO allocations can be found by solving

$$\max_{\tau_0^k, \{c_t^1, k_t, l_t^1\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left[u(c_{1,t}) + v(l_{1,t}) \right]$$

$$\text{s.t.} \qquad \sum_{t=0}^{\infty} \beta^t \left[u(\lambda c_{1,t}) + v(\mathcal{K}(\lambda) l_{1,t}) \right] \geq \underline{U}_2, \tag{34}$$

and subject to feasibility (3), implementability (10), the equation determining λ (22), tax limits (13) and (14), and consumption limits (15) for a given level of utility $\underline{U}_2$.

The Lagrangian is

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

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

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

$$\left. + \mu_t \left[F(k_{t-1}, e_t) + (1 - \delta) k_{t-1} - k_t - \frac{1 + \lambda}{2} c_{1,t} - g \right] \right\} - \psi \underline{U}_2 - \mathbf{W},$$

where $\mathbf{W} = u'(c_{1,0}) \, \Delta_1 k_{1,-1} \, (1 + (r_0 - \delta)(1 - \tau_0^k))$ and $\lambda = g(\mathbf{C}, \mathbf{L}, \mathbf{K})$.

We use two FOCs in the proof: for capital and for labor. The FOC for capital is unchanged relative to the previous Lagrangian. The FOC for labor at $t > 0$ with the utility function (6) is given by:

$$- \omega (l_{1,t})^{\sigma_l} \left[1 + \psi \mathcal{K}(\lambda)^{\sigma_l} (\mathcal{K}(\lambda) + \mathcal{K}'(\lambda) g_{\mathbf{L}} \mathbf{L}_t) + \Delta_1 (1 + \sigma_l) \right] + \psi (\lambda c_{1,t})^{-\sigma_c} g_{\mathbf{L}} \mathbf{L}_t c_{1,t}$$

$$- \gamma_{t-1} (c_{1,t})^{-\sigma_c} F_{ke}(k_{t-1}, e_t) \, \frac{1}{2}(\phi_1 + \phi_2 [\mathcal{K}(\lambda) + \mathcal{K}'(\lambda) g_{\mathbf{L}} \mathbf{L}_t]) \, (1 - \widetilde{\tau}) \tag{35}$$

$$= -F_e(k_{t-1}, e_t) \, \frac{1}{2}(\phi_1 + \phi_2 [\mathcal{K}(\lambda) + \mathcal{K}'(\lambda) g_{\mathbf{L}} \mathbf{L}_t]) \, \mu_t,$$

where

$$\mathbf{L}_t \equiv \frac{\partial \mathbf{L}}{\partial l_{1,t}} = -\omega \left(1 + \sigma_l\right) \beta^t \left(l_{1,t}\right)^{\sigma_l}, \tag{36}$$

and, implicit differentiating (21) with respect to $\mathbf{L}$,

$$g_{\mathbf{L}} = \frac{-\mathcal{K}(\lambda)}{\mathbf{C} - \frac{\sigma_c}{\sigma_l}\mathcal{K}(\lambda)\lambda^{-1}\mathbf{L}}.$$

C 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 τ_t^k is satisfied. If yes, impose $\gamma_t = 0$. Otherwise, the constraint on capital taxes has to be satisfied as equality, i.e., $\tau_t^k = \widetilde{\tau}$. ($T$ equations)

   - The remaining 3 equations come from the present-value budget constraints (PVBC) and the FOC for λ. The discounted sums in the PVBCs are calculated using the time paths 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.

D 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 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 $\phi_w/\phi_c = 0.95$ 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.

Table 2: Sensitivity analysis

Calibration	Workers gain as much as possible		Capitalists gain as much as possible	
	duration of transition (years)	revenue share of $\tau^k(\%)$	duration of transition (years)	revenue share of $\tau^k(\%)$
benchmark	26	21.7	11	12.7
$\tau_{SQ}^k = 0.3$	35	25.5	17	16.8
$\tau_{SQ}^k = 0.57$	17	15.4	7	7.7
$\tau_{SQ}^l = 0.15$	30	36.1	13	21.6
$\tau_{SQ}^l = 0.3$	14	7.2	8	4.6
$\lambda_{SQ} = 0.5$	25	21.5	12	13.4
$\lambda_{SQ} = 0.6$	25	21.3	10	11.4
quintiles	26	25.1	16	19.1

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

E   A Model of Bassetto and Benhabib

Until recently it was thought that in most dynamic models of factor taxation that economists use one would obtain $\tau_\infty^k = 0$. A string of papers, including Lansing (1999), Bassetto and Benhabib (2006), Reinhorn (2014), and Straub and Werning (2015), have shattered this view, as they show that when a proof allows, correctly, for unbounded Lagrange multipliers, the result fails in many versions of standard models. Our Proposition 1 recovers $\tau_\infty^k = 0$ even if we allow for unbounded Lagrange multipliers. We contend that the reason is that we introduce a minimum consumption level as an additional constrain in the planner's problem. We cannot compare in detail our result to all the papers mentioned above as there is no space, but in this appendix we compare carefully our result to Bassetto and Benhabib (2006) (BB hereafter). We show in detail the reason that we obtain a different result.

BB show in their Corollary 1 that $\tau_\infty^k > 0$ in a model similar to ours. This provided a counterexample to previous results arguing that $\tau_\infty^k = 0$ in heterogeneous-agents models. This appendix uses our proof strategy to show that a small modification of BB leads to the same result as in our paper, namely $\tau_\infty^k = 0$.

Another role of this appendix is to show a more elegant version of our proof strategy, since the model below is simpler and the proof shorter than in the main text. Although in this appendix we refer to some derivations written in the main text, this appendix is self-contained provided that the relevant parts of the main text are included.

Although BB is very close to our model there are still a number of differences: BB assume *i)* linear production function, *ii)* no consumption limit (15), i.e., they use $\widetilde{c} = 0$, *iii)* government can set a lump-sum redistributive transfer each period, *iv)* utility restricted to log() for the zero-tax result of Corollary 1 in BB, *v)* they focus on one RPO: the point that maximizes the utility of the median voter, *vi)* labor supply is inelastic, *vii)* agents are equally productive, i.e., $\phi_j = \phi$, $\forall j$, *viii)* a continuum of agents. It is of interest to study which of these differences leads to a different long-run result. Below we prove $\tau_\infty^k = 0$ when we maintain *iii)* to *vii)* as in BB but we introduce a strictly concave production function and a consumption limit with $\widetilde{c} > 0$, in other words we modify *i)* and *ii)*. We do not have a continuum of agents in the model below for simplicity, but this assumption is immaterial.

We now modify and specialize the model in the main text of our paper to introduce features *iii)* to *vii)*. In this way the model becomes a version of BB. We still differ from BB in *i)* and *ii)*. Assume utility depends only on consumption so $v(l) = 0$ and $l = 1$.[36] We restrict $\phi_1 = \phi_2 = 1$. Agent of type 1 has relative mass ω. In addition to capital and labor

---

[36]BB actually consider a utility $u(c) = (c - B)^{\gamma_c+1}/(\gamma_c + 1)$, so we take $B = 0$, as in their Corollary 1.

taxes, there is a lump-sum transfer T_t, common to all agents, so that the consumer's budget constraint is

$$c_{j,t} + k_{j,t} = w_t l_{j,t}(1 - \tau_t^l) + T_t + k_{j,t-1}\left[1 + (r_t - \delta)(1 - \tau_t^k)\right], \text{ for } j = 1, 2. \quad (37)$$

Aggregate consumption is denoted $c_t = \omega c_{1,t} + (1 - \omega)c_{2,t}$. Assume $\omega > 0.5$ so that $i = 1$ is the median voter. All the other features of the environment remain as in the main text.

Importantly, production is equal to $F(k_{t-1}, 1)$ and, unlike in BB, $F(\cdot, 1)$ is strictly concave, i.e., $F_{kk}(k, 1) < \infty$ for all $k > 0$.

The presence of fixed labor supply and transfers means that, as shown in BB, sufficient conditions for aggregate quantities to be a CE are feasibility

$$F(k_{t-1}, 1) + (1 - \delta)k_{t-1} \geq k_t + c_t + g, \quad (38)$$

and that the tax limit (13) holds for all periods. Also, the limit on consumption that we add relative to BB has to be included in the set of sufficient equilibrium conditions. In other words, the implementability constraints (10) and (11) that are part of the set of sufficient conditions in the main text do not constrain aggregate quantitites in the equilibrium of the model in this appendix.

There are two further minor differences with BB: their capital tax multiplies old capital too, and we have assumed $\omega = 1/2$ in the main text. These assumptions are immaterial.

E.1 The planner's problem

We find the policy that is optimal for the median voter, agent 1. Since $c_{1,t}$ appears in the objective function of the planner we have to take into account the relationship between $c_{1,t}$ and c_t imposed by equilibrium conditions.

From the budget constraints for the aggregate and for agent 1 we have

$$\sum_{t=0}^{\infty} \beta^t \frac{u'(c_t)}{u'(c_0)}\left[c_t - w_t(1 - \tau_t^l) - T_t\right] = \left(k_{-1} - k_{-1}^g\right)\left[1 + (r_0 - \delta)(1 - \tau_0^k)\right]$$

$$\sum_{t=0}^{\infty} \beta^t \frac{u'(c_t)}{u'(c_0)}\left[c_{1,t} - w_t(1 - \tau_t^l) - T_t\right] = k_{1,-1}\left[1 + (r_0 - \delta)(1 - \tau_0^k)\right]$$

where $-k_{-1}^g$ is initial government debt.

As in the main text, $\lambda c_{1,t} = c_{2,t}$ for all t for some $\lambda > 0$ in equilibrium. Therefore $c_{1,t} = \varkappa c_t$ for $\varkappa = 1/(\omega + (1 - \omega)\lambda)$. Combining these equations we have that in equilibrium

$$\varkappa = 1 - u'(c_0)\frac{(k_{-1} - k_{1,-1} - k_{-1}^g)\left[1 + (r_0 - \delta)(1 - \tau_0^k)\right]}{\sum_{t=0}^{\infty} \beta^t u'(c_t)c_t}. \quad (39)$$

45

The planner's problem is to maximize the utility of the median voter

$$\max_{\{c_t, k_t\}} \sum_{t=0}^{\infty} \beta^t \, u\left(\varkappa c_t\right),$$

subject to (3), (13), (15) and when $\varkappa$ (or, equivalently, λ) is related to aggregate variables by (39). The Lagrangian of the policy-maker's problem is

$$
\begin{aligned}
\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \, \Big\{ & u\left(\varkappa c_t\right) + \xi_t(c_t - \widetilde{c}) \\
& + u'\left(c_t\right)\left[\gamma_t - \gamma_{t-1}\left(1 + (r_t - \delta)(1 - \widetilde{\tau})\right)\right] \\
& + \mu_t\left[F\left(k_{t-1}\right) + (1 - \delta)k_{t-1} - k_t - c_t - g\right]\Big\}
\end{aligned}
\tag{40}
$$

where $\mu_t, \gamma_t, \xi_t \geq 0$, are the Lagrange multipliers of (3), (13), (15), $\forall t$, with the usual complementary slackness conditions, and $\gamma_{-1} = 0$.

Using that $u()$, is CRRA the FOCs for the planner are:

- for consumption at $t > 0$:

$$
\begin{aligned}
& \varkappa^{-\sigma_c} u'\left(c_t\right)\left(\varkappa + \varkappa'_t c_t\right) + \xi_t \\
& + u''\left(c_t\right)\left[\gamma_t - \gamma_{t-1}\left(1 + (r_t - \delta)(1 - \widetilde{\tau})\right)\right] = \mu_t
\end{aligned}
\tag{41}
$$

where $\varkappa'_t$ is the derivative of the right side of (39) with respect to c_t, namely,

$$\varkappa'_t = \beta^t \frac{1 - \varkappa}{\sum_{j=0}^{\infty} \beta^j u'(c_j)c_j}(1 - \sigma_c)u'(c_t)\tag{42}$$

- for capital at $t \geq 0$:

$$\mu_t \frac{1}{\beta\left(1 - \delta + F_k\left(k_t, 1\right)\right)} + \gamma_t \frac{u'\left(c_{t+1}\right)F_{kk}\left(k_t, 1\right)(1 - \widetilde{\tau})}{1 - \delta + F_k\left(k_t, 1\right)} = \mu_{t+1}.\tag{43}$$

As in Straub and Werning (2015) and as in the main text, we take for granted that there is a steady state for capital and consumption, i.e., in the planner's solution $c_t, k_t \to c^{ss}, k^{ss} < \infty$.

Proposition 3. *In the model of this appendix,*

(a) assume $\widetilde{c} = 0$. Then either $c^{ss} = 0$ or $\tau^{k,ss} = 0$.

(b) assume $\widetilde{c} > 0$. Then $\tau^{k,ss} = 0$.

Proof. Assume, towards a contradiction, that $\tau^{k,ss} > 0$ and $c^{ss} > 0$. The portion of the proof of Proposition 1 in the main text starting from "The Euler equation of the consumer..." until equation (28) applies to the current model literally and this proves $\gamma_t \to 0$.

Equation (42) implies $\varkappa'_t \to 0$, and $\gamma_t \to 0$ implies $\gamma_t u''(c_t) \to 0$. Then, given that $c^{ss} > 0$, we have $\xi_t \geq 0$, $\mu_t \to 0$, and $u'(c^{ss}) > 0$. If we take limits on (41) and we use the features mentioned in the previous sentence, we see that the left-hand side of (41) converges to a number larger than $\varkappa^{-\sigma_c+1} u'(c^{ss}) > 0$. Therefore, (41) cannot hold. This shows that we cannot have $\tau^{k,ss} > 0$ and $c^{ss} > 0$.

Then, if $\tilde{c} = 0$, this gives part *(a)*. If $\tilde{c} > 0$, then $c^{ss} > 0$. Hence we must have $\tau^{k.ss} = 0$. This proves part *(b)*. $\square$

Notice that this result is formally compatible with Corollary 1 of BB. We obtain $\tau^k_\infty = 0$ in part *b)* because the model in this appendix differs from theirs in two aspects: we have a strictly-concave F and $\tilde{c} > 0$. The reader can check how introducing a linear F or $\tilde{c} = 0$ would invalidate our proof of $\tau^k_\infty = 0$.

In the case $\tilde{c} = 0$, Proposition 3 allows two possibilities:

1. $\tau^{k,ss} > 0$ and $c^{ss} = 0$, or

2. $\tau^{k,ss} = 0$ and $c^{ss} > 0$.

We are silent about which one occurs with a strictly-concave production function. BB provide a stronger result in that, under linear production function, $\tau^{k,ss} > 0$.

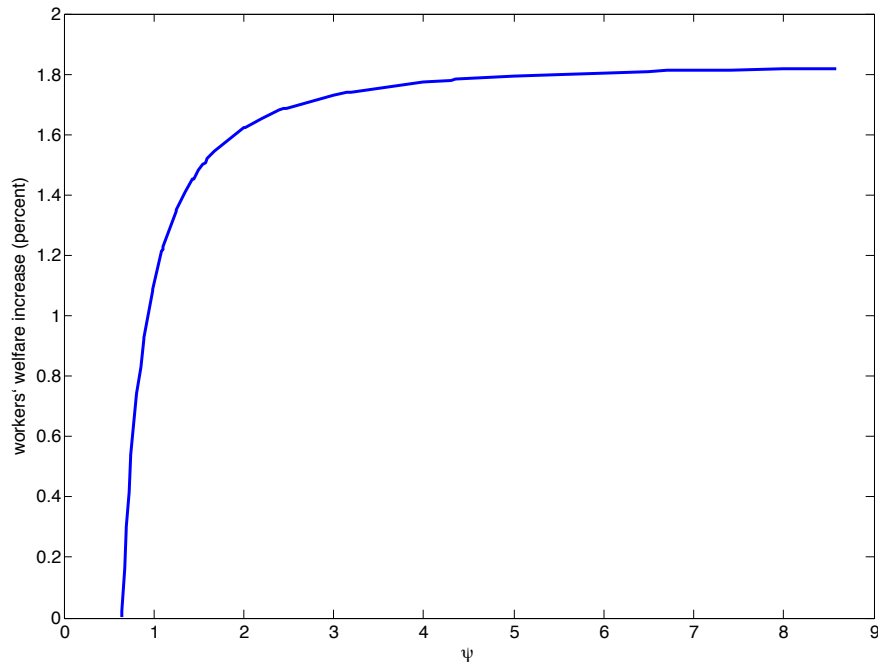Figure 1: The Ramsey Pareto frontier of Pareto-improving equilibria with fixed labor supply
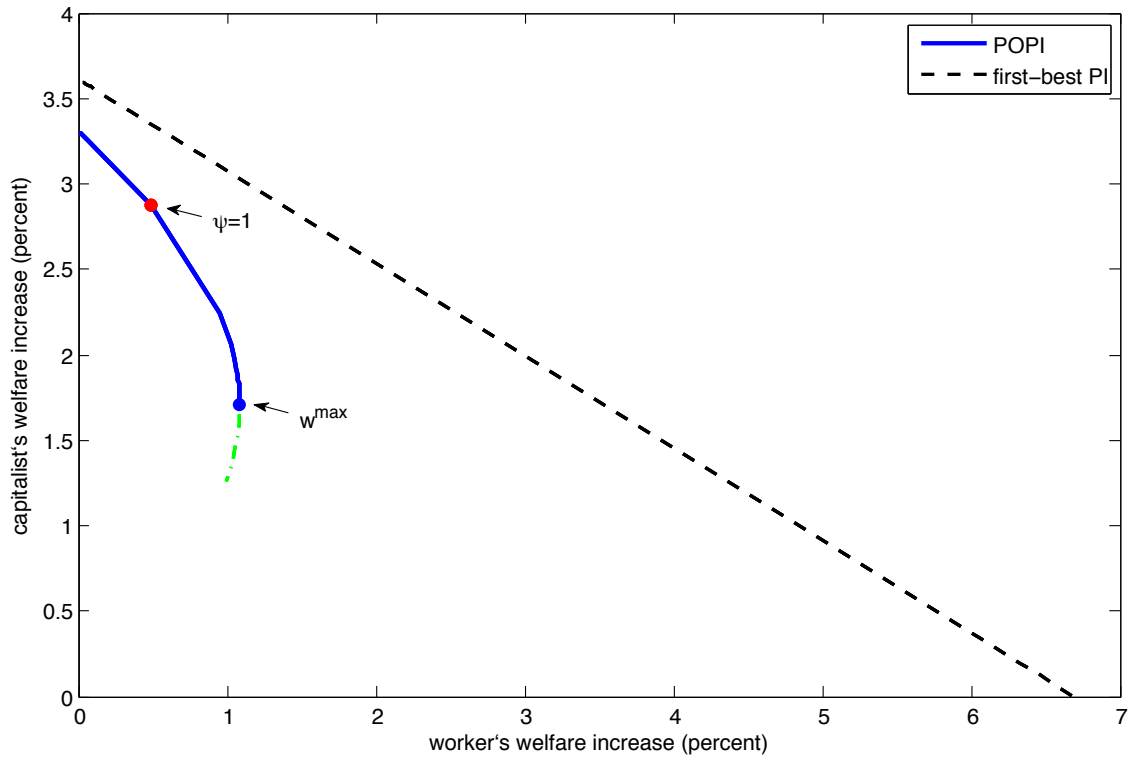
Notes: Welfare is measured as the percentage increase in status-quo consumption that would give the consumers the same lifetime utility as the optimal tax reform. The point $\psi = 1$ corresponds to the policy under 'the veil of ignorance,' and the point w^{max} represents the case where workers' utility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

Figure 2: The Ramsey Pareto frontier of Pareto-improving equilibria in the baseline model
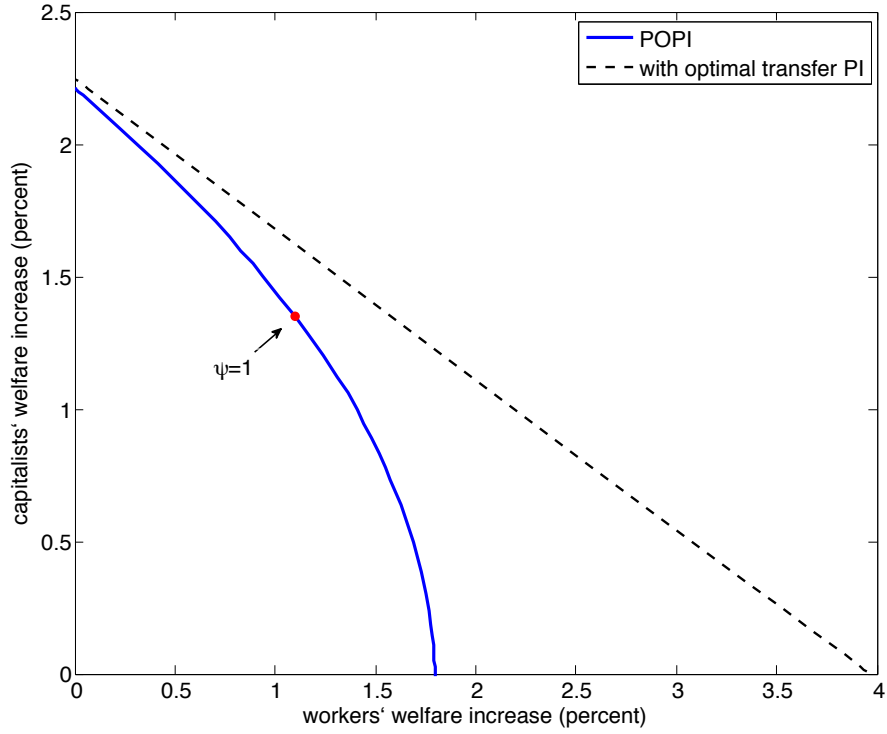
Notes: Welfare is measured as the percentage increase in status-quo consumption that would give the consumers the same lifetime utility as the optimal tax reform. The point $\psi = 1$ corresponds to the policy under 'the veil of ignorance.'

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

Notes: Welfare is measured as the percentage increase in status-quo consumption that would give the workers the same lifetime utility as the optimal tax reform.

49

Figure 4: A typical time path for capital taxes
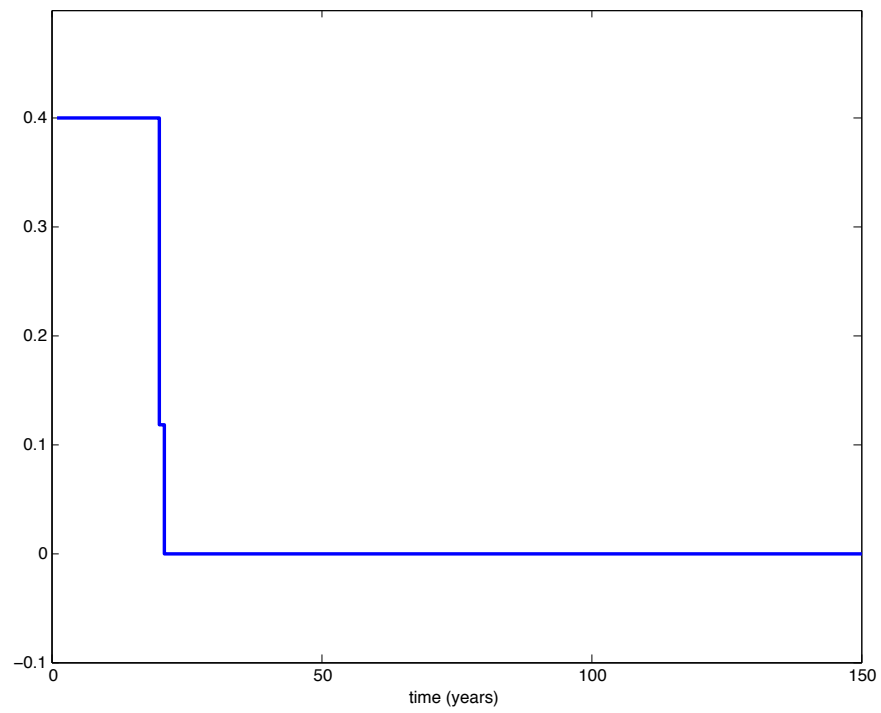
Figure 5: Properties of POPI programs in baseline model

Duration of transition (years)

Share of capital taxes in goverment revenues

Workers' relative Pareto weight and consumption share (normalized)

$\lambda/(1+\lambda)$
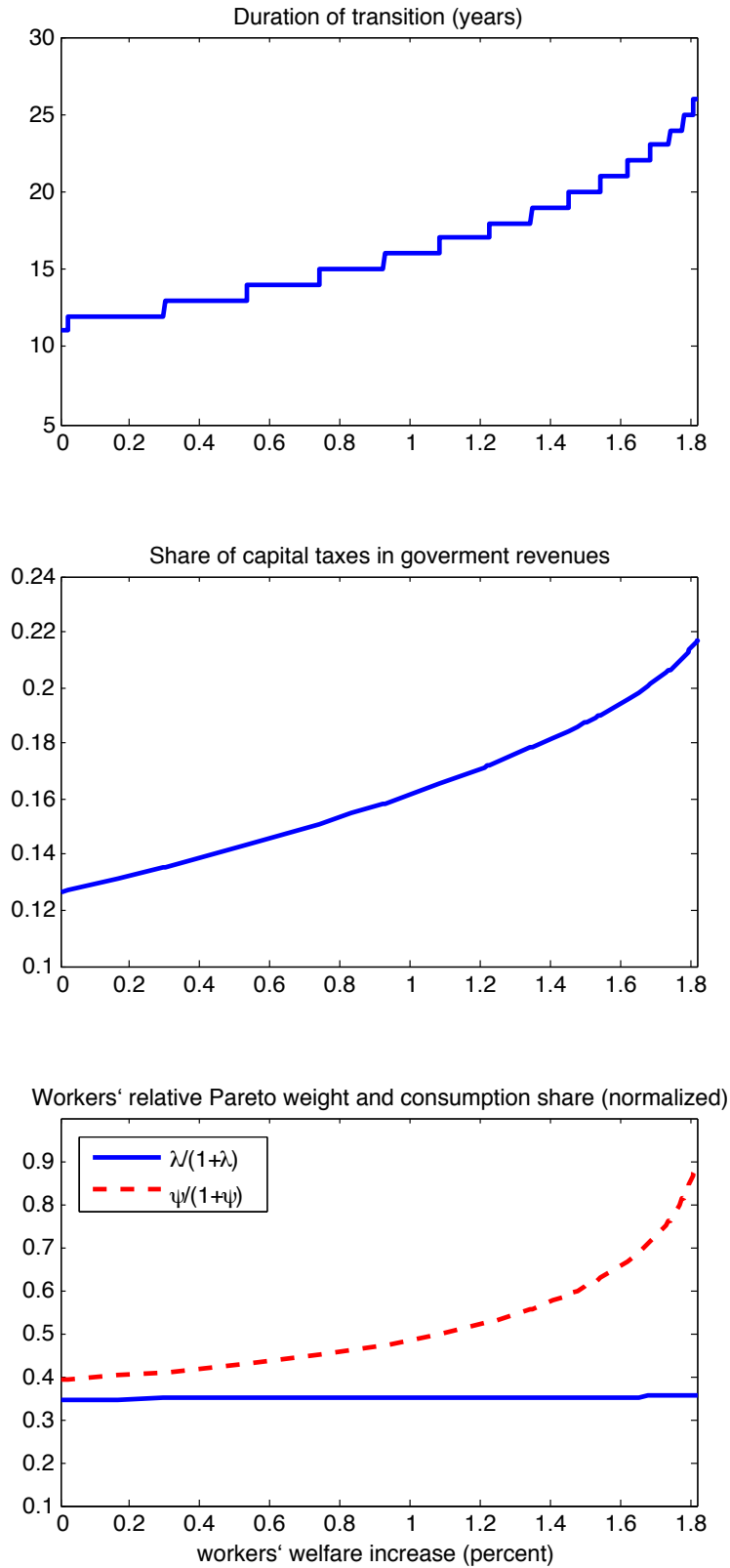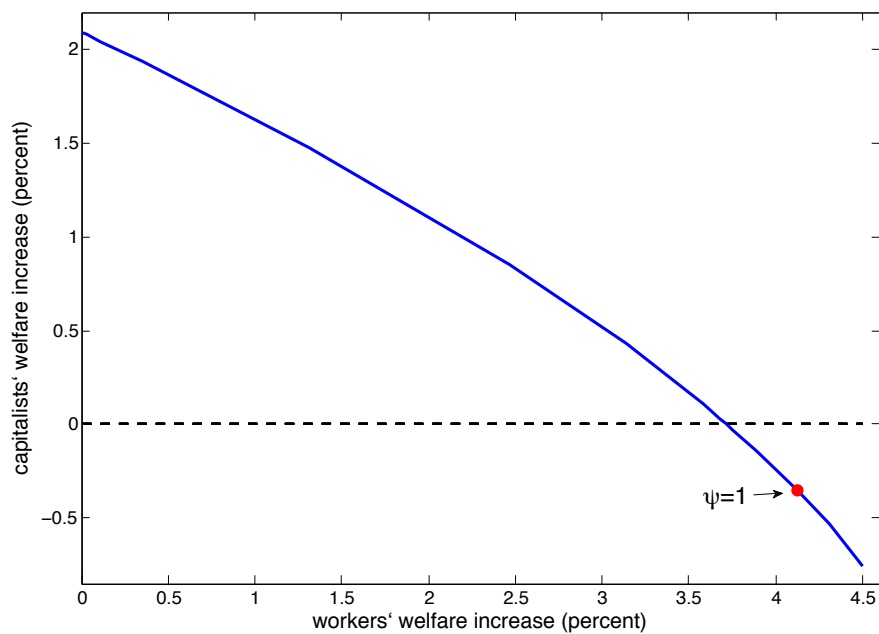$\psi/(1+\psi)$

workers' welfare increase (percent)

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 lifetime utility as the optimal tax reform. The point $\psi = 1$ corresponds to the policy under 'the veil of ignorance.'

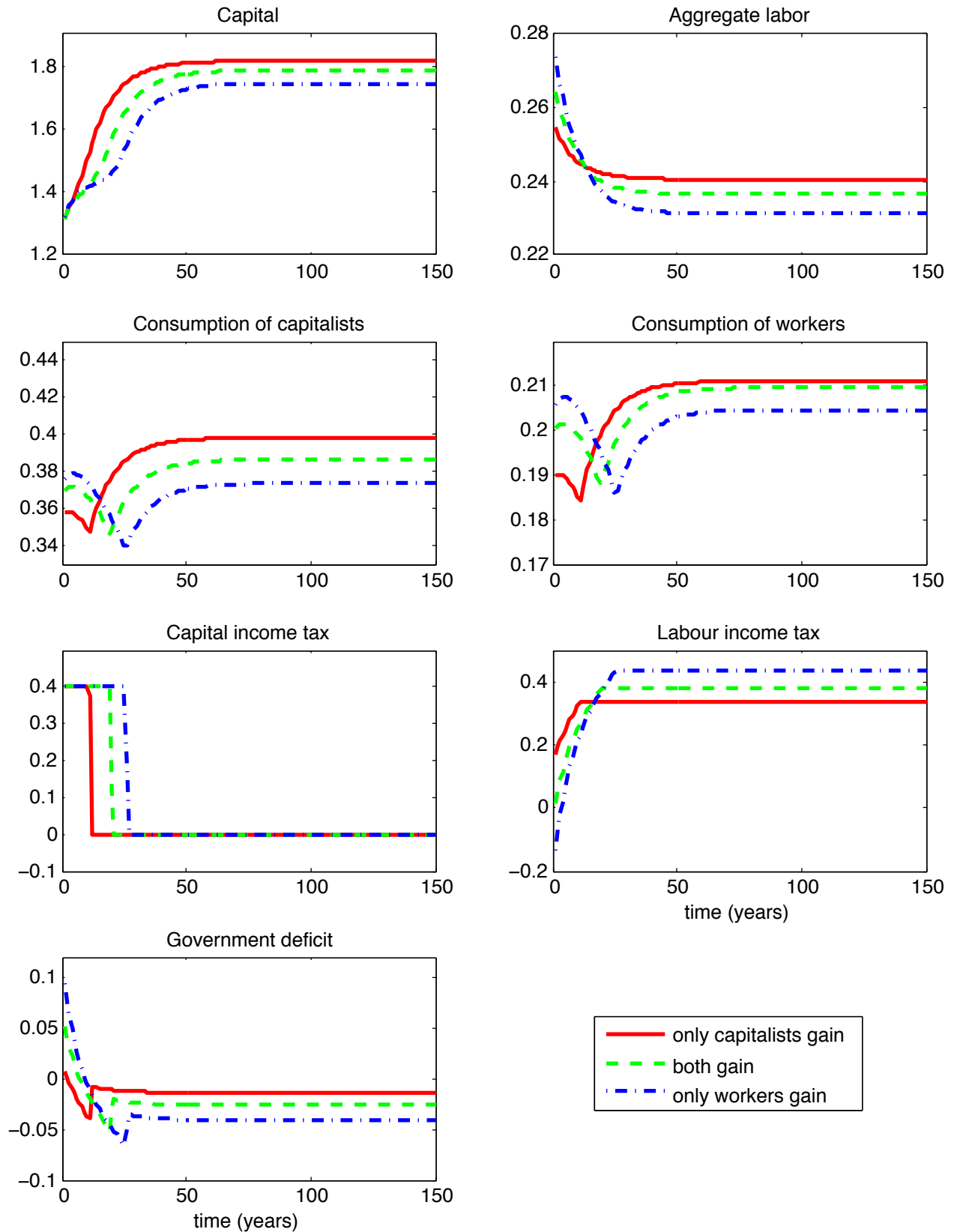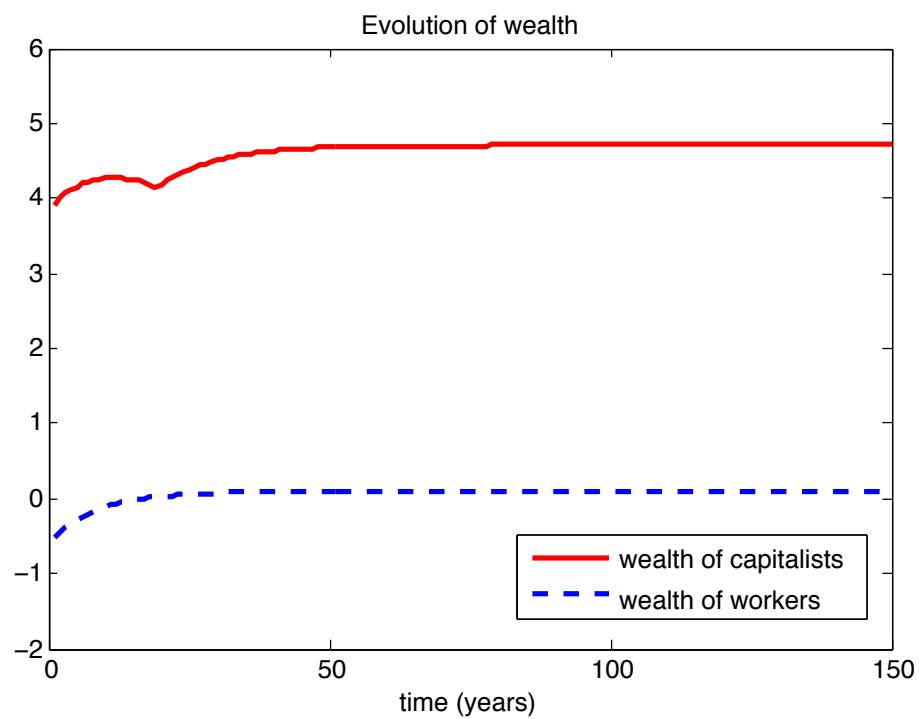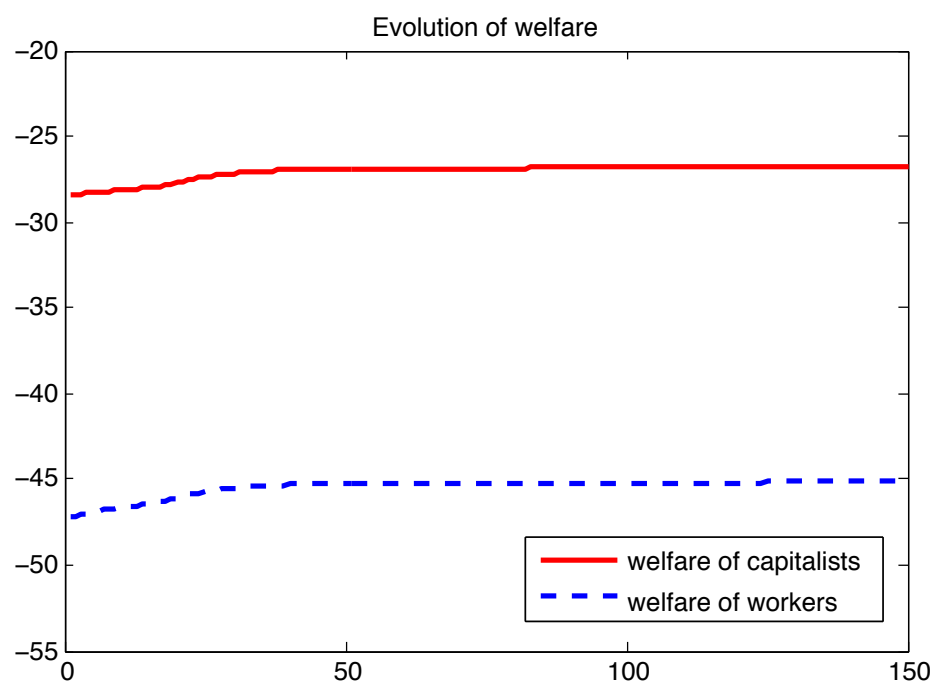Figure 7: The time paths of selected variables for three POPI plans in the baseline model

Figure 8: Typical time paths for consumers' welfare and wealth

Pareto-Improving Optimal Capital and Labor Taxes

by Katharina Greulich, Sarolta Laczó, and Albert Marcet

1 Alternative solution strategies for RPO allocations, ACK and Flodén

The closest paper to ours is Flodén (2009). It is important to clarify the differences. Flodén solves a planner's problem that maximizes the utility of one agent (the 'optimized' agent). Then Proposition 5 in his paper argues 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 λ). These differences are important and we examine them carefully below as, in our view, finding the 'optimized-agent' solution 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 Flodén's approach and review his contribution. We reinterpret his setup and show that it can be seen as a way of searching over competitive equilibria in a systematic and simple manner.

Flodén writes the planner's problem as Atkeson, Chari, and Kehoe (1999), ACK hereafter. This approach keeps consumption of all agents in the equilibrium conditions, instead of summarizing the allocations of other agents using (7) and (8), see main text, and λ, as we do. Although this makes computations different, it should give the correct allocations, the same as we find. We describe ACK in detail below, compare it with our approach, and use it to comment on the results of Flodén (2009).

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., utility of agent j is

$$U_{j,t} = \frac{1}{1-\mu} \left(c_{j,t} - \frac{\zeta}{1+1/\gamma} l_{j,t}^{1+1/\gamma} \right)^{1-\mu}.$$

This is a non-separable utility function, unlike ours, but it is immediate to extend our computational approach to this case. In addition, Flodén (2009) considers a general measure $\widetilde{\lambda}(j)$ ($\lambda(j)$ in Flodén, 2009) of agents of type $j \in [0,1]$. 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.

We use the notation

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

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

1.1 Using an ACK Lagrangian

Instead of representing equilibrium conditions with (7) and (8), ACK keep equilibrium conditions

$$\frac{u_{1c,t}}{u_{1c,t+1}} = \frac{u_{jc,t}}{u_{jc,t+1}} \quad \text{and} \quad \frac{u_{1l,t}}{u_{1c,t}\phi_1} = \frac{u_{jl,t}}{u_{jc,t}\phi_j}, \quad \forall j, \tag{1}$$

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 online appendix.

We focus on the case where $\widetilde{\lambda}$ is a discrete measure with J types of agents, where J is a finite integer, and agent j has mass $\widetilde{\lambda}_j$. This is the case of our main text with $J = 2$ and $\widetilde{\lambda}_1 = \widetilde{\lambda}_2 = 1/2$. 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 of this online appendix.

The Lagrangian to find the RPO allocations using this approach is

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \left\{ \left(\sum_{j=1}^{J} \psi_j U_{j,t} + \Delta_j \left[U_{j,t}(1-\mu) + \frac{u_{1c,t}\zeta l_{j,t}^{1/\gamma+1}}{\gamma+1} \right] \right. \right.$$
$$\left. + \rho_{jt} \left[u_{1c,t} u_{jc,t+1} - u_{jc,t} u_{1c,t+1} \right] + \xi_{jt} \left[u_{1c,t} u_{jl,t} \phi_1 - u_{1l,t} u_{jc,t} \phi_j \right] \right) \tag{2}$$
$$\left. + \mu_t \left(\sum_{j=1}^{J} \widetilde{\lambda}_j c_{j,t} + g + k_t - (1-\delta) k_{t-1} - F(k_{t-1}, e_t) \right) \right\} + \sum_{j=1}^{J} \Delta_j W_{j,-1}.$$

We use Flodén's notation except that we use ψ instead of his agent weights ω, Δ_j for the multipliers of individual implementability constraints instead of λ_j, and μ_t for the multiplier of the feasibility constraint instead of Flodén's $-\nu_t$.

Notice that this approach calls for solving for many variables, namely,

$$\left\{ (c_{j,t}, l_{j,t})_{j=1}^{J}, (\rho_{j,t}, \xi_{j,t})_{j=2}^{J}, \mu_t, k_t \right\}_{t=0}^{\infty}.$$

We prefer representing CE in the main text using (7) and (8) to substitute out agent $2,...,J$'s consumption and labor because then the planner's problem can be written as a maximization over $\tau_0^k, (\lambda_j)_{j=2}^{J}, \{c_t^1, \mu_t, k_t, l_t^1\}_{t=0}^{\infty}$. This reduces enormously the number of variables and

2

multipliers to be computed, and it is much more convenient for computation. More precisely, given the algorithm described in Appendix C, the number of variables to solve for with J agents would be $4J \times T + 2 + J$ using the ACK approach, while using our approach the number of variables to compute is only $4T + 2 + J$.

Of course solving the Lagrangian (2) is equally valid, and it should give the same solution as we find.

1.2 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 $\phi^{RA} \equiv \left(\sum_{j=1}^{J} \widetilde{\lambda}_j \phi_j^{1+\gamma} \right)^{\frac{1}{1+\gamma}}$ and initial wealth $\sum_{j=1}^{J} \widetilde{\lambda}_j k_{j,-1} = k_{-1} - k_{-1}^g$. This RA consumes $C_t^{RA} = \sum_{j=1}^{J} \widetilde{\lambda}_j c_{j,t}$. His Proposition 3 shows that as long as a CE satisfies

$$\sum_{t=0}^{\infty} \beta^t \left[u_{C^{RA},t} C_t^{RA} + u_{l^{RA},t} l_t^{RA} \right] = W_{-1}^{RA}, \tag{3}$$

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

Flodén finds equilibria that arise from the FOCs of the Lagrangian on page 300 in Flodén (2009). The reader can check that one can go from the above Lagrangian (2) to Flodén's with the following three modifications:

1. Equation (3) is introduced in the planner's problem as an additional constraint.

2. The competitive equilibrium conditions (1) 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., $\sum_{j=1}^{J} \widetilde{\lambda}_j c_{j,t}$ 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 (3) is guaranteed to hold. Therefore, modification 1 is redundant. All this means is that the multipliers λ_j and $\wedge$ (in Flodén's notation) are not uniquely defined, but the FOCs obtained from introducing modification 1 should give the same allocations as (2).

Modification 2 is also correct, indeed it implies and is implied by (1).

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

$$\sum_{j=1}^{J} \widetilde{\lambda}_j c_{j,t} = C_t^{RA}, \tag{4}$$

one could put only C_t^{RA} in the feasibility constraint. A similar point applies to aggregate labor.

As it is written, the Lagrangian on page 300 in Flodén (2009) 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 (4) 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 RPO allocation.

That the FOCs of Flodén's Lagrangian do not give the correct solution can be seen in the following way. Let $\mathcal{L}^2$ represent the expression in the first two lines of (2). The correct FOC with respect to $c_{j,t}$ from (2) is

$$\frac{\partial \mathcal{L}^2}{\partial c_{j,t}} = -\mu_t \widetilde{\lambda}_j. \tag{5}$$

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

$$\frac{\partial \mathcal{L}^2}{\partial c_{j,t}} = 0, \tag{6}$$

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

$$\sum_{j=1}^{J} \frac{\partial \mathcal{L}^2}{\partial c_{j,t}} = -\mu_t = \frac{\partial \mathcal{L}^2}{\partial C_t^{RA}},$$

as should hold in the optimmum. Instead his solution has $\sum_{j=1}^{J} \frac{\partial \mathcal{L}^2}{\partial c_{j,t}} = 0$. 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.

The only case where (6) is correct is when an agent has $\widetilde{\lambda}_j = 0$. In other words, it seems that the case where the FOCs are valid is where the planner gives full measure in her objective function to agents who have zero measure in the market.

If we considered a measure $\widetilde{\lambda}(.)$ with a continuous density $\widetilde{\lambda}'$ (where $\widetilde{\lambda}$ represents the measure of agents denoted λ on page 283 in Flodén (2009)), we would have the same problem.

Then the solution computed by Flodén (2009) amounts to giving full weight to an agent with density $\widetilde{\lambda}'(j) = 0$

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 $\bar{s}$. The proof of that proposition shows that the FOCs for this modified problem coincide with the FOCs on page 300 which are as (6). But if (as we think) the latter do not give an RPO 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 (5) has to hold instead of (6). Therefore, it is not true that all RPO solutions can be found by selecting an optimized agent even with GHH utility.

1.3 A rationale for Flodén's solution

The above discussion means that Flodén's results have a consistent and interesting interpretation as follows.

Imagine we consider optimizing a weighted sum of utilities of J' agents (where J' is a discrete number) and that these agents have mass zero in the economy. For this RPO allocation the planner's FOCs are indeed (6). But this is only a very small share of RPO solutions. Any RPO arising from a welfare function that gives positive weight to all agents, as is implied by most RPO, (6) 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 agents that have zero mass or zero density. 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.

References

Atkeson, A., V. Chari, and P. Kehoe (1999). Taxing Capital Income: A Bad Idea. *Federal Reserve Bank of Minneapolis Quarterly Review 23*(3), 3–17.

Flodén, M. (2009). Why Are Capital Income Taxes So High? *Macroeconomic Dynamics 13*(3), 279–304.