Pareto-Improving Optimal Capital and Labor Taxes

Katharina Greulich
Sarolta Laczó
Albert Marcet

This version: November 2019
(March 2016)

Barcelona GSE Working Paper Series

Working Paper nº 887

Pareto-Improving Optimal Capital and Labor Taxes[*]

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

November 2019

Abstract

We study optimal Pareto-improving fiscal policy in a model where agents are heterogeneous in their labor productivity and wealth and markets are complete. We first argue that recent results that find positive long-run capital taxes in the Ramsey equilibrium in standard models obtain for special parameter values. If the government is prevented from immiserating future generations the Chamley-Judd result reemerges. In addition, we question the traditional focus on long-run taxes. We show that a gradual reform is crucial: labor taxes should be cut and capital taxes should remain high for a very long time in order to achieve a Pareto improvement. Therefore, the long-run optimal tax mix is the opposite of the short- and medium-run one. The length of the transition determines who benefits more from the tax reform. The initial labor tax cut causes early deficits which lead to a positive level of government debt in the long run. Further, we show that a Benthamite policy (equal weights for all agents) is often not Pareto improving, and that, given significant heterogeneity, the optimal fiscal policy is time-consistent if a Pareto improvement is required at the time of reoptimization. We address a number of technical issues: sufficiency of first-order conditions for the Ramsey optimum, asymptotic behavior, and solution algorithms.

JEL classification: E62, H21

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

[*]We wish to thank the editor, three anonymous referees, Marco Bassetto, Jess Benhabib, Charles Brendon, Jordi Caballé, Begoña Domínguez, Joan M. Esteban, Giulio Fella, Michael Golosov, Andrea Lanteri, Andreu Mas-Colell, Claudio Michelacci, Sujoy Mukerji, Michael Reiter, Sevi Rodríguez, Raffaele Rossi, Kjetil Storesletten, Ludwig Straub, Jaume Ventura, Iván Werning, Philippe Weil, Fabrizio Zilibotti, and seminar and conference participants at various places for useful comments and suggestions. Michael Reiter provided the implementation of Broyden's algorithm. Greulich acknowledges support from the National Centre of Competence in Research (NCCR) FINRISK and the Research Priority Program on Finance and Financial Markets of the University of Zürich. Laczó acknowledges funding from the JAE-Doc grant co-financed by the European Social Fund. Marcet acknowledges funding from the Axa Foundation, the Excellence Program of Banco de España, the European Research Council under grants FP/2007-2013, n. 324048 and 788547, and AGAUR (Generalitat de Catalunya). The views expressed in this paper are not those of Swiss Re.

[†]Swiss Re

[‡]Queen Mary University of London and CEPR; School of Economics and Finance, Mile End Road, London, E1 4NS, United Kingdom. Email: s.laczo@qmul.ac.uk.

[§]University College London, Barcelona GSE, and CEPR; Drayton House, 30 Gordon Street, London, WC1H 0AX. Email: a.marcet@ucl.ac.uk.

1 Introduction

We study optimal policy with heterogeneous agents when the government chooses between labor taxes, capital taxes, and debt, focusing on Pareto-improving policies. Previous related studies leave many open issues. Recent works challenge the traditional result that long-run capital taxes should be zero, denoted $\tau^k_\infty = 0$, showing that τ^k_∞ might be positive and large. We first argue that if a reasonable constraint on policies is included $\tau^k_\infty = 0$ reemerges.

However, we find that even if $\tau^k_\infty = 0$ there is a need to redistribute along the transition. Hence the standard focus in the literature on long-run results using welfare functions with fixed weights can be misleading. Our aim is to put these issues in context and provide a unified story about redistribution and efficiency in factor taxation.

A large literature argued that the original $\tau^k_\infty = 0$ result in Chamley (1986) and Judd (1985) is robust to many extensions, as it efficiently promotes investment. Lowering capital taxes in practice is controversial, as it lowers taxes for richer taxpayers, apparently favoring efficiency over equity. However, some papers argued that $\tau^k_\infty = 0$ even with heterogeneous agents, for example, Judd (1985) and Atkeson, Chari, and Kehoe (1999), assuming the government maximized a welfare function with fixed weights. This suggested that taxing capital is a 'bad idea': there is no equity-efficiency trade-off, large capital taxes observed in practice must be a failure of fiscal policy-making, and lowering capital taxes should benefit everybody. However, this view clashes with the results in Correia (1999), Domeij and Heathcote (2004), Flodén (2009), and Garcia-Milà, Marcet, and Ventura (2010) (GMV hereafter), showing that in similar models as those considered above a large part of the population would suffer a large utility loss if capital taxes were abolished.

Furthermore, some recent results by Reinhorn (2019) and Straub and Werning (2019) (SW hereafter) show that previous proofs treated Lagrange multipliers incorrectly, and that a correct proof delivers $\tau^k_\infty > 0$ in some cases. Lansing (1999) and Bassetto and Benhabib (2006) (BB hereafter) provide more examples with $\tau^k_\infty > 0$. SW find a discontinuity in long-run taxes: small changes in the parameters of the model can cause τ^k_∞ to switch from zero to hundred percent. When BB and Section 2 of SW find a large τ^k_∞ in a heterogeneous-agent model, the authors motivate the result by the need for redistribution.

This possibly leaves a confusing picture. It seems difficult to make any general recommendation about labor and capital taxation. Should we expect τ^k_∞ to be large? Is the size of τ^k_∞ related to redistribution? Are the long-run results on optimal policy a good guidance for policy in the short and medium run? We answer no to all these questions in the context of a standard model. We assume full commitment, complete markets, agents that are heterogeneous in their labor productivity and wealth, an upper bound on capital taxes, and no agent-specific lump-sum transfers.

We first show that if the government cannot commit to immiserating future generations, then $\tau^k_\infty = 0$ reemerges in Ramsey-Pareto-optimal (RPO) policies. However, even $\tau^k_\infty = 0$ along all the points on the Pareto frontier that we examine, an equity-efficiency trade-off exists: RPO policies include a very long transition of high capital taxes and low labor taxes to ensure all agents gain from the policy (in our main calibration capital taxes should be high for 16 to 24 years). Therefore, tax policies are the opposite of the long run for many years. Those high capital taxes lower total investment and output, but they are required to redistribute wealth in favor of workers and, therefore, to achieve a Pareto improvement. In addition, the period of high capital taxes is longer for points on the frontier that favor more the workers. These results show that steady state analysis hides issues of redistribution, the transition is crucial to understand RPO policies.[1] Further, there is no discontinuity: the length of the period of high capital taxes increases gradually to achieve a larger redistribution toward workers as we move along the Pareto frontier.

The size of the equity-efficiency trade-off depends on the elasticity of labor supply. If labor is elastic, as in our main calibration, a long transition of low labor taxes is optimal, as it efficiently promotes growth and redistribution simultaneously, and welfare losses from redistribution are small. If labor is inelastic, an even longer period of high capital taxes is needed, optimal policy can barely promote growth, and the losses from redistribution are large.

We also show that as a result of low initial labor taxes the government initially accumulates deficits, leading to a positive long-run level of debt. Thus a theory of long-run debt can arise from a need to run deficits early on to fund a tax reform.

The literature on heterogeneous-agent macro models is now abundant and mainstream but it rarely addresses optimal policy. When it does, it tends to use a Benthamite welfare function with equal weights for all agents. Jeremy Bentham made his contributions in the early 19th century, when economics was in its infancy, closer to our time Paul Samuelson promoted the view that there is no such a thing as a 'correct' or 'fair' welfare function. Most textbooks in microeconomics emphasize that utility functions are only useful for ordering possible choices, their cardinal value is irrelevant. Therefore economists should be content describing policies along the Pareto frontier without arguing that a particular point on that

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

frontier is 'the best'.[2] We find that this point matters in practice: asymptotic results using welfare functions with fixed weights have obscured the equity-efficiency trade-off in factor taxation for decades, since an optimal transition involves high capital taxes, low labor taxes, and redistribution. Furthermore, the location of the Benthamite policy on the Pareto frontier is more or less arbitrary, and it can be far from Pareto improving. Furthermore, fixed welfare weights can be misleading when studying time-consistency.

Our focus on Pareto improvements speaks to the issue of gradualism in implementing policy reforms, as has been discussed in the political economy literature. In our case, all rational voters would be in favor of an optimal reform where capital taxes take a long time to be lowered before they reach $\tau^k_\infty = 0$.

Solving our model gives rise to a number of technical issues. Welfare weights should be chosen endogenously as a function of the point on the frontier to be analyzed. The ratio of consumption has to be chosen optimally, it is not directly given by welfare weights as in the absence of distortions. A further difficulty arises because the set of competitive equilibria is potentially not convex. In this case there might be many solutions to the first-order conditions (FOCs). We reduce analytically the set of possible solutions to FOCs to be sure that our computations pick the maximum.[3] In addition, non-convexities may lead to a duality gap. We show a way to check that the duality gap is empty or very small.

The above results are robust to various parameter changes and even to the possibility of progressive taxation. If the government can introduce a universal deductible (as considered in many papers on dynamic taxation), it is optimal to set the deductible to zero. That is, a flat-rate tax schedule is preferred over a progressive one. This is because a positive deductible increases the marginal tax rate and exacerbates total distortions.

Finally, we find that if policy can only be overturned by consensus and there is sufficient heterogeneity optimal policy is time-consistent, thus consensus builds commitment.[4] However, under fixed welfare weights we have generically time-inconsistency.

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 some analytical

[2] These comments also apply to any fixed welfare weights. These are sometimes justified by appealing to probabilistic voting or Nash bargaining. This interpretation poses some issues of its own, as it is not clear under what conditions the game that justifies these interpretations implies fixed welfare weights. We do not address this issue in this paper, as we limit ourselves to optimal allocations on the frontier.

[3] The issue of multiple solutions to FOCs is often ignored in models of optimal policy. An exception is Bassetto (2014), Section 3.1, showing how heterogeneity may lead to situations in which the FOCs are not sufficient. SW show, in a representative-agent model, that the Ramsey problem is convex when the upper bound on the capital income tax is 100 percent. Convexity ensures that they pick the optimum.

[4] Armenter (2007) found a similar result in a model close to that of Judd (1985), Section 3, where markets are not complete.

results, including $\tau^k_\infty = 0$, some properties of the transition, and sufficient conditions to find all RPO solutions. Our numerical results are in Section 4, including those on progressive taxation. Section 5 explores time-consistency. Section 6 concludes. The Appendix contains some algebraic details and proofs. The Online Appendix contains a description of our computational approach, a sensitivity analysis, and it gives details on the relation of our solution method to other approaches in the literature.

2 The model

2.1 The environment

We consider a production economy with heterogeneous consumers, complete markets, and certainty. Firms produce according to a production function $F(k_{t-1}, e_t)$, where k is total capital and e is total efficiency units of labor.

We consider two types of consumers, $j = 1, 2$, for simplicity.[5] Consumers differ in their initial wealth $k_{j,-1}$ and labor productivity ϕ_j. Agent j obtains income in period t from renting their capital at the rental price r_t and from selling their labor for a wage $w_t\phi_j$. Agents pay taxes at rate τ^l_t on labor income and τ^k_t on capital income net of a depreciation allowance at each time t. The period-t budget constraint of consumer j is

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

For comparison, below we also consider lump-sum taxes under some restrictions.

Consumer j has utility function $\sum_{t=0}^{\infty} \beta^t (u(c_{j,t}) + v(l_{j,t}))$, where $c_{j,t}$ is consumption and $l_{j,t} \in [0, 1]$ is labor (fraction of time spent working) of consumer j in period t. We assume $u_c > 0$, $v_l < 0$, and the usual Inada and concavity conditions. For many of our results we use the following assumption:

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

$$u(c) = \frac{c^{1-\sigma_c}}{1-\sigma_c} \quad \textit{and} \quad v(l) = -\omega\frac{l^{1+\sigma_l}}{1+\sigma_l}. \tag{1}$$

The production function F is strictly concave and increasing in both arguments, twice 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.

[5]It is immediate to extend our analysis to many types of consumers.

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

The government chooses capital and labor taxes, has to spend $g \geq 0$ in every period, saves in capital, and has initial capital k^g_{-1} (debt if $k^g_{-1} < 0$). Ponzi schemes for consumers and the government are ruled out.

We normalize the mass of each group to 1/2. Capital depreciates at a rate $\delta < 1$. Market clearing conditions for all t are

$$e_t = \frac{1}{2}\sum_{j=1}^{2} \phi_j l_{j,t}, \quad k_t = k^g_t + \frac{1}{2}\sum_{j=1}^{2} k_{j,t}, \quad \text{and}$$
$$\frac{1}{2}\sum_{j=1}^{2} c_{j,t} + g + k_t - (1-\delta)\, k_{t-1} = F\left(k_{t-1}, e_t\right). \tag{2}$$

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 CE allocations satisfy.

Consumers' FOCs with respect to consumption and labor yield

$$u'\left(c_{j,t}\right) = \beta u'\left(c_{j,t+1}\right)\left[1 + \left(r_{t+1} - \delta\right)\left(1 - \tau^k_{t+1}\right)\right], \ \forall t, \tag{3}$$
$$-\frac{v'\left(l_{j,t}\right)}{u'\left(c_{j,t}\right)} = w_t\left(1 - \tau^l_t\right)\phi_j, \ \forall t, \tag{4}$$

i.e., the Euler equation and the consumption-labor optimality condition, respectively, for $j = 1, 2$. Using equation (3), the budget constraints of consumer j for all $t = 0, 1, ...$ can be summarized in the present-value budget constraint

$$\sum_{t=0}^{\infty} \beta^t \frac{u'\left(c_{j,t}\right)}{u'\left(c_{j,0}\right)}\left[c_{j,t} - w_t \phi_j l_{j,t}\left(1 - \tau^l_t\right)\right] = k_{j,-1}\left[1 + \left(r_0 - \delta\right)\left(1 - \tau^k_0\right)\right], \ \text{for } j = 1, 2. \tag{5}$$

Assumption **A1** simplifies our characterization as follows. It is clear that (3) for $j = 2$ can be replaced by the condition

$$c_{2,t} = \lambda c_{1,t}, \ \forall t, \tag{6}$$

for some constant λ to be determined in equilibrium. Further, (4) for $j = 2$ can then be replaced by

$$l_{2,t} = \mathcal{K}(\lambda) l_{1,t}, \ \forall t, \tag{7}$$

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 primitives, namely, the functions $u()$ and $v()$ and labor productivities ϕ_j.[6]

Firms behave competitively, hence equilibrium factor prices equal marginal products, i.e.,

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

Therefore, factor prices can be substituted out in the CE conditions.

Using (4) and rearranging, (5) for consumer 1 becomes

$$\sum_{t=0}^{\infty} \beta^t \left(u'\left(c_{1,t}\right) c_{1,t} + v'\left(l_{1,t}\right) l_{1,t}\right) = u'\left(c_{1,0}\right) k_{1,-1} \left[1 + \left(r_0 - \delta\right)\left(1 - \tau_0^k\right)\right]. \tag{8}$$

Using (3), (4), (6), and (7), we can write (5) for consumer 2 as

$$\sum_{t=0}^{\infty} \beta^t \left(u'\left(c_{1,t}\right) \lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'\left(l_{1,t}\right) \mathcal{K}(\lambda) l_{1,t}\right) = u'\left(c_{1,0}\right) k_{2,-1} \left[1 + \left(r_0 - \delta\right)\left(1 - \tau_0^k\right)\right]. \tag{9}$$

The implementability conditions (8) and (9) involve only consumption and labor of type-1 consumers, initial wealth of the two types, and λ, which is sufficient to capture the sharing rule between the two groups, given that markets are complete. Werning (2007) and GMV provide the same key characterization.

It is easy to show that the necessary and sufficient conditions for a CE allocation are feasibility, the sharing rules for consumption and labor, and the present-value budget (or, implementability) constraints. Formally, sequences $\{(c_{j,t}, l_{j,t})_{j=1,2}, k_t\}_{t=0}^{\infty}$ are a CE, for given initial conditions on capital, if they satisfy (2), (6), (7), (8), and (9) for some λ to be determined consistent with all equilibrium conditions.[7] Given a set of CE allocations, taxes are backed out from (3) and (4), and individual capital is backed out from (8) for each t.

2.3 The policy problem

Now we describe in detail the policy problem, and we introduce some additional constraints on policies. 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 for all future periods, both the government and the agents have rational expectations, and the government understands the mapping between policy actions and equilibrium outcomes.

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

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

2.3.1 Additional constraints on policy

We assume that, in addition to allocations being a CE, the government faces further constraints. First, the government cannot impose capital taxes above a certain upper bound:

Constraint on Policy 1. *Capital taxes satisfy* $\tau_t^k \leq \widetilde{\tau}$, $\forall t$, *for a given* $\widetilde{\tau} \in (0, 1]$.

Many papers in the optimal factor taxation literature assume a bound only at $t = 1$. Some papers consider the above constraint $\forall t$ for the special case $\widetilde{\tau} = 1$, for example, Chamley (1986), Atkeson, Chari, and Kehoe (1999), and SW.

The case $\widetilde{\tau} < 1$ adds difficulties, as the feasible set for the government is non-convex, but it is of interest for various reasons. Capital flight in an open economy, or tax evasion in any economy, would be massive for very high τ_t^k, even if less than 100 percent. Another motivation is credibility: optimal policies under rational expectations involve taxes at the upper bound ($\tau_t^k = \widetilde{\tau}$) for a few initial periods before τ_t^k goes to zero in the long run. This initial tax hike could have devastating effects on investment in a world with partial credibility of government policy, or if agents form their expectations by learning from past experience.[8]

Combining this limit with (3) for $j = 1$, (1), and (6), it is easy to see that the tax limit is satisfied in a competitive equilibrium if and only if $\tau_0^k \leq \widetilde{\tau}$ and

$$u'(c_{1,t}) \geq \beta u'(c_{1,t+1}) \left[1 + (r_{t+1} - \delta)(1 - \widetilde{\tau})\right], \ \ \forall t \geq 0. \tag{10}$$

Adding (10) to the constraints of the government introduces Constraint in Policy 1 using the primal approach hence τ_t^k do not appear explicitly.

We also introduce the following constraint on consumption.

Constraint on Policy 2.

$$c_{1,t} \geq \widetilde{c}, \ \forall t, \ \textit{for some } \widetilde{c} \geq 0. \tag{11}$$

Given (6) this is equivalent with a lower bound on consumption for both consumers.

We focus on the case $\widetilde{c} > 0$ where the planner is constrained to choose policies where consumption is uniformly bounded away from zero. The interpretation is that the government cannot credibly commit today to 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 our successors. Alternatively, considering that very low levels of utility would eventually lead to a revolt or social conflict as in Benhabib and Rustichini

[8] Lucas (1990) offered a similar reasoning to motivate his study of a tax reform that abolishes capital taxes immediately. Ideally issues of credibility and learning would be introduced explicitly. An analysis of capital taxes under learning can be found in Giannitsarou (2006). We study time-consistency later in the paper.

(1996), a $\widetilde{c} > 0$ is a simple reduced-form version of a no-revolt constraint. In Section 3.3 we impose explicitly a minimum constraint on utility.

Although this constraint is stated in terms of consumption allocations, given that we use the primal approach, it is indirectly a constraint on tax policy. Consumers never see themselves as facing a lower bound (11), but they face taxes that induce them to act in such a way that (11) always holds.

2.3.2 The Ramsey problem

We study Ramsey-Pareto-optimal (RPO) allocations, and we focus on Pareto-improving allocations relative to a status quo. Let

$$\mathcal{S} \equiv \left\{ \text{sequences } \{(c_{j,t}, l_{j,t})_{j=1,2}, k_t\}_{t=0}^{\infty} \text{ which are a CE and satisfy (10) and (11)} \right\}.$$

We define a RPO allocation as an element of $\mathcal{S}$ such that the utility of one or more agents cannot be improved within the set $\mathcal{S}$ without hurting other agents. 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

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

where minimum utility $\underline{U}_2$ varies along all possible utilities that consumer 2 can attain in $\mathcal{S}$.

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

$$\begin{aligned} \max_{\tau_0^k, \lambda, \left\{c_t^1, k_t, l_t^1\right\}_{t=0}^{\infty}} \quad & \sum_{t=0}^{\infty} \beta^t \left(u\left(c_{1,t}\right) + v\left(l_{1,t}\right) \right) \\ \text{s.t.} \quad & \sum_{t=0}^{\infty} \beta^t \left(u\left(\lambda c_{1,t}\right) + v\left(\mathcal{K}(\lambda) l_{1,t}\right) \right) \geq \underline{U}_2, \end{aligned} \tag{12}$$

for $\underline{U}_2$ attainable in $\mathcal{S}$, subject to feasibility (2), implementability (8) and (9), tax limits (10) and $\tau_0^k \leq \widetilde{\tau}$, and consumption limits (11). We have used (6) and (7) to substitute for c_2 and l_2 to obtain (12).

We focus on RPO allocations which are also Pareto improving relative to a status-quo CE allocation where taxes are set as in the past.[9] We call these POPI allocations. Let the utilities attained by agent j at the status quo be U_j^{SQ}.[10] POPI allocations can be found by considering only minimum utility values $\underline{U}_2$ such that $\underline{U}_2 \geq U_2^{SQ}$ and such that

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

[9]The benchmark policies and allocations also satisfy the constraints on policy, (10) and (11).

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

where * denotes the RPO optimized value of each variable for a given $\underline{U}_2$.

Let ψ be the Lagrange multiplier of the minimum-utility constraint (12), let Δ_1 and Δ_2 be the multipliers of the implementability constraints (8) and (9), respectively, and μ_t, γ_t, and ξ_t be the multipliers of the feasibility constraint (2), the tax limit (10), and the consumption limit (11), respectively, at time t. The Lagrangian for the government's problem is

$$\begin{aligned}\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \Bigg\{ & u\left(c_{1,t}\right) + v\left(l_{1,t}\right) + \psi\left(u\left(\lambda c_{1,t}\right) + v\left(\mathcal{K}(\lambda) l_{1,t}\right)\right) \\ & + \xi_t (c_{1,t} - \widetilde{c}) \\ & + \Delta_1 \left(u'\left(c_{1,t}\right) c_{1,t} + v'\left(l_{1,t}\right) l_{1,t}\right) \\ & + \Delta_2 \left(u'\left(c_{1,t}\right) \lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'\left(l_{1,t}\right) \mathcal{K}(\lambda) l_{1,t}\right) \\ & + \gamma_t \left\{u'\left(c_{1,t}\right) - \beta u'\left(c_{1,t+1}\right)\left[1 + \left(r_{t+1} - \delta\right)\left(1 - \widetilde{\tau}\right)\right]\right\} \\ & + \mu_t \left[F\left(k_{t-1}, e_t\right) + (1-\delta) k_{t-1} - k_t - \frac{1+\lambda}{2} c_{1,t} - g\right]\Bigg\} - \psi \underline{U}_2 - \mathbf{W},\end{aligned} \tag{13}$$

where $\mathbf{W} = u'\left(c_{1,0}\right)\left(\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}\right)\left[1 + (r_0 - \delta)(1 - \tau_0^k)\right]$ with $\tau_0^k \leq \widetilde{\tau}$. Further, ξ_t, γ_t, $\mu_t \geq 0$, $\forall t$, and $\psi \geq 0$, with complementary slackness conditions.[11]

The first line of this Lagrangian has the usual interpretation: finding a Pareto-efficient allocation amounts to maximizing a welfare function. The weight of consumer 1 is normalized to one, the 'weight' ψ of consumer 2 is the Lagrange multiplier of the minimum-utility constraint. The next three lines in (13) correspond to the minimum consumption and the equilibrium deficits of consumers. The fifth line ensures that $\tau_t^k \leq \widetilde{\tau}$ for all $t > 0$. The last line includes the feasibility constraint. The term $\mathbf{W}$ collects the terms on the right side of (8) and (9).

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 (2019) is easier to use under these circumstances. Appendix A shows the recursive Lagrangian and the FOCs.

In a model with lump-sum taxes, the ratio of consumptions would be immediately given by $\lambda = \psi^{1/\sigma_c}$. Key to our approach is the fact that λ has to be chosen optimally, which leads

[11] Strictly speaking, in models of taxation it is not impossible that $\mu_t < 0$ if the government is very wealthy and taxes are very negative. We contend that this is not the case when our model is calibrated to actual economies. In any case, $\mu_t \geq 0$ can always be guaranteed if the government can choose g_t endogenously with a constraint $g_t \geq g$ and free disposal of g_t.

to a non-trivial FOC, see Appendix A. The fact that λ is a choice for the government reflects the fact that the government can vary consumers' relative wealth by varying the total tax burden of labor and capital in discounted present value. We further demonstrate and discuss how λ behaves differently from Pareto weights around Figure 3 in Section 4.3.1.[12]

As is often the case in optimal-taxation models, the feasible set of sequences for the planner is non-convex so that FOCs derived from the Lagrangian are necessary but not sufficient. We address this in detail in Section 3.4.

For the government's problem to be well defined, we should ensure that $\mathcal{S}$ is non-empty and that initial government debt is sustainable. This is guaranteed if $\underline{U}_2$ is achievable in $\mathcal{S}$ and if there is a status-quo equilibrium, and if $\widetilde{c}$ is lower that status-quo consumption. Since $\mathcal{S}$ is compact and the objective function is continuous and bounded, existence of a Ramsey optimum will be taken for granted in the rest of the paper.

3 Characterization of equilibria

In this section we describe some analytical results, including $\tau^k_\infty = 0$, treatment of dynamic participation constraints, and sufficiency of FOCs.

3.1 Zero capital taxes in the long run

We now examine under what conditions $\tau^k_\infty = 0$ in our model. This steady-state result is of independent interest given some recent developments in the literature, and it will be helpful in characterizing and interpreting the transition.

The result $\tau^k_\infty = 0$ was proved traditionally under the assumption that Lagrange multipliers of the feasibility constraint in the Ramsey problem have a finite steady state. But Reinhorn (2019) and SW show that these multipliers diverge under some conditions and $\tau^k_\infty \neq 0$. In addition, SW find $\tau^k_\infty > 0$ and that consumption goes to zero if initial government debt is above a certain level, see their Section 3, and they emphasize that this is not a knife-edged case. This reinforces a previous result by Lansing (1999) and BB showing that $\tau^k_\infty > 0$ is possible with heterogeneous agents.[13]

[12] As far as we know, no other paper has implemented the optimal choice of λ. Werning (2007) mentioned that λ (called 'market weights') had to be chosen optimally but did not use this in his paper. Flodén (2009) considers a model with many labor productivity-wealth types and capital/labor taxation, as 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. In the Online Appendix, we argue that the approach in Flodén (2009) does not find all RPO allocations, although it does provide a useful method to search over competitive equilibria.

[13] The framework of BB is quite close to ours, we discuss below how our results relate to theirs.

We share with the literature just described a preoccupation with using the FOCs of the Ramsey problem appropriately, and we do not bound Lagrange multipliers. However, our Proposition 1 below resuscitates the Chamley-Judd result, as we show $\tau^k_\infty = 0$ in a very general setup for most parameter values. While there are various differences with the models in the previous paragraph, we show that those results use parameter values or assumptions that are of measure zero in the parameters of our model and that are excluded for our result. In a sense to be made precise below, our result lends support to the view that $\tau^k_\infty > 0$ is actually knife-edged in the complete markets case.[14]

We proceed as follows. We take for granted the existence of a steady state for allocations:

A2. *Ramsey Optimal allocations have a finite steady state, namely,*

$$(c_{1,t}, k_t, e_t) \to (c^{ss}, k^{ss}, e^{ss}) < \infty.$$

Limits in this statement and in the rest of the paper are taken as $t \to \infty$. This is a reasonable way to proceed, because real variables have natural bounds. But as mentioned before, a proper proof cannot restrict multipliers to be unbounded or to have a limit.

Clearly, under this assumption and if $c^{ss} > 0$, capital taxes have a finite limit, i.e., $\tau^k_t \to \tau^k_\infty < \infty$.[15] The proof uses that a familiar argument in growth theory guarantees

$$F_k(k^{ss}, e^{ss}) > \delta. \tag{14}$$

We now provide a sequence of results leading to $\tau^k_\infty = 0$.

Lemma 1. *Assume **A1** and **A2**, and consider the case where $c^{ss} > 0$ and $\tau^k_\infty > 0$. Then $\mu_t \to 0$. If in addition $\tilde{\tau} < 1$ then $\gamma_t \to 0$.*

Proof. In Appendix B. □

Lemma 1 in itself does not say anything directly about fiscal policy, but it shows that the key difference between our approach and SW is the different asymptotic behavior of μ. Lemma 1 says that if $\widetilde{c} > 0$ and $\tilde{\tau} < 1$ then $\tau^k_\infty > 0$ can only happen if $\mu^{ss}, \gamma^{ss} = 0$ while SW show that if $\tilde{\tau} = 1$ and $c^{ss} = 0$ it can happen that $\tau^k_\infty > 0$ and $\mu^{ss} = \infty$.

[14] More specifically, the homogeneous-agent environment in Section 3 of SW is a special case of our paper. However, they assume $\widetilde{c} = 0$ and $\widetilde{\tau} = 1$, while we derive our results for any $\widetilde{c} > 0$ and $\widetilde{\tau} < 1$. The remaining papers also make special assumptions that we rule out such as no debt, no need to raise public revenue, etc.

[15] For a formal proof, note that the Euler equation of consumer 1 implies

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

This equation, **A2**, (14), and the fact that $\infty > u'(c^{ss}) > 0$ imply that if $c^{ss} > 0$ then $\tau^k_t \to \tau^k_\infty < \infty$.

Let

$$\Omega^l \equiv 1 + \psi \mathcal{K}(\lambda)^{1+\sigma_l} + \left(\Delta_1 + \frac{\phi_2}{\phi_1}\mathcal{K}(\lambda)\Delta_2\right)(1+\sigma_l),$$
$$\Omega^c \equiv 1 + \psi \lambda^{1-\sigma_c} + (\Delta_1 + \lambda\Delta_2)(1-\sigma_c).$$

Proposition 1. *Assume* ***A1****,* ***A2****, and* $\widetilde{\tau} < 1$*. Assume* $\Omega^l \neq 0$ *or* $\Omega^c > 0$*.*

a) Then either $c^{ss} = 0$ *or* $\tau^k_\infty = 0$*.*

b) If in addition $\widetilde{c} > 0$*, then* $\tau^k_\infty = 0$*.*

c) Furthermore, in the case $c^{ss} > \widetilde{c}$*, there is an integer* $N < \infty$ *such that*

$$\tau^k_t = 0, \ \ \forall t \geq N+1. \tag{15}$$

Proof. In Appendix B.

Proposition 1 says that $\tau^k_\infty > 0$ can only happen if $\Omega^l = 0$, suggesting this is a knife-edged case. The assumption $\Omega^l \neq 0$ should cover most cases of interest, and it was satisfied in all our computations. Proposition 1 also highlights that the capital tax may not converge to zero in the case (seemingly of measure-zero) where $\Omega^l = 0$. The alternative requirement $\Omega^c > 0$ echoes that of SW.[16]

In the standard case where agents are identical, a positive marginal cost of taxation, i.e., $\frac{\partial \mathcal{L}}{\partial \tau^k_0} > 0$, would imply that $\Delta_1 = \Delta_2 > 0$, and, therefore, that $\Omega^l > 0$.[17] In our model with two heterogeneous consumers we still expect $\frac{\partial \mathcal{L}}{\partial \tau^k_0} = u'(c_{1,0})(\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1})(r_0 - \delta) > 0$, but, interestingly, this allows for one of the Δ_j's to be negative. As a matter of fact, in our baseline calibration we find $\Delta_2 < 0$ for most Pareto-improving allocations. This has some implications for redistribution that we discuss at the end of Section 4.4. It also means that generically the sign of Ω^l is undetermined.

Although we are mainly interested in the case where lump-sum taxes are not available, it is useful to consider agent-specific lump-sum taxes T_j to each agent. Note that if lump-sum taxes satisfy $T_2 = T_1 \frac{\phi_2}{\phi_1}\mathcal{K}(\lambda)$ they are 'labor-income-neutral'.[18] Similarly, a 'total-income-neutral' lump-sum tax sets $T_2 = T_1 \lambda$. We now make the following assumption.

[16] Their condition can be written as $1 + \Delta(1-\sigma_c) > 0$, where Δ (μ in their notation) is the Lagrange multiplier of the lifetime-budget constraint of the representative household, see their Proposition 7. In our case the condition contains additional heterogeneity terms, therefore ψ and λ play a role as well.

[17] In models with distortionary taxes it is usually welfare enhancing that private agents are initially poorer or, equivalently, that τ^k_0 is high, leading to positive Δ's.

[18] This because in equilibrium $\frac{\phi_2}{\phi_1}\mathcal{K}(\lambda) = \frac{l_{2,t} w_t \phi_2}{l_{1,t} w_t \phi_1}$, hence the distribution of non-capital income is unchanged in this case.

A3. *For agent-specific lump-sum taxes satisfying either*

a) $T_2 = T_1 \frac{\phi_2}{\phi_1} \mathcal{K}(\lambda)$ *or*

b) $T_2 = T_1 \lambda$ *when* $\sigma_c < 1$,

a marginal increase of lump-sum taxes above $T_1 = 0$ *is welfare improving.*

A3 fails only if there are no gains from raising lump-sum taxes keeping the distribution of non-capital or total income unchanged. It is unlikely that **A3** fails in reasonably calibrated models. It would fail, for example, if the government is so rich and has such high initial savings that it has to set negative distortionary tax rates, and hence lump-sum taxes would only exacerbate the distortions.[19]

Corollary 1. *Assume* ***A1-A3*** *and* $\widetilde{\tau} < 1$*. Then Proposition 1 holds.*

Proof. In Appendix B.

Proposition 1 and Corollary 1 suggest that, in the current model, $\tau_t^k > 0$ can only occur in knife-edged cases where either $\Omega^l \neq 0$ or $\widetilde{c} = 0$ or $\widetilde{\tau} = 1$, these cases being excluded in various steps of the proof. The examples of BB and SW use $\widetilde{c} = 0$ and/or $\widetilde{\tau} = 1$. As a matter of fact, all their examples have $c_t \to 0$.

3.2 Sufficient conditions for a solution

The results in Section 3.1 relied on the fact that the FOCs are necessary for a Ramsey solution, therefore multiplicity of critical points was not an issue. But if we obtain numerical simulations by solving FOCs, multiplicity needs to be addressed. In this section we address multiplicity given a weight ψ. Formally, for a fixed constant $\psi \in [-\infty, \infty]$, consider the following modified model (MM).

$$\max_{\tau_0^k, \lambda, \left\{c_t^1, k_t, l_t^1\right\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left[u\left(c_{1,t}\right) + v\left(l_{1,t}\right) + \psi\left(u\left(\lambda c_{1,t}\right) + v\left(\mathcal{K}(\lambda) l_{1,t}\right)\right)\right], \tag{16}$$

subject to $\mathcal{S}$. Notice that we allow for negative ψ's, and $\psi = \infty$ means that consumer 1 receives zero weight. The FOCs of this problem coincide with the conditions of RPOs.

Strictly speaking $\Omega^l = 0$ cannot be ruled out from the outset, hence we cover this case in our simulations.[20] The following proposition characterizes further the solution if $\Omega^l = 0$.

[19] SW make the "assumption of positive initial private wealth, $k_0 + b_0 > 0$" to show that Δ is positive, see footnote 23 and their Appendix I.

[20] We said above that $\Omega^l = 0$ 'appears' to be a measure-zero case. But this statement should be taken with a grain of salt. But in a non-convex feasible set it could happen that a set of positive measure on, say, the heterogeneity parameters $\phi_j, k_{j,-1}$, $j = 1, 2$, gives rise to the same multipliers and $\Omega^l = 0$. Also, **A3** is useful for intuition but since it is an assumption about the behavior of an optimum it can not be guaranteed ex-ante.

Proposition 2. *Assume **A**1-**A**2, $\widetilde{\tau} < 1$ and $\widetilde{c} > 0$. Assume in addition that $\mu_t > 0$ for some t. If $\Omega^l = 0$ then $\tau_t^k = \widetilde{\tau}$ for all $t = 0, 1, ...$ and $c_t = \widetilde{c}$ for t large enough.*

Proof. In Appendix B.

The assumption that $\mu_t > 0$ for some t will hold in any reasonable calibration.[21]

As mentioned in Albanesi and Armenter (2012), "[...] the set of admissible allocations is not convex for many second-best problems. [...] Often, sufficiency of the first-order conditions is verified numerically or strong conditions on primitives are imposed." But in an infinite-dimensional problem exploring numerically all possible solutions can be difficult. Propositions 1 and 2 help for this task as they narrow down τ_∞^k to only be 0 or $\widetilde{\tau}$, and in this last case we know that $c_t = \widetilde{c}$ for t large. This is made precise in the following.

Algorithm

Step 1 *Look for solutions that satisfy Proposition 1 as follows: for each candidate N, compute the infinite 'tail' of the sequence imposing (15), checking that all Lagrange multipliers have the correct sign, and taking a $\widetilde{c}$ sufficiently small. If such an allocation can be found and it has $\Omega^l \neq 0$, this solution is a candidate for the optimal solution.*[22]

Step 2. *Find a solution with $\tau_t^k = \widetilde{\tau}$ for all t. If $\Omega^l = 0$ this solution is a candidate for the optimal solution.*[23]

In each step we have to check numerically if there are several solutions with the stated properties, as is done in scores of papers in economics.[24]

If Step 1 delivers only one solution and we find no solution or solutions only with $\Omega^l \neq 0$ in Step 2, we are done. If we find more than one candidate solution, either because Step 1 has more than one solution or because Step 2 satisfies $\Omega^l = 0$, then the algorithm ends as follows.

Step 3. *Compute the utility corresponding to each candidate solution found in Steps 1 and 2 and pick the solution with the highest utility.*

[21]This is because μ_t measures the increase in welfare from lowering g in a given period. Lower g means more consumption, more leisure *and* lower distortionary taxes, thus it is almost certain to increase welfare.

[22]See Online Appendix A for more details on the computations.

[23]Strictly speaking there is a third possibility where $\Omega^l = 0$ and $\mu_t = 0$ for all t, just in case the condition $\mu_t > 0$ in Proposition 2 is violated. But this is trivially ruled out in the computations. Actually, $\mu_t = 0$ for all t implies many other unlikely restrictions in addition to $\Omega^l = 0$, such as $1 + \psi\lambda^{1-\sigma_c} + (\Delta_1 + \lambda\Delta_2)(1 - \sigma_c) = 0$ (from the FOC for consumption at $t > 0$), $k_{-1}^1\Delta_1 + k_{-1}^2\Delta_2 = 0$ (from the FOC for labor at $t = 0$), etc.

[24]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 FOCs in non-convex maximization problems, show a 'solution' which is only guaranteed to be a maximum if one searches for all possible critical points.

Since Propositions 1 and 2 exhaust all possible steady states this algorithm is certain to give the correct solution. In all the optimal allocations we computed in Section 4, the computer could not find a solution with all the properties of Step 2, and found one solution in Step 1 with $\Omega^l \neq 0$, hence $\tau_t^k = 0$ in all the calculations shown below.

3.3 Dynamic participation constraints

Constraint on Policy 2 captures the idea that a government may be unable to fully commit to a policy that leads to immiseration. As argued in Benhabib and Rustichini (1996), such a policy would lead to very low welfare and social unrest, which would likely prevent the policy from being continued. To introduce this idea more explicitly, we now replace Constraint on Policy 2 by the following dynamic participation constraints (PCs).

Constraint on Policy 3.

$$\sum_{i=0}^{\infty} \beta^i \left(u\left(c_{j,t+i}\right) + v\left(l_{j,t+i}\right)\right) \geq \underline{U}, \ \forall t, \ \ j = 1, 2, \ \textit{for some finite } \underline{U}. \tag{17}$$

This implies a relatively minor change in the analysis. In particular, CE are still summarized by the same equations as before, namely, (2), (6), (7), (8) and (9) for some λ constant through time. The Ramsey problem is as before with (17) replacing (11). Using the results in Marcet and Marimon (2019), the first two lines of the Lagrangian for the government's problem (13) are replaced by

$$\begin{aligned}\mathcal{L} = \sum_{t=0}^{\infty} \beta^t & \left(u\left(c_{1,t}\right) + v\left(l_{1,t}\right)\right)\left(1 + M_{1,t}\right) + \left(u\left(\lambda c_{1,t}\right) + v\left(\mathcal{K}(\lambda) l_{1,t}\right)\right)\left(\psi + M_{2,t}\right) \\ & - \left(\nu_{1,t} + \nu_{2,t}\right)\underline{U},\end{aligned}$$

while the remaining lines in (13) stay unchanged. Here, $\nu_{j,t} \geq 0$ are the Lagrange multipliers of (17), $M_{j,t} = M_{j,t-1} + \nu_{j,t}$ for all $t \geq 0$ and $M_{j,-1} = 0$ for $j = 1, 2$.

A large literature has introduced PCs in models of risk sharing with partial commitment, for example, Marcet and Marimon (1992, 2019), Kocherlakota (1996), and Ábrahám and Laczó (2018). This literature exploits the fact that the terms $(1 + M_{1,t})$ and $(\psi + M_{2,t})$ act as time-varying Pareto weights: the weight of agent j increases in periods when the PC of j becomes binding, and it stays constant otherwise.[25] This increase in the welfare weight ensures that the PC holds for the corresponding agent, avoiding default in the risk-sharing literature, or avoiding social conflict in our application. The ratio $u'(c_{2,t})/u'(c_{1,t})$

[25] In our case only the participation constraint of one agent can ever be binding. If, say, $\lambda^* < 1$, then $M_{1,t} = 0$ for all t.

is time-varying and equal to $(1+M_{1,t})/(\psi+M_{2,t})$.[26] Instead in our model $u'(c_{2,t})/u'(c_{1,t})$ is constant through time according to (6), and the dynamics of $(1+M_{1,t})/(\psi+M_{2,t})$ only determine the dynamics of distortionary taxes. This is not surprising given that, as mentioned in Section 2.3.2, even in our baseline model $u'(c_{2,t})/u'(c_{1,t})$ is not directly given by the Pareto weights.

While studying the effect that PCs may have on the dynamics of taxes is of interest, we leave a detailed analysis of this issue for future research. Here we focus only on the implications on the limiting behavior of capital taxes to guarantee that Lemma 1 and Propositions 1 and 2 apply. We only give an outline of the proof.[27]

The key difference is that the FOCs for consumption and labor hold with $\xi_t = 0$, and Ω^l and Ω^c are replaced by

$$\begin{aligned}\Omega_t^l &\equiv 1+M_{1,t}+(\psi+M_{2,t})\mathcal{K}(\lambda)^{1+\sigma_l}+\left(\Delta_1+\frac{\phi_2}{\phi_1}\mathcal{K}(\lambda)\Delta_2\right)(1+\sigma_l) \quad \text{and}\\ \Omega_t^c &\equiv 1+M_{1,t}+(\psi+M_{2,t})\lambda^{1-\sigma_c}+(\Delta_1+\lambda\Delta_2)(1-\sigma_c).\end{aligned}$$

Now, if we strengthen slightly our assumption to guarantee that $v(l^{ss}) < \overline{v}$ for some $\overline{v} < \infty$, then, taking limits in (17), it is clear that $c^{ss} > 0$ for any $\underline{U} > (u(0)+\overline{v})/(1-\beta)$. Since the proofs of Lemma 1 and Propositions 1 and 2 hinge on $c^{ss} > 0$, it is easy to check that these results obtain under Constraint on Policy 3 as long as the conditions on Ω^l and Ω^c are replaced by the same conditions on Ω^l_∞ and Ω^c_∞, respectively, and the requirement $\underline{U} > (u(0)+\overline{v})/(1-\beta)$ replaces $\widetilde{c} > 0$.[28]

Therefore, the numerical results in Section 4 can be interpreted as solving the model in the current section with a $\underline{U}$ sufficiently low for PCs to never be binding.

3.4 The frontier of the equilibrium set

Since the set of feasible allocations for the government $\mathcal{S}$ is not necessarily convex, a Lagrangian approach is not guaranteed to give all the RPO allocations. In Section 3.2 we already discussed how to address the issue of multiple solutions to the FOCs for a given welfare weight ψ. A second concern arises in the determination of ψ: since we trace out the

[26] Alvarez and Jermann (2000) and Ábrahám and Cárceles-Poveda (2006) consider a continuum of agents without and with capital, respectively, and show that the equilibrium in such an environment can be decentralized with endogenous borrowing limits. Park (2014) studies optimal taxation in this model.

[27] Notice that Constraint on Policy 3 does not imply Constraint on Policy 2: given $\widetilde{c} > 0$ there are consumption allocations satisfying (17) for which, say, $c_0 < \widetilde{c}$. Therefore, the feasible set for the planner is different in the model of this subsection, and the above propositions need to be proved separately.

[28] In models with PCs it can happen that $M_{j,t} \to \infty$. Note that the contradiction that sustains the proofs of Propositions 1 and 2 can be obtained even if $\Omega_t^l \to \infty$.

frontier of utilities, the duality gap (i.e., the set of RPO solutions that are not a saddle point of the corresponding Lagrangian for some welfare weight ψ) might be non-empty. In this case we would ignore some RPO allocations as we trace out the Ramsey Pareto frontier by varying ψ. To be precise, let the feasible set of utilities

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

and let $\mathcal{F}$ be the boundary (or 'frontier') of $\mathcal{S}^U$. Without distortions and a concave utility function $\mathcal{F}$ 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 cannot be found by maximizing a welfare function for some fixed weight ψ. Furthermore, 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, we are finding all RPO equilibria using a welfare function in our model. We will check this condition numerically in our application.

Let $U_j(\psi)$ be the utility of consumer $j = 1, 2$ at the solution to the MM problem defined in (16).

A4. *MM has a unique solution for all $\psi \geq 0$. Furthermore, $U_2(\cdot)$ is invertible on $[0, \infty]$.*

Proposition 3. *Assume **A4**. Then*

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

2. *Every RPO 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. In Appendix B.

Part 2 of Proposition 3 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 for this $\psi < 0$,[29]

[29] Notice that if we had a standard model without distortions and $u(0) = -\infty$, then there exists no solution for MM with $\psi < 0$. In that case part 3 would, of course, not apply, and it would not define a point on $\mathcal{F}$.

these points are not Pareto optimal, since both consumers' utilities could be increased along the frontier. More points on the frontier can be found if the consumers switch places in the objective function of 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, **A4** may not hold for some parameterizations. But it can be checked numerically whether it holds in a given application. We apply the Algorithm of Section 3.2 for each ψ 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. These checks can only be done approximately, as they rely on numerical approximations, but to the extent that invertibility is verified for a very fine grid of ψ's, a duality gap is unlikely to exist or is very small, as it would have to sneak in between grid points. We have done this for our numerical examples below, and, as required, $U_2(\psi)$ appears invertible, therefore MM fully characterizes all RPO solutions.

The POPI plans can be found with $\psi \in [0, \infty]$ such that $(U_1(\psi), U_2(\psi))$ are larger than the status-quo utilities of both consumers.

4 Numerical results

Most of the literature on optimal factor taxation has focused on long-run results, including the recent results on $\tau_\infty^k > 0$ and the previous section. We now turn to the analysis of the transition. We find that capital taxes have to be high for a large number of periods before becoming zero at $t = N + 1$. High capital taxes are needed to redistribute wealth in favor of workers and achieve a Pareto improvement. 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.

Further, N is larger for RPO allocations that favor more the workers, and it is very large for all POPI allocations. Recent results suggested a discontinuity for taxes depending on small changes in parameter values, for example, in SW small changes in parameter values may cause optimal τ_∞^k to jump from zero to its highest possible value. But we find that when taking into account the transition there is no discontinuity: small changes in parameters cause small changes in N.

We now present and discuss our numerical results in detail relying on the long-run results and the Algorithm described in Section 3. More details on our computational strategy are in Online Appendix A. We first explain how we calibrate the model. Then we examine the model with fixed labor supply. Section 4.3 shows the results for our baseline model.

4.1 Calibration

We calibrate the model at a yearly frequency. The parameter values are summarized 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, including effective tax rates, are taken from the dataset provided by Trabandt and Uhlig (2012).[30] We compute averages for the period 2001-2010. 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 the status-quo rate by assumption, i.e., we set $\widetilde{\tau} = 0.401$.

Table 1: Parameter values

	Parameter	
Preference parameters	β	0.96
	σ_c	2
	σ_l	3
	ω	845.4
Heterogeneity parameters	ϕ_w/ϕ_c	0.91
	$k_{c,-1}$	4.356
	$k_{w,-1}$	-1.136
Production parameters	α	0.394
	δ	0.074
Public sector	g	0.094
	k^g_{-1}	-0.315
	$\tilde{\tau}$	0.401

We set some preference parameters a priori. We use the usual discount factor β=0.96. Risk aversion is larger than log utility: $\sigma_c = 2$. The choice of $\sigma_l = 3$ generates an elastic supply of labor, and it prevents hours from greatly differing across consumers with different wealth.[31] Hence Frisch elasticity of labor supply is lower than in many real-business-cycle applications but is more in line with micro estimates.

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

[31]See for example GMV for a discussion of the trade-offs in choosing σ_l.

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

Our two types of consumers are heterogeneous in labor efficiency ϕ_j and initial wealth $k_{j,-1}$. GMV show that the relevant aspect of heterogeneity when studying proportional labor and capital income taxation is agents' wage-wealth ratio, a fact also used in Correia (2010). In our calibration we follow the calculations of GMV using 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. That is, capitalists are wealthier relative to their labor earnings potential, while both types of consumers work and save. Given this split of the population, the calibration proceeds as follows: (i) ϕ_w/ϕ_c is calibrated to the ratio that places in the numerator (denominator) the average wage of workers (capitalists), 0.91, and (ii) λ is calibrated to the ratio of consumptions, 0.54.[32]

Finally, we find ω, δ, g, k^g_{-1}, and the initial wealth of private agents in the model, $k_{c,-1}$ and $k_{w,-1}$, that are consistent with all chosen parameters, including ϕ_w/ϕ_c, and that status-quo equilibrium satisfies that (i) aggregate hours equal the fraction of time worked for the working age population, 0.245, (ii) the consumption ratio satisfies $\lambda = c_w/c_c = 0.54$, (iii) g over output equals 0.2, and (iv) k^g_{-1} over output matches the average public assets-GDP ratio from the data, -66.8 percent of GDP.[33]

4.2 Results with fixed labor supply

In the baseline case POPI plans differ from the first best for two reasons. First, as is standard in models of factor taxation, the need to raise tax revenue generates inefficiencies. Second, Pareto improvements may require redistribution and a further distortion. We first analyze a model with fixed labor supply, since in this version of the model distortions could be entirely avoided, hence it shows in a clean way the trade-off between efficiency and redistribution.

[32] The consumption ratio is measured by ratio of average total labour and capital income of each type, given actual asset holdings and their returns, see GMV for more details. This is reasonable because at steady state the ratio of incomes is equal to the consumption ratio. GMV reported the ratios for five quintiles. For our calibration we average out the numbers they report for each half of the population.

[33] As Table 1 shows, the initial wealth of workers turns out to be negative, i.e., workers are borrowers. Figure 5 confirms they stay as borrowers in the main calibration. Given our capital tax formulation, this means that workers receive a subsidy τ^k_t on their interest payments. One could argue that this is not a good way to model actual capital taxes, as subsidies to borrowing are limited. Removing the subsidy to borrowers would complicate somewhat the analysis: the feasible set of workers would have a kink, the ratio of consumptions would no longer be constant, and the subsidy would now depend on net borrowing taking into account ownership of assets, including real estate. This could cause a larger departure from the standard Chamley model, so we leave it for future research.

Formally, in this section we take $v(l) = 1$ and $l_{j,t} \leq \bar{l} = 0.245$, matching the fraction of hours worked. All parameters unrelated to the utility from leisure are as in Table 1.[34]

Under homogeneous agents and fixed labor supply the policy-maker would set $\tau_t^k = 0$, $\forall t$, collect all revenues from taxes on labor, and thus implement the first-best allocation. In a model with heterogeneous agents, this policy would avoid distortions but would pick a specific point on the frontier that is not necessarily a Pareto improvement, instead it might make workers worse off. The first best can only be implemented if the government in addition can stipulate agent-specific lump-sum transfers at time 0 (denoted T_w, T_c). But since we focus on the case $T_w = T_c = 0$, deviations from the first-best policy are necessary for distributive reasons if a Pareto improvement is to be achieved.

In Figure 1 we compare the set of POPI plans to the first best. Units in this graph are consumption-equivalent welfare gains.[35] The dashed (black) line labeled 'first-best PI' represents allocations with $\tau_t^k = 0$ for all t and optimal redistributive lump-sum transfers $T_w = -T_c$.[36] 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. This frontier is non-standard as it has an increasing part depicted with a dot-dashed (green) line, these points are not Pareto optimal, the POPI allocations coincide with the decreasing part of $\mathcal{F}$ depicted with a solid (blue) line.

Using Proposition 3 part 1, the decreasing part of $\mathcal{F}$ is found with $\psi > 0$ in MM, higher ψ corresponding to points further to the right along the solid (blue) line. Higher ψ 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 U^w max in Figure 1. At that point capital taxes are above zero for 41 years. 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 3 part 3. These equilibria are so inefficient that both agents' stance is worse than at the point U^w max.

Figure 1 clearly shows that the absence of lump-sum transfers generates large losses in efficiency. The worker has almost nothing to gain, even at the point U^w max, which requires $N = 41$. The utility loss is smaller if we give all the benefits of the reform to the capitalist, this requires $N = 26$ years.

This model shows in a clean way the trade-off between efficiency and redistribution that

[34]Notice that in the case of a fixed labor supply, the evolution of labor taxes is undetermined, only the net present value of labor taxes is determined.

[35]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.

[36]BB derive asymptotic results for fixed labor supply and lump sum universal taxes $T_w = T_c$.

we mentioned in the introduction: even though there is a policy that avoid all distortions, a period of high capital taxes is necessary for redistribution and to achieve a Pareto improvement. Because the need for redistribution is so high, N is very large for all POPI tax reforms. High capital taxes induce less investment for many periods, and the Pareto frontier is significantly below the first best.

4.3 Main results

We now return to our baseline model, which features elastic labor supply.

4.3.1 The welfare frontier and capital taxes

Figure 2 reports the set of POPI plans. The units in the axes are as in the previous figure. Again we contrast our main model with the case of redistributive lump-sum transfers $T_w = -T_c$. Note that the first best is not attained even with $T_w = -T_c$, because distortionary capital and/or labor taxes are still needed to raise tax revenue. First-best allocations would only be achieved with unconstrained T_w and T_c.

As with fixed labor supply, the absence of redistributive transfers clearly reduces the welfare gains achievable by POPI allocations, and capital taxes need to be high for a long time. However, the equilibrium frontier $\mathcal{F}$, the solid (blue) line in Figure 2, is now decreasing in the range of Pareto-improving allocations, it is now feasible to leave either the worker or the capitalist indifferent relative to the status quo. Furthermore, the total welfare loss relative to the case with transfers is now much lower, the two frontiers are relatively close to each other. In Section 4.3.3 we highlight that labor taxes play a crucial role for this.

The solutions behind the Pareto frontier in Figure 2 are all according to Step 1 for each ψ, the algorithm failed to find a solution when we tried to impose constraints $\Omega^l = 0$ and $c_t = \widetilde{c}$ for t large. SW find equilibria with $\tau^k_\infty > 0$ when debt is high, so in order to look for some solution according to Step 2 we explore what happens if initial government debt is higher than in our calibration. We have looked for solutions according to Step 2 fixing $\psi = 0.4$ and increasing the initial level of government debt, letting the algorithm find Ω^l.[37] In all the cases we found that $\Omega^l > 0$ always, and it is in fact increasing with debt, thus a solution according to Step 2 was also not found for high debt.

Now we compare some key characteristics of different points on the frontier. The length of the transition increases as welfare gains are shifted towards the worker. This is illustrated in the first panel of Figure 3 showing the duration of the transition, N, on the vertical axis for

[37] We impose asset market clearing, hence we decrease the initial capital stock at the same time.

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 16 to 24 years as we increase the welfare gain of the worker from zero (i.e., leaving the worker indifferent with the status quo) to 2.4 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 16.2 to 20.8 percent, as the second panel in Figure 3 reveals.[38] This shows that 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. The longer the period of high capital taxes, the less revenue has to be raised from labor taxes in present value, and the lower the relative tax burden of the worker.

More generally, our paper speaks to the issue of implementing 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 be considered unfair, and it certainly acts as an obstacle for the actual implementation of such reforms. Considering Pareto improvements addresses these issues. The above results show that a gradual reform towards $\tau^k_\infty = 0$ ensures that all consumers benefit and hence support the reform. This is in line with the literature on gradualism of political reforms, which has been at the center of some policy debates.[39] In light of this, 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.

The final panel in Figure 3 compares ψ and λ^{σ_c}, both normalized. Recall that $\lambda^{\sigma_c} = \psi$ would hold in a first-best situation without distortionary taxation or distributive conflict ($\Delta_1 = \Delta_2 = \gamma_t = 0, \forall t$), while in our second-best world the optimal choice of the consumption ratio λ is non-trivial, see Section 2.3.2. Figure 3 shows that as we increase the welfare of the worker, the marginal cost of doing so (as measured by ψ) increases rapidly, while λ^{σ_c} increases only mildly. This shows that it is very difficult to alter the ratio of consumptions even if the planner favors one type of consumers, given that the government only has access to proportional taxes to resolve issues of efficiency and redistribution.

If optimal lump-sum redistributive transfers across consumers are possible, the graphs in Figure 3 would look very different. In that case capital taxes are suppressed after 11 years for all ψ, and the share of capital taxes is always 12.5 percent. The multiplier ψ increases very

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

[39] 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.

little as the utility promise to the worker increases, while λ rises much more than without transfers. This is because shifting welfare gains and consumption between agents is much easier with redistributive lump-sum transfers, hence the planner lowers quickly capital taxes to achieve more efficiency. The policies and the path of aggregate variables is very similar along the Pareto frontier.

In Online Appendix B, we show that the main features described here are robust to some changes in parameter values. In particular, we consider two different measurements for the relevant tax rates and consumption inequality at the status quo. We also consider a case with higher inequality, calibrating $\phi_j/k_{j,-1}$ to the top and bottom quintiles of wage-wealth ratios. In addition, we consider all these scenarios for log utility ($\sigma_c = 1$). In all these cases the results are similar to the ones for the baseline calibration.

4.3.2 Welfare weights

Optimal policy with heterogeneous agents is often studied with fixed welfare weights, ψ. Some papers interpret ψ as arising from probabilistic voting or as the bias of the planner in favor of some agents. Most papers focus on the Benthamite case of $\psi = 1$, justified by a moral choice under the 'veil of ignorance'. Given our focus on Pareto-improving allocations, the value of ψ is determined in equilibrium, and there is no reason why $\psi = 1$ should reflect an equitable reform.

The focus of the literature on fixed welfare weights is not innocuous. Our results show how even if $\tau^k_\infty = 0$ holds at all RPO that we report, the interaction between redistribution and efficiency is a key issue. High capital taxes are optimal for a very long time, and the length of the transition increases gradually as the government redistributes more in favor of workers, as the first panel of Figure 3 shows. These features would be hidden by studying optimal policy in steady state with fixed ψ.

We now discuss the relationship between ψ and equity. We dub 'equitable reform' a RPO solution which implies that both agents gain more or less equally,[40] that is, points on the frontiers of Figures 1 and 2 which are near the 45° line. Figure 1 shows that with fixed labor supply the Benthamite policy is Pareto improving but gives most of the welfare gains to the capitalist. Even $\psi = \infty$ (corresponding to U^w max) does not achieve an equitable reform. This shows that a very large relative Pareto weight might be required in order to achieve an equitable reform. In the case of Figure 2 where labor supply is flexible, optimal policy for $\psi = 1$ is not even Pareto improving, a weight $\psi \in [0.35, 0.49]$ is needed for a Pareto

[40]Such a reform could be the outcome of a Nash bargaining game played by agents at $t = 0$ when both agents have a similar bargaining power and the outside option is the status quo.

improvement. This shows that $\psi = 1$ is not related to an equitable reform or even to a Pareto improvement. Benthamite policies can be located at arbitrary points on the frontier depending on the model and the calibration.

We further discuss how fixing ψ matters for time-consistency in Section 5.

4.3.3 The time path of the economy

The evolution of aggregate capital and labor, individual consumptions, tax rates, and government deficit are pictured in Figure 4. The three different paths in each panel show different policies along the POPI frontier, for $\psi = 0.3467$, 0.4000, and 0.4861. For $\psi = 0.3467$ ($\psi = 0.4861$), capitalists (workers) get all the benefits of the tax reform and workers (capitalists) are indifferent between the reform and the status quo, while $\psi = 0.4000$ is presented as an intermediate value.

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.

It is interesting to note that if labor supply is elastic low labor taxes weaken the efficiency-redistribution trade-off. Low taxes increase labor supply causing the return on capital to go up, increasing investment and achieving higher efficiency, while at the same time this policy redistributes wealth toward workers so as to achieve a Pareto improvement. Thus low initial labor taxes promote both efficiency and redistribution.[41] This explains why with flexible labor supply the POPI frontier is closer to the frontier with optimal lump-sum redistributive transfers than it is with fixed labor supply, compare Figures 1 and 2.

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 3 shows.

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 tax revenues fall again, the government deficit quickly reaches its long-run value, which can be positive or negative. We can also see

[41] Section III of Jones, Manuelli, and Rossi (1993) finds that in a model with homogeneous agents labor taxes should be very negative and capital taxes very high in the first period.

from Figure 4 that POPI policies imply that the government runs a primary surplus, hence is indebted in the long run. 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 Progressive taxation

Given that redistribution is a main theme of the paper, it might strike the reader as restrictive to allow only for flat-rate taxes. After all, one of the prime instruments of redistribution in the real world is progressive taxation. We now introduce progressive taxes in a simple way.

We assume that the planner can choose a uniform deductible D_t so that labor taxes paid at time t by agent j are given by $\tau_t^l(w_t\phi_l l_{j,t} - D_t)$, and similarly for capital taxes. As is well known, under complete markets any path for such deductibles is equivalent to a universal lump-sum transfer $\mathcal{D}$ in period 0. Using the notation in Section 3.1, this amounts to $-\mathcal{D} \equiv T_w = T_c$. Progressive taxation requires $\mathcal{D} \geq 0$. This tax scheme has been used extensively by the literature on taxation and by Werning (2007) and BB in models of optimal policy. Ramsey policy in this case is found by adding the term $u'(c_{1,0})(\Delta_1 + \Delta_2)\mathcal{D}$ to the $\mathbf{W}$-term in (13), and let the planner maximize over $\mathcal{D}$ additionally.

We find that if we restrict our attention to $\mathcal{D} \geq 0$ (progressive taxation), the optimal choice is to set $\mathcal{D} = 0$. Therefore, access to progressive taxation does not change any of our conclusions: optimal policy implies 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. On the other hand, productive efficiency recommends a negative $\mathcal{D}$, as this allows to raise revenue in a distortion-free manner. In the standard case of a representative-agent model only this second force is present, and it is well known that the first best can be achieved by choosing a negative $\mathcal{D}$ large 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. If the government set $\mathcal{D} > 0$ then marginal tax rates would have to increase, leading to more productive distortions.

If we remove the progressivity constraint, the government would set $\mathcal{D} < 0$. How can this 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 5 and 14 years at the two extremes of the POPI frontier. Welfare gains are larger than in the case with optimal lump-sum redistributive transfers: capitalists can gain maximum 4.0 percent and workers 6.2 percent in welfare-equivalent consumption units. Therefore, the optimal tax scheme would be extremely regressive. We think such a tax scheme is not implementable for reasons outside the model.[42]

This speaks to previous work on $\tau^k_\infty > 0$. BB find positive capital taxes for a case where $\mathcal{D}$ is positive and serves to redistribute toward wealth-poor agents (and the capital tax serves to raise revenue). But this means that BB consider a case where the total cost of distortions with $T_w = T_c = 0$ is negative, and therefore $\Delta_1 + \Delta_2 < 0$. While it may be possible to find such cases, in our calibrated model the scenario considered in BB does not occur.

A different scenario would occur if the government can set agent-specific transfers but is still restricted to progressive taxes, i.e., $\mathcal{D}_c, \mathcal{D}_w \geq 0$. As we mentioned after Proposition 1, we find $\Delta_2 < 0$ for most POPI allocations, in particular, whenever the worker's welfare gains are larger than 0.762 in Figure 2. That is, when the planner wishes to redistribute a lot toward workers, it needs a very long period of high capital taxes. It would be beneficial for all to just increase the worker's initial wealth, as this would make it easier to achieve a Pareto improvement. It is obvious that if $\Delta_2 < 0$, the government would choose $\mathcal{D}_w > 0 = \mathcal{D}_c$. Interestingly, the deductible is removed for high incomes in some modern income tax codes (the UK's, for example), which somewhat resembles this scheme. This raises a lot of interesting issues that we do not address any further in this paper.

To summarize, if the government is restricted to use progressive taxation with a common deductible for all agents, the government would set $\mathcal{D} = 0$, and optimal policy is as in our baseline model.

4.5 The evolution of wealth and welfare

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 by exploring the evolution of welfare and wealth. In the next section we address issues of time-consistency more formally.

[42]Recall that we have calibrated our model according to wage-wealth ratios, because, as shown in GMV, this is the appropriate criterion with flat rate taxes. In the real world some consumers with a high wage-wealth ratio are rich (young stockbrokers) and some consumers with a low wage-wealth ratio are poor (farmers in economically depressed areas). For the analysis of progressive taxation, the population should be classified also according to total income. We leave this issue for future research.

The time paths of consumers' welfare from period t onwards and wealth are plotted in Figure 5. 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 (6) and (7), 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. In equilibrium this happens because of the permanent income hypothesis: consumers anticipate future tax changes, therefore they save in early periods to pay for higher labor taxes in the long run and smooth consumption and labor.

5 Time-consistency

We now study formally time-consistency. In particular, we study whether the planner would reoptimize if a new plan is required to be approved by *consensus*, that is, if the new plan from period t onwards is Pareto improving relative to the original RPO chosen in period 0. We show that in this case time-consistency is restored if agents are sufficiently different. This contrasts with the well known result that under homogeneous agents, or with heterogeneous agents and fixed welfare weights, Ramsey tax policies are time-inconsistent.

Formally, we define 'consensus time-consistency' (CTC) as follows. Denote by $x^* = \left\{\tau_0^{k,*}, \lambda^*, \left\{c_{1,t}^*, k_t^*, l_{1,t}^*\right\}_{t=0}^{\infty}\right\}$ the RPO allocation given initial capital holdings and a certain value for $\underline{U}_2$. Assume that the optimal plan is followed for $Q-1$ periods and then in period Q the planner 'reoptimizes' taking as given initial conditions. A 'consensus reoptimized policy' in period Q, denoted $x^Q = \left\{\tau_Q^{k,Q}, \lambda^Q, \left\{c_{1,t}^Q, k_t^Q, l_{1,t}^Q\right\}_{t=Q}^{\infty}\right\}$, is a RPO with initial conditions $(k_{-1}, k_{1,-1}, k_{2,-1}) = (k_{Q-1}^*, k_{1,Q-1}^*, k_{2,Q-1}^*)$, the same consumption and tax limits $(\widetilde{c}, \widetilde{\tau})$, and for some value $\underline{U}_2$ such that both agents are better off from period Q onwards in the reoptimized policy, that is,

$$\sum_{t=0}^{\infty} \beta^t \left[u\left(c_{j,t+Q}^*\right) + v\left(l_{j,t+Q}^*\right)\right] \leq \sum_{t=0}^{\infty} \beta^t \left[u\left(c_{j,t+Q}^Q\right) + v\left(l_{j,t+Q}^Q\right)\right] \text{ for } j = 1, 2.$$

Let $x^{Q,*}$ denote the continuation of the RPO allocation from period $t = Q$ onwards, that is, $x^{Q,*} = \left\{\tau_Q^{k,*}, \lambda^*, \left\{c_{1,t}^*, k_t^*, l_{1,t}^*\right\}_{t=Q}^{\infty}\right\}$.

Definition 1. *A RPO policy x^* is consensus time-consistent (CTC) if, for all Q, the only consensus reoptimized policy is the continuation of the RPO, that is, $x^Q = x^{Q,*}$.*

In words, CTC holds if, in the future, there is no consensus in changing the policy.

We now show that the continuation $x^{Q,*}$ satisfies the FOCs of the reoptimization problem. Using superscripts * and ^{Q} to denote the variables and multipliers of the RPO and the reoptimized solution, respectively, define Δ_1^Q, Δ_2^Q, and ψ^Q (and the corresponding $\Omega^{c,Q}$ and $\Omega^{l,Q}$) such that

$$\begin{aligned} \frac{\Omega^{c,Q}}{\Omega^{l,Q}} &= \frac{\Omega^{c,*}}{\Omega^{l,*}}, \qquad (18) \\ \Delta_1^Q k_{1,Q-1}^* + \Delta_2^Q k_{2,Q-1}^* &= \frac{\Omega^{c,Q}}{\Omega^{c,*}} \gamma_{Q-1}^*, \qquad (19) \end{aligned}$$

and such that the FOC with respect to λ is satisfied from Q onwards for x^Q and the same λ, that is,

$$\begin{aligned} &\sum_{t=Q}^{\infty} \beta^t \Bigg[\left(\psi^Q \left(\lambda^*\right)^{-\sigma_c} + \Delta_2^Q \right) \left(c_{1,t}^*\right)^{1-\sigma_c} - \left(\psi^Q \mathcal{K}(\lambda^*)^{\sigma_l} + \Delta_2^Q \frac{\phi_2}{\phi_1} \right) \mathcal{K}'(\lambda^*) \omega \left(l_{1,t}^*\right)^{1+\sigma_l} \\ &- \frac{\Omega^{c,Q}}{\Omega^{c,*}} \gamma_{t-1}^* \left(c_{1,t}^*\right)^{-\sigma_c} F_{ke}\left(k_{t-1}^*, e_t^*\right) \frac{\phi_2}{2} \mathcal{K}'(\lambda^*) l_{1,t}^* \left(1 - \widetilde{\tau}\right) \qquad (20) \\ &- \frac{\Omega^{c,Q}}{\Omega^{c,*}} \frac{\mu_t^*}{2} \left(c_{1,t}^* - F_e\left(k_{t-1}^*, e_t^*\right) \phi_2 \mathcal{K}'(\lambda^*) l_{1,t}^* \right) \Bigg] = 0. \end{aligned}$$

These relations give three equations that the three unknowns Δ_1^Q, Δ_2^Q, and ψ^Q should satisfy given the known quantities x^*. If there is indeed such values Δ_1^Q, Δ_2^Q, and ψ^Q, we have the following.

Lemma 2. *Assume the conditions of Proposition 1. Fix a period Q. Assume that equations (18)-(20) have a solution for some Δ_1^Q, Δ_2^Q, and $\psi^Q \geq 0$. Then the RPO continuation $x^{Q,*}$ satisfies the FOCs of the consensus reoptimization problem.*

Proof. In Appendix B. □

Even though we have three equations to determineΔ_1^Q, Δ_2^Q, ψ^Q, there are some important cases when a solution for these multipliers will not exist. Barring these cases, we have the following.

Corollary 2. *Assume the conditions of Lemma 2 hold for all Q. Consider a RPO solution such that the consensus reoptimization problem has only one critical point for all Q, and that the Pareto frontier for the reoptimized problem is strictly decreasing at all periods. Then the RPO solution is CTC.*

Corollary 2 suggests that in order to sustain the full commitment tax reform, it is enough to require that the constitution can only be changed under wide consensus.[43]

[43]The result in Armenter (2007) is for a setup where workers cannot save, therefore it does not apply to

In some relevant cases the requirement in Lemma 2 that there is one solution for Δ_1^Q, Δ_2^Q, and $\psi^Q \geq 0$ is not satisfied. In particular, consider the case where the tax limit is not binding at Q ($\tau_Q^{k,*} < \widetilde{\tau}$) and agents are homogeneous. In this case it can be seen that for any ψ^Q we have $\frac{\Omega^{c,Q}}{\Omega^{l,Q}} = 1$, hence (18) cannot hold.[44] This recovers the well known result that the full commitment policy is time-inconsistent under homogeneous agents: a reoptimized policy would modify the full commitment solution by increasing capital taxes all the way to the limit and reset $\tau_Q^{k,Q} = \widetilde{\tau}$ in order to tax capital k_{Q-1}^*, which is supplied inelastically at $t = Q$. Therefore, as expected, in this case Lemma 2 does not apply.

However, if agents are sufficiently heterogeneous $\frac{\Omega^{c,Q}}{\Omega^{l,Q}}$ changes with ψ^Q.[45] In general, in this case $\psi^Q \neq \psi^*$ in order to satisfy (18)-(20).

It is interesting to point out that the time-consistency result in Corollary 2 obtains only because we express the consensus requirement in terms of utilities, i.e., both agents need to achieve a higher utility in case of reoptimization. If optimal policy was set with fixed welfare weights, we would instead find the traditional result that the planner prefers to reoptimize by raising capital taxes as much as possible. However, with weights adjusting endogenously under CTC, if both agents are rational voters, we have time-consistency.

This shows another reason why using fixed welfare weights is not innocuous. One needs to take into account how future reexamination of policy may change the welfare weights. We have checked that Lemma 2 applies in our model: a positive ψ^Q could always be found and it always turned out to be smaller than ψ in the cases we studied. For instance, in the case of $\psi = 0.40$ and reoptimization in period $Q = 5$, the continuation utilities are respected if $\psi^Q = 0.33$. If reoptimization occurs at the steady state, $\psi^Q = 0.30$. These results mean that consensus amounts to lowering the influence of the worker on the welfare function at the time of reoptimization. The latter result also means that in our calibrated model there

our paper. Armenter's result is stronger in the sense that it does not require the caveats of our Corollary 2. Unfortunately, the proof strategy in that paper cannot be adapted to our model. Armenter argues that one can piece together the reoptimized policy at Q with the RPO allocation in previous periods. This is the standard approach to prove time-consistency in dynamic programming. But since λ^Q is in principle different from λ^*, the pieced-together allocation would imply a jump in λ in period Q, and, therefore, it is not a competitive equilibrium from the point of view of period zero.

[44] To complete the argument, note that homogeneous agents implies $\lambda = 1$, and hence (under the very mild requirement that $k_{j,Q-1} \neq 0$) (19) implies $0 = \Delta_1^Q k_{1,Q-1}^* + \Delta_2^Q k_{1,Q-1}^* = \Delta_1^Q + \lambda \Delta_2^Q$, and the Δ's drop out from the Ω's. Then, for any ψ^Q we have $\frac{\Omega^{c,Q}}{\Omega^{l,Q}} = 1$ and if $\gamma_{Q-1} = 0$ this generically cannot equal

$$\frac{\Omega^{c,*}}{\Omega^{l,*}} = \frac{1 + \psi^* + 2\Delta_1^* (1 - \sigma_c)}{1 + \psi^* + 2\Delta_1^* (1 + \sigma_l)}$$

as is required by (18).

[45] With heterogeneity and $\frac{k_{2,-1}}{k_{1,-1}} \neq \frac{\phi_2}{\phi_1} \neq \lambda$, even if $\Delta_1^Q k_{1,Q-1}^* + \Delta_2^Q k_{1,Q-1}^* = 0$, we still have $\Delta_1^Q + \lambda \Delta_2^Q \neq 0$, and we may find a ψ^Q satisfying (18).

is sufficient heterogeneity so that the solution is time-consistent even when the capital tax has dropped to zero.

6 Conclusion

We study the efficiency-equity trade-off in setting capital and labor taxes. We first show that the traditional result $\tau^k_\infty = 0$ reemerges if one imposes reasonable constraints on policy, in particular, if the government is prevented from immiserating consumers and with a tax limit below 100 percent. Hence $\tau^k_\infty = 0$ seems a more robust result than recent papers suggest.

However, $\tau^k_\infty = 0$ does not mean that generically low capital taxes are good for all agents. To achieve an optimal Pareto-improving policy, capital taxes should be high (and labor taxes low) for a very long time before they become zero (high), thus an equity-efficiency trade-off is resolved during the transition. With an elastic labor supply the efficiency-equity trade-off is less pronounced and the loss from redistribution is lower. This is because lower labor taxes during the transition both promote wealth redistribution and boost investment. The results are robust to variations in parameter values and even to the introduction of progressive taxes. 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.

We also find that results with fixed welfare weights can be misleading. In addition, Benthamite policies can be far from equitable, and they can hurt large parts of the population. The solution is time-consistent if consensus is required at the time of reoptimization and if agents are sufficiently different. The welfare weight should adjust to avoid time-inconsistency. Therefore, if policies can only be overturned by consensus, the optimal tax reform is credible.

Our analysis suggests that issues of redistribution are crucial in designing optimal policies involving capital and labor taxes, even when $\tau^k_\infty = 0$. 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 workers for the elimination of capital taxes that are less costly in terms of efficiency, for example, promoting certain types of government spending, cuts to other taxes, or introducing other types of progressivity. The transition in our model is very long, therefore partial credibility on the veto power of all groups or the absence of rational expectations might render this policy ineffective in practice. Introducing partial credibility, learning about expectations, and political economy in the determination of optimal taxes would therefore be of interest and might influence optimal policy.

References

Ábrahám, Á. and E. Cárceles-Poveda (2006). Endogenous Incomplete Markets, Enforcement Constraints, and Intermediation. *Theoretical Economics 1*(4), 439–459.

Ábrahám, Á. and S. Laczó (2018). Efficient Risk Sharing with Limited Commitment and Storage. *Review of Economic Studies 85*(3), 1389–1424.

Albanesi, S. and R. Armenter (2012). Intertemporal Distortions in the Second Best. *Review of Economic Studies 79*(4), 1271–1307.

Alvarez, F. and U. J. Jermann (2000). Efficiency, Equilibrium, and Asset Pricing with Risk of Default. *Econometrica 68*(4), 775–797.

Armenter, R. (2007). Time-Consistent Fiscal Policy and Heterogeneous Agents. *Review of Economic Dynamics 10*(1), 31–54.

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.

Benhabib, J. and A. Rustichini (1996). Social Conflict and Growth. *Journal of Economic Growth 1*(1), 125–142.

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

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

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

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.

Kocherlakota, N. R. (1996). Implications of Efficient Risk Sharing without Commitment. *Review of Economic Studies 63*(4), 595–609.

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.

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

Marcet, A. and R. Marimon (1992). Communication, Commitment, and Growth. *Journal of Economic Theory 58*(2), 219–249.

Marcet, A. and R. Marimon (2019). Recursive Contracts. *Econometrica 87*(5), 1589–1631.

Park, Y. (2014). Optimal Taxation in a Limited Commitment Economy. *The Review of Economic Studies 81*(2), 884–918.

Reinhorn, L. J. (2019). On Optimal Redistributive Capital Taxation. *Journal of Public Economic Theory 21*(3), 460–487.

Straub, L. and I. Werning (2019). Positive Long-Run Capital Taxation: Chamley-Judd Revisited. *American Economic Review* (forthcoming).

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 can be written as

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

given $\gamma_{-1} = 0$ and with $\xi_t,\ \gamma_t,\ \mu_t \geq 0,\ \forall t$, and $\psi \geq 0$, with complementary slackness conditions.

The FOCs are:

- for consumption at $t > 0$:

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

- for consumption at $t = 0$ γ_{t-1} is replaced by $\left(\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}\right)$ and $\widetilde{\tau}$ by τ_0^k

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

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

- for labor at $t = 0$ γ_{t-1} is replaced by $\left(\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}\right)$ and $\widetilde{\tau}$ by τ_0^k in the third line

- for capital at $t \geq 0$:

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

- for the multiplier of the promise-keeping constraint:

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

- for relative consumption, λ:

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

- for γ_t at $t \geq 0$:

$$\text{either} \quad \gamma_t > 0 \quad \text{and} \quad u'(c_{1,t}) = \beta u'(c_{1,t+1}) [1 + (r_{t+1} - \delta)(1-\widetilde{\tau})],$$
$$\text{or} \quad \gamma_t = 0 \quad \text{and} \quad u'(c_{1,t}) \geq \beta u'(c_{1,t+1}) [1 + (r_{t+1} - \delta)(1-\widetilde{\tau})].$$

- for Δ_j: the corresponding lifetime budget constraint.

B Proofs

Proof of Lemma 1. Assume that $\tau_\infty^k > 0$. Taking limits in (3) gives

$$\beta \left[1 + (F_k(k^{ss}, e^{ss}) - \delta)(1 - \tau_\infty^k) \right] = 1.$$

Then using (14) we have $\beta(1 + F_k(k^{ss}, e^{ss}) - \delta) > 1$, hence there is a constant A such that $1 > A > \frac{1}{\beta(1-\delta+F_k(k^{ss}, e^{ss}))}$. Obviously,

$$1 > A > \frac{1}{\beta(1-\delta + F_k(k_t, e_{t+1}))} \quad \text{for } t \text{ large enough.} \tag{21}$$

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

$$\mu_t \frac{1}{\beta(1-\delta + F_k(k_t, e_{t+1}))} + \gamma_t \frac{u'(c_{1,t+1}) F_{kk}(k_t, e_{t+1})(1-\widetilde{\tau})}{1-\delta + F_k(k_t, e_{t+1})} = \mu_{t+1}. \tag{22}$$

We have $F_{kk}(k,e) \leq 0$ by concavity and $\gamma_t \geq 0$, hence the second term on the left-hand side is non-positive. This, together with $\mu_t \geq 0$ and (21), implies that for t large enough

$$\mu_t A \geq \mu_{t+1}.$$

Since $A < 1$ and $\mu_t \geq 0$, this proves that $\mu_t \to 0$.

To prove $\gamma_t \to 0$ when $\tilde{\tau} < 1$ we plug $\mu_t \to 0$ into (22) to obtain

$$\gamma_t \frac{u'(c_{1,t+1}) F_{kk}(k_t, e_{t+1})(1-\widetilde{\tau})}{1-\delta+F_k(k_t, e_{t+1})} \to 0. \tag{23}$$

Now we show that the term multiplying γ_t in (23) cannot go to zero. First we prove that the denominator cannot go to infinity: feasibility and $c^{ss} > 0$ imply $\frac{1+\lambda}{2}c^{ss} + g + \delta k^{ss} = F(k^{ss}, e^{ss}) > 0$, hence by **A1** $e^{ss}, k^{ss} > 0$. Therefore, $F_k(k^{ss}, e^{ss}) < \infty$ and the denominator of the term multiplying γ_t is finite. To prove that the numerator cannot go to zero, note that we also need $F_{kk}(k^{ss}, e^{ss}) < 0$. Even if F is strictly concave we could have $F_{kk}(k^{ss}, e^{ss}) = 0$ for $e^{ss} = 0$. But we have already proved $e^{ss}, k^{ss} > 0$, therefore $F_{kk}(k^{ss}, e^{ss}) < 0$. Then, **A2**, $c^{ss} > 0$, and $\widetilde{\tau} < 1$ give

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

hence (23) implies $\gamma_t \to 0$. □

Proof of Proposition 1. For the utility function in **A1** the FOC for labor at $t > 0$ is

$$\begin{aligned} &\Omega^l \omega (l_{1,t})^{\sigma_l} + \gamma_{t-1} (c_{1,t})^{-\sigma_c} F_{ke}(k_{t-1}, e_t) \frac{1}{2} (\phi_1 + \phi_2 \mathcal{K}(\lambda))(1-\widetilde{\tau}) \\ &= F_e(k_{t-1}, e_t) \frac{1}{2} (\phi_1 + \phi_2 \mathcal{K}(\lambda)) \mu_t. \end{aligned} \tag{24}$$

Assume $c^{ss} > 0$. Lemma 1 guarantees that $\mu_t, \gamma_t \to 0$. In the proof of Lemma 1 we already showed $e^{ss}, k^{ss} > 0$, therefore $F_e(k^{ss}, e^{ss}) < \infty$. Also, differentiating both sides of $F_k k + F_e e = F$ with respect to e gives $F_{ke}(k^{ss}, e^{ss}) < \infty$. Putting all this together, taking limits on both sides of (24), we have $\Omega^l \omega (l_1^{ss})^{\sigma_l} \to 0$. Then, given that $\Omega^l \neq 0$, this implies $e^{ss} = F(k^{ss}, e^{ss}) = 0$, which is impossible since it violates feasibility.

For the case $\Omega^c > 0$ the FOC for consumption and Lemma 1 imply $\lim \xi_t < 0$ which is impossible since $\xi_t \geq 0$.

Therefore it is impossible that $c^{ss} > 0$ and $\tau^k_\infty > 0$. This proves part a).

Part b) is a corollary of part a).

Now we prove part c). Given $c^{ss} > \widetilde{c} \geq 0$, part a) implies $\tau^k_t \to 0$, hence there is a finite integer N such that

$$\tau^k_t < \widetilde{\tau} \textit{ and } c_{1,t+1} > \widetilde{c}, \ \forall t \geq N.$$

Pick the smallest N for which these inequalities both occur, ie. $\gamma_t = \xi_{t+1} = 0$ for all $t \geq N$. Then the FOC with respect to consumption (see Appendix A) with the utility function in **A1** gives

$$\Omega^c \left(c_{1,t}\right)^{-\sigma_c} = \mu_t \frac{1+\lambda}{2}, \ \ \forall t \geq N+1.^{46} \tag{25}$$

Plugging (25) into the FOC with respect to capital and using $\gamma_t = 0$ again, we get

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

It is clear that this equation is only compatible with the Euler equation of the consumer, (3), if (15) holds. □

Proof of Corollary 1. It is trivial to see that under **A3**a) we have $\Delta_1 + \frac{\phi_2}{\phi_1}\mathcal{K}(\lambda)\Delta_2 > 0$ so that $\Omega^l > 0$, and under **A3**b) we have $\Delta_1 + \lambda\Delta_2 > 0$ so that $\Omega^c > 0$, hence the conditions of Proposition 1 are satisfied. □

Proof of Proposition 2. Constant returns to scale implies $F_k k + F_e e = F$. Differentiating both sides with respect to k gives $F_{kk}k + F_{ek}e = 0$. Also, we have $e_t, k_t > 0$ for all t, otherwise we would have $c_t = 0$ for some t and utility would be $-\infty$. Therefore, strict concavity of F gives $F_{ke}(e_t, k_t) > 0$ for all t. Therefore, if $\Omega^l = 0$ the FOC for labor (24) implies that either $\gamma_{t-1} = \mu_t = 0$ or both $\gamma_{t-1}, \mu_t > 0$ for any $t > 0$. Consider now two cases for μ_0.

Case 1: $\mu_0 > 0$. If $\mu_{t-1} > 0$ the FOC for capital at $t-1$ would be violated if $\gamma_{t-1} = \mu_t = 0$. Therefore, if $\mu_{t-1} > 0$ then $\gamma_{t-1}, \mu_t > 0$. By induction we have that $\gamma_{t-1}, \mu_t > 0$ for all $t > 0$, hence $\tau_t^k = \widetilde{\tau}$ for all t.

Case 2: $\mu_0 = 0$. In the FOC for capital at t the term multiplying γ_t is negative, while the term multiplying μ_{t+1} is positive. Therefore, if $\mu_t = 0$ then $\gamma_t = \mu_{t+1} = 0$. This implies by induction that $\mu_t = 0$ for all t, which is incompatible with the assumption stated.

Therefore $\tau_t^k = \widetilde{\tau}$ for all t. In this case Lemma 1 applies so that $\mu_t, \gamma_t \to 0$. Then, the case $\Omega^c > 0$ can be discarded as the consumption FOC would imply $\xi_t < 0$ for t large which is impossible. If $\Omega^c \leq 0$ then $\xi_t = 0$ for t large can also be discarded because the FOC for consumption and leisure impose a singularity in the steady states. If $\Omega^c < 0$ consumption FOC implies $\xi_t \to> 0$. In the last two cases we conclude that $c_t = \widetilde{c}$ for t large. □

Proof of Proposition 3. The proof of part 1 is obvious, it is only stated for future reference. Part 2 is less obvious as there could be a duality gap. Consider a pair of utilities $\left(\overline{U}_1, \overline{U}_2\right) \in \mathcal{S}^U$ that correspond to a RPO allocation. Invertibility in **A4** guarantees that there is a $\bar{\psi}$ such that $\overline{U}_2 = U_2\left(\bar{\psi}\right)$. If $\bar{\psi}$ is finite we have

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

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

since the equilibrium that gives rise to $\left(\overline{U}_1, \overline{U}_2\right)$ 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 $\overline{U}_2 = U_2\left(\bar{\psi}\right)$, the above inequality implies $\overline{U}_1 \leq U_1\left(\bar{\psi}\right)$. But the fact that $\left(\overline{U}_1, \overline{U}_2\right)$ is the utility of a RPO allocation implies $\overline{U}_1 \geq U_1\left(\overline{\psi}\right)$. Therefore, the RPO allocation with utilities $\left(\overline{U}_1, \overline{U}_2\right)$ 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.

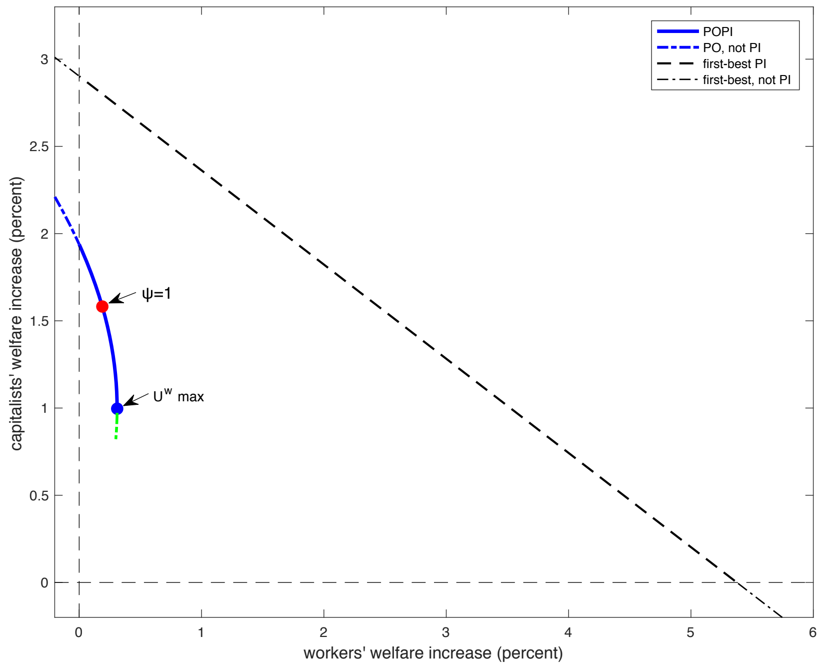
Let us now consider part 3. If $\psi \geq 0$ part 3 follows from part 2. Consider now a given $\psi < 0$. We can find points in $\mathbb{R}^2$ outside $\mathcal{S}^U$ which are arbitrarily close to $\left(U_1\left(\psi\right), U_2\left(\psi\right)\right)$ as follows: for any $\varepsilon > 0$ we have $\left(U_1\left(\psi\right) + \varepsilon, U_2\left(\psi\right) - \varepsilon\right) \notin \mathcal{S}^U$, since this point achieves a higher value of the objective function of MM than its maximum. Since $\left(U_1\left(\psi\right) + \varepsilon, U_2\left(\psi\right) - \varepsilon\right)$ can be made arbitrarily close to $\left(U_1\left(\psi\right), U_2\left(\psi\right)\right)$, this last point is on the frontier $\mathcal{F}$. $\square$

Proof of Lemma 2. Set the multipliers

$$(\gamma_t^Q, \mu_t^Q, \xi_t^Q) = \frac{\Omega^{c,Q}}{\Omega^{c,*}}(\gamma_t^*, \mu_t^*, \xi_t^*) \text{ for } t \geq Q.$$

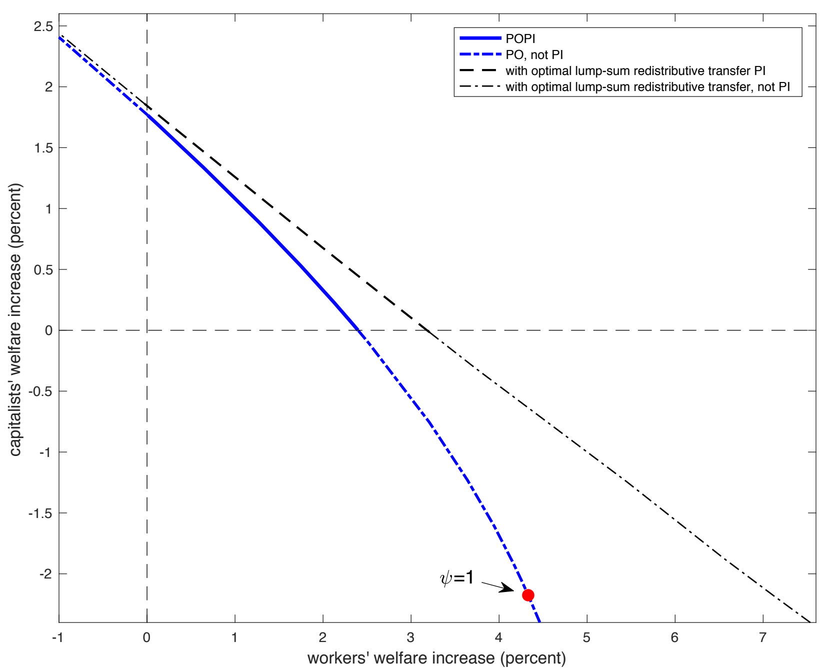
It can be checked routinely that these multipliers and the continuation x_Q^* satisfy the FOCs of the consensus reoptimized problem. This is because multiplying both sides of all FOCs in the RPO problem by $\frac{\Omega^{c,Q}}{\Omega^{c,*}}$ gives the FOCs for the consensus reoptimization problem for the above ^{Q} multipliers and for the continuation of the Ramsey allocation. Furthermore, it is clear that all resource and budget constraints, plus the consumption and tax limits are satisfied. The consensus requirement obviously holds for x_Q^*. $\square$

Figure 1: The Ramsey Pareto frontier of Pareto-improving equilibria with fixed labor supply



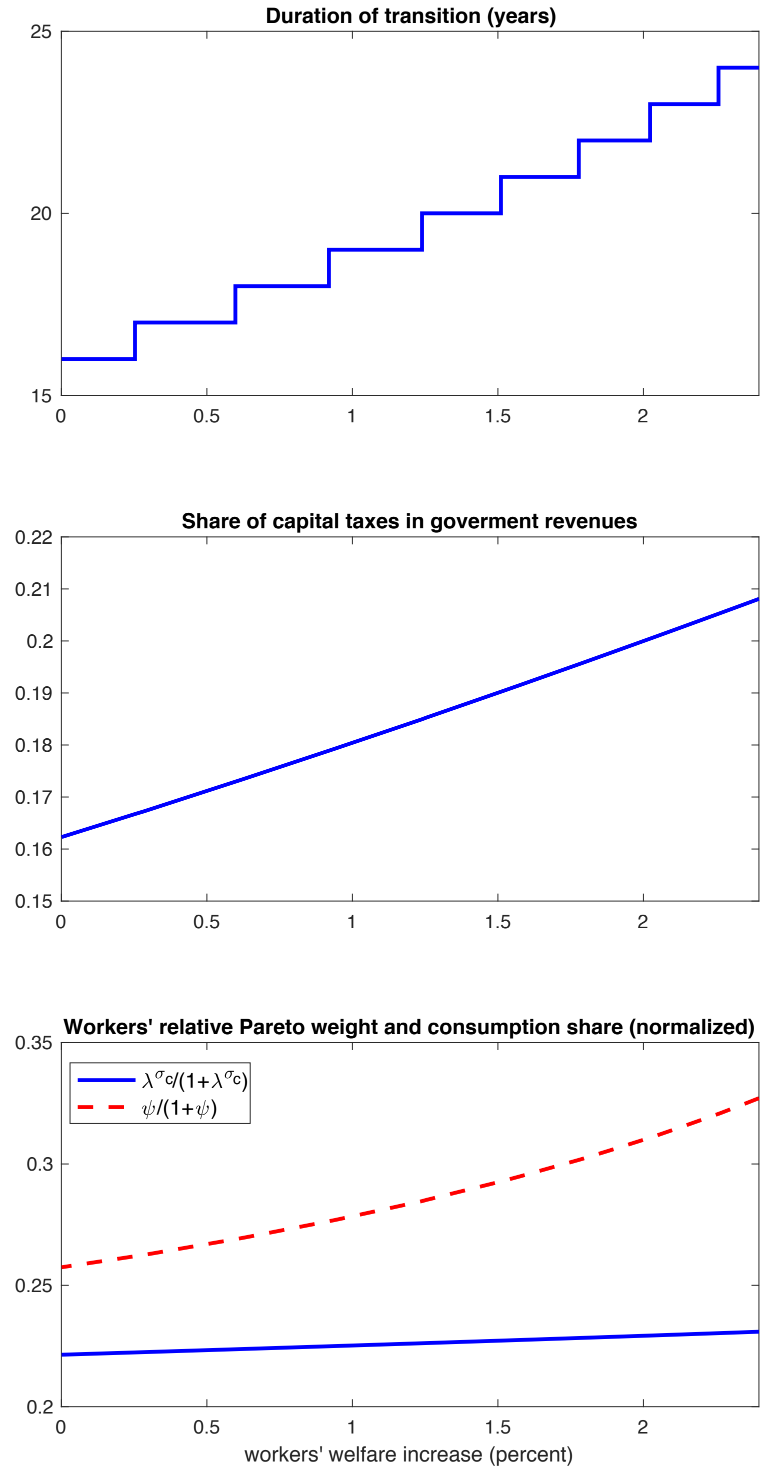
Notes: Welfare is measured as the percentage increase in status-quo consumption that would give the consumers the same lifetime utility as the optimal tax reform. The point $\psi = 1$ corresponds to the Benthamite policy, and the point U^w max represents the case where workers' utility is highest, i.e., $\psi \to \infty$.

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 Benthamite policy.

Figure 3: Properties of POPI tax reforms in the baseline model



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.

Figure 4: The time paths of selected variables for three POPI plans in the baseline model

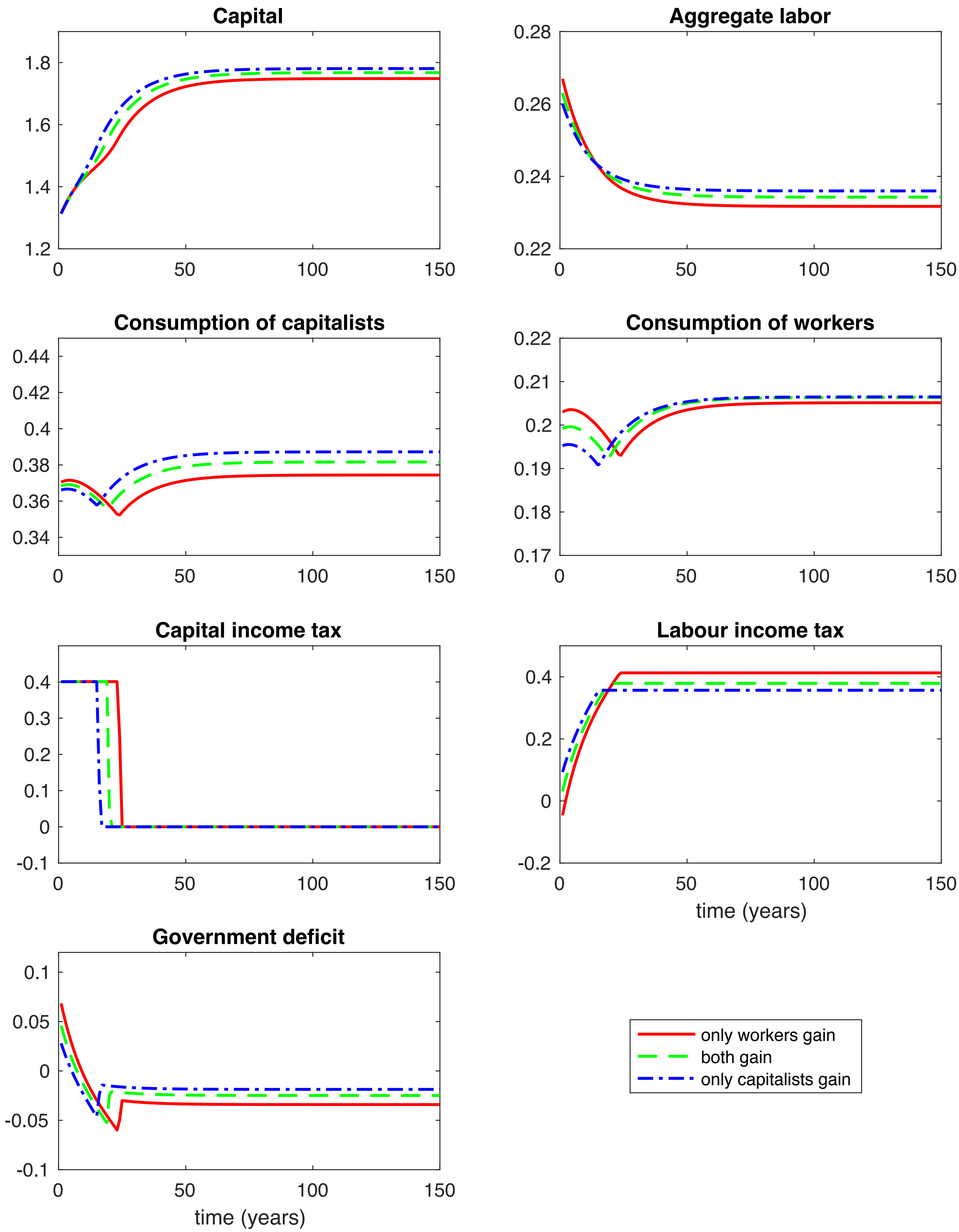
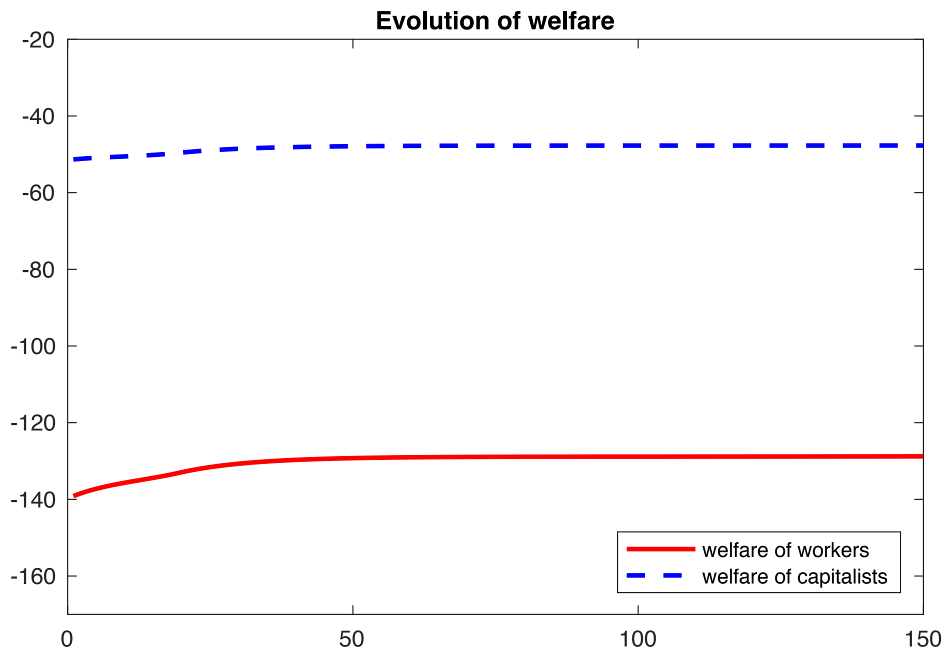
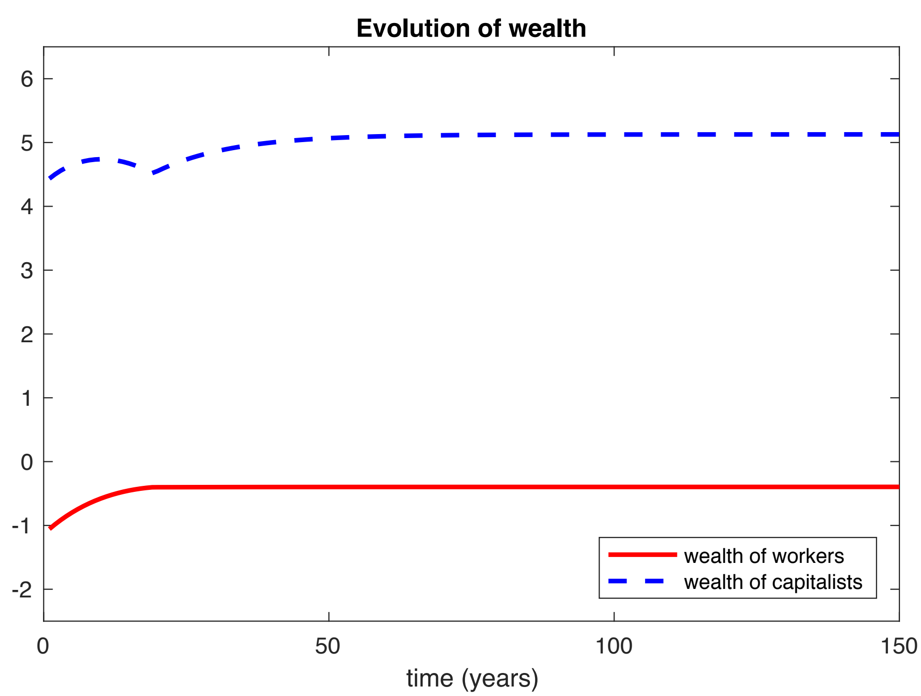


Figure 5: Typical time paths for consumers' welfare and wealth





Online Appendices

for "Pareto-Improving Optimal Capital and Labor Taxes"

by Katharina Greulich, Sarolta Laczó, and Albert Marcet

1 Computational strategy

In this online appendix, we detail our computational strategy. Whenever we mention a "FOC with respect to..." in this appendix, we refer to the FOCs of the Ramsey planner's problem, which are in Appendix A in the paper.

1. Choose T, the number of periods after which the steady state is assumed to be reached. (We use $T = 150$.) All parameter values, initial conditions, and functional forms are taken as given. Fix ψ.

2. Propose as a candidate solution a $(3T + 3)$-dimensional vector

   $X = \{k_0, ..., k_{T-1}, e_0, ..., e_{T-1}, \gamma_0, ..., \gamma_{T-1}, \Delta_1, \Delta_2, \lambda\}$.[1]

3. For each $(\Delta_1, \Delta_2, \lambda)$, find the steady state with either $\tau^k_\infty = 0$ or $\tau^k_\infty = \widetilde{\tau}$. In the first case, we set $\gamma^{ss} = 0$ and find $(k^{ss}, c^{ss}, e^{ss}, \mu^{ss})$ using the FOCs for consumption, labor, and capital, and the resource constraint. In the second case, we set $c^{ss} = \widetilde{c}$ and $\gamma^{ss} = \mu^{ss} = 0$ and find (k^{ss}, e^{ss}) using the consumer's Euler equation with $\tau^k_\infty = \widetilde{\tau}$ and the resource constraint. We then set time-T variables to these steady-state values.

4. For each candidate solution X, we compute c_t from the resource constraint and μ_t from the FOC for labor for $t = 0, ..., T-1$. Thus the resource constraint and FOC for labor always hold as equality. Obviously $\{r_t, w_t, F_{kl,t}, F_{kk,t}\}$ in the FOCs are found using the production function.

5. Then, given X, we set up a system of $3T + 3$ equations to solve for X that satisfies

   $$G(X) = 0,$$

   where the function G captures that FOCs with respect to capital and consumption, Kuhn-Tucker conditions, and discounted sums equations have to hold.

[1] Note that this is not the minimal number of variables we could find solving a fixed point problem. $2T+3$ would be sufficient if we solved out μ_t. However, convergence is better if the approximation errors are spread over a larger number of variables.

More precisely, we take $G = (G_1, G_2, G_3, G_4)$, where $G_i : R^{3T+3} \to R^T$ for $i = 1, 2, 3$, and $G_4 : R^{3T+3} \to R^3$. Letting $\widetilde{\gamma}_t(X)$ be the value of γ_t that solves exactly the FOC for consumption given $(k_{t-1}, l_t, \gamma_{t-1}, \Delta_1, \Delta_2, \lambda)$ in the candidate solution X, the elements of G are defined as follows:

- $G_{1,t} = \widetilde{\gamma}_t(X) - \gamma_t$, $t = 0, ..., T-1$.
- Let $I_+(x)$ be the indicator function of $[0,\infty)$ and $IND_t(X)$ be defined, as a function of the candidate solution, by

$$IND_t(X) = I_+\left(\frac{u'(c_{1,t})}{u'(c_{1,t+1})} - \beta\left[1 + (r_{t+1} - \delta)(1 - \widetilde{\tau})\right]\right).$$

  Then

$$G_{2,t}(X) = IND_t(X)\gamma_t + (1 - IND_t(X))\left\{\frac{u'(c_{1,t})}{u'(c_{1,t+1})} - \beta\left[1 + (r_{t+1} - \delta)(1 - \widetilde{\tau})\right]\right\},$$

  $t = 0, ..., T-1$.
- $G_{3,t}$ sets the FOC for capital to zero when $\widetilde{\gamma}_t(X)$ is introduced in the FOC, $t = 0, ..., T-1$, i.e.,

$$G_{3,t} = \mu_t + \widetilde{\gamma}_t(X)\beta(c_{1,t+1})^{-\sigma_c} F_{kk}(k_t, e_{t+1})(1 - \widetilde{\tau}) - \beta\mu_{t+1}(1 - \delta + F_k(k_t, e_{t+1})).$$

- G_4 sets the life-time budget constraints of both agents and the FOC with respect to λ to zero.

Note that the FOCs for consumption and labor at $t = 0$ differ from the FOC in later periods, see Appendix A.

We use a trust-region dogleg algorithm and Broyden's algorithm, repeatedly when necessary, to solve this system of $(3T + 3)$ equations. We thank Michael Reiter for providing us his implementation of Broyden's algorithm.

Notice that the above algorithm imposes (up to the precision of the solution) that $\tau_t^k \leq \widetilde{\tau}$, but it does not impose $\gamma_t \geq 0$. We check ex post that the last inequality holds for all t. It did for all cases when we found a solution to this system of equations.

2 Sensitivity analysis

We check the sensitivity of our results to the measurement of relevant tax rates and 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. Note that, given the calibration strategy that we use (described in Section 4.1), considering different values of λ_{SQ} and taxes in effect changes the distribution of private wealth of agents $k_{j,-1}$ in all of the alternative calibrations considered in this robustness exercise.

The benchmark case splits the observed population into two groups that have above or below median wage-wealth ratio. In the real world there is a very large heterogeneity of wage-wealth ratios even within each of these groups. Therefore, the Pareto-improving allocations that we compute in the text could worsen the welfare of agents further out in the distribution. As another robustness check, we recalibrate the heterogeneity parameters in our model to the top and bottom quintiles of the wage-wealth distribution in the PSID. That is, now agent w (c) represents families in the group of 20% highest (lowest) wage-wealth ratios, rather than the top and bottom half as in the main text. This affects the calibration of wages as well as initial wealth. For bottom and top quintiles GMV report that ${}^{\phi_w}/_{\phi_c} = 0.95$ and $\lambda_{SQ} = 0.31$, see their Tables 2 and 3.

Table 1 summarizes some aspects of the simulations for these alternative parameters, changing the values of parameters one at a time relative to the baseline calibration. It reports 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 policies as for the baseline calibration described in Section 4, and in some cases the results are reinforced, as the transition is even longer.

We have also solved our model with $u() = \log()$ for a baseline calibration and these parameter changes. The results are presented in Table 2.

Table 1: Sensitivity analysis, $\sigma_c = 2$

Calibration	Workers gain as much as possible		Capitalists gain as much as possible	
	duration of transition (years)	revenue share of τ^k(%)	duration of transition (years)	revenue share of τ^k(%)
Baseline	24	20.8	16	16.2
$\tau^k_{SQ} = 0.3$	33	25.1	22	20.7
$\tau^k_{SQ} = 0.57$	26	30.7	12	20.1
$\tau^l_{SQ} = 0.15$	35	38.9	20	29.7
$\tau^l_{SQ} = 0.3$	22	15.0	14	11.6
$\lambda_{SQ} = 0.5$	24	20.6	16	16.5
$\lambda_{SQ} = 0.6$	25	21.2	15	15.2
High inequality	26	19.4	16	14.8

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

Table 2: Sensitivity analysis, $\sigma_c = 1$

Calibration	Workers gain as much as possible: duration of transition (years)	Workers gain as much as possible: revenue share of τ^k(%)	Capitalists gain as much as possible: duration of transition (years)	Capitalists gain as much as possible: revenue share of τ^k(%)
Baseline	26	21.7	11	12.7
$\tau^k_{SQ} = 0.3$	35	25.5	17	16.8
$\tau^k_{SQ} = 0.57$	17	15.4	7	7.7
$\tau^l_{SQ} = 0.15$	30	36.1	13	21.6
$\tau^l_{SQ} = 0.3$	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
High inequality	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.

3 Alternative solution strategies for RPO allocations

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

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

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

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

For reference, we repeat here two equilibrium conditions we use, equations (6) and (7) in the main text:

$$c_{2,t} = \lambda c_{1,t},\ \forall t, \tag{1}$$

and

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

Flodén writes the planner's problem as Atkeson, Chari, and Kehoe (1999), ACK hereafter, by keeping consumption of all agents in the equilibrium conditions, instead of summarizing the allocations of other agents using (1), (2), and λ, as we do. Although this makes computations different, it should not affect the allocations found. We describe this approach in detail below.

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

We use the notation

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

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

Using an ACK Lagrangian

Instead of representing equilibrium conditions with (1) and (2), as we do, Flodén follows ACK and the keeps equilibrium conditions

$$\frac{u_{1c,t}}{u_{1c,t+1}} = \frac{u_{jc,t}}{u_{jc,t+1}} \text{ and } \frac{u_{1l,t}}{u_{1c,t}\phi_1} = \frac{u_{jl,t}}{u_{jc,t}\phi_j}, \ \forall j, \tag{3}$$

as separate constraints in the planner's problem. Feasibility, firm behavior, and budget constraints are as in the main text of our paper. For simplicity we do not consider consumption limits or tax limits in this appendix.

We focus on the case where $\widetilde{\lambda}$ 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.

The Lagrangian to find the RPO allocations using this approach is

$$\begin{aligned}\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \Bigg\{ \Bigg(\sum_{j=1}^{J} \psi_j U_{j,t} + \Delta_j \left[U_{j,t}(1-\mu) + \frac{u_{1c,t}\zeta l_{j,t}^{1/\gamma+1}}{\gamma+1} \right] \\ + \rho_{jt}\left[u_{1c,t}u_{jc,t+1} - u_{jc,t}u_{1c,t+1}\right] + \xi_{jt}\left[u_{1c,t}u_{jl,t}\phi_1 - u_{1l,t}u_{jc,t}\phi_j\right] \Bigg) \\ + \mu_t \left(\sum_{j=1}^{J} \widetilde{\lambda}_j c_{j,t} + g + k_t - (1-\delta) k_{t-1} - F(k_{t-1}, e_t) \right) \Bigg\} + \sum_{j=1}^{J} \Delta_j W_{j,-1}.\end{aligned} \tag{4}$$

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

We prefer representing CE in the main text using (1) and (2) to substitute out agent 2's consumption and labor because then the planner's problem can be written as a maximization over $\tau_0^k, \lambda, \{c_t^1, k_t, l_t^1\}_{t=0}^{\infty}$. This reduces enormously the number of variables and multipliers to be computed, and it is much more convenient for computation. More precisely, given the algorithm described in Appendix 1, the number of variables to solve for with J agents would be $(2+J) \times T + 2 + J$ using the ACK approach, while using our approach the number of variables to compute is only $3T + 2 + J$. But solving the Lagrangian (4) is equally valid, and it should give the same solution as we find. Hence in this appendix we characterize RPO solutions to (4), as is done in Flodén (2009).

Using a representative agent

Flodén actually uses a modification of the above Lagrangian applying his Proposition 3. This proposition says that CE constraints can be summarized in an implementability constraint of a representative agent (RA) who has productivity $\phi^{RA} \equiv \left(\sum_{j=1}^{J} \widetilde{\lambda}_j \phi_j^{1+\gamma}\right)^{\frac{1}{1+\gamma}}$ and $\sum_{j=1}^{J} \widetilde{\lambda}_j k_{j,-1} = k_{-1} - k_{-1}^g$. 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

$$\sum_{t=0}^{\infty} \beta^t \left[u_{C^{RA},t} C_t^{RA} + u_{l^{RA},t} l_t^{RA}\right] = W_{-1}^{RA}, \tag{5}$$

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. The reader can check that one can go from the above Lagrangian (4) to Flodén's with the following three modifications:

1. Equation (5) is introduced in the planner's problem as an additional constraint.

2. The competitive equilibrium conditions (3) 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 (5) 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 (4).

Modification 2 is also correct, indeed it implies and is implied by (3).

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

$$\sum_{j=1}^{J} \widetilde{\lambda}_j c_{j,t} = C_t^{RA}, \tag{6}$$

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 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 (6) 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 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 (4). The correct FOC with respect to $c_{j,t}$ from (4) is

$$\frac{\partial \mathcal{L}^2}{\partial c_{j,t}} = -\mu_t \widetilde{\lambda}_j. \tag{7}$$

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 one can see that he is using the FOC

$$\frac{\partial \mathcal{L}^2}{\partial c_{j,t}} = 0, \tag{8}$$

which are not compatible with optimality. Therefore, the FOCs in **?** do not give a RPO solution. In particular, his solution does not insure that

$$\frac{\partial \mathcal{L}^2}{\partial c_{j,t}} = \frac{\partial \mathcal{L}^2}{\partial c_{1,t}} \frac{\widetilde{\lambda}_j}{\widetilde{\lambda}_1},$$

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

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

$$C_t^{RA} = \int c(j) d\widetilde{\lambda}(j),$$

we would find the FOC

$$\frac{\partial \mathcal{L}^2}{\partial c_{j,t}} = -\widetilde{\lambda}'(j)\mu_t, \ \forall j \in [0,1]. \tag{9}$$

This is incompatible with (8). The correct solution would imply $\int_I \frac{\partial \mathcal{L}^2}{\partial c_{s,t}} d\widetilde{\lambda}(s) = -\mu_t \int_I d\widetilde{\lambda}(s)$ for any subset of agents I, but Flodén's FOCs give $\int_I \frac{\partial \mathcal{L}^2}{\partial c_t(s)} d\widetilde{\lambda}(s) = 0$.

The only case where (8) is correct is when an agent has $\widetilde{\lambda}_j = 0$ in the discrete case or $\widetilde{\lambda}'(j) = 0$ in the continuous case. 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.

Later on Proposition 5 in Flodén argues that all RPO solutions can be traced out by maximizing the utility with respect to one 'optimized' agent, whose initial state is denoted $\overline{s}$. The proof of that proposition shows that the FOCs for this modified problem coincide with the FOCs on page 300 which are as (8). 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 (7) has to hold instead of (8). Therefore, it is not true that all RPO solutions can be found by selecting an optimized agent even with GHH utility.

In our opinion one can only find all RPO allocations by taking properly into account the utilities of all agents in the economy, as we do in the main text, or as a direct application of (4) would do.

A rationale for Flodén's solution

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

Imagine we consider optimizing a weighted sum of utilities of J' agents (where J' is a discrete number) and that these agents have mass zero in the economy. This can either be because $\widetilde{\lambda}_j = 0$ for all $j = 1, ..., J'$ and $J' < J$ in the discrete case or because we consider only a discrete number of agents in the continuous case. For this RPO allocation the planner's

FOCs are indeed (8). But this is only a very small share of RPO solutions. For any welfare function that gives positive weight to all agents, (8) does not work.

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