Pareto-Improving Optimal Capital and Labor Taxes*

Katharina Greulich   Sarolta Laczó†   Albert Marcet‡

August 2022

Abstract

We study optimal Pareto-improving factor taxation when agents are heterogeneous in their labor productivity and wealth and markets are complete. Pareto-improving policies require a gradual reform: labor taxes should be cut, and capital taxes should remain high for a long time before reaching the limit. This policy redistributes wealth in favor of workers, promotes growth, and causes early deficits and government debt in the long run. We address several technical issues, such as sufficiency of Lagrangian solutions in a Ramsey problem, their relation to welfare functions, and solution algorithms. We also provide a proof that long-run capital taxes are zero.

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 (APMPAL) and Horizon2020 GA n. 788547 (APMPAL-HET), AGAUR (Generalitat de Catalunya), and the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S).

†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.

‡Centre de Recerca en Economia Internacional (CREI), ICREA, BSE and UPF; Ramon Trias Fargas, 25-27, 08005-Barcelona, Spain. Email: albert.marcet@crei.cat.

1 Introduction

We study optimal policy with heterogeneous agents when the government chooses 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 optimal long-run capital taxes are zero, (we denote this property as $\tau_\infty^k = 0$), showing that τ_∞^k might be positive and large. We first argue that if a reasonable constraint on policies is included and cases where the government would prefer to waste consumption are excluded, then $\tau_\infty^k = 0$ reemerges.

However, even if $\tau_\infty^k = 0$ 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 $\tau_\infty^k = 0$ 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. But some papers argued that $\tau_\infty^k = 0$ even with heterogeneous agents, for example, Judd (1985) and Atkeson, Chari, and Kehoe (1999). This may suggest that taxing capital is a 'bad idea' for everyone: there is no equity-efficiency trade-off and the large capital taxes observed in practice must be a failure of fiscal policy-making, since 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 (2020) (SW hereafter) show that previous proofs treated Lagrange multipliers incorrectly, and that a correct proof delivers $\tau_\infty^k > 0$ for some parameter values. Lansing (1999), Bassetto and Benhabib (2006) (BB hereafter), and Benhabib and Szőke (2021) (BSz hereafter) provide more examples with $\tau_\infty^k > 0$. SW find a discontinuity in long-run optimal taxes: small changes in the parameters of the model can cause τ_∞^k to switch from 0 to 100 percent. When BB, BSz, 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? Is there a discontinuity in the total amount of optimal capital taxes? We argue that in the context of a standard model and for a reasonable calibration the answer to all these questions is no. We assume full commitment, complete markets, agents that are heterogeneous in their labor productivity and wealth, an upper bound on capital taxes, a strictly concave production function and no agent-specific lump-sum transfers.

We reexamine optimal policy under the following two elements: (i) we introduce a constraint on government policy that prevents immiserating future generations (no immiseration), and (ii) we consider environments where wasteful government spending is undesirable (absence of optimal waste). We prove that under these conditions $\tau^k_\infty = 0$ reemerges in the model we consider. We find that (ii) is indeed an assumption, as it requires that certain endogenous Lagrange multipliers are positive. To our knowledge the possibility that optimal waste can arise in this type of models had been ignored in the literature, and it reconciles our results with those in BSz.

Even though $\tau^k_\infty = 0$ along all the points on the Pareto frontier that we examine, an equity-efficiency trade-off still exists: Ramsey Pareto optimal (PO) policies include a very long transition of high capital taxes and low labor taxes if all agents are to 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 a very long time. Those high capital taxes reduce total investment and output, but they are required in order to redistribute wealth in favor of workers[1] 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 PO policies and it is a crucial element in order to generate sizeable welfare gains.[2] Further, there is no discontinuity: the length of the period of high capital taxes increases gradually to achieve a larger redistribution toward workers, therefore the share of capital tax revenue moves slowly 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

[1] Even though all our agents work and have some wealth, throughout the paper we refer to 'Workers' as the group with a higher ratio of labor productivity to initial wealth. We call the other group 'capitalists'.

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

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 when economics was in its infancy, but closer to our time Kenneth Arrow promoted the view that there is no such a thing as a 'correct' or 'fair' welfare function. Most textbooks in microeconomics take this view and suggest that economists should be content describing policies along the Pareto frontier without arguing that a particular point on that frontier is 'the best'.[3] In our approach the welfare weights are endogenous, they just index different Pareto-optimal allocations. We find that this point matters in practice. First, asymptotic results using welfare functions with fixed weights have obscured the equity-efficiency trade-off in factor taxation for decades, as the redistribution needed for a Pareto-improvement is resolved along a very long transition. Furthermore, the location of the Benthamite policy on the Pareto frontier is more or less arbitrary, and it can be far from Pareto improving.[4]

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 order for all rational voters to be in favor of an optimal reform capital taxes need to be high for a long time 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 relative consumption of different individuals 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, so the first-order conditions (FOCs) may have multiple solutions. We reduce analytically the set of possible solutions to the FOCs to be sure that our computations pick the maximum.[5] In addition, non-convexities may lead to a duality gap. We check that the duality gap is empty or very small.

[3]These comments also apply to any fixed welfare weights. These are sometimes justified by appealing to probabilistic voting or Nash bargaining, but this interpretation poses some issues of its own. We do not address this issue in this paper.

[4]A companion paper argues that fixed weights also matter for time consistency.

[5]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.

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 would increase the marginal tax rate and exacerbate total distortions. As it turns out, a longer transition is a more efficient way to redistribute.

The rest of the paper is organized as follows. In Section 2 we lay out our baseline model. Section 3 proves some analytical results, including $\tau^k_\infty = 0$, some properties of the transition, and sufficient conditions for solutions to the FOCs. Our numerical results are in Section 4, including those on progressive taxation. Section 5 concludes. The Appendix contains some algebraic details and proofs. The Online Appendix contains a description of our computational approach, sensitivity analyses, it gives details on the relation of our results and solution method to other approaches in the literature, and it discusses in detail why optimal waste can arise in the model considered.

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. 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.[6]

We consider two types of consumers, $j = 1, 2$.[7] Consumers differ in their initial wealth $k_{j,-1}$ and labor productivity ϕ_j. Agent j obtains income in period t from renting out 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

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

For comparison, below we also consider lump-sum taxes or transfers.

[6] BB and BSz consider some examples with a linear production function. Note that we exclude this knife-edged case.

[7] This is for simplicity, it is immediate to extend our analysis to many types of consumers.

Consumer j has utility function $\sum_{t=0}^{\infty} \beta^t \left(u\left(c_{j,t}\right)+v\left(l_{j,t}\right)\right)$, 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*

$$u\left(c\right)=\frac{c^{1-\sigma_c}}{1-\sigma_c} \quad \text{and} \quad v\left(l\right)=-\omega \frac{l^{1+\sigma_l}}{1+\sigma_l}, \tag{2}$$

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

The government chooses capital and labor taxes, has to spend $g \geq 0$ in every period, saves in capital, and has initial capital k_{-1}^g (debt if $k_{-1}^g < 0$). Ponzi schemes for consumers and the government are ruled out. The two types of consumers have equal mass. Capital depreciates at a rate $\delta < 1$. Market clearing conditions for all t are

$$e_t = \frac{1}{2} \sum_{j=1}^{2} \phi_j l_{j,t}, \quad k_t = k_t^g + \frac{1}{2}\sum_{j=1}^{2} k_{j,t}, \quad \text{and}$$

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

2.2 Conditions of competitive equilibria

Our competitive-equilibrium (CE) concept is standard: consumers (firms) maximize utility (profits) taking sequences of prices and taxes as given, markets clear, and the budget constraint of the government is satisfied. We now find a set of necessary and sufficient conditions for a CE allocation.

Consumers' FOCs with respect to consumption and labor yield

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

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

i.e., the Euler equation and the consumption-labor optimality condition, respectively, for $j = 1, 2$. Using a standard argument, (1) and (4), for all t and $j = 1, 2$, can be summarized in the present-value budget constraint

$$\sum_{t=0}^{\infty} \beta^t \frac{u'\left(c_{j,t}\right)}{u'\left(c_{j,0}\right)}\left[c_{j,t} - w_t \phi_j l_{j,t}\left(1-\tau_t^l\right)\right] = k_{j,-1}\left[1+\left(r_0-\delta\right)\left(1-\tau_0^k\right)\right]. \tag{6}$$

Using (5) and rearranging for consumer 1 this becomes

$$\sum_{t=0}^{\infty} \beta^t \left(u'\left(c_{1,t}\right) c_{1,t} + v'\left(l_{1,t}\right) l_{1,t}\right) = u'\left(c_{1,0}\right) k_{1,-1} \left[1 + \left(r_0 - \delta\right)\left(1 - \tau_0^k\right)\right]. \tag{7}$$

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

$$c_{2,t} = \lambda c_{1,t}, \ \forall t, \tag{8}$$

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

$$l_{2,t} = \mathcal{K}(\lambda) l_{1,t}, \ \forall t, \tag{9}$$

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 $\mathcal{K}(\cdot)$ depends only on the primitives σ_c, σ_l, and ϕ_j, $j = 1, 2$.[8]

Using (4), (5), (8), and (9), we can write (6) for consumer 2 as

$$\sum_{t=0}^{\infty} \beta^t \left(u'\left(c_{1,t}\right) \lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'\left(l_{1,t}\right) \mathcal{K}(\lambda) l_{1,t}\right) = u'\left(c_{1,0}\right) k_{2,-1} \left[1 + \left(r_0 - \delta\right)\left(1 - \tau_0^k\right)\right]. \tag{10}$$

The implementability conditions (7) and (10) 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.

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.

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 (3), (8), (9), (7), and (10), respectively, for some λ to be determined consistent with all equilibrium conditions.[9] Given a set of CE allocations, taxes are backed out from (4) and (5), and $k_{j,t}$ from the analog of (7) at t.

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

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

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.

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 needed in our proofs and it seems more relevant: capital flight in an open economy or tax evasion would be massive for τ_t^k close to 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.[10]

It is easy to see that, combining (4) for $j = 1$, (2), and (8), the tax limit holds if and only if $\tau_0^k \leq \widetilde{\tau}$ and

$$u'(c_{1,t}) \geq \beta u'(c_{1,t+1}) \left[1 + (r_{t+1} - \delta)(1 - \widetilde{\tau})\right], \quad \forall t \geq 0. \tag{11}$$

Adding (11) to the constraints guarantees that Constraint in Policy 1 holds, allowing us to use the primal approach, as τ_t^k for any $t \geq 1$ does not appear explicitly in the optimization problem.

We also introduce the following constraint on consumption.

Constraint on Policy 2.

$$c_{1,t} \geq \widetilde{c}, \; \forall t, \; \textit{for some } \; \widetilde{c} \geq 0. \tag{12}$$

[10] 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 in models of optimal policy. A study of capital taxes in a model of learning can be found in Giannitsarou (2006).

Given (8) 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 motivation is that government cannot credibly commit today to policies that immiserate future generations, because of either morall or practical concerns about how to treat those who come after us. A related interpretation is that very low levels of utility in the future will be blocked by the political system, or eventually lead to revolt or social conflict, as in Benhabib and Rustichini (1996). In Section 3.4 we impose explicitly such a minimum constraint on utility. The above constraint can be seen as a simple reduced form of that case.

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 (12), but they face taxes that induce them to act in such a way that (12) always holds.

2.3.2 The Ramsey problem

It follows from the previous discussion that the choice set of the government is

$$\mathcal{S} \equiv \left\{ \text{sequences } \{(c_{j,t}, l_{j,t})_{j=1,2}, k_t\}_{t=0}^{\infty} \text{ which are a CE and satisfy (11) and (12)} \right\}.$$

We define a Ramsey Pareto Optimal (PO) 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 PO 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 PO allocations can be found by solving

$$\begin{aligned} \max_{\tau_0^k, \lambda, \left\{c_t^1, k_t, l_t^1\right\}_{t=0}^{\infty}} & \sum_{t=0}^{\infty} \beta^t \left(u\left(c_{1,t}\right) + v\left(l_{1,t}\right) \right) \\ \text{s.t.} \quad & \sum_{t=0}^{\infty} \beta^t \left(u\left(\lambda c_{1,t}\right) + v\left(\mathcal{K}(\lambda) l_{1,t}\right) \right) \geq \underline{U}_2, \end{aligned} \tag{13}$$

for $\underline{U}_2$ attainable in $\mathcal{S}$, subject to feasibility (3), implementability (7) and (10), tax limits (11) and $\tau_0^k \leq \widetilde{\tau}$, and consumption limits (12). We have used (8) and (9) to substitute for c_2 and l_2 to obtain (13).

We focus on PO allocations which are also Pareto improving relative to a status-quo CE allocation where taxes are set as in the past. We call these POPI allocations. Let the utilities attained by agent j at the status quo be U_j^{SQ}.[11] 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

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

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

Let ψ be the Lagrange multiplier of the minimum-utility constraint (13), let Δ_1 and Δ_2 be the multipliers of the implementability constraints (7) and (10), respectively, and μ_t, γ_t, and ξ_t be the multipliers of the feasibility constraint (3), the tax limit (11), and the consumption limit (12), respectively, at time t. The Lagrangian for the government's problem is

$$\begin{aligned}
\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \Bigg\{ & u\left(c_{1,t}\right) + v\left(l_{1,t}\right) + \psi\left(u\left(\lambda c_{1,t}\right) + v\left(\mathcal{K}(\lambda) l_{1,t}\right)\right) \\
& + \xi_t (c_{1,t} - \widetilde{c}) \\
& + \Delta_1 \left(u'\left(c_{1,t}\right) c_{1,t} + v'\left(l_{1,t}\right) l_{1,t}\right) \\
& + \Delta_2 \left(u'\left(c_{1,t}\right) \lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'\left(l_{1,t}\right) \mathcal{K}(\lambda) l_{1,t}\right) \\
& + \gamma_t \left\{u'\left(c_{1,t}\right) - \beta u'\left(c_{1,t+1}\right)\left[1 + \left(r_{t+1} - \delta\right)\left(1 - \widetilde{\tau}\right)\right]\right\} \\
& + \mu_t \left[F\left(k_{t-1}, e_t\right) + (1-\delta) k_{t-1} - k_t - \frac{1+\lambda}{2} c_{1,t} - g\right] \Bigg\} - \psi \underline{U}_2 - \mathbf{W},
\end{aligned} \tag{14}$$

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 and $\gamma_t \geq 0$, $\forall t$, and $\psi \geq 0$.

The first line of this Lagrangian has the usual interpretation: a Pareto-efficient allocation maximizes a welfare function. The weight of consumer 1 is normalized to one, the 'weight' ψ of consumer 2 is the Lagrange multiplier to be found endogenously. The next three lines in (14) 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 sides of (7) and (10).

The tax limit is a forward-looking constraint, therefore standard dynamic programming does not apply. Appendix A shows how to obtain a recursive formulation using recursive contracts (Marcet and Marimon, 2019). That Appendix also gives the first-order conditions (FOCs).

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

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 and that this equality does not hold. The optimal choice of λ leads to a non-trivial FOC shown in Appendix A. The fact that λ is a choice for the government reflects the fact that the government can vary consumers' relative wealth by its policy choice, in particular, varying the total tax burden of labor and capital in discounted present value. We further demonstrate and discuss how λ behaves differently from ψ^{1/σ_c} 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 the FOCs derived from the Lagrangian are necessary but not sufficient. We address this in detail in Section 3.5.

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 $\mathcal{S}$, if there is a status-quo equilibrium, as we require in our calibration, and if $\widetilde{c}$ is lower than status-quo consumption. Since $\mathcal{S}$ is compact and the objective function is continuous and bounded above for feasible allocations, 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 our result $\tau^k_\infty = 0$, the treatment of dynamic participation constraints, and sufficiency of FOCs.

3.1 Zero capital taxes in the long run

We now examine under what conditions $\tau^k_\infty = 0$ obtains 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 in that case $\tau^k_\infty > 0$. In addition, SW emphasize that this is not a knife-edged case as $\tau^k_\infty > 0$ and consumption goes to zero if initial government debt is above a certain level, see their

[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 optimal choice in his paper. Flodén (2009) considers a model similar to ours. In Online Appendix E, we argue that his approach does not find all PO allocations, although it does provide a useful method to search over competitive equilibria.

Section 3. Lansing (1999), BB, and BSz show heterogeneous-agent models where $\tau^k_\infty > 0$. [13]

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 $\tau^k_\infty = 0$ except in a set of parameter values that, in our version of the model, has measure zero.

We proceed as follows. We take for granted the existence of a steady state for allocations.[14]

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 $t \to \infty$. As discussed in SW, 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

$$F_k(k^{ss}, e^{ss}) > \delta. \tag{15}$$

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 $\tau^k_\infty > 0$. Assume that $\mu_t \geq 0$ for all t large. Then $\mu_t \to 0$. If in addition $\tilde{\tau} < 1$ then $\gamma_t \to 0$.*

Proof. In Appendix B. □

The requirement that $\mu_t \geq 0$ is routinely taken for granted in the literature. We will show in Section 3.2 that, perhaps surprisingly, this fails in some models. Therefore $\mu_t \geq 0$ is indeed an assumption.

Lemma 1 suggests that the key difference between our results and SW is the different asymptotic behavior of μ: SW show that if $\tilde{\tau} = 1$ and $c^{ss} = 0$, it can happen that $\tau^k_\infty > 0$

[13]The frameworks of BB and BSz are quite close to ours. We discuss in footnote 17 how our approach and results relate to BB, and in Section 3.2 and Online Appendix D the relation with BSz.

[14]Therefore in our paper we do not consider the example with perpetual growth of Section III in BSz.

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

and $\mu^{ss} = \infty$, but Lemma 1 says that if $\widetilde{c} > 0$ and $\tilde{\tau} < 1$ then $\tau_\infty^k > 0$ is incompatible with $\mu^{ss} > 0$.

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**, $\widetilde{\tau} < 1$ and $\mu_t \geq 0$ for t large. Assume for parts a), b), and c) that either $\Omega^l \neq 0$ or $\Omega^c > 0$.*

a) Then either $c^{ss} = 0$ or $\tau_\infty^k = 0$.

Assume for parts b), c), and d) that $\widetilde{c} > 0$.

b) $\tau_\infty^k = 0$ and $\Omega^l \geq 0$.

c) Furthermore, if $c^{ss} > \widetilde{c}$ and $\Omega^c \neq 0$, then there is an integer $N < \infty$ such that

$$\tau_t^k = 0 \text{ for all } t > N. \tag{16}$$

If in addition $c_t > \widetilde{c}$ for all t, then there is an N such that in addition to (16) we have

$$0 \leq \tau_N^k \leq \widetilde{\tau} \text{ and} \tag{17}$$
$$\tau_t^k = \widetilde{\tau} \text{ for all } t < N. \tag{18}$$

In words, capital taxes transition to the steady state in two periods.

d) If $\Omega^l = 0$ and $\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1} > 0$, then $\tau_t^k = \widetilde{\tau}$ for all t and $\Omega^c \leq 0$. If $c^{ss} > \widetilde{c}$ then $\Omega^c = 0$.

Proof. In Appendix B. □

Note that this proposition characterizes all cases. Given the minor requirements on multipliers, parts a), b), and c) ensure zero long-run capital taxes for the case $\Omega^l \neq 0$, while the case $\Omega^l = 0$ is covered in part d). The case $\Omega^l \neq 0$ was satisfied in all our computations. The alternative requirement for zero taxes $\Omega^c > 0$ echoes that of SW.[16] Note also that part b) determines $\Omega^c \geq 0$. Part c) shows a familiar result that the transition to zero taxes occurs

[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.

in two periods, the proof does not use uniqueness of critical points, this part highlights that $c^{ss} > \widetilde{c}$ is needed in order to obtain this 'bang-bang' result in our model.[17]

The requirement $\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1} > 0$ is satisfied in the standard case when the government wishes to tax capital in the initial period as much as possible, that is, it sets $\tau_0^k = \widetilde{\tau}$, hence generically $u'(c_{1,0})(\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1})(r_0 - \delta) > 0$ (see $\frac{\partial \mathcal{L}}{\partial \tau_0^k}$ in Appendix A). Under homogeneous agents, this would also imply that $\Delta_1 = \Delta_2 > 0$.[18] Our model with two heterogeneous consumers, interestingly, 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.

Since the behavior of long-run taxes depends on Ω^l or Ω^c and these are endogenous objects, one may wonder in what situations we can ensure that the requirements that lead to parts a), b), and c) are satisfied. 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, $j = 1, 2$. If lump-sum taxes satisfy $T_2 = T_1 \frac{\phi_2}{\phi_1} \mathcal{K}(\lambda)$ we can call them 'labor-income-neutral', since the relative labor income of the two agents is the same before and after tax.[19]

Corollary 1. *Assume **A1**, **A2**, $\widetilde{\tau} < 1$, and $\mu_t \geq 0$ for t large. If agent-specific lump-sum taxes are labor-income-neutral and a marginal increase of lump-sum taxes above $T_1 = 0$ is welfare-improving, then parts a), b), and c) of Proposition 1 hold.*

Proof. In Appendix B. □

The requirement that increasing lump-sum taxes is welfare improving is likely to hold 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. But for most models and calibrations in the literature, the government finds it hard to collect enough taxes.

Proposition 1 and Corollary 1 suggest that (excluding the case $\Omega^l = 0$, seemingly of measure zero), within the context of our model, $\tau_\infty^k > 0$ can only occur in knife-edged cases. For example, the homogeneous-agent environment in Section 3 of SW, where $c_t \to 0$, is a special case of our paper with $\widetilde{\tau} = 1$ and $\widetilde{c} = 0$ (and $\phi_1 = \phi_2$ and $k_{1,-1} = k_{2,-1}$). Corollary 1

[17] BB also do not use uniqueness of critical points to prove this bang-bang result, but their approach of 'piecing together' a potentially better policy in the future cannot be easily applied here, because the share of consumption λ has to remain constant through time, and the potentially better policy would in general imply a different λ, so the 'pieced-together' allocation is not an equilibrium.

[18] In models with distortionary taxes, it is usually welfare enhancing that private agents are initially poorer or, equivalently, that τ_0^k is high, leading to positive Δ's.

[19] 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.

of BB also assumes $\widetilde{c} = 0$ and a linear production function, while our results hold for $\widetilde{c} > 0$ and any strictly concave F. The cases $\tau_\infty^k > 0$ shown in BSz are not knife-edged, but they do not satisfy the assumption '$\mu_t \geq 0$ for t large', see the next subsection.

3.2 The multiplier μ and optimal waste

The recent paper BSz considers a model very similar to ours. The authors provide conditions on endogenous objects guaranteeing that optimal taxes satisfy $\tau_t^k = \widetilde{\tau} < 1$ for all t and $c_t \to c^{ss} > 0$. In Section III they show how these conditions are satisfied for some parameter values. This result apparently contradicts our Proposition 1.

It turns out that the driving force behind the two results is the sign of the multipliers μ_t on the feasibility constraint. Our $\tau_t^k \to 0$ result is derived under the assumption that $\mu_t \geq 0$ for t large, but, as we show in Online Appendix D, in the example of Section IIIA of BSz $\mu_t \to \mu^{ss} < 0$. This is why our results do not apply to their case.[20] Furthermore, Proposition 3 in Online Appendix D shows a partial converse: essentially this result states that $\tau_t^k = \widetilde{\tau} < 1$ for all t only if μ_t is negative in the long run.

A negative μ_t in our model would imply that throwing away consumption in some periods is welfare enhancing. This may seem like a mistake when social welfare is an increasing function of consumption. But it is not, since the current model amounts to imposing feasibility as *equality*, and equality constraints can have multipliers of either sign. Equivalently, the government has to set $g_t = g$ for a fixed g. If instead we allow for free disposal $g_t \geq g$, the government could implement consumption waste by setting $g_t > g$. We demonstrate in Online Appendix D that in the example of BSz Section IIIA, the objective function of the government is indeed increased by setting $g_t > g$ in periods where $\mu_t < 0$.

Optimal waste arises here because even though setting $g_t > g$ in the long run lowers aggregate consumption, it also increases the stochastic discount factor.[21] This increases the discounted value of capital tax revenue collected, redistributing wealth in favor of agents who have little wealth. This may increase the utility of agents the policy-maker cares about (the median voter as in BB or only poor agents as in BSz) if they are sufficiently poor relative to aggregate wealth.[22]

[20] BSz contains a discussion of some results in the previous version of our paper. They do not show the values of μ_t as they use an alternative dual approach.

[21] A related mechanism is described in the recent paper Debortoli, Nunes, and Yared (2021). They show that time-inconsistency arises in the Lucas and Stokey (1983) economy, because future wasteful tax rates may be desirable as they lower current equilibrium interest rates.

[22] Note that we find that increasing wealth inequality in our model does not give rise to $\tau_t^k = \widetilde{\tau}$ for all t, even if the policy-maker only cares about the worker, see Online Appendix C. In addition, optimal policy can be far from Pareto improving if the planner ignores some agents, as in BB and BSz. We come back to

The result of BSz is useful because it alerts to the fact that negative μ_t's may arise in standard models, a possibility previously ignored in the literature.[23] This raises the question: is our assumption $\mu_t \geq 0$ for t large (i.e., absence of optimal waste) in Proposition 1 likely to hold for reasonable parameter configurations?

Recall that $\mu_t \geq 0$ fails when less wealthy agents benefit from 'consumption waste'. Intuitively, this is more likely to occur when the tax system causes low *aggregate* efficiency distortions. In the jargon of fiscal policy, optimal waste is more likely to arise when tax rates are far from those corresponding to the 'peak of the Laffer curve'. It is clear that this is the case in the example of BSz Section IIIA, because (i) $g = 0$ and $k^g_{-1} = 0$, hence there is little need to raise public revenues, (ii) a universal lump-sum transfer $\mathcal{D}$ (same as in our Section 4.4) implies that higher g_t's do not lead to higher distortionary taxes,[24] and (iii) $\widetilde{\tau} = 10\%$ is low, so that setting $\tau^k_t = \widetilde{\tau}$ for all t is not highly distortionary, while the weight of capital in production is high ($\rho = 0.95$ in their CES function). Most applied work using DSGE models similar to ours suggest that existing tax distortions are very high,[25] hence the efficiency loss is large, and no agent is likely to benefit from wasting consumption. Indeed, moving BSz parameters slightly in the direction of increasing total tax distortions causes negative μ's to disappear.[26] Therefore, we think the case $\mu_t \geq 0$ is likely to be relevant for reasonable calibrations.

3.3 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 those results are valid even if there are multiple critical points. But multiplicity is an issue once we rely on numerical simulations obtained from solutions to FOCs. 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).

$$\max_{\tau^k_0, \lambda, \left\{c^1_t, k_t, l^1_t\right\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left[u\left(c_{1,t}\right) + v\left(l_{1,t}\right) + \psi\left(u\left(\lambda c_{1,t}\right) + v\left(\mathcal{K}(\lambda) l_{1,t}\right)\right)\right], \tag{19}$$

this in Section 4.3.1

[23]For example SW (pages 9 and 25) take for granted that this multiplier, denoted λ in their paper, is non-negative.

[24]More precisely, $g = 0$ and high capital taxes imply that $\mathcal{D} > 0$ in BSz. Therefore a higher g_t does not call for larger distortionary taxes. In Section 4.4 and Online Appendix C, we show that things change when g is calibrated to the data. In particular, progressive taxation then leads to $\mathcal{D} = 0$, so that higher g_t is likely to imply higher distortionary taxes.

[25]See, for example, a suite of calibrated DSGE models in Trabandt and Uhlig (2011).

[26]We have found that if, ceteris paribus, $\widetilde{\tau} = 12\%$ or $\rho = 0.94$, it is no longer optimal to keep capital taxes at their upper bound indefinitely.

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 POs.

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 exploring numerically all possible solutions in an infinite-dimensional problem can be difficult. Proposition 1 is useful for this task, because it covers all cases for Ω^l, and it narrows down the possible values of τ^k_∞ to 0 and $\widetilde{\tau}$. Then we have the following.

Algorithm to find optimal solutions to MM

Step 1 *For each candidate N, compute the infinite 'tail' of the sequence imposing (16), 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 is a candidate solution.*[27]

Step 2. *Find a solution with $\tau^k_t = \widetilde{\tau}$ for all t. If $\Omega^l = 0$ this a candidate solution.*

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, each step involves a finite-dimensional problem.

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

Step 3. *Compute the utility corresponding to each candidate solution and pick the solution with the highest utility.*

Since, according to Proposition 1, this algorithm exhausts all possible steady states, it is certain to give the correct solution. In all the optimal allocations we computed in Section 4 there was no candidate solution with all the properties of Step 2, and we found one candidate solution in Step 1 with $\Omega^l \neq 0$, hence $\tau^k_\infty = 0$ in all the calculations shown below.

3.4 Dynamic participation constraints

Constraint on Policy 2 is a simple way to capture the idea that a policy entailing $c^{ss} = 0$ will be blocked by some political mechanism or social conflict because agents' future welfare will

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

be so low. To introduce this idea more explicitly, we now replace Constraint on Policy 2 by the following dynamic participation constraints (PCs).[28]

Constraint on Policy 3.

$$\sum_{i=0}^{\infty} \beta^i \left(u\left(c_{j,t+i}\right) + v\left(l_{j,t+i}\right)\right) \geq \underline{U},\ \forall t,\ \ j = 1, 2,\ \ \textit{for some finite } \underline{U}. \tag{20}$$

This implies a relatively minor change in the analysis. The Ramsey problem is as before only with (20) replacing (12). Using the results in Marcet and Marimon (2019), the first two lines of the Lagrangian for the government's problem (14) are replaced by

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

while the remaining lines in (14) stay unchanged. Here, $\nu_{j,t} \geq 0$ are the Lagrange multipliers of (20), $M_{j,t} = M_{j,t-1} + \nu_{j,t}$ 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.[29] 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. In those models the ratio $u'(c_{2,t})/u'(c_{1,t})$ is time-varying and equal to $(1 + M_{1,t})/(\psi + M_{2,t})$.[30] Instead in the model of this section, $u'(c_{2,t})/u'(c_{1,t})$ is constant through time according to (8), 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 asymptotic

[28] Ideally the right hand side of (20) would be derived from an explicit model of political economy or social conflict. For example Benhabib and Rustichini (1996) derive a similar constraint from a mechanism of social conflict, or Kocherlakota (1996) from assuming that there is an outside option of autarky. We leave endogenizing $\underline{U}$ for future research.

[29] 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.

[30] 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.

results analogous to Lemma 1 and Proposition 1.[31] We only give an outline of the proof.

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, choose some $\underline{U} > 0$ for the case $\sigma_c < 1$, or $\underline{U} > -\infty$ for the case $\sigma_c \geq 1$. Given the functional form (2), taking limits in (20), it is clear that for these choices of $\underline{U}$ if (20) holds then $c^{ss} > 0$. Since the proofs of Lemma 1 and Proposition 1 hinge on $c^{ss} > 0$, it is easy to check that the same limiting 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.[32]

Therefore, the analogous asymptotic results obtain and 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.

3.5 The Pareto frontier

Since the set of feasible equilibrium allocations $\mathcal{S}$ is not necessarily convex, a Lagrangian approach is not guaranteed to give all the PO allocations. We have already discussed in Section 3.3 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: the duality gap (i.e., the set of PO 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 PO 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 with a concave utility function, $\mathcal{F}$ corresponds to the PO 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 some fixed weight ψ. But if $\mathcal{S}^U$ is not convex, its frontier may have

[31] Notice that Constraint on Policy 3 does not imply Constraint on Policy 2: given $\tilde{c} > 0$ there are consumption allocations satisfying (20) for which, say, $c_0 < \tilde{c}$. Therefore, Lemma 1 and Proposition 1 do not apply immediately to this case.

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

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 PO 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 PO equilibria. 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 (19).

A3. *MM has a unique solution for all $\psi \geq 0$. Furthermore, $U_2(\cdot)$ is invertible on $[0, \infty]$.*

Proposition 2. *Assume **A3**. Then the following statements hold.*

a) A solution to MM for any $\psi \in [0, \infty]$ is a PO allocation.

b) Every PO allocation is also the solution of MM for some $\psi \in [0, \infty]$.

c) 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 b) of Proposition 2 implies that we can find all PO allocations by solving MM varying ψ from zero to infinity. Part c) 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 $\psi < 0$,[33] 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 c) to find an increasing part of the frontier $\mathcal{F}$ which is not Pareto optimal.

Since the feasible set is non-convex, **A3** may not hold for some parameterizations. But it can be checked numerically for a given application. We record all utilities for a fine grid of ψ's, applying the Algorithm of Section 3.3 for each ψ, 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

[33]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 $\mathcal{F}$.

points. Figures 1 and 2 show the utility pairs $(U_1(\psi), U_2(\psi))$ for a grid of ψ's. The function $U_2(\psi)$ appears invertible on these figures, therefore MM fully characterizes all PO 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 $\tau_\infty^k > 0$ in 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 for redistribution to 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 PO 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 in Section 4.2, we examine the model with fixed labor supply. Section 4.3 shows the results for our baseline model. We discuss progressive taxation in Section 4.4.

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-to-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).[34] We compute averages for

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

the period 2001-2010. The average effective tax rates are: $\tau^l = 0.214$ and $\tau^k = 0.401$. Note that tax rates at the status quo matter in several ways. Firstly, they influence the status-quo, and hence the 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

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
	$\widetilde{\tau}$	0.401

We set some preference parameters a priori. We use the usual discount factor β=0.96. The coefficient of relative risk aversion is $\sigma_c = 2$. The choice of $\sigma_l = 3$ generates an elastic supply of labor, and it prevents hours from greatly differing across consumers with different wealth. Hence Frisch elasticity of labor supply is lower than in many real-business-cycle applications but is more in line with micro estimates.[35]

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' (type-2 consumers), indexed w in the calibrated model, and (ii) those with below the median wage-wealth ratio, called 'capitalists' (type-1 consumers), indexed c. That is, capitalists are wealthier relative to their

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

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.[36]

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, such that the 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.[37]

4.2 Results with fixed labor supply

In our baseline model 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. 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.[38]

Under homogeneous agents and fixed labor supply, the policy-maker would set $\tau^k_t = 0$, $\forall t$, 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 and T_c. But since we focus

[36] 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.

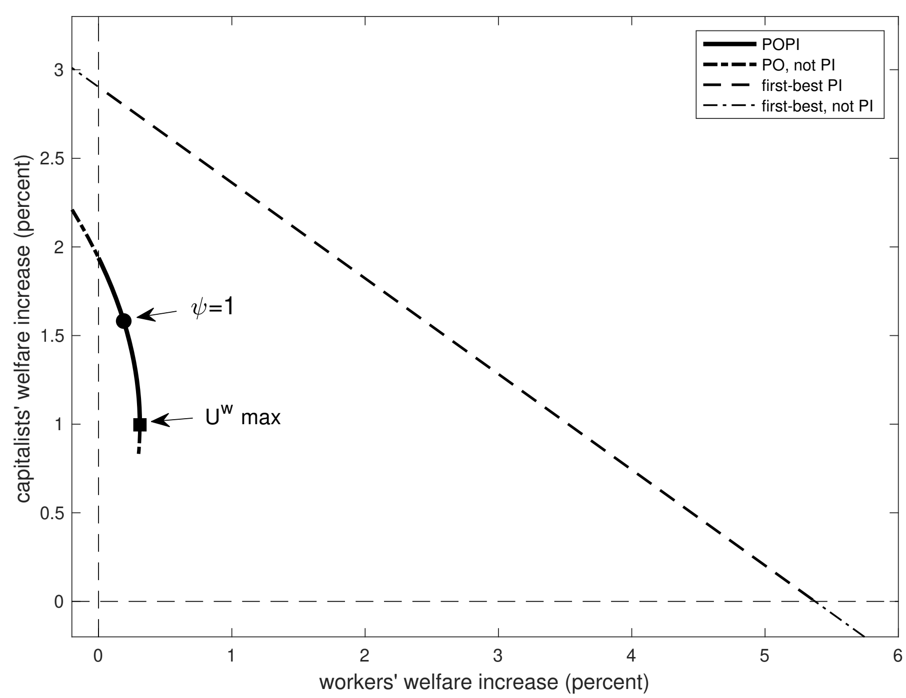
[37] As Table 1 shows, the initial wealth of workers turns out to be negative, i.e., workers are borrowers, and we find that they stay borrowers in the main calibration. Given our capital tax formulation, this means that workers receive a subsidy τ^k_t on their interest payments. One could argue that this is not a good way to model actual capital taxes, as subsidies to borrowing are limited. Removing the subsidy to borrowers would complicate the analysis somewhat: 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.

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

on the case $T_w = T_c = 0$, deviations from the first-best policy are necessary for distributive reasons to achieve a Pareto improvement.

In Figure 1 we compare the set of POPI plans to the first best. Units in this graph are consumption-equivalent welfare gains.[39] The dashed (black) line labeled 'first-best PI' represents allocations with $\tau_t^k = 0$ for all t and optimal redistributive lump-sum transfers $T_w = -T_c$.[40] 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.

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., $\psi \to \infty$.

Using Proposition 2 part a), the decreasing part of $\mathcal{F}$ is found with $\psi > 0$ in MM, higher

[39]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.

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

ψ corresponding to points further to the right along the solid (blue) line. Higher ψ's 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 2 part c). 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 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

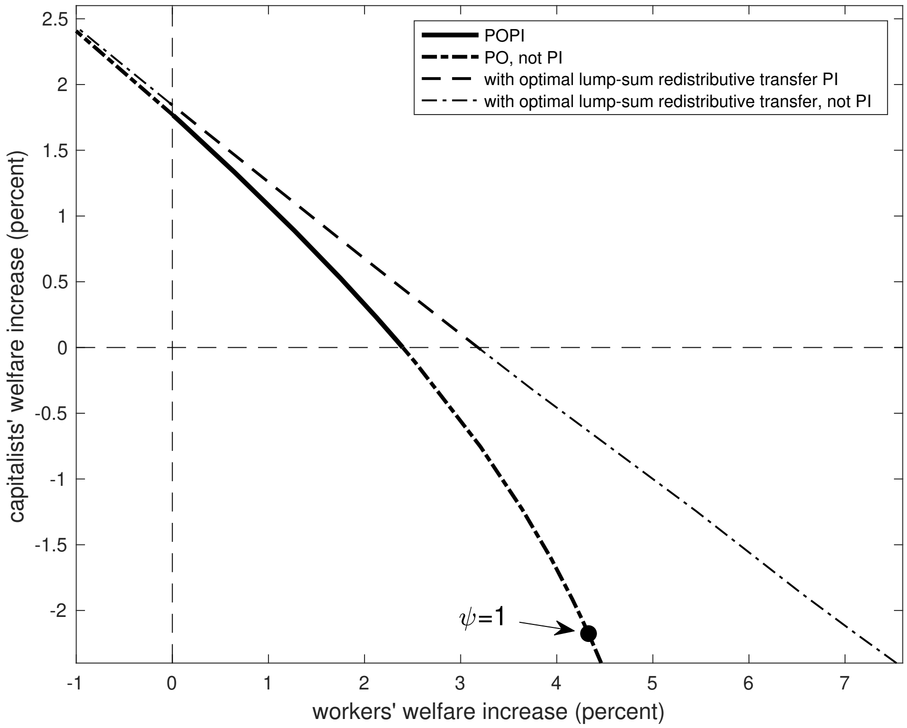
We now return to our baseline model, which features elastic labor supply.

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 whole 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.

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.

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.[41] 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 not found for high debt either.

Now we compare some key characteristics of different points on the frontier. The length of the transition increases as welfare gains are shifted toward the worker. This is illustrated in the first panel of Figure 3 showing the duration of the transition, N, on the vertical axis for each POPI allocation, indexed by the welfare gain of the worker on the horizontal axis. We see that the number of periods before capital taxes drop to zero increases from 16 to 24 years

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

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.[42] 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 toward $\tau^k_\infty = 0$ 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.[43] 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 of 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 little as the utility promise to the worker increases, while λ rises much more than without

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

[43]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.

transfers. This is because shifting welfare gains and consumption between agents is much easier with redistributive lump-sum transfers, hence the planner lowers capital taxes quickly to increase efficiency. The policies and the paths 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 Endogenous 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 PO 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 with fixed ψ.[44]

We now discuss the relationship between ψ and equity. We dub 'equitable reform' a PO solution which implies that both agents gain equally,[45] that is, points on the frontiers of Figures 1 and 2 which are on the 45º 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

[44] Furthermore, in a companion paper we show that optimal policy is 'consensus time-consistent'. This result would also be hidden if only fixed welfare weights were considered.

[45] 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.

even Pareto improving, a weight $\psi \in [0.35, 0.49]$ is needed for a Pareto 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.

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 case.

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 upper bound. The kinks in the paths of labor taxes and government deficit occur precisely in the intermediate period when capital taxes transit from the maximum to zero.

It is interesting to note that if labor supply is elastic, low labor taxes weaken the efficiency-redistribution trade-off. Low labor 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.[46] 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

[46] 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.

4.3.4 High capital taxes

We now compare the optimal solution with the one that would arise if capital taxes are kept at the upper bound forever. This is of interest per se, and it is Step 2 of the algorithm, to check if the solution is as in part d) of Proposition 1.

In this formulation the government faces the restriction $\tau_t^k = \widetilde{\tau}$ for all t, but it chooses labor taxes. The Pareto frontier for this policy problem is shown as the dashed line in Figure 5, while the Pareto frontier for the baseline model of Section 4.3.1 is the solid line. The frontiers show a larger range of Pareto optimal allocations, from ψ small to $\psi = \infty$.

In all cases welfare is now lower, therefore the optimal solution has $\tau_\infty^k = 0$, as in Section 4.3.1. The value of Ω^l reaches its minimum of 1.9 when the weight of workers is small, which implies the allocation computed in Step 2 of the algorithm is not optimal. The multipliers μ_t are always positive.

The solid line achieves much higher utility gains at the left of the graph, but the welfare gain becomes negligible when the benefits of the reform are more targeted to the worker. This is not surprising: as we saw earlier the transition to zero capital taxes takes longer as we move to the right of Figure 5, therefore the welfare gain from eventually lowering capital taxes is less significant. The rightmost points of these Pareto frontiers correspond to $\psi = \infty$, i.e., the case where the planner only cares about workers, as in BSz. At that point the welfare gains of the worker are almost the same under the two policies.

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), BB, and BSz 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 (14), and letting 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 $\mathcal{D} = 0$, including in the case where the Pareto weight of the capitalist is 0. Therefore, access to progressive taxation does not change any of our conclusions: optimal policy implies not to use progressivity, hence the computations in Section 4.3.1 are also valid for the case of progressive taxation.

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 distortions.

If we remove the progressivity constraint, the government would choose a regressive tax scheme with $D < 0$. How can this be Pareto improving in a model where, given the results in Sections 4.2 and 4.3.1, redistributive concerns are a key issue? The reason is that the government now redistributes by choosing 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. The solid line in Figure 6 is the resulting Pareto frontier. Capitalists can gain maximum 4.0 percent and workers 6.2 percent in welfare-equivalent consumption units considering Pareto-improving policies. Welfare gains are larger than in the case with optimal lump-sum redistributive transfers $T_w = -T_c$. We think such a regressive tax scheme would not be POPI if we considered a richer form of heterogeneity, so we do not pursue this analysis further in this paper.[47]

[47] 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.

This speaks to previous work on $\tau_\infty^k > 0$. BB and BSz find positive long-run capital taxes for calibrations where $\mathcal{D}$ is optimally positive and serves to redistribute toward wealth-poor agents, while the capital tax serves to raise revenue. But this means that they consider a case where the total cost of distortions with $T_w = T_c = 0$ is negative. We discuss this issue analytically in detail in Online Appendix D.

We have also computed optimal policies combining the features of this section and Section 4.3.4, that is, with a constraint $\tau_t^k = \widetilde{\tau}$ and an optimal $\mathcal{D}$, positive or negative. Figure 6 shows the resulting Pareto frontier as a dashed line. Just as in Section 4.3.4, welfare losses are large (minor) for allocations that favor the capitalist (worker). The optimal $\mathcal{D}$ is always negative.

In addition, in Online Appendix C, we further examine the role of wealth inequality in determining optimal policy allowing for $\mathcal{D} \neq 0$, bringing our calibration closer to the parameters considered in BSz, where tax distortions are very small. We consider six combinations of parameter values and levels of inequality. We find that even when the government only cares about wealth-poor agents, optimal tax policies involve $\tau_\infty^k = 0$. A negative labour income tax, combined with a lump-sum tax and zero capital tax in the long run, serves to promote equity better than a high capital tax combined with a lump-sum transfer.

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. It is obvious that if $\Delta_2 = \mathcal{D}_w < 0$, 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.

5 Conclusion

We study the efficiency-equity trade-off in setting capital and labor taxes when markets are complete. We first show that the traditional result $\tau_\infty^k = 0$ reemerges in our model if one imposes reasonable constraints on policy, in particular, if the government is prevented from immiserating consumers, and the government would not prefer to waste consumption. Hence $\tau_\infty^k = 0$ seems a more robust result than some recent papers suggest.[48] It will be interesting

[48]The literature has identified some cases where $\tau_\infty^k > 0$ without immiseration in stationary models: (i) the log case of Lansing (1999), Reinhorn (2019), and Section I.B. in SW, and (ii) the $\beta r(1 - \widetilde{\tau}) = 1$ case of BB. Both of these cases are knife-edged.

to see if similar results are found in other models.

However, $\tau^k_\infty = 0$ does not mean that low capital taxes are good for all agents. In a calibrated version of the model, we find that in order 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) in the long run, 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. We explore variations in parameter values and model specification and find that results are robust 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 demonstrate how results with fixed welfare weights can be misleading. We use welfare weights as an artefact to compute a whole array of Pareto-optimal policies. In this way we can study a number of issues, such as the speed of the transition and how it relates to redistribution, and the importance of gradual reforms in order to achieve Pareto improvements. In addition, Benthamite policies can be far from equitable, and they can hurt large parts of the population.

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 optimal policy in heterogeneous-agent models, both from an empirical and a theoretical point of view, when policies are not selected by a certain arbitrary set of weights. 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.

Finally, understanding the role of negative μ_t's (optimal waste) could open interesting avenues for future research, such as establishing conditions under which negative μ's occur more generally, and solving for optimal policy allowing for free disposal in government spending, i.e., $g_t \geq g$ for all t.

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.

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.

Benhabib, J. and B. Szőke (2021). Optimal Positive Capital Taxes at Interior Steady States. *American Economic Journal: Macroeconomics 13*(1), 114–50.

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

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

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

Debortoli, D., R. Nunes, and P. Yared (2021). Optimal Fiscal Policy without Commitment: Revisiting Lucas-Stokey. *Journal of Political Economy 129*(5), 1640–1665.

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.

Lucas, R. E. and N. L. Stokey (1983). Optimal Fiscal and Monetary Policy in an Economy without Capital. *Journal of Monetary Economics 12*(1), 55–93.

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 (2020). Positive Long-Run Capital Taxation: Chamley-Judd Revisited. *American Economic Review 110*(1), 86–119.

Trabandt, M. and H. Uhlig (2011). The Laffer Curve Revisited. *Journal of Monetary Economics 58*(4), 305–327.

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 First-order conditions and recursive Lagrangian

Using the derivations in Section 2, the functional form u, v in **A1** and Marcet and Marimon (2019), the Lagrangian of the policy-maker's problem can be written as

$$\begin{aligned}\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \Big\{ & \Omega^c \; u\left(c_{1,t}\right) + \Omega^c \; v\left(l_{1,t}\right) + \\ & + \xi_t (c_{1,t} - \widetilde{c}) + u'\left(c_{1,t}\right) \left\{\gamma_t - \gamma_{t-1}\left[1 + \left(r_t - \delta\right)\left(1 - \widetilde{\tau}\right)\right]\right\} \\ & + \mu_t \left[F\left(k_{t-1}, e_t\right) + (1-\delta) k_{t-1} - k_t - \frac{1+\lambda}{2} c_{1,t} - g \right] \Big\} - \psi \underline{U}_2 - \mathbf{W}\end{aligned}$$

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

The FOCs, using the functional form u, v in **A1,** are:

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

$$\Omega^c u'\left(c_{1,t}\right) + \xi_t + u''\left(c_{1,t}\right)\left\{\gamma_t - \gamma_{t-1}\left[1 + \left(r_t - \delta\right)\left(1 - \widetilde{\tau}\right)\right]\right\} = \mu_t \frac{1+\lambda}{2} \tag{21}$$

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

$$\begin{aligned} & \Omega^l v'\left(l_{1,t}\right) - \gamma_{t-1} u'\left(c_{1,t}\right) F_{ke}\left(k_{t-1}, e_t\right) \frac{1}{2}\left(\phi_1 + \phi_2 \mathcal{K}(\lambda)\right)\left(1 - \widetilde{\tau}\right) \\ & = -F_e\left(k_{t-1}, e_t\right) \frac{1}{2}\left(\phi_1 + \phi_2 \mathcal{K}(\lambda)\right) \mu_t \end{aligned} \tag{22}$$

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

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

- for the multiplier of the promise-keeping constraint:

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

- for relative consumption, λ, using (22) to simplify:

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

- for γ_t at $t \geq 0$:

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

- for Δ_j: the corresponding lifetime budget constraint.
- for τ_0^k:

$$\begin{aligned}\text{either}\quad \Delta_1 k_{1,-1}+\Delta_2 k_{2,-1}=0 \quad \text{and} \quad \tau_0^k\leq\widetilde{\tau},\\ \text{or}\quad \Delta_1 k_{1,-1}+\Delta_2 k_{2,-1}\geq 0 \quad \text{and} \quad \tau_0^k=\widetilde{\tau}.\end{aligned}$$

To obtain a recursive formulation, for simplicity, consider the standard case where $\tau_0^k=\widetilde{\tau}$. In this case $\mathcal{L}$ is unchanged if we delete $\mathbf{W}$ and set $\gamma_{-1}=\Delta_1 k_{1,-1}+\Delta_2 k_{2,-1}$. Then, for given (Δ_1,Δ_2,ψ), the Lagrangian is of the form considered in Marcet and Marimon (2019), and optimal allocations satisfy $(c_{1,t},l_{1,t},k_t,\gamma_t)=\mathcal{P}(k_{t-1},\gamma_{t-1})$, for all $t\geq 0$ and for a time-invariant policy function $\mathcal{P}$ and the above γ_{-1}.

B Proofs

Proof of Lemma 1. Assume that $\tau_\infty^k>0$. Taking limits in (4) gives

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

Then using (15) we have $\beta\left(1+F_k\left(k^{ss},e^{ss}\right)-\delta\right)>1$, hence there is a constant A such that $1>A>\frac{1}{\beta(1-\delta+F_k(k^{ss},e^{ss}))}$. Obviously,

$$1>A>\frac{1}{\beta\left(1-\delta+F_k\left(k_t,e_{t+1}\right)\right)}\quad\text{for } t \text{ large enough.} \tag{23}$$

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

$$\mu_t\frac{1}{\beta\left(1-\delta+F_k\left(k_t,e_{t+1}\right)\right)}+\gamma_t\frac{u'\left(c_{1,t+1}\right)F_{kk}\left(k_t,e_{t+1}\right)\left(1-\widetilde{\tau}\right)}{1-\delta+F_k\left(k_t,e_{t+1}\right)}=\mu_{t+1}. \tag{24}$$

We have $F_{kk}\left(k,e\right)\leq 0$ 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 (23), 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 (24) to obtain

$$\gamma_t \frac{u'(c_{1,t+1}) F_{kk}(k_t, e_{t+1})(1-\widetilde{\tau})}{1-\delta+F_k(k_t, e_{t+1})} \to 0. \tag{25}$$

Now we show that the term multiplying γ_t in (25) cannot go to zero. First we prove that the denominator cannot go to infinity: feasibility and $c_1^{ss} > 0$ 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 $u'(c_1^{ss}) \frac{F_{kk}(k^{ss},e^{ss})(1-\widetilde{\tau})}{1-\delta+F_k(k^{ss},e^{ss})} < 0$. Hence (25) implies $\gamma_t \to 0$. □

Proof of Proposition 1. Part a). Assume towards a contradiction that $c^{ss} > 0$ and $\tau_\infty^k > 0$. Lemma 1 guarantees that $\mu_t, \gamma_t \to 0$. In the proof of Lemma 1 we have already showed $e^{ss}, k^{ss} > 0$, therefore $F_e(k^{ss}, e^{ss}) < \infty$. Differentiating both sides of $F_k k + F_e e = F$ with respect to k gives $F_{kk}k + F_{ek}e = 0$, hence $0 \leq F_{ke}(k^{ss}, e^{ss}) < \infty$. Putting all this together, taking limits on both sides of (22), 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.

Furthermore, since whenever $\Omega^c > 0$, the FOC for consumption (see Appendix A) 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 $\tau_\infty^k > 0$. This proves part a).

Part b). That $\tau_\infty^k = 0$ is a corollary of part a). Given $\tau_\infty^k = 0$ and $\widetilde{\tau} \leq 1$ we have that $\gamma_t = 0$ for t large enough so that (22) implies $\Omega^l \omega (l_{1,t})^{\sigma_l} = F_e(k_{t-1}, e_t) \frac{1}{2}(\phi_1 + \phi_2 \mathcal{K}(\lambda)) \mu_t$ and $\mu_t \geq 0$ implies $\Omega^l \geq 0$.

Part c). We first prove (16). Given part b) there is a finite integer such that $\gamma_{t-1} = \xi_t = 0$ for all $t \geq N$. Plugging this in (21) implies $\Omega^c (c_{1,t})^{-\sigma_c} = \mu_t \frac{1+\lambda}{2}$ for all $t \geq N$. Plugging this in the FOC for capital for all $t \geq N$ gives

$$(c_{1,t})^{-\sigma_c} = \beta (c_{1,t+1})^{-\sigma_c} (1 - \delta + F_k(k_t, e_{t+1})), \tag{26}$$

which together with (4) implies (16).

In the previous paragraph we only used $\gamma_t = 0$ for t sufficiently large. To prove the remainder of part c) we need to show that once $\gamma_t = 0$ it stays at this value. Formally, there is a finite N such that

$$\gamma_t > 0 \text{ for all } t < N-1 \text{ and } \gamma_t = 0 \text{ for all } t \geq N-1. \tag{27}$$

For this purpose we first show that $\tau_t^k \geq 0$ for all t. If $\gamma_{t-1} > 0$ then $\tau_t^k = \widetilde{\tau} > 0$, while if $\gamma_{t-1} = 0$ then (21) gives $\Omega^c (c_{1,t-1})^{-\sigma_c} \leq \mu_{t-1} \frac{1+\lambda}{2}$ and $\Omega^c (c_{1,t})^{-\sigma_c} \geq \mu_t \frac{1+\lambda}{2}$. Plugging all this in the FOC for capital at $t-1$ and using $\gamma_{t-1} = 0$ again, we have

$$(c_{1,t-1})^{-\sigma_c} \leq \beta (c_{1,t})^{-\sigma_c} (1 - \delta + F_k(k_{t-1}, e_t)).$$

Together with (4) this implies $\tau_t^k \geq 0$ for all t.

Now we show that if $\gamma_{t-1} = 0$ then $\gamma_t = 0$. Notice first that, using the Kuhn-Tucker conditions, $\gamma_{t-1}\left[1 + (r_t - \delta)(1 - \widetilde{\tau})\right] = \gamma_{t-1}\frac{(c_{1,t-1})^{-\sigma_c}}{\beta(c_{1,t})^{-\sigma_c}}$. Therefore (21) can be rewritten as

$$\gamma_{t-1}\left(c_{1,t-1}\right)^{-\sigma_c} = \left(\mu_t \frac{1+\lambda}{2} - \Omega^c \left(c_{1,t}\right)^{-\sigma_c}\right)\frac{c_{1,t}}{\sigma_c}\beta + \gamma_t \left(c_{1,t}\right)^{-\sigma_c}\beta.$$

Substituting forward the term $\gamma_t \left(c_{1,t}\right)^{-\sigma_c}$, using the fact that the transversality condition requires $\beta^t \mu_t \to 0$ and other boundedness conditions, we find

$$\gamma_t \left(c_{1,t}\right)^{-\sigma_c} = \sum_{i=1}^{\infty} \beta^i \frac{c_{1,t+i}}{\sigma_c}\left(\mu_{t+i}\frac{1+\lambda}{2} - \Omega^c \ \left(c_{1,t+i}\right)^{-\sigma_c}\right), \text{ for all } t \geq 1. \tag{28}$$

This implies that if $\gamma_t > 0$ for a given t, then

$$\mu_{t+i}\frac{1+\lambda}{2} > \Omega^c \left(c_{1,t+i}\right)^{-\sigma_c} \text{ for some } i \geq 1. \tag{29}$$

Using the FOC for capital, $\gamma_t \geq 0$, and $F_{kk} \leq 0$, we have $\mu_t \geq \beta\mu_{t+1}\left(1 - \delta + F_k\left(k_t, e_{t+1}\right)\right)$, for all t. Iterating we have

$$\mu_t \geq \mu_{t+i}\beta^i \prod_{h=1}^{i}\left(1 - \delta + F_k\left(k_{t+h-1}, e_{t+h}\right)\right).$$

Assume, toward a contradiction, that $\gamma_{t-1} = 0$ and $\gamma_t > 0$ for some t. Then (21) implies that $\Omega^c \left(c_{1,t}\right)^{-\sigma_c} > \mu_t \frac{1+\lambda}{2}$. Together with the previous two inequalities this implies

$$\left(c_{1,t}\right)^{-\sigma_c} > \left(c_{1,t+i}\right)^{-\sigma_c}\beta^i \prod_{h=1}^{i}\left(1 - \delta + F_k\left(k_{t+h-1}, e_{t+h}\right)\right).$$

But using (4), $\tau^k_{t+1} = \widetilde{\tau} > 0$, and since we have showed that $\tau^k_t \geq 0$ for all t, we have

$$\left(c_{1,t}\right)^{-\sigma_c} < \left(c_{1,t+i}\right)^{-\sigma_c}\beta^i \prod_{h=1}^{i}\left(1 - \delta + F_k\left(k_{t+h-1}, e_{t+h}\right)\right),$$

which is a contradiction. Therefore if $\gamma_{t-1} = 0$, then $\gamma_t = 0$.

Take the smallest N for which (18) holds. Given part b) $N < \infty$. Since $\gamma_{N-1} = 0$, the last paragraph implies (27) by induction. The same argument we used to prove (16) now holds for the same N in (18). We have already proved that (17) holds for all t so the proof of part c) is complete.

Part d). We have already argued that $F_{kk}k + F_{ek}e = 0$ for all t. We have $k_t > 0$ for all t, otherwise c_t would equal 0 for some t, and utility would be $-\infty$. Since the status-quo policy is feasible, $k_t = 0$ cannot happen in an optimum. Therefore, strict concavity of F gives $F_{kk}k < 0$. This implies $F_{ke}(e_t, k_t)e_t > 0$ for all t, hence $F_{ke}(e_t, k_t) > 0$ for all t. This means that combining $\Omega^l = 0$ with (22), we have $\mathcal{A}_t\gamma_{t-1} = \mu_t$ for $\mathcal{A}_t = \frac{(c_{1,t})^{-\sigma_c}F_{ke}(k_{t-1},e_t)(1-\widetilde{\tau})}{F_e(k_{t-1},e_t)} > 0$, for all t.

If $\mu_t > 0$ at any t, substituting out $\mu_{t+1} = \mathcal{A}_{t+1}\gamma_t \geq 0$ in the FOC for capital (24) we have that $\gamma_t > 0$. Therefore $\mathcal{A}_{t+1}\gamma_t = \mu_{t+1} > 0$. Furthermore the assumption that $\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1} > 0$, the FOC for labor at time 0, and $\Omega^l = 0$ imply $\mu_0 > 0$. Therefore, by induction $\mu_t > 0$ for all $t \geq 0$, $\gamma_t = \mathcal{A}_{t+1}\mu_{t+1} > 0$, hence $\tau_t^k = \widetilde{\tau}$ for all t.

Lemma 1 implies that $\mu_t, \gamma_t \to 0$. Since $\xi_t \geq 0$, taking limits in the consumption FOC, we have that $\Omega^c \leq 0$. □

Proof of Corollary 1. It is trivial that we have $\Delta_1 + \frac{\phi_2}{\phi_1}\mathcal{K}(\lambda)\Delta_2 > 0$, hence $\Omega^l > 0$. □

Proof of Proposition 2. The proof of part a) is obvious, the result is only stated for reference. Part b) 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 PO allocation. Invertibility in **A3** 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

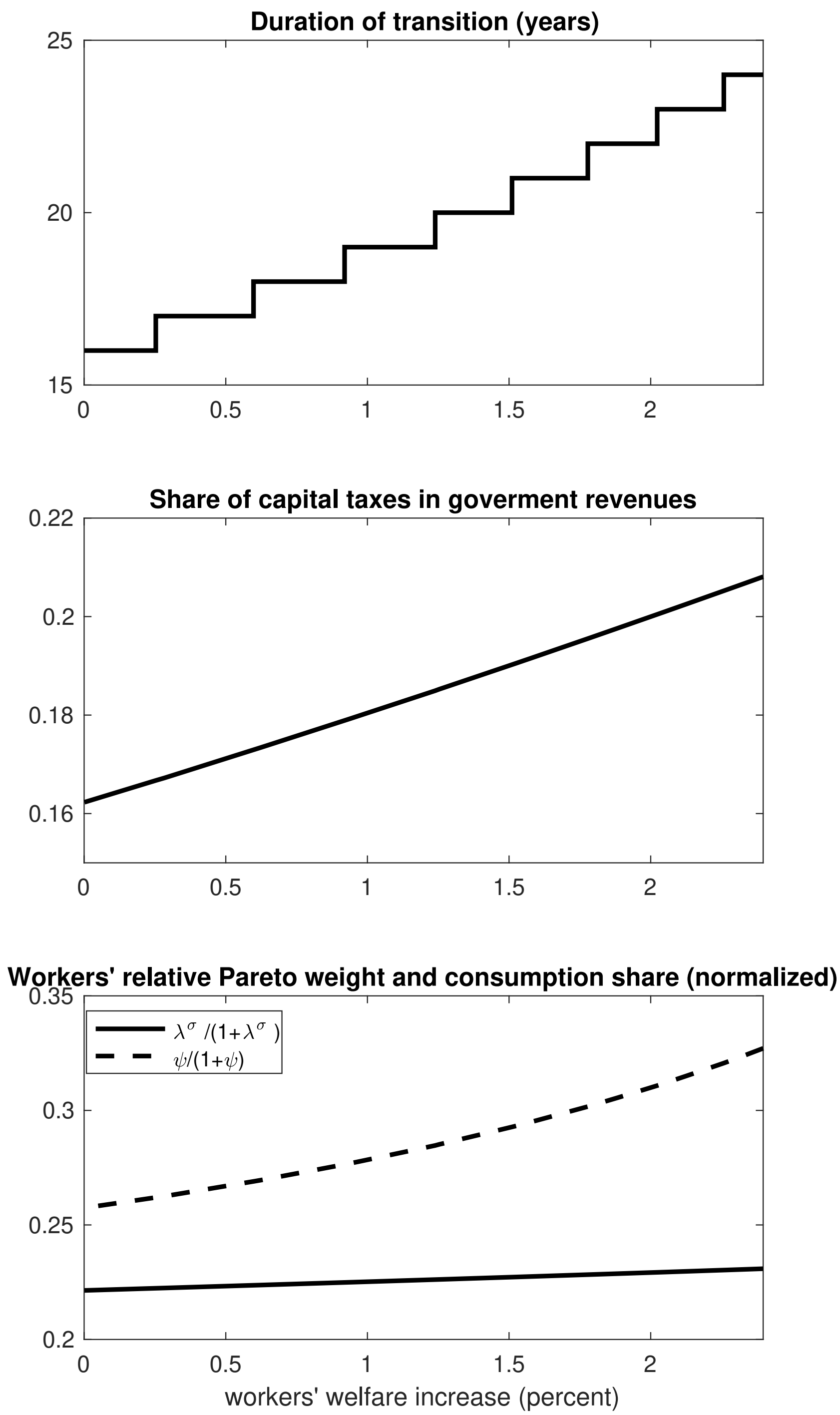
$$\overline{U}_1 + \bar{\psi}\overline{U}_2 \leq U_1\left(\bar{\psi}\right) + \bar{\psi}U_2\left(\bar{\psi}\right),$$

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 PO allocation implies $\overline{U}_1 \geq U_1\left(\overline{\psi}\right)$. Therefore, the PO allocation with utilities $\left(\overline{U}_1, \overline{U}_2\right)$ attains the maximum of MM with $\bar{\psi}$. Uniqueness implies that this PO 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 c). If $\psi \geq 0$ then part c) follows from part b). Consider now a given $\psi < 0$. We can find points in $\mathbb{R}^2$ outside $\mathcal{S}^U$ which are arbitrarily close to $\left(U_1(\psi), U_2(\psi)\right)$ as follows: for any $\varepsilon > 0$ we have $\left(U_1(\psi) + \varepsilon, U_2(\psi) - \varepsilon\right) \notin \mathcal{S}^U$, since this point achieves a higher value of the objective function of MM than its maximum. Since $\left(U_1(\psi) + \varepsilon, U_2(\psi) - \varepsilon\right)$ can be made arbitrarily close to $\left(U_1(\psi), U_2(\psi)\right)$, this last point is on the frontier $\mathcal{F}$. □

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

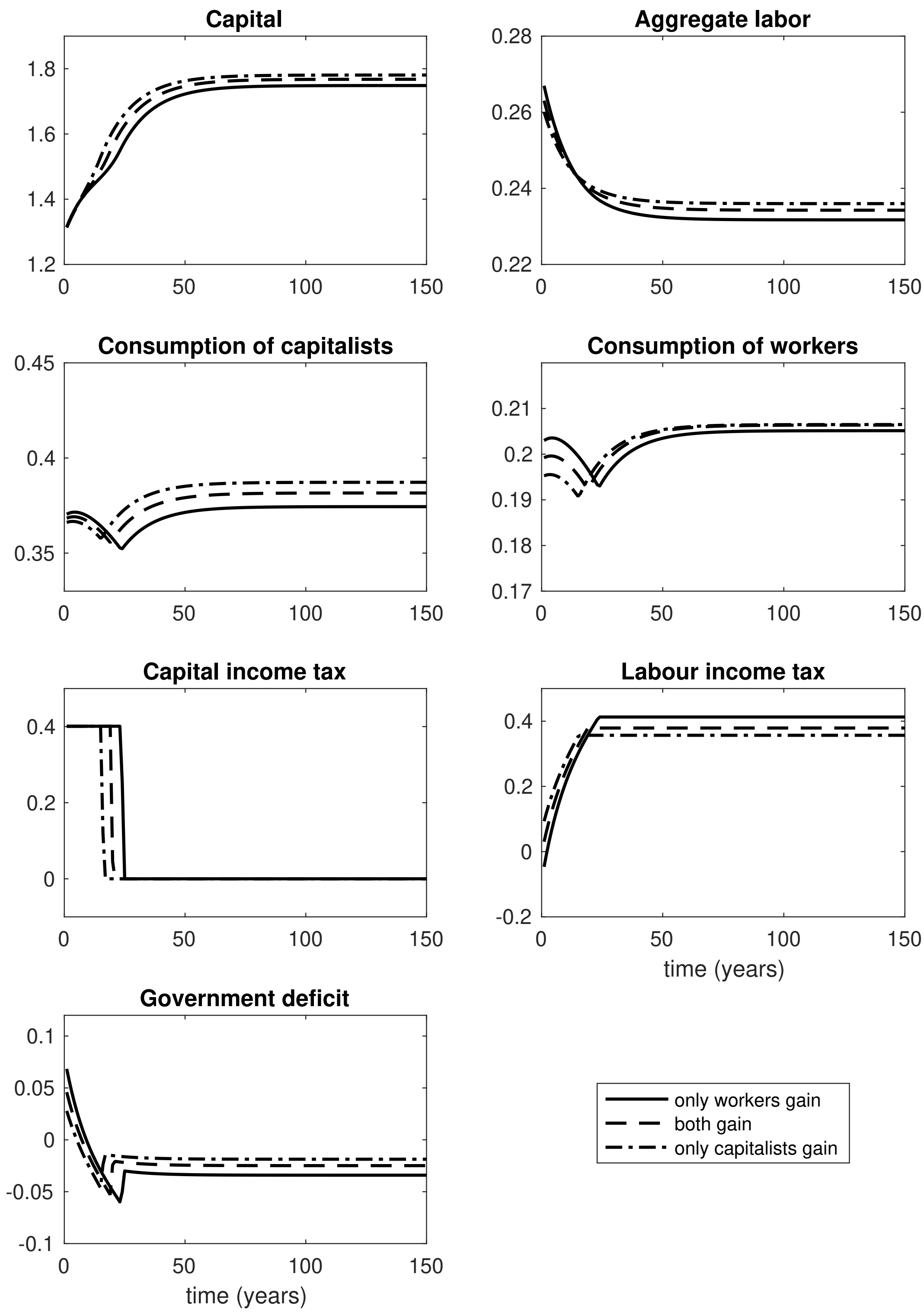
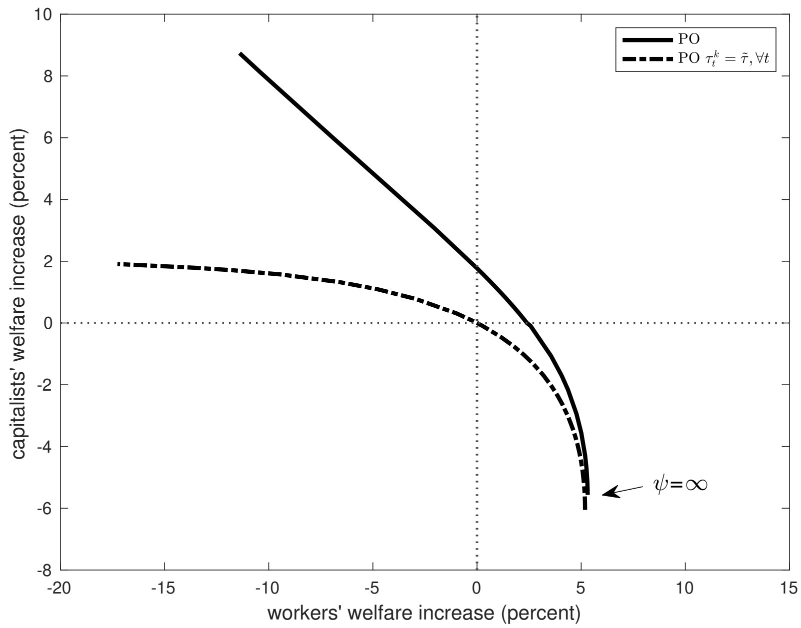
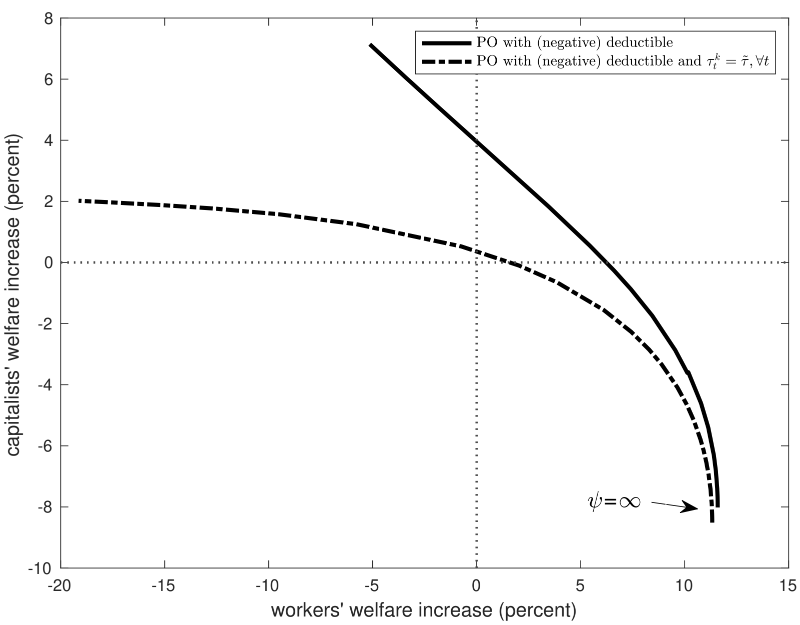


Figure 5: Comparison of Ramsey Pareto frontiers without a deductible



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.

Figure 6: Comparison of Ramsey Pareto frontiers with a deductible



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.