

Pareto-Improving Optimal Capital and Labor Taxes^{*}

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

January 2015

Abstract

We study optimal fiscal policy in a model with agents who are heterogeneous in their labor productivity and wealth, and there is an upper bound on the capital tax rate each period. We focus on Pareto-improving plans. We show that the optimal tax reform is to cut labor taxes and leave capital taxes high in the short and medium run. Only in the very long run would capital taxes be zero. For our calibration labor taxes should be low for the first eleven to twenty-six years, while capital taxes should be at their maximum. This policy ensures that all agents benefit from the tax reform and that capital grows quickly after the reform. Therefore, the long-run optimal tax mix is the opposite of the short- and medium-run one. The initial labor tax cut is financed by deficits, which lead to a positive level of government debt in the long run, reversing the standard prediction that the government accumulates savings in models with optimal capital taxes. The welfare benefits from the tax reform are high and can be shifted entirely to capitalists or workers by varying the length of the transition.

JEL classification: E62, H21

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

^{*}We wish to thank Jordi Caballé, Begoña Domínguez, Joan M. Esteban, Andreu Mas-Colell, Michael Reiter, Sevi Rodríguez, Raffaele Rossi, Kjetil Storesletten, Jaume Ventura, Iván Werning, Philippe Weil, Fabrizio Zilibotti, and seminar participants in various institutions for useful comments and suggestions. Michael Reiter provided the implementation of Broyden's algorithm used in this paper. Greulich acknowledges support from the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK) and the Research Priority Program on Finance and Financial Markets of the University of Zurich. Laczó acknowledges funding from the JAE-Doc grant co-financed by the European Social Fund. Marcet acknowledges CREI, DGES, CIRIT (SGR2005-00097) and Barcelona GSE Research and the XREA network for their support.

[†]Swiss Re

[‡]Bank of England, Threadneedle Street, EC2R 8NH, London, United Kingdom. Email: sarolta.laczo@bankofengland.co.uk.

[§]Institut d'Anàlisi Econòmica (CSIC), ICREA, UAB, MOVE, Barcelona GSE & CEPR, Campus UAB, 08193 Bellaterra, Barcelona, Spain. Email: albert.marcet@iae.csic.es

1 Introduction

A large literature on optimal dynamic taxation concludes that long-run capital taxes should be zero. This result, which originally goes back to [Chamley \(1986\)](#) and [Judd \(1985\)](#), has been very resilient to many modifications of the basic model.¹ That capital taxes should be so low is a controversial policy recommendation. Given the highly skewed distribution of wealth, one expects that lowering capital taxes and increasing labor taxes instead would necessarily hurt less wealthy taxpayers. However, in standard models capital taxes should be zero in the long run even with heterogeneous agents, and even if the government only considers policy allocations that improve the welfare of consumers with very little wealth ([Atkeson, Chari, and Kehoe, 1999](#), for example).² One interpretation of this result has been that there is no equity-efficiency trade-off involved in lowering capital taxes.

We consider a model in which agents are heterogeneous in their labor productivity-wealth ratio ([Flodén, 2009](#), [Correia, 2010](#), [Garcia-Milà, Marcet, and Ventura, 2010](#)), excluding lump-sum transfers, and with an upper bound on capital taxes below 100%. We focus on Pareto-improving tax reforms. These features, first of all, create a meaningful equity-efficiency trade-off. Moreover, the restriction to Pareto-improving allocations is natural because the surprising part of the Chamley-Judd result is that long-run capital taxes should be zero even if the government improves all consumers' welfare. Aside from this literature-driven motivation, it seems that a sufficiently large and angry minority can block a tax reform, or it may credibly threaten to overturn the reform in a future vote, so that in order to change the taxation status quo, a sufficiently large part of the population should agree. Further, we deviate from much of the optimal policy literature in explicitly studying the entire path of optimal capital and labor taxes, not only the steady state.

Recent literature on optimal policy in models with heterogeneous, infinitely-lived agents includes [Niepelt \(2004\)](#) and [Bassetto \(2014\)](#), who study how taxes affect taxpayers of different wealth in stochastic models without capital. [Werning \(2007\)](#) studies redistribution with progressive taxation. [Flodén \(2009\)](#) considers a model with many labor productivity-wealth

¹The literature is very large, a fair summary would be quite long. A very incomplete summary of the literature is that in the few cases where steady-state optimal capital taxes are not zero, they are often small or even negative, or capital approaches zero in the long run ([Straub and Werning, 2014](#)).

²[Aiyagari \(1995\)](#) shows that capital taxes should be positive in the long run due to capital overaccumulation in a model with heterogeneous agents and incomplete markets. We do not focus on this implication of heterogeneous agents for two reasons. First, because the result is tenuous: [Chamley \(2001\)](#) shows that depending on the stochastic form of income shocks the long-run capital tax could be negative, and [Marcet, Obiols-Homs, and Weil \(2007\)](#) argue that a similar result holds when introducing endogenous labor supply. Second, the result is specifically for 'the veil of ignorance' welfare function, and it may not hold for other welfare functions.

types but with Gorman-aggregable preferences,³ and studies optimal policies for different welfare functions, which is equivalent to maximizing the welfare of one agent. The policies are typically not Pareto-improving. Our aim is to find optimal tax reforms from which all groups of consumers benefit.

We first show analytically that in our model optimal capital taxes are still zero in the long run. However, they will be at the upper bound for some periods, and then transit to zero in two periods. To find the length of the transition period and the effects of the tax reform on allocations and welfare, we turn to numerical methods.

To demonstrate the effects of heterogeneity in isolation we first study a model with completely inelastic labor supply. In this case the first best is achieved with homogeneous agents by setting capital taxes to zero in all periods. However, with heterogeneous agents capital taxes should be positive for a very long time, between 13 and 25 years for our calibration, before they are abolished. This long period of high capital taxes is needed in order to raise more tax revenues from the capitalists and less from the workers. Only then all consumers benefit from the tax reform. Therefore, even though the planner has access to non-distortive labor taxes, she has to resort to distortive capital taxation to lower the workers' tax burden. The resulting total welfare losses are quite large, but they are needed to ensure a Pareto improvement.

If labor supply is somewhat elastic, we also find that redistributive concerns cause the transition to be very long: capital taxes should be high for 11 to 26 years before they are set to zero. Now we find that labor taxes should be lower than at the status quo during the transition. Lower initial labor taxes increase labor supply, thus promoting growth in the early periods. This implies that optimal factor taxation depends very much on heterogeneity. In addition, the long-run properties of optimal policies should not be used for policy recommendations. Zero capital taxes in the long run are only Pareto optimal and Pareto improving if they go along with high capital taxes and low labor taxes during the transition. Furthermore, in light of this result, high capital taxes can be part of an optimal reform, and they are not necessarily a failure of a political system or a result of frequent voting, as some papers in the political economy literature suggest. The puzzle now would become, why are labor taxes so high?

Our results are complementary to some papers which establish that large parts of the

³Gorman-aggregable preferences are attractive for tractability, however, assuming that each consumer spends the same share of any additional income on each good independent of wealth is restrictive. Further, the distribution is not allowed to play any role for aggregate variables. Finally, such preferences are not consistent with balanced growth.

population would suffer a large utility loss if capital taxes were suddenly abolished. Relevant references are [Correia \(1999\)](#) (some analytic results), [Domeij and Heathcote \(2004\)](#) (a model with incomplete markets), [Conesa and Krueger \(2006\)](#) (with overlapping generations), [Flodén \(2009\)](#) (a model with many labor productivity-wealth types but with Gorman-aggregable preferences), and [Garcia-Milà, Marcet, and Ventura \(2010\)](#). The results of these contributions stand in stark contrast to [Lucas \(1990\)](#), who showed that the welfare of a representative agent would increase if capital taxes were abolished immediately and all tax revenue were obtained by taxing only labor. Hence, while designing the transition of capital and labor taxes optimally may not be very important with homogeneous agents, with heterogeneous agents there is indeed an important equity-efficiency trade-off. Our results are further in line with the literature on gradualism of political reforms, which has been at the center of some policy debates.⁴ The very long period of high capital taxes we find can be seen as a gradual reform designed to ensure that all consumers' welfare improves.

In our main model government debt is positive in the long run. This is because the government initially runs a deficit to finance the initial drop in labor taxes. The behavior of long-run debt is, therefore, the opposite from the standard case under capital taxation, where the government often accumulates savings. In the face of the recently-renewed interest in studying the determinants of the optimal level of debt,⁵ this shows that a positive level of government debt can be a by-product of an optimal reform.

The results are robust to various parameter changes. We also explore if progressive taxation might achieve redistribution to ensure Pareto improvements and avoid the distortions associated with high capital taxes. We find that this is not the case. Finally, we also investigate numerically the time consistency of the solution. We find that the tax reform is time consistent if it can only be overturned by consensus, therefore, heterogeneity and the requirement of Pareto improvement build in some time consistency. This is in line with [Armenter \(2004\)](#), who finds the same result analytically in a simpler model.

The paper is organized as follows. In [Section 2](#) we lay out our baseline model and discuss further the motivation for our assumptions. [Section 3](#) presents some properties of the model obtained analytically, including a proof that capital taxes are zero at the steady state and about the form of the transition. Our numerical results are in [Section 4](#). [Section 5](#) concludes.

⁴For 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 show a gradual reform that improves all consumers' welfare.

⁵See [Faraglia, Marcet, and Scott \(2010\)](#) and the references therein.

2 The Model

We consider an economy with two heterogeneous consumers, discrete time, capital accumulation, but without uncertainty. Our emphasis changes in four aspects relative to most of the Ramsey taxation literature.

- i) We study the whole path for taxes.
- ii) We preclude agent-specific redistributive lump-sum transfers. It is well known that the Chamley-Judd result survives even in this case.⁶ This assumption seems reasonable in a literature that has focused on the effects of distortive taxation, and because most tax codes (and constitutions) stipulate that all individuals are equal in front of the law.
- iii) We search for allocations that improve the welfare all groups in the population, i.e., we study plans that are Pareto improvements.
- iv) We impose an upper bound on the capital tax rate in each period. [Chamley \(1986\)](#) and [Atkeson, Chari, and Kehoe \(1999\)](#) assume an upper bound of 100 percent for capital taxes in all periods. Many other papers in the optimal taxation literature assume a bound only in the initial period. Optimal policies under these constraints imply that capital taxes should be very high in the first few periods, much higher than current actual capital taxes which, by all measures, are already high. The initial tax hike recommended by these models could have devastating effects on investment in the real world if there is partial credibility of government policy, or if agents form their expectations by learning from past experience.⁷ To avoid this tax hike in the initial periods we use a capital tax ceiling lower than 100 percent. In particular, in all of our computational exercises we fix this ceiling to the status-quo capital tax rate. Alternatively, this bound can be interpreted as a value that avoids massive capital flight in an open economy with partial mobility of capital.

⁶To our knowledge the first complete proof that capital taxes are zero in the long run even in the absence of agent-specific lump sum redistribution is in Section B1 of [Chari and Kehoe \(1999\)](#) also described in [Atkeson, Chari, and Kehoe \(1999\)](#). [Chamley \(1986\)](#) discusses the case of heterogeneous agents by considering a government with a welfare function that weighs all consumers linearly, this implicitly assumes there are redistributive lump-sum transfers. [Judd \(1985\)](#) considered a model where an consumer has only capital income and another consumer has only wage income, so the distorting taxes in that paper are, in a way, agent-specific.

⁷[Lucas \(1990\)](#) offered a similar reasoning to motivate his study of a tax reform that abolishes capital taxes immediately. Of course, one could infer from our discussion that issues such as credibility and learning should be introduced explicitly in the analysis, instead of indirectly with the tax limit. Doing so would imply deviating very much from the benchmark models in the dynamic Ramsey taxation literature. In this paper we prefer to stay as close as possible to those models in order to understand the reasons for our results. The time consistency literature deals, in a way, with the credibility issue. An analysis of capital taxes under learning can be found in [Giannitsarou \(2006\)](#).

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

2.1 The environment

There are two consumers, $j = 1, 2$, with utility $\sum_{t=0}^{\infty} \beta^t [u(c_{j,t}) + v(l_{j,t})]$, where $c_{j,t}$ is consumption and $l_{j,t}$ is labor of consumer j in period t . We assume $u_c > 0$, $v_l < 0$, and usual Inada and concavity conditions. Agents differ in their initial wealth $k_{j,-1}$ and their labor productivity ϕ_j . Agent j obtains income in period t from renting his/her capital at the rental price r_t and from selling his/her labor for a wage $w_t \phi_j$. Agents pay taxes at rate τ_t^l on labor income and τ_t^k on capital income net of depreciation allowances. Therefore, the period- t budget constraint of consumer j is given by

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

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

The government chooses capital and labor taxes, has to spend g in every period, saves in capital, and has initial capital k_{-1}^g . The government can get indebted, that is, k^g can be negative. Ponzi schemes for consumers and the government are ruled out.

We focus on the case with two types of consumers, and normalize the mass of each group to $1/2$. Then the market clearing conditions are

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

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

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

for all t .

2.2 Conditions of competitive equilibria

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

Combining the first-order conditions (FOCs) with respect to consumption and labor for consumer j yields

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

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

the Euler equation and the consumption-labor optimality condition, respectively, for all j . We assume that the current utility function is

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

where ω is the relative utility weight of hours, σ_c is the coefficient of relative risk aversion, and σ_l is the inverse of the (constant) Frisch elasticity of labor supply. In this case the above FOCs imply

$$\frac{c_{2,t}}{c_{1,t}} = \lambda \quad \text{and} \quad \frac{l_{2,t}}{l_{1,t}} = \lambda^{-\frac{\sigma_c}{\sigma_l}} \left(\frac{\phi_2}{\phi_1} \right)^{\frac{1}{\sigma_l}}, \quad \forall t, \quad (7)$$

for some λ constant through time.⁸

Using equation (4) the budget constraints of consumer j for all $t = 0, 1, \dots$ can be summarized in the present-value budget constraint

$$\sum_{t=0}^{\infty} \beta^t \frac{u'(c_{j,t})}{u'(c_{j,0})} (c_{j,t} - w_t \phi_j l_{j,t} (1 - \tau_t^l)) = k_{j,-1} (1 + (r_0 - \delta) (1 - \tau_0^k)), \quad \text{for } j = 1, 2. \quad (8)$$

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

$$\sum_{t=0}^{\infty} \beta^t (u'(c_{1,t}) c_{1,t} + v'(l_{1,t}) l_{1,t}) = u'(c_{1,0}) k_{1,-1} (1 + (r_0 - \delta) (1 - \tau_0^k)). \quad (9)$$

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

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

⁸Note that labor supply depends not only on labor productivity but also on the distribution of consumption/wealth through λ , unlike under Gorman-aggregation.

where

$$f(\lambda, l_{1,t}) \equiv \lambda^{-\frac{\sigma_c}{\sigma_l}} \left(\frac{\phi_2}{\phi_1} \right)^{\frac{1}{\sigma_l}} l_{1,t}, \quad (11)$$

which gives $l_{2,t}$ for each possible value of the endogenous variables λ and l_t^1 from (7).⁹

One can show that necessary and sufficient conditions for a sequence $\{c_t^1, k_t, l_t^1\}$ and a constant λ to be a CE are feasibility (3) and (9) and (10). In other words, as long as these conditions hold capital and labor taxes can be found that ensure that all FOCs of consumers hold. Note that the presence of heterogeneous agents implies that the ratio λ has to be found optimally subject to the constraints (9) and (10).

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

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

Using these conditions we can eliminate factor prices from the characterization of competitive equilibria.

Once we have solved for the optimal allocations, taxes can be found from (4) and (5). Consumption and labor of consumer 2 are found from (7), and individual capital is backed out from the budget constraint period by period. Finally, factor prices are found using the firms' optimality conditions.

2.3 The policy problem

We assume that the planner chooses *Pareto-optimal* allocations. A standard argument justifies that this is equivalent to assuming that the planner maximizes the utility of, say, consumer 1, subject to the constraint that the utility of consumer 2 has a minimum value of $\underline{U}^2$, i.e.,

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

where $\underline{U}^2$ is restricted so that the set of feasible competitive equilibria satisfying this constraint is non-empty. Varying the value of the minimum utility $\underline{U}^2$ along all possible utilities that can be achieved in equilibrium for consumer 2, we can trace out the whole set of Pareto-efficient allocations.

We will concentrate our attention on allocations that are *Pareto improving* relative to a certain status quo. Let U_{SQ}^j be the status-quo utility obtained by consumer j , achieved with

⁹Walras' law guarantees that the budget constraint of the government is implied by the above equations plus feasibility so it can be ignored.

some taxation scheme that is already in place.¹⁰ The Pareto-improving allocations can be found by considering only minimum utility values for consumer 2, $\underline{U}^2$, such that $\underline{U}^2 \geq U_{SQ}^2$ and that the planner's objective at the maximum satisfies

$$\sum_{t=0}^{\infty} \beta^t [u(c_{1,t}^*) + v(l_{1,t}^*)] \geq U_{SQ}^1, \quad (13)$$

where $*$ denotes the optimized value of each variable for a given $\underline{U}^2$. Hereafter we refer to these Pareto-optimal and Pareto-improving plans as 'POPI' allocations. Proposition 2 below provides a way to compute all the utility values on the frontier and to select the POPI allocations.

Finally, we introduce a *tax limit*, denoted $\tilde{\tau}$, and impose $\tau_t^k \leq \tilde{\tau}$ for all $t = 0, 1, \dots$, i.e., that capital taxes are never above a certain constant $\tilde{\tau}$ exogenously given. Combining this limit with the Euler equation of consumer 1, it is easy to see that the tax limit is satisfied in equilibrium if and only if

$$u'(c_{1,t}) \geq \beta u'(c_{1,t+1}) (1 + (r_{t+1} - \delta)(1 - \tilde{\tau})), \quad \forall t > 0, \quad (14)$$

$$\tau_0^k \leq \tilde{\tau}. \quad (15)$$

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

We look for a Ramsey equilibrium where the government chooses an optimal sequence of tax rates and deficits to maximize the utility of consumer 1, subject to the constraint that taxes and prices have to be compatible with CE and subject to the above additional constraints. As is standard in the Ramsey taxation literature, we assume that the government has full credibility, i.e., full commitment to the announced policies, and both the government and the agents have rational expectations.

The government/planner solves

$$\begin{aligned} \max_{\tau_0^k, \lambda, \{c_t^1, k_t, l_t^1\}_{t=0}^{\infty}} \quad & \sum_{t=0}^{\infty} \beta^t [u(c_{1,t}) + v(l_{1,t})] \\ \text{s.t.} \quad & \sum_{t=0}^{\infty} \beta^t [u(\lambda c_{1,t}) + v(f(\lambda, l_{1,t}))] \geq \underline{U}^2, \end{aligned} \quad (16)$$

feasibility (3), the implementability constraints (9) and (10), and the tax limits, (14) and (15). We have used (7) and (11) to substitute for c_2 and l_2 to obtain (16). $\underline{U}^2$ has to satisfy

¹⁰The status-quo utility, in general, depends on the distribution of capital at period -1 , but we leave this dependence implicit.

the requirements discussed above to achieve a Pareto improvement. Notice that a special feature of this problem is that the constant λ has to be determined as a part of welfare maximization, therefore it appears as an argument in the optimization problem.

Let ψ be the Lagrange multiplier of the minimum utility constraint (16), let Δ_1 and Δ_2 be the multipliers of (9) and (10), respectively, and let γ_t be the multiplier of (14). The Lagrangian for the government's problem is

$$\begin{aligned} \mathcal{L} = \sum_{t=0}^{\infty} \beta^t & \left\{ u(c_{1,t}) + v(l_{1,t}) + \psi [u(\lambda c_{1,t}) + v(f(\lambda, l_{1,t}))] \right. \\ & + \Delta_1 [u'(c_{1,t}) c_{1,t} + v'(l_{1,t}) l_{1,t}] \\ & + \Delta_2 \left[u'(c_{1,t}) \lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'(l_{1,t}) f(\lambda, l_{1,t}) \right] \\ & + \gamma_t [u'(c_{1,t}) - \beta u'(c_{1,t+1}) (1 + (r_{t+1} - \delta)(1 - \tilde{\tau}))] \\ & \left. + \mu_t \left[F(k_{t-1}, e_t) + (1 - \delta)k_{t-1} - k_t - \frac{1 + \lambda}{2} c_{1,t} - g \right] \right\} - \psi \underline{U} - \mathbf{W}, \end{aligned} \quad (17)$$

where $\mathbf{W} = u'(c_{1,0}) (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}) (1 + (r_0 - \delta)(1 - \tau_0^k))$. Further, $\gamma_t \geq 0$ and $\mu_t \geq 0$, $\forall t$, and $\psi \geq 0$, with complementary slackness conditions. We know that the resource constraint binds in each period, hence, $\mu_t > 0$, $\forall t$.

The first line of this Lagrangian has the usual interpretation: finding a Pareto-efficient allocation amounts to maximizing a welfare function where the planner weights linearly the utility of the two consumers. The weight of consumer 1 is normalized to one and the weight of consumer 2 is the Lagrange multiplier of the minimum utility constraint. However, it is important that this ψ is not chosen arbitrarily in our setup. Instead it has to satisfy the Pareto-improving constraints. The next two lines correspond to some terms in the budget constraints of consumers. The fourth line ensures that $\tau_t^k \leq \tilde{\tau}$ for all $t > 0$. The last line is the feasibility constraint. The term $\mathbf{W}$ collects the period-0 terms in the budget constraints of the consumers.

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

The tax limit is a forward-looking constraint, therefore standard dynamic programming does not apply. Using a promised-utility approach would be complicated because of the appearance of a state variable (marginal utility of consumption) that has to be bounded to stay in the set of feasible marginal utilities, and, since there is also a natural state variable, k , characterizing this set would be quite difficult. The Lagrangian approach of [Marcet and Marimon \(2011\)](#) is easier to use under these circumstances. Appendix A shows the recursive

Lagrangian and the FOCs with respect to consumption, labor, and capital. In the rest of this section we comment on features of the remaining FOCs which differ from other papers on dynamic taxation.

Since the relative consumption of consumers, λ , is a choice variable, we need to set the derivative of $\mathcal{L}$ with respect to λ equal to zero. This gives

$$\begin{aligned} & \sum_{t=0}^{\infty} \beta^t \left\{ \psi [u'(\lambda c_{1,t}) c_{1,t} + v'(f(\lambda, l_{1,t})) f_{\lambda}(\lambda, l_{1,t})] \right. \\ & + \Delta_2 \left(u'(c_{1,t}) c_{1,t} + \frac{\phi_2}{\phi_1} v'(l_{1,t}) f_{\lambda}(\lambda, l_{1,t}) \right) \\ & \left. - \gamma_{t-1} u'(c_{1,t}) F_{ke}(k_{t-1}, e_t) \frac{\phi_2}{2} f_{\lambda}(\lambda, l_{1,t}) (1 - \tilde{\tau}) - \frac{\mu_t}{2} (c_{1,t} - F_e(k_{t-1}, e_t) \phi_2 f_{\lambda}(\lambda, l_{1,t})) \right\} \\ & - u'(c_{1,0}) (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}) F_{ke}(k_{-1}, e_0) \frac{\phi_2}{2} f_{\lambda}(\lambda, l_{1,0}) (1 - \tau_0^k) = 0. \end{aligned} \quad (18)$$

The fact that λ has to be chosen is a reflection of the fact that the government can vary the ratio of consumptions of the consumers by varying the total tax burden of labor and capital in discounted present value.

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

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

In other words, the minimum utility constraint may or may not be binding. In the first case, the Lagrangian amounts to maximizing the weighted sum of utilities of consumers 1 and 2 with weight 1 and ψ , respectively. If the minimum utility constraint is not binding, the planner gives zero weight to consumer 2. The latter case would only occur in models without frictions if the planner would be willing to give a very low utility to consumer 2. We show below that in our case it occurs even if the lower bound $\underline{U}^2$ is the status-quo utility. This is because even if $\psi = 0$, consumer 2 has to consume due to the fact that the allocations are determined in equilibrium, which implies that the budget constraint of consumer 2 has to be satisfied, insuring consumer 2 some revenue for any tax policies.

Similarly, for γ_t for each t we have that

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

We use these conditions below to characterize the path of capital taxes.

It turns out that the Δ_j 's may be positive or negative, since the corresponding present-value budget constraints have to be satisfied as equality. This becomes clear by looking at the following interpretation. With two consumers the marginal utility cost of distortive taxation is $\frac{\partial \mathcal{L}}{\partial \tau_0^k} = u'(c_{1,0}) (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}) (r_0 - \delta)$. Hence,

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

with the inequality being strict as long as any taxes are raised after the initial period. This allows for one of the Δ_j 's to be negative, which will indeed be the case whenever the constraints on redistribution that are imposed by the CE conditions and the Pareto-improvement requirement are sufficiently severe. To see this, consider a slightly modified model in which the social planner is allowed to redistribute initial wealth between consumers by means of lump-sum transfers T^j , $j = 1, 2$, such that $T^1 = -T^2$. All this modification does to the Lagrangian is to change the implementability constraints, in particular, the term $u'(c_{1,0}) (\Delta_1 - \Delta_2) T^1$ is added to **W**. Now, the derivative of the Lagrangian with respect to the lump-sum transfer between consumers is $\frac{\partial \mathcal{L}}{\partial T^1} = u'(c_{1,0}) (\Delta_1 - \Delta_2)$. For any given T^1 , and in particular for $T^1 = 0$ as in our baseline model, this expression is a measure of the marginal utility cost of the transfer not being optimal. If the planner were free to choose T^1 optimally, we would have $\Delta_1 = \Delta_2 > 0$. If the planner would like to redistribute more towards consumer 2, then $\Delta_1 - \Delta_2 > 0$, and vice versa. If the transfer is much too low (high), the derivative will be large in absolute value and Δ_2 (Δ_1) will be negative. In sum, while the weighted sum of the multipliers on the present-value budget constraints is related to the cost of distortive taxation, their difference indicates the cost of not being able to redistribute using lump-sum transfers. Hence, these multipliers capture in a simple way the two forces which drive the solution of our model away from the first best: the absence of lump-sum taxes and of agent-specific lump-sum transfers.

For the government's problem to be well defined we should ensure that the set of feasible equilibria is non-empty. This is guaranteed, for example, by the existence of a status-quo equilibrium, i.e., if $\tilde{\tau}$ is larger or equal to the status-quo capital tax, and if $\underline{U}^2$ is close to the status-quo utility.

3 Characterization of equilibria

Here we describe some analytical results.

3.1 Qualitative behavior of capital taxes

To the best of our knowledge there is no previous proof of zero long-run capital taxes which applies to our model.¹¹

Proposition 1. *Assume log utility of consumption ($\sigma_c = 1$), $c > 0$ at the steady state, $F(k, 0) = F(0, e) = 0$, and $0 < \tilde{\tau} < 1$. Then the optimal capital tax rate jumps from the tax limit to zero in two periods. Formally, there is a finite N such that*

$$\tau_t^k = \begin{cases} \tilde{\tau}, & \forall t \leq N, \\ 0, & \forall t \geq N + 2. \end{cases}$$

Proof. We proceed in two steps. First, we show that it is not possible for the tax limit to be binding forever in the optimal allocation. Then we show that capital taxes go from the limit to zero in two periods.

Now we prove that the tax limit cannot be binding in all periods. Let variables without a lower index t denote steady-state values. First of all, notice that in the log case the FOC with respect to consumption for $t > 0$ (see Appendix A) becomes

$$c_1^{-1} (1 + \psi\lambda) - c_1^{-2} [\gamma_t - \gamma_{t-1} (1 + (F_k(k, e) - \delta)(1 - \tilde{\tau}))] = \mu_t \frac{1 + \lambda}{2} \quad (19)$$

If indeed the solution had $\tau_t^k = \tilde{\tau}$ for all t , then (4) would be

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

Evaluating (19) at the steady state and plugging the last equation into the one above we have

$$\mathbf{A} - c_1^{-2} \left(\gamma_t - \frac{\gamma_{t-1}}{\beta} \right) = \mu_t \frac{1 + \lambda}{2}, \quad (20)$$

where $\mathbf{A} = c_1^{-1}(1 + \psi\lambda)$. Notice that we are only imposing that variables are constant, not multipliers. This is the right way to proceed, because real variables have natural bounds and existence of a steady state can be expected. On the other hand, the multipliers should not have bounds, otherwise there is no sense in which the Lagrangian is guaranteed to give a maximum, and a steady state in the variables could be compatible with multipliers that go to infinity.

¹¹The results in Chari and Kehoe (1999) and Atkeson, Chari, and Kehoe (1999) are similar, they also prove the tax limit cannot be binding forever and that the transition takes two periods. But the results in those papers are not directly applicable here. They do not consider a tax limit and heterogeneity at the same time, and, more importantly, their proof is for the case of a capital tax limit of 100 percent. For this particular bound if the tax limit were binding forever feasibility would be violated. In our case, where $\tilde{\tau}$ is the status-quo tax rate, the same line of argument cannot be used, because the economy could stay at the status quo forever. This is why a more-involved argument is needed.

The FOC for labor for $t > 0$ (see Appendix A) at steady state is

$$\begin{aligned} & -\omega l_1^{\sigma_l} \left[1 + \frac{\psi}{\lambda} \frac{\phi_2}{\phi_1} f_l(\lambda, l_1) + \left(\Delta_1 + \Delta_2 \frac{\phi_2}{\phi_1} f_l(\lambda, l_1) \right) (1 + \sigma_l) \right] \\ & - \gamma_{t-1} c_1^{-1} F_{ke}(k, e) \frac{1}{2} (\phi_1 + \phi_2 f_l(\lambda, l_1)) (1 - \tilde{\tau}) = -F_e(k, e) \frac{1}{2} (\phi_1 + \phi_2 f_l(\lambda, l_1)) \mu_t. \end{aligned} \quad (21)$$

Collecting the terms which do not depend on the multipliers γ or μ we have

$$\mathbf{B} + \mathbf{C} \gamma_{t-1} = \mu_t, \quad (22)$$

where

$$\begin{aligned} \mathbf{B} &= \frac{\omega l_1^{\sigma_l} \left[1 + \frac{\psi}{\lambda} \frac{\phi_2}{\phi_1} f_l(\lambda, l_1) + \left(\Delta_1 + \Delta_2 \frac{\phi_2}{\phi_1} f_l(\lambda, l_1) \right) (1 + \sigma_l) \right]}{F_e(k, e) \frac{1}{2} (\phi_1 + \phi_2 f_l(\lambda, l_1))}, \\ \mathbf{C} &= c_1^{-1} \frac{F_{ke}(k, e)}{F_e(k, e)} (1 - \tilde{\tau}). \end{aligned}$$

Notice that the terms $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ do not depend on the multipliers γ or μ . Note also that given our assumption that the steady state involves $c_1 > 0$, assumptions $F(k, 0) = F(0, e) = 0$ imply that $k > 0$ and $e > 0$. This implies that $0 < F_e(k, e) < \infty$. The constant returns to scale assumption and concavity imply that $F_{ke}(k, e)k = -F_{ee}(k, e)e > 0$. All these observations imply that

$$\frac{F_{ke}(k, e)}{F_e(k, e)} > 0, \quad (23)$$

Further, again given $c_1 > 0$ and in turn $k > 0$, μ has to be finite and hence constant at the steady state.

Now, let us rearrange (20) as

$$\gamma_t = \frac{\gamma_{t-1}}{\beta} + c_1^2 \mathbf{A} - c_1^2 \mu \frac{1 + \lambda}{2}.$$

Since $1/\beta > 1$, this is an unstable difference equation for all $\gamma \neq 0$. We know that $\gamma_t \geq 0$, $\forall t$, therefore, unless $\gamma = 0$ at the steady state, it converges to infinity. But then from (22) μ would also have to be infinite, a contradiction. Therefore, the tax limit cannot be binding in all periods, there has to be a period t where $\tau_t^k < \tilde{\tau}$.

Now we show that capital taxes go from the limit to zero in two periods. The previous argument implies that there is a finite $N + 1$ which is the first period where the tax limit is not binding, so that $\tau_{N+1}^k < \tilde{\tau}$ and $\tau_t^k = \tilde{\tau}$ for all $t \leq N$ in the optimum. Given N , consider the following modification to the baseline model. Assume that instead of the uniform tax limit in all periods, we impose

$$\tau_t^k \leq \tilde{\tau}, \quad \forall t \neq N + 1,$$

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

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

$$\tau_t^k \leq \tilde{\tau}, \quad \forall t \leq N,$$

but τ_t^k is unconstrained for all $t > N$. Let us denote with a $\hat{\cdot}$ the solution to MM2. Clearly, the FOCs for this model are the same as for the baseline model except that

$$\hat{\gamma}_t = 0, \quad \forall t \geq N. \quad (24)$$

Notice that γ_t is the multiplier associated with the constraint on τ_{t+1}^k , so that τ_{N+1}^k being the first unconstrained tax means γ_N is the first multiplier that must be 0.

Combining (24) with (19) implies¹²

$$\hat{c}_{1,t}^{-1} \left(1 + \psi \hat{\lambda} \right) = \hat{\mu}_t \frac{1 + \hat{\lambda}}{2}, \quad \forall t \geq N + 1. \quad (25)$$

This last equation does not hold for $t = N$, because $\hat{\gamma}_{N-1} \neq 0$ appears in (19). Plugging (24) in the FOC with respect to capital (see Appendix A) we get

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

Then, using (25) we have

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

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

$$\hat{\tau}_t^k = 0, \quad \forall t \geq N + 2. \quad (26)$$

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

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

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

¹²Notice that in order to obtain the following equation we absolutely need log utility. This equation would not hold for different risk aversions, because the term u'' in the FOC for consumption would not disappear, if capital is not exactly at steady state. Therefore, log utility is necessary in order obtain the proof.

Since we have already argued that the solution to MM1 was equal to the solution of the baseline model, this completes the proof.¹³ $\square$

3.2 The frontier of the equilibrium set

We now study the frontier of equilibrium utilities. Formally, let $\mathcal{F}$ be the frontier of the set

$$\left\{ (U_1, U_2) \in R^2 : U_j = \sum_{t=0}^{\infty} \beta^t [u(c_{j,t}) + v(l_{j,t})] \text{ for some } \{(c_{j,t}, l_{j,t})_{j=1,2}, k_t\} \text{ a CE} \right\}. \quad (27)$$

In the standard case without distortions and a concave utility function, it is well known that $\mathcal{F}$ coincides with the Pareto frontier, and it defines U_1 as a decreasing function of U_2 , or vice versa. In that case all these allocations can be traced out by optimizing welfare functions which give different weights to each consumer. Instead, given the distortions introduced by proportional taxes, we cannot be sure that the set (27) is convex and if considering a welfare function allows us to trace out the frontier of equilibria. Further, now the frontier of equilibria may not coincide with the set of Pareto-optimal allocations. There are indeed models where the frontier of equilibria has a convex part that cannot be found by maximizing welfare functions. However, we can still find sufficient conditions guaranteeing that by maximizing a welfare function in the standard way we obtain points on the frontier, and we can be confident that some of these points are Pareto optimal while others are not. Furthermore, we can give sufficient conditions for finding all Pareto-optimal allocations.

We consider another two minor modifications of the baseline model. In particular, first, we replace the minimum utility constraint (12) by an equality constraint

$$\sum_{t=0}^{\infty} \beta^t [u(c_{2,t}) + v(l_{2,t})] = \underline{U}^2, \quad (28)$$

where $\underline{U}^2$ is restricted so that the set of feasible competitive equilibria satisfying this constraint is non-empty. Let us call this model MM3. Second, MM4 consists of solving, for a given $\psi \in [-\infty, \infty]$,

$$\max \sum_{t=0}^{\infty} \beta^t \{u(c_{1,t}) + v(l_{1,t}) + \psi [u(\lambda c_{1,t}) + v(f(\lambda, l_{1,t}))]\}, \quad (29)$$

subject to all CE constraints and the tax limit. Notice that we allow for negative ψ 's and that we consider the case $\psi = \pm\infty$ as a convention to denote the case where consumer 1 or 2 receive no weight.

¹³Notice that for the proof to work we do need to consider the two modified models MM1 and MM2. If we tried to compare MM2 to the solution of the baseline model directly, we would not be able to rule out that $\tau_{N+1}^k > \tilde{\tau}$. The solution to MM2 would then be unfeasible in the baseline model and could not be compared to it.

We show that MM4 can be used to trace out a large part of the frontier $\mathcal{F}$ and the Pareto-optimal allocations on it. Given $\psi \in [-\infty, \infty]$, let $U_j(\psi)$ be the utility of consumer $j = 1, 2$ in the solution to MM4. We need two assumptions.

A1 *A solution to MM4 exists for all $\psi \in [-\infty, \infty]$. Also, $U_j(\psi)$ is well defined for $j = 1, 2$.*

That a solution exists is guaranteed by standard requirements such as that the set of equilibria is non-empty and that the utility functions are bounded. That $U_j(\psi)$ is well defined amounts to assuming that each ψ gives a unique utility level for each consumer or, equivalently, that $\mathcal{F}$ does not have a linear part.

A2 *$U_2(\cdot)$ is monotone decreasing and invertible for $\psi \in [0, \infty]$.*

Proposition 2. *Assume A1.*

1. *Given $\psi \in [-\infty, \infty]$, the solution of MM4 also solves MM3 for $\underline{U}_2 = U_2(\psi)$.*
2. *Given $\psi \in [-\infty, \infty]$, the solution of MM4 defines a point on the frontier:*

$$(U_1(\psi), U_2(\psi)) \in \mathcal{F}.$$

3. *Given $\psi \geq 0$, the solution to MM4 is Pareto optimal.*
4. *Assume, in addition, A2. Then, every Pareto-optimal allocation is also the solution of MM4 for some $\psi \geq 0$.*

Proof. Fix $\psi \in [-\infty, \infty]$. To show part 1, let $U_1^{MM3}(\underline{U}_2)$ be the value of the maximum of MM3. By definition, $U_1(\psi) + \psi U_2(\psi)$ is the value of the maximum of MM4. Since the solution to MM3 is feasible in MM4 we have

$$U_1(\psi) + \psi U_2(\psi) \geq U_1^{MM3}(U_2(\psi)) + \psi U_2(\psi)$$

so that $U_1(\psi) \geq U_1^{MM3}(U_2(\psi))$. Also, the solution to MM4 is feasible in MM3 for $\underline{U}_2 = U_2(\psi)$, therefore $U_1(\psi) \leq U_1^{MM3}(U_2(\psi))$. This shows that $U_1(\psi) = U_1^{MM3}(U_2(\psi))$, i.e., that the maxima of MM3 and MM4 coincide when $\underline{U}_2 = U_2(\psi)$

Part 2: for any $\psi \in [-\infty, \infty]$ we just need to find pairs of utilities which do not belong to the set (27) but are arbitrarily close to $(U_1(\psi), U_2(\psi))$. Consider any $\varepsilon > 0$. For $\psi \geq 0$ the pair of utilities $(U_1(\psi) + \varepsilon, U_2(\psi) + \varepsilon)$ is outside the set (27), otherwise it would have been chosen over the optimum in MM4, since it achieves a higher value of the objective, and can be made arbitrarily close by considering arbitrarily small ε . Therefore $(U_1(\psi), U_2(\psi))$

is on the frontier for $\psi \geq 0$. For $\psi \leq 0$ a similar argument shows that points such as $(U_1(\psi) + \varepsilon, U_2(\psi) - \varepsilon)$ are outside the feasible set and can be made arbitrarily close to $(U_1(\psi), U_2(\psi))$.

Part 3: if there were a feasible combination of utilities $(\tilde{U}_1, \tilde{U}_2)$ that Pareto dominates $(U_1(\psi), U_2(\psi))$ the optimum of MM4 would not be attained at $(U_1(\psi), U_2(\psi))$.

For part 4, note that if $(\hat{U}_1, \hat{U}_2)$ is a Pareto-optimal allocation, assumption A2 guarantees that there is a $\hat{\psi}$ such that $\hat{U}_2 = U_2(\hat{\psi})$. This and the fact that $(\hat{U}_1, \hat{U}_2)$ is Pareto optimal means that $\hat{U}_1 \geq U_1(\hat{\psi})$, hence $\hat{U}_1 + \hat{\psi}\hat{U}_2 \geq U_1(\hat{\psi}) + \hat{\psi}U_2(\hat{\psi})$. But the fact that the solution to MM4 is attained at $(U_1(\hat{\psi}), U_2(\hat{\psi}))$ means that the reverse inequality also holds, i.e., $\hat{U}_1 + \hat{\psi}\hat{U}_2 \geq U_1(\hat{\psi}) + \hat{\psi}U_2(\hat{\psi})$. Hence, the maximum of MM4 for $\hat{\psi}$ is attained at $(\hat{U}_1, \hat{U}_2)$. $\square$

Proposition 2 implies that by varying ψ from plus to minus infinity and maximizing (29), we can trace out points on the frontier of equilibria, $\mathcal{F}$, and all points $(U_1(\psi), U_2(\psi))$ for positive ψ are Pareto optimal. Furthermore, under A2 we are sure that we will find all Pareto-optimal allocations in this fashion. The points on $\mathcal{F}$ corresponding to negative ψ solve MM3 for $\underline{U}^2 = U_2(\psi)$, but these equilibria are not Pareto optimal, since, as indicated by the negative Lagrange multiplier ψ , the first consumer's utility could be increased in MM3 while increasing $\underline{U}^2$ as well.

More points on the frontier can be found if the consumers switch places in the objective function (29), that is, if ψ multiplies the utility of consumer 1. Then, by varying ψ (now the weight of consumer 1) from zero to negative infinity again, we could find points on the equilibrium frontier that are not Pareto optimal and that are obtained by forcing the planner to give a certain utility to consumer 1.

There is a caveat: assumptions A1 and A2 need to be checked. Since the feasible set is non-convex, the only way to check A1 is to check that there is only one solution to MM4. This can be done numerically by searching for more solutions to the FOCs, as in scores of papers where the maximum is found by searching for all critical points, and if more than one is found, the values of the objective function are compared. We can explore numerically if A2 holds by recording all utilities for a fine grid of ψ 's and checking that $U_2(\psi)$ is increasing. In this way we can be confident that we have found all Pareto-optimal competitive equilibria and that we traced out a "large part" of the frontier $\mathcal{F}$. We did this check for our numerical results shown below.

The Pareto-optimal and Pareto-improving (POPI) plans can be found as those points of

$\mathcal{F}$ which have a non-negative ψ and provide utilities which are larger than the status quo utilities. Note that non-optimal points on $\mathcal{F}$, i.e., points where $\psi < 0$, may also be Pareto improving relative to the status quo. In this case, $\mathcal{F}$ has an increasing part, where a plan strictly improves the utility of one of the consumers. i.e., it is not possible to shift all the gains to the other consumer in a Pareto-optimal way. This will be illustrated in a version of our model with fixed labor supply in Section 4.2, where the frontier features such a part.

4 Numerical Results

We now present and discuss our numerical results. Details on our computational strategy are in Appendix B. In the next subsection, we discuss how we calibrate the model. Afterwards, we first analyze the case in which labor supply is fixed. Second, in Section 4.3, we study optimal policy in an economy with flexible labor supply. We also contrast the results with those from an extension of our model in which lump-sum transfers are permitted, in order to gain intuition for the forces at work.

4.1 Calibration

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

We calibrate our parameters using the private sector’s FOCs at steady state, taking as given average effective tax rates and government debt, to match United States macro and micro data for the period 2001-2010. The macro variables are computed using the data provided by Trabandt and Uhlig (2012), who collected data from the OECD and other sources.¹⁴ We use the effective tax rates they have computed, in particular, $\tau^h = 0.214$ and $\tau^k = 0.401$. The average level of government debt was 66.8 percent of GDP. Note that the choice of tax rates and government indebtedness in the status quo matters for several reasons. First of all, they influence the status-quo steady-state (and hence initial) capital stock. Secondly, status-quo utilities depend on these variable, and thus restrict the scope for Pareto improvements. Thirdly, we assume that during the reform the capital tax rate can never increase above its initial level, i.e., we set $\tilde{\tau}$ equal to the status-quo capital income tax rate, 0.401.

We set some preference parameters a priori. The utility function is as stated in Section 2. We set the annual discount factor to the most-commonly-used value, $\beta = 0.96$. We choose

¹⁴<https://sites.google.com/site/mathiastrabandt/home/downloads/LafferNberDataMatlabCode.zip>

Table 1: Parameter values of the baseline economy

	Parameter	Value
Preference parameters	β	0.96
	σ_c	1
	σ_l	3
	ω	221.1
Heterogeneity parameters	ϕ_c/ϕ_w	1.1
	$k_{c,-1}$	3.831
	$k_{w,-1}$	-0.611
Production parameters	α	0.394
	δ	0.074
Public sector	g	0.094
	k_{-1}^g	-0.315
	τ^l	0.214
	τ^k	0.401

$\sigma_c = 1$ in keeping with a large part of the literature on taxation and in order to use Proposition 1. The choice of $\sigma_l = 3$ is for the case of an elastic supply of labor, which prevents hours from greatly differing across consumers with different wealth.¹⁵ Note that this implies a lower Frisch elasticity of labor supply than many applications of real business cycles but is in line with micro estimates.

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

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

Finally, we find ω , δ , g , k_{-1}^g , and the initial wealth of each group in the model, $k_{c,-1}$ and $k_{w,-1}$, consistent with the steady state given the status-quo tax rates, using (i) the

¹⁵See [Garcia-Milà, Marcet, and Ventura \(2010\)](#) for a discussion of the tradeoffs in choosing σ_l .

consumption-labor FOC of capitalists and that average hours should match the fraction of time worked for the working age population, $h = 0.245$,¹⁶ (ii) that capital held by the agents and the government has to equal steady-state aggregate capital, (iii) that g over output has to equal government consumption over GDP (iv) that k_{-1}^g over output has to match the public assets-GDP ratio from the data, and (v-vi) the present-value budget constraint of both groups.

4.2 Pareto-optimal and Pareto-improving plans with fixed labor supply

In general, the set of POPI plans deviates from the first best for two reasons. One is the standard reason in models of factor taxation: the need to raise tax revenue discourages the supply of capital and/or labor. The second reason is the lack of non-distortive means of redistribution between types of consumers. Since our paper is mostly about the latter, we first analyze a case in which only the redistributive effect is present. To do so, we assume fixed labor supply. In a model with homogeneous agents, the policy-maker would abolish capital taxes immediately and collect all revenues free of distortions from taxes on labor, and thus implement the first-best allocation. In a model with heterogeneous agents, if the government could stipulate agent-specific lump-sum transfers at time zero (with $T^w = -T^c$, as introduced at the end of Section 2), the problem of how to redistribute wealth would be resolved. Then the first-best could be achieved for any distribution of welfare gains. But in the case of interest where lump-sum redistribution is not possible, deviations from the first-best policy with zero capital taxes at all times are necessary for distributive reasons.

We set hours worked to 0.245 to match the data, see above. All parameters unrelated to the utility from leisure are as in Table 1.

In Figure 1 we compare the set of POPI plans to the first best in terms of welfare gains.¹⁷ First of all, note that A2 is satisfied and that by lowering ψ the utility of consumer 2 goes down. Therefore Proposition 2 can be used to trace out the Pareto-optimal and additional points on the frontier $\mathcal{F}$. The line labeled ‘first-best PI’ represents those allocations where agent-specific lump-sum transfers are available and which are Pareto improving. Clearly, the absence of transfers significantly reduces the scope for Pareto improvements. All POPI plans depicted on the solid line are inferior to the first best. Why? It turns out that the first-best

¹⁶Hours to be allocated between work and leisure: 13.64.

¹⁷In all the figures reporting results on welfare, the welfare gains for each consumer are measured as the percentage, 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 utility, and the positive orthant contains Pareto-improving allocations.

plans that are Pareto improvements upon the status quo would all involve positive transfers to the worker, $T^w > 0$. Absent these transfers, the immediate abolition of capital taxes would severely hurt the worker, as has been shown previously in a number of contributions.¹⁸ This is because capital taxes in the status quo are disproportionately borne by capitalists, and when they are abolished, labor taxes have to rise in order for the government to meet its budget constraint. This increase in labor taxes due to an immediate reform has a strong redistributive effect and, from the perspective of the worker, it would overcompensate the welfare gains arising from increased efficiency. The only thing the planner can do to make the abolition of capital taxes palatable for the worker is to keep capital taxes as high as possible for a long time (the N periods of Proposition 1) before setting capital taxes to zero from time $N + 2$ onwards.

Capital taxes have to be at the upper bound for 12 years (in the POPI plan where the worker gains nothing) to 24 years (when the worker gains the most possible). Some revenue is then still raised from capital taxes so that labor taxes need not provide all revenue. This implies a cost in efficiency, because the economy remains distorted for a long time, while it would be non-distorted if lump-sum redistribution were feasible. This is why POPI plans are second best even though taxation would be entirely non-distortive in a homogeneous-agent model or if lump-sum redistribution were available.¹⁹

The absence of a lump-sum redistributive instrument not only drives the set of POPI plans away from the first best, it also limits the degree to which welfare gains can be shifted towards the worker. This can be seen from Figure 1. By forcing the capitalist onto a lower utility level than the lowest point of the POPI curve, the planner would also harm the worker.

It is worthwhile noting that even though the utility loss relative to the first best is small if we only focus on equilibria which leave the worker indifferent and give all the benefits of the reform to the capitalist (i.e., if we focus on points where the frontiers cross the vertical axis of Figure 1), the utility loss is very large if we try to give some of the benefits to the worker. The most we can give to the worker is a 1.08% improvement, which is about one-seventh of the most the worker could gain with lump-sum redistribution. There is little to be gained from cutting capital taxes if the worker must enjoy most of the benefits. Finally, the optimal policy under ‘the veil of ignorance,’ i.e., when $\psi = 1$, gives welfare gains of 2.87 and 0.48 percent for capitalists and workers, respectively.

¹⁸See Correia (1999), Domeij and Heathcote (2004), Conesa and Krueger (2006), and Garcia-Milà, Marcet, and Ventura (2010).

¹⁹Notice that in the case of a fixed labor supply the evolution of labor taxes is undetermined, all that matters is that the net present value of labor taxes balances the government’s budget given the optimal path for capital taxes found.

4.3 Main model

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

4.3.1 Welfare frontier and capital tax

Figure 2 reports the set of POPI plans in terms of welfare gains. Again, we contrast our main model with the case with optimal agent-specific transfers, $T^w = -T^c$. Note that even with the access transfers, the first best is no longer attained, because positive distortive capital and/or labor taxes are needed to raise tax revenue to finance government spending.

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

As the distribution of welfare gains varies along the frontier of POPI plans, so do the corresponding capital tax schedule and relative consumption of agents. Qualitatively the properties of capital taxes over time are always the same: capital taxes are at their upper bound for all but the last period of the transition and then they stay at zero, as we know from Proposition 1. A typical time path for capital taxes is drawn in Figure 3.

The length of the transition increases as welfare gains are shifted towards the worker. This is illustrated in the first panel of Figure 4 showing the duration of the transition in the vertical axis for each POPI allocation indexed by the welfare gain of the worker on the horizontal axis. We see that the number of periods before capital taxes drop to zero increases from eleven to twenty-six years as we increase the welfare gain of the worker from zero (i.e., leaving the worker indifferent with status quo) to 1.8%, which leaves the capitalist indifferent with status quo. Along with the duration of the transition, the present value share of capital

taxes in government revenues increases from 12.7% to 21.7%, as the second panel in Figure 4 reveals.²⁰ This is the clue to why a longer period of high capital taxes is beneficial for the worker: the worker contributes to the public coffers primarily through labor taxes, which means that his burden in the long run stands to increase through the reform, while the capitalist's long-run burden decreases. The earlier capital taxes are suppressed, the more revenue has to be raised from labor taxes, and the bigger is the relative tax burden of the worker.

The final panel in Figure 4 depicts ψ , the multiplier on the minimum utility constraint (or, the relative Pareto weight) of the worker, and λ , the ratio of the worker's consumption to the capitalist's in equilibrium. We put these two variables in the same picture, because $\psi = \lambda$ would hold in a first-best situation, without distortive taxation or distributive conflict ($\Delta_1 = \Delta_2 = 0$), and if the upper bound on capital taxes never binds ($\gamma_t = 0, \forall t$). In our second-best world, by contrast, as we increase the welfare of the worker, the marginal cost of doing so (as measured by ψ) explodes, while his consumption share increases only mildly. In fact, it always remains very close to its value in the status quo, which is 0.54. This shows that it is very difficult to alter the ratio of consumptions, λ , even if the planner cares very differently about the two types of consumers, given that she has access only to proportional taxes.

If optimal lump-sum transfers were possible, the graphs in Figure 4 would look very different. We find that for all Pareto-optimal allocations capital taxes would be suppressed after 9 years, and the share of capital taxes would always be 10.3%. The multiplier on the worker's utility constraint ψ would increase very little with U^w , while λ would rise much more than without transfers. This is because in this case the redistribution can be achieved with agent-specific lump-sum taxes independently of the fact that the planner lowers quickly capital taxes to achieve aggregate efficiency. The policies and the path of the economy would hardly depend on the distribution of the gains from reform. Shifting welfare gains and consumption between agents would be much easier, as indicated by the behavior of ψ and λ .²¹

Focusing on Pareto-improving allocations means that the unit of interest is the utility that each consumer achieves through various tax reforms. Under this view, the weight ψ is just a Lagrange multiplier determined in equilibrium, and it measures the cost of enforcing the minimum utility constraint. The fact that ψ has to increase so much to achieve a small

²⁰For comparison, the share of capital taxes in revenues is about 37.1% in the status quo.

²¹Note that even with lump-sum transfers, we do not obtain $\psi = \lambda$, which only holds in optimal allocations if there is no distortionary taxation.

redistribution is a reflection of the difficulties the planner faces in redistributing wealth from one consumer to the other when only proportional capital and labor taxes are available.

Another way of looking at ψ is as the relative Pareto weight of the worker in the welfare function of the government. This suggests an interpretation of ψ as a measure of the bias of the social planner in favor of the worker. In particular, if one were to focus on the optimal allocation under ‘the veil of ignorance,’ the relevant policy would be the one corresponding to $\psi = 1$, since both types of consumers are equally abundant in the economy. This also corresponds to a model of probabilistic voting where both agents are equally influential. Many recent papers on dynamic optimal policy with heterogeneous agents use this welfare function. As can be seen in Figure 2, the optimal policy for $\psi = 1$ corresponds to a welfare gain by the worker of about 1.10%. The optimal reform under ‘the veil of ignorance’ happens to be Pareto improving. It gives more benefit to the capitalist (1.35%) than the worker. For this welfare function capital taxes are zero after 17 years.

In Appendix C we show that the main features described here are robust to changes in parameter values. In particular, we consider measurement issues for the relevant tax rates and consumption inequality at the status quo. We recalibrate and solve our baseline model considering both a lower and a higher value for each of the three data moments.

4.3.2 The time path of the economy

The evolution of capital, aggregate labor, consumptions, capital and labor tax rates, and government deficit are pictured in Figure 5. First, note that qualitatively the paths are very similar across the set of POPI plans. The horizontal shifts in the graphs occur because the more a plan benefits the worker, the longer capital taxes remain at their initial level. The kinks in the paths of labor taxes and government deficit occur precisely in the intermediate period when capital taxes transit from their maximum to zero.

The most surprising observation is, perhaps, that labor taxes should be initially lowered, and they should remain low for a long time. The reason for this behavior is the following: the planner wants to frontload capital taxes for the usual reason described at length in the literature that early capital taxes imply taxing capital that is already in place.²² Therefore, it is optimal to keep capital taxes at the upper limit in the first few periods and then let them go to zero. With such high capital taxes investors would not invest much. However, the government has another instrument which can be used to boost output and capital accumulation in the early periods. The government can lower labor taxes, inducing an

²²For example, Jones, Manuelli, and Rossi (1993) describe in detail this issue in several models with homogeneous agents.

increase in labor supply, causing the return on capital to go up, increasing investment in the initial periods, and achieving a faster convergence to the optimal long-run capital-labor ratio compatible with zero capital taxes. The upper right panel in Figure 5 shows that aggregate labor supply is very high in the early periods.²³ Note that the accumulation of capital accelerates around the period when capital taxes become zero, as can be seen by comparing the kink in the graph for labor taxes with the capital accumulation graph. Therefore, eventually the zero capital tax is the one promoting growth and helping the economy converge to the new steady state. Absent this backloading of labor taxes, capital would initially grow only to the extent that the expectation of low capital taxes in the distant future raises incentives to save early on. In this case capital accumulation would be much slower, as in the fixed-labor-supply case of Section 4.2. Therefore, low early labor taxes are an instrument to induce investment in the early periods in the case of elastic labor supply.

A similar result of low early labor taxes has been found in models of homogeneous agents.²⁴ The same pattern can be observed in our model if optimal transfers are allowed. We have computed that in the case with agent-specific lump-sum transfers, the period of low labor taxes would be much shorter, about 5-6 years, matching the lower duration of the transition to zero capital taxes. However, implementing this policy without lump-sum transfers would leave the workers worse than the status quo. Redistributive concerns lengthen the transition up to more than four times, as described in the previous paragraph.

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

A somewhat surprising pattern which emerges from the figures is that the long-run labor tax rate is higher for the policy that favors the worker. 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 for these policies, so that the share of labor taxes in the total present value of government revenues is lower for policies that favor more the worker, as the second panel of Figure 4 shows. This suggests that the long-run labor tax rate is high for two reasons. First, when capital taxation is abandoned late, the initial boost to capital accumulation comes mainly from extremely low initial labor

²³Capitalists, who also have higher labor productivity, always work less than workers.

²⁴For example, Section III of Jones, Manuelli, and Rossi (1993) shows a model where labor taxes should be very negative and capital taxes should be very high in the first period only.

taxes. That is, the backloading of labor taxes is strongest in these cases. Second, long-run labor supply is lower the later capital taxes are suppressed, while the gross wage is always the same.²⁵

Since government expenditures are constant, the low initial labor taxes translate into government deficits. Only as labor taxes rise and output grows the government budget turns into surplus. Once capital taxes are suppressed and revenues fall again, the government deficit quickly reaches its long-run value, which can be positive or negative depending on whether during the transition the government accumulated wealth or not. We can see from Figure 5 that most POPI policies imply that the government runs a primary surplus in the long run. This implies that the government is in debt in the long run, because the primary surplus is needed to pay the interest on debt. Therefore, for most POPI tax reforms low taxes in the initial periods generate a positive level of long-run government debt.

This feature of the model is quite different from that of Chamley (1986), where the government accumulated savings in the early periods to lower the labor tax bill in the long run. Here, the early drop in labor taxes is financed in part with long-run government debt, showing that one possible reason for government debt is to finance the initial stages of a reform.

4.4 Extentions

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

4.4.1 Progressive taxes

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

We assume that the planner can choose a lump-sum payment $\mathcal{D}$ which is paid in period zero uniformly across all consumers. Following Werning (2007), under complete mar-

²⁵Since the long-run real return on capital is determined by the rates of time preference and depreciation and the production function is Cobb-Douglas, the long-run capital-labor ratio and wage are independent of the policy, as long as capital taxes are zero eventually.

kets this is equivalent to a fixed deductible from the tax base in each period. A positive $\mathcal{D}$ means progressive taxation. Introducing this in the model means that we need to add $u'(c_{1,0}) [\Delta_1 + \Delta_2] \mathcal{D}$ to the $\mathbf{W}$ -term in equation (17). We then let the planner maximize over $\mathcal{D}$ additionally.

We find that if we restrict our attention to non-negative $\mathcal{D}$ (progressive taxation), the optimal choice is to set $\mathcal{D}=0$. Therefore, the government will choose *not* to use this progressive instrument. 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, redistributive concerns would advise the government to choose a positive $\mathcal{D}$, since capitalists are richer. But a negative $\mathcal{D}$ is equivalent to a lump-sum tax, and it allows to raise revenue in a distortion-free manner. In the standard case of a representative-agent model, where only this second force is present, the first best can be achieved by choosing a negative $\mathcal{D}$ big enough (in absolute value) to raise all government revenue ever needed. In our model with heterogeneous agents it turns out that the second force is stronger.

How can a negative $\mathcal{D}$ be Pareto improving? The government now redistributes by choosing very negative labor taxes for many periods. In fact, the present value of revenues from labor taxes is not only negative but even bigger in absolute value than the revenue from capital taxes. The transition is 6 and 25 years at the two extremes of the POPI frontier. Welfare gains are larger than in the case with optimal transfers: capitalists can gain maximum 5.0 percent and workers 3.7 percent in welfare-equivalent consumption units.

This optimal tax scheme (negative $\mathcal{D}$ and negative labor taxes) is Pareto improving only because we did not consider consumers of different wealth within each group of consumers in our calibration. In the real world some consumers have a high wage-wealth ratio who are rich (say, some young stockbrokers) and consumers with a low wage-wealth ratio who are poor (say, some farmers in economically depressed areas). We calibrated according to wage-wealth ratios, because, following [Garcia-Milà, Marcet, and Ventura \(2010\)](#), this is appropriate when only proportional taxes are allowed. Once progressive taxation is considered, the total income of the consumer is also relevant. Therefore, a careful study of progressive taxation should introduce total income in the calibration. In that case the optimal scheme described above would unlikely be Pareto improving. This is left for future research. However, the results in this subsection show that progressive taxation may have difficulties in solving the redistribution problem.

4.4.2 The evolution of wealth and welfare and time consistency

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

The time paths of consumers' welfare and wealth are plotted in Figure 6. Welfare increases along with the accumulation of capital, and, contrary to the conjecture, both consumers' welfare evolves more or less in lock step. The reason is that, by the CE conditions (7), both relative consumption and relative leisure are roughly constant over time. Therefore, it is not the case that workers will lose dramatically when capital taxes finally drop to zero.

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

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

We assume that the optimal plan is followed for $M - 1$ periods and then in period M the planner reoptimizes taking $k_{g,M-1}$, $k_{w,M-1}$, and $k_{c,M-1}$ as given. We then check whether the reoptimized solution differs from the remaining path under the original solution when consensus is required. That is, a reoptimization takes place only if a Pareto-improving allocation can be found relative to the consumers' continuation utilities at the period of reoptimization.

From our numerical experiments it seems impossible to make one consumer strictly better off without hurting the other, so that reoptimizing with consensus always leads to the confirmation of the original plan in terms of taxes and allocations. The time-invariant multipliers

ψ , Δ_1 , and Δ_2 change, and the time-variant multipliers μ_t and γ_t are rescaled by a factor $\frac{1+\tilde{\psi}}{1+\psi}$, where the tilde indicates the reoptimized solution. Moreover, we have the relationship $\gamma_{M-1} = \frac{1+\psi}{1+\tilde{\psi}} \left(\tilde{\Delta}_1 k_{1,M-1} + \tilde{\Delta}_2 k_{2,M-1} \right)$. Inspection of the FOCs reveals that the remainder of the original optimal plan satisfies the FOCs of the reoptimization problem, as long as these relationships between the multipliers hold and $\tilde{\psi}$ and the Δ_j 's are appropriately chosen. Interestingly, $\tilde{\psi}$ always turns out to be smaller than ψ . For instance, in the case of $\psi = 1$ and reoptimization in $M = 5$, the continuation utilities are respected if $\psi = 0.63$. Hence, the influence of the worker on the solution under consensus reform (measured by his relative Pareto weight ψ) has to be lower at the point of reoptimization.

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

5 Conclusion

We find that there is an equity-efficiency trade-off in the determination of capital and labor taxes. Capital taxes should be zero in the long run, but this is an optimal Pareto-improving policy only if capital taxes are very high (and labor taxes very low) for a very long time after the reform starts. The government typically accumulates debt in the long run in order to finance the initial cut in labor taxes. Lower initial labor taxes are necessary for two reasons: first, to redistribute wealth in favor of workers in order to ensure that they also gain from the reform, and second, to boost investment in the initial periods. These features of the optimal policy remain in the special case where the planner has a welfare function which weights all consumers equally, as if under ‘the veil of ignorance.’

Many of our results are numerical, for a given calibration of heterogeneity according to wage-wealth ratios. The results are robust to variations in parameter values and even to the introduction of progressive taxation. If labor supply is perfectly inelastic, it is very costly to make the workers enjoy significant benefits from the capital tax cut. The solution is time consistent if consensus is required at the time of reoptimization, suggesting that the tax reform is credible if it can only be overturned when all agents agree.

We find that issues of redistribution are crucial in designing optimal policies involving capital and labor taxes, even though the Chamley-Judd result survives with heterogeneous agents. Therefore, research on these issues should be encouraged, both from an empirical

and a theoretical point of view. One avenue for research is to study other policy instruments which may be used to compensate the workers for the elimination of capital taxes. For example, certain types of government spending or other tax cuts could play this role. More empirical work on the relevant aspects of heterogeneity so that issues of progressivity can be addressed carefully is certainly needed. The transition in our model is very long. Less-than-full credibility and less-than-fully-rational expectations might render this policy ineffective in practice. Introducing issues of partial credibility, learning about expectations, and political economy would therefore be of interest and might influence the picture on what an optimal policy should do.

References

- Aiyagari, S. (1995). Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting. *Journal of Political Economy* 103(6), 1158–75.
- Armenter, R. (2004). Redistribution, Time-Consistent Fiscal Policy and Representative Democracy. Mimeo, Federal Reserve Bank of New York.
- Atkeson, A., V. Chari, and P. Kehoe (1999). Taxing capital income: A bad idea. *Federal Reserve Bank of Minneapolis Quarterly Review* 23(3), 3–17.
- Bassetto, M. (2014). Optimal Fiscal Policy with Heterogeneous Agents. *Quantitative Economics* 5(3), 675–704.
- Chamley, C. (1986). Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. *Econometrica* 54, 607–622.
- Chamley, C. (2001). Capital Income Taxation, Wealth Distribution and Borrowing Constraints. *Journal of Public Economics* 79(1), 55–69.
- Chari, V. and P. Kehoe (1999). Optimal Fiscal and Monetary Policy. In J. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1, Chapter 26, pp. 1671–1745. Elsevier.
- Conesa, J. and D. Krueger (2006). On the Optimal Progressivity of the Income Tax Code. *Journal of Monetary Economics* 53(7), 1425–1450.
- Correia, I. (2010). Consumption Taxes and Redistribution. *American Economic Review* 100(4), 1673–1694.
- Correia, I. H. (1999). On the Efficiency and Equity Trade-off. *Journal of Monetary Economics* 44(3), 581 – 603.
- Domeij, D. and J. Heathcote (2004). On the Distributional Effects of Reducing Capital Taxes. *International Economic Review* 45(2), 523–554.
- Faraglia, E., A. Marcet, and A. Scott (2010). In Search of a Theory of Debt Management. *Journal of Monetary Economics* 57(7), 821–836.
- 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., R. Manuelli, and P. 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, 59–83.
- Lau, L., Y. Qian, and G. Roland (2001). Reform without Losers: An Interpretation of China’s Dual-Track Approach to Transition. *Journal of Political Economy* 108(1), 120–143.
- Ljungqvist, L. and T. J. Sargent (2012). *Recursive Macroeconomic Theory*. Third edition, MIT Press, Cambridge, Massachusetts.
- Lucas, R. (1990). Supply Side Economics: An Analytical Review. *Oxford Economic Papers* 42, 293–316.
- Marcet, A. and R. Marimon (2011). Recursive Contracts. Mimeo.
- Marcet, A., F. Obiols-Homs, and P. Weil (2007). Incomplete Markets, Labor Supply and Capital Accumulation. *Journal of Monetary Economics* 54, 2621–2635.
- Niepelt, D. (2004). Tax Smoothing versus Tax Shifting. *Review of Economic Dynamics* 7(1), 27–51.
- Straub, L. and I. Werning (2014). Positive Long Run Capital Taxation: Chamley-Judd Revisited. Working Paper 20441, National Bureau of Economic Research.
- 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 of the policy-maker's problem

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

$$\begin{aligned} \mathcal{L} = \sum_{t=0}^{\infty} \beta^t & \left\{ u(c_{1,t}) + v(l_{1,t}) + \psi [u(\lambda c_{1,t}) + v(f(\lambda, l_{1,t}))] \right. \\ & + \Delta_1 [u'(c_{1,t}) c_{1,t} + v'(l_{1,t}) l_{1,t}] \\ & + \Delta_2 \left[u'(c_{1,t}) \lambda c_{1,t} + \frac{\phi_2}{\phi_1} v'(l_{1,t}) f(\lambda, l_{1,t}) \right] \\ & + \gamma_t u'(c_{1,t}) - \gamma_{t-1} u'(c_{1,t}) (1 + (r_t - \delta)(1 - \tilde{\tau})) \\ & + \mu_t \left[F(k_{t-1}, e_t) + (1 - \delta)k_{t-1} - k_t - \frac{1 + \lambda}{2} c_{1,t} - g \right] \Big\} - \psi \underline{U}^2 \\ & - u'(c_{1,0}) (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}) (1 + (r_0 - \delta)(1 - \tau_0^k)), \end{aligned} \quad (30)$$

with $\psi \geq 0$ and $\gamma_t \geq 0$, with the usual complementary slackness conditions, and $\gamma_{-1} = 0$. The FOCs for the Lagrangian are:

- for consumption at $t > 0$:

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

- for consumption at $t = 0$:

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

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

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

- for labor at $t = 0$:

$$\begin{aligned}
& v'(l_{1,0}) + \psi v'(f(\lambda, l_{1,0})) f_l(\lambda, l_{1,0}) \\
& + \Delta_1[v'(l_{1,0}) + v''(l_{1,0}) l_{1,0}] + \Delta_2 \frac{\phi_2}{\phi_1} [v'(l_{1,0}) f_l(\lambda, l_{1,0}) + v''(l_{1,0}) f(\lambda, l_{1,0})] \\
& - u'(c_{1,0}) (\Delta_1 k_{1,-1} + \Delta_2 k_{2,-1}) F_{ke}(k_{-1}, e_0) \frac{1}{2} (\phi_1 + \phi_2 f_l(\lambda, l_{1,0})) (1 - \tau_0^k) \\
& = -F_e(k_{t-1}, e_t) \frac{1}{2} (\phi_1 + \phi_2 f_l(\lambda, l_{1,t})) \mu_0
\end{aligned}$$

- for capital at $t \geq 0$:

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

B Computational strategy: Approximation of the time path

1. Fix T as the number of periods after which the steady state is assumed to have been reached. (We use $T = 150$.)
2. Propose a $3T + 3$ -dimensional vector $X = \{k_0, \dots, k_{T-1}, l_0, \dots, l_{T-1}, \gamma_0, \dots, \gamma_{T-1}, \Delta_1, \Delta_2, \lambda\}$. Note that this is not the minimal number of variables we could find solving a fixed point problem. $2T + 3$ would be sufficient. However, convergence is better if the approximation errors are spread over a larger number of variables.
3. With k_{-1} and g known, find $\{c_t, F_{k,t}, F_{l,t}, F_{kl,t}, F_{kk,t}\}$ from the resource constraint and the production function.
4. Calculate $\{\mu_t\}$ from the FOC for labor.
5. Calculate $\{\gamma_t\}$ from the FOC for consumption, making use of $\{\mu_t\}$ and the guess for $\{\gamma_{t-1}\}$ from the X-vector.
6. Form the $3T + 3$ residual equations to be set to 0:
 - The FOC for capital (Euler equation) has to be satisfied. (T equations)
 - The vector $\{\gamma_t\}$ has to converge, i.e., old and new guesses have to be equal. (T equations)
 - Check for each period whether the constraint on τ^k is satisfied. If yes, impose $\gamma_t = 0$. Otherwise, the constraint on capital taxes has to be satisfied with equality. (T equations)

- The remaining 3 equations come from the present-value budget constraints (PVBC) and the FOC for λ . The discounted sums in the PVBCs are calculated using the time path of the variables for the first T periods and adding the net present value of staying at the steady state thereafter.
7. Iterate on X to set the residuals to 0. We use a trust-region dogleg algorithm and Broyden’s algorithm, repeatedly when necessary, to solve this $(3T + 3)$ -dimensional fixed point problem.

C Sensitivity analysis

To check the sensitivity of our results to the measurement of relevant tax rates and consumption inequality at the status quo, we recalibrate and solve our baseline model considering both a lower and a higher value for each the three data moments. Table 2 summarizes the results by reporting the duration of the transition and the revenue share of capital taxes for the two extreme points of the set of POPI plans. We always find the same qualitative properties of the optimal policy as for the baseline calibration described in Section 4.

Table 2: Sensitivity analysis

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

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

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

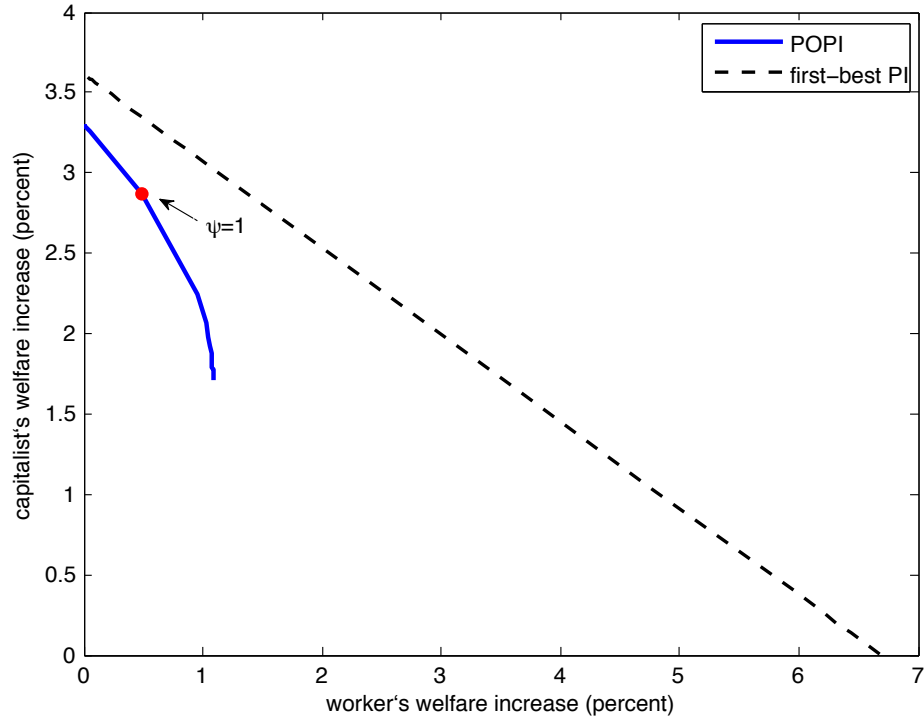
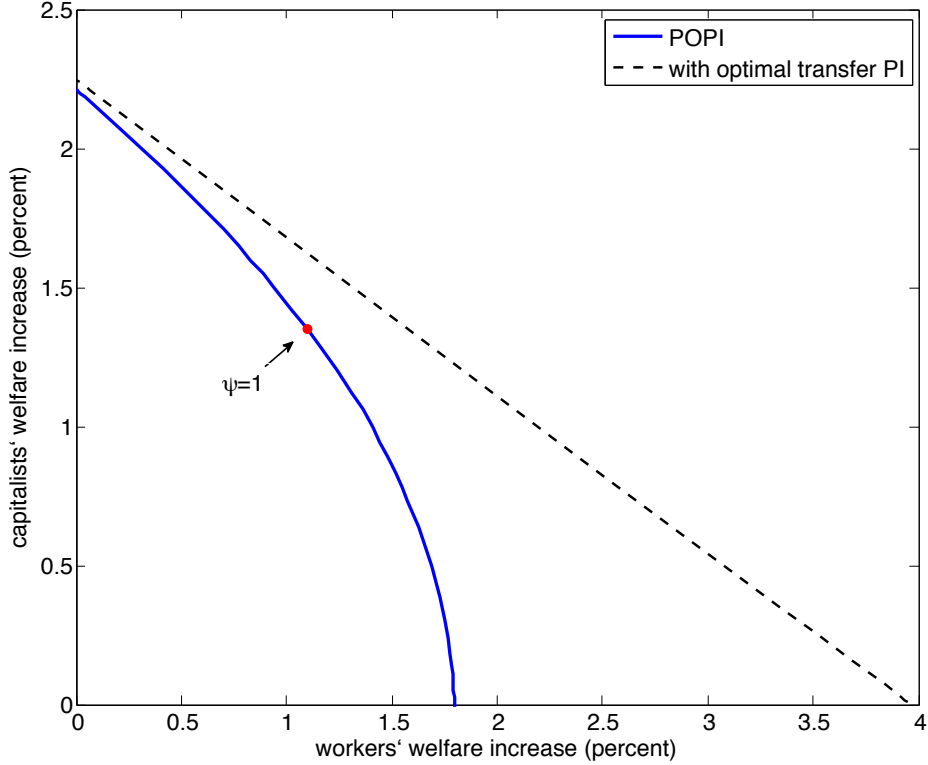


Figure 2: The Pareto frontier of Pareto-improving equilibria in the baseline model



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

Figure 3: A typical time path for capital taxes

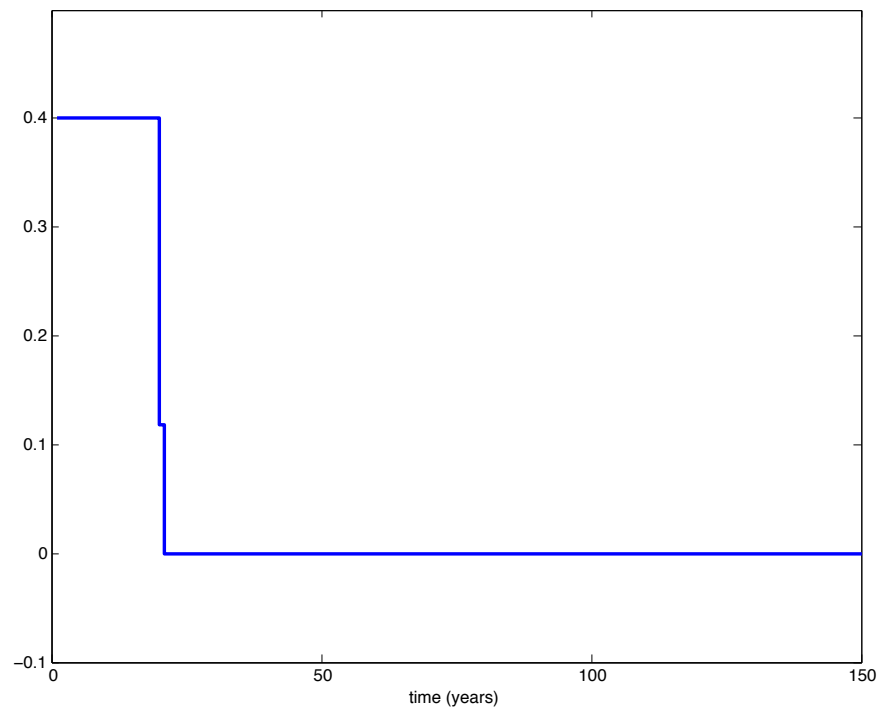


Figure 4: Properties of POPI programs in baseline model

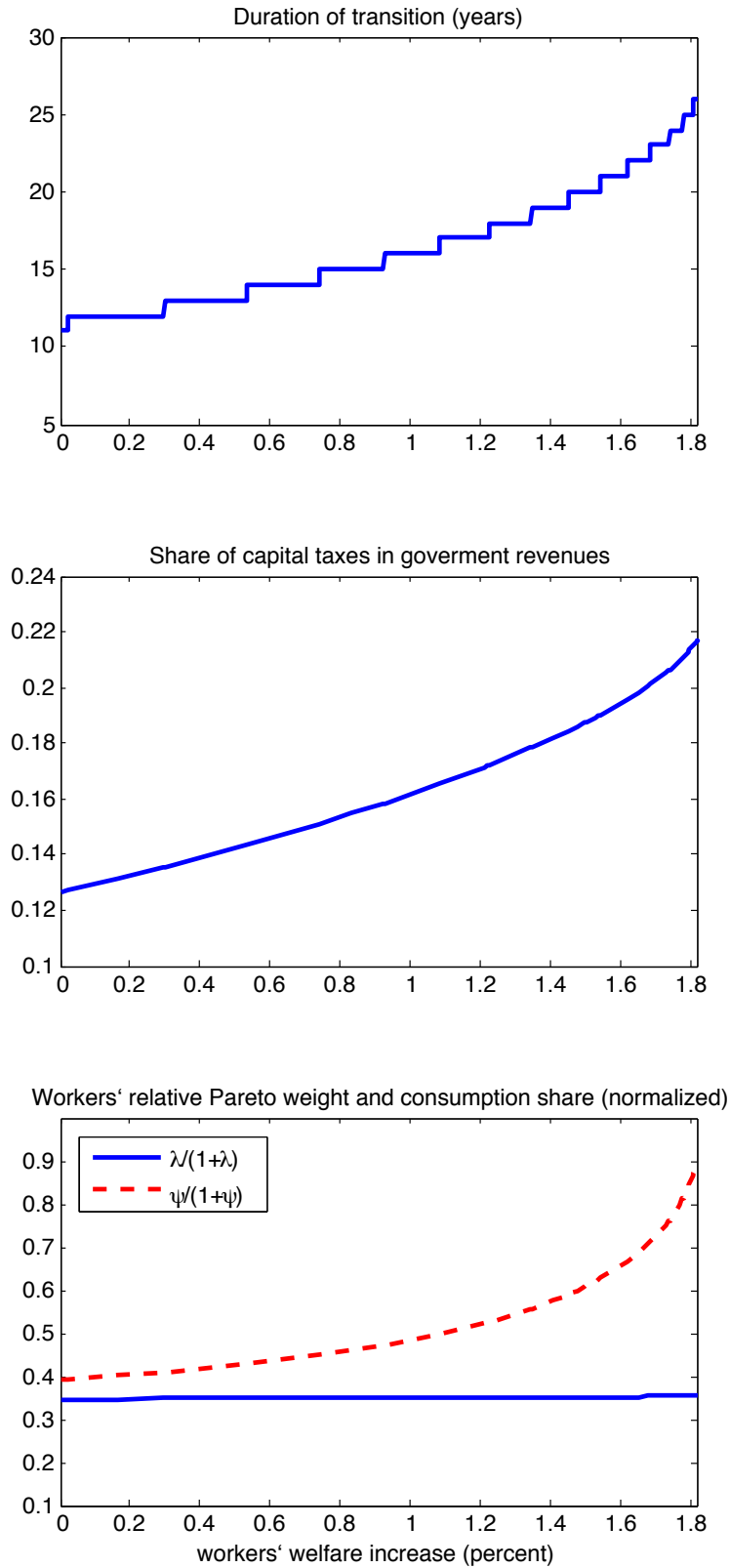


Figure 5: The time paths of selected variables for three POPI plans in the baseline model

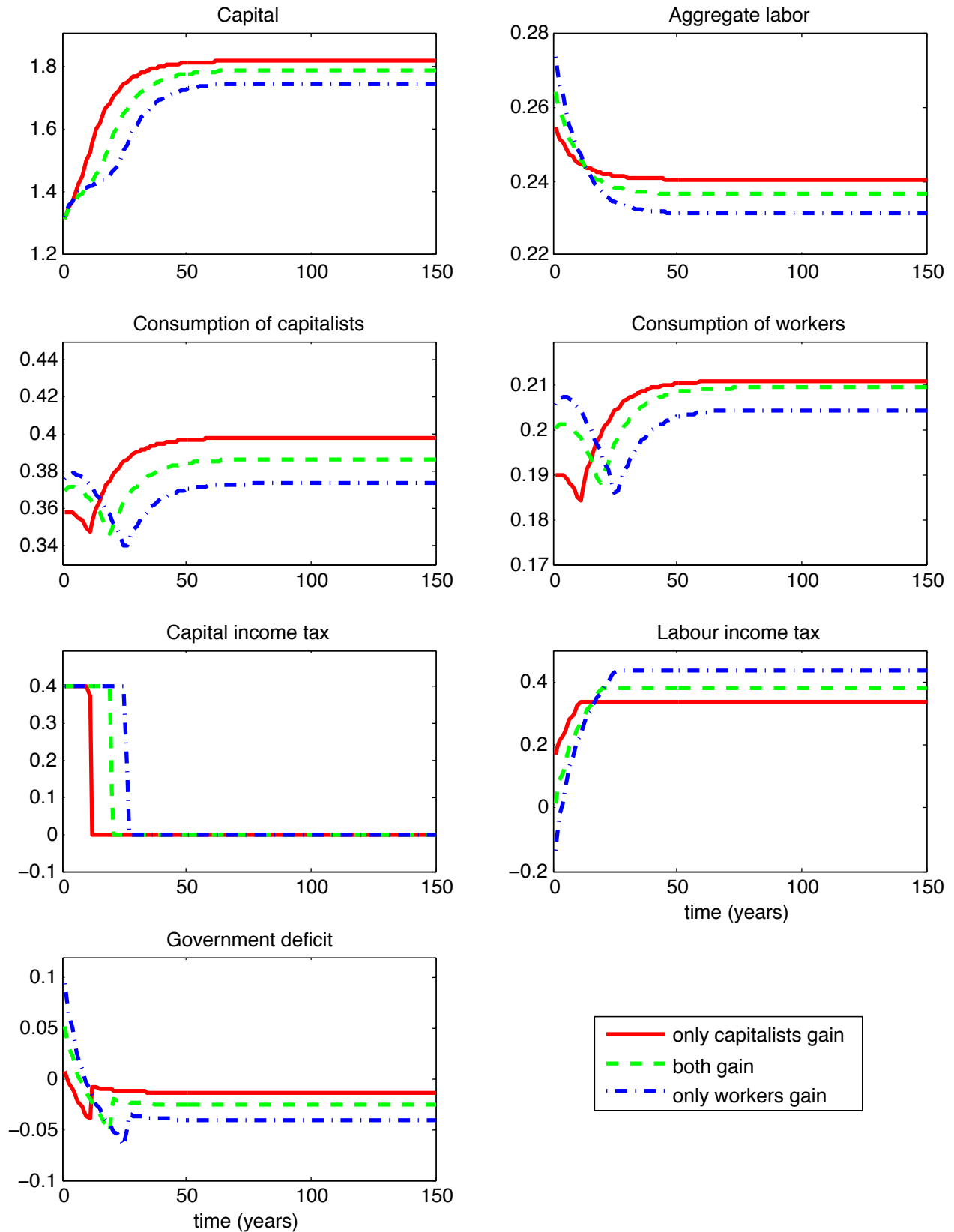


Figure 6: Typical time paths for consumers' welfare and wealth

