Online Appendices

for "Pareto-Improving Optimal Capital and Labor Taxes"

by Katharina Greulich, Sarolta Laczó, and Albert Marcet

A Computational strategy

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

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

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

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

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

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

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

   $$G(X) = 0,$$

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

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

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

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

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

  Then

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

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

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

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

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

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

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

B Sensitivity analysis for our baseline model

We check the sensitivity of our results to the measurement of relevant tax rates and inequality at the status quo. We recalibrate and solve our baseline model considering both a lower and a higher value for each the three data moments. Note that, given the calibration strategy that we use (described in Section 4.1), considering different values of λ_{SQ} and taxes in effect changes the distribution of private wealth of agents $k_{j,-1}$ in all of the alternative calibrations considered in this robustness exercise.

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

Table 1 summarizes some aspects of the simulations for these alternative parameters, changing the values of parameters one at a time relative to the baseline calibration. It reports the duration of the transition and the revenue share of capital taxes for the two extreme points of the set of POPI plans. We always find the same qualitative properties of the optimal policies as for the baseline calibration described in Section 4, and in some cases the results are reinforced, as the transition is even longer.

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

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

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

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

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

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

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

C Inequality and deductible

In this section we explore in more detail how initial wealth inequality affects optimal policy in our model, in particular at one point of the Pareto frontier: the social planner cares only about the welfare of the worker (the wealth-poor agent). The aim is to inspect the claim in Benhabib and Szőke (2021), BSz hereafter, that a combination of high inequality and maximising the welfare of a wealth-poor agent is the reason for the $\tau^k_\infty > 0$ result that they find. We allow for a deductible, as in Section 4.4 and in BSz.

We consider 6 scenarios, combining two possible calibrations of inequality with three different sets of parameter values.

For the calibration of inequality we start with 'baseline inequality' as in our baseline calibration in the main text. Remember that we have calibrated the initial wealth of workers and capitalists by taking the consumption ratio of the two groups from the data $c^w_{\text{SQ}}/c^c_{\text{SQ}} = 0.54$, initial wealth is found from the lifetime budget constraint under status quo taxation. The case of 'high inequality' sets $c^w_{\text{SQ}}/c^c_{\text{SQ}} = 0.2$, i.e., the planner only cares about workers who consume a fifth as much as the capitalists. The resulting initial wealth values are in Table 3 below.

The three sets for parameter values are as follows. The first is our baseline calibration. Second, we adapt parameterisation in part towards BSz by assuming homogenous labor productivity ($\phi^w = \phi^c = 1$), zero labor tax, no government spending, and no initial government debt. We will call this 'BSz calibration'. Our third case is the same as our BSz calibration but with a lower upper bound on the capital income tax rate that BSz use, $\tilde{\tau} = 0.1$.

Table 3 shows some key features of optimal tax policies and the resulting allocations in the 6 scenarios, as well as the optimal allocations when we impose restrictions on tax policies from BSz, namely, $\tau_t^k = \tilde{\tau}$ and $\tau_t^l = 0$, for all t, which we dub 'BSz tax policies'.

Table 3: Wealth inequality and optimal policy

Calibration:	Baseline		'BSz'		'BSz' $\tilde{\tau} = 0.1$	
			Baseline inequality			
k_0^w	-1.136		-2.741		-2.553	
k_0^c	4.356		5.935		10.67	
τ_∞^k	0	**0.401**	0	**0.401**	0	**0.1**
Duration	44	∞	43	∞	96	∞
τ_∞^l	0.243	**0**	-0.036	**0**	-0.050	**0**
Deductible	-6.438	-1.702	-8.802	1.064	-13.12	0.461
$\lambda = {}^{c^w}/_{c^c}$	0.645	0.571	0.691	0.559	0.668	0.549
V^w	-152.64		-103.58	-111.33	-69.01	-72.772
			High inequality			
k_0^w	-5.167		-8.228		-10.95	
k_0^c	8.387		11.62		19.26	
τ_∞^k	0	**0.401**	0	**0.401**	0	**0.1**
Duration	69	∞	62	∞	129	∞
τ_∞^l	0.271	**0**	-0.171	**0**	-0.240	**0**
Deductible	-25.36	-1.763	-34.17	1.060	-53.09	0.369
$\lambda = {}^{c^w}/_{c^c}$	0.545	0.245	0.615	0.238	0.598	0.229
V^w	-270.82	-419.39	-169.05	-259.38	-116.36	-209.66

Table 3 shows that our optimal tax policies give higher lifetime utility to the wealth-poor than the BSz tax policies. In addition, Table 3 shows that at the optimum (i) the deductible is always negative, i.e., a lump-sum tax is always optimal, and (ii) in the BSz calibrations the labour income tax is negative; and both are increasing (in absolute value) with initial wealth inequality. That is, increasing wealth inequality and caring only about the poor agent does not imply that BSz policies are optimal for these parameter values, even if we minimise the need to raise public revenue. A negative labour income tax rate is progressive in the BSz calibrations and, combined with a lump-sum tax and zero capital tax in the long run, serves to promote equity better than a high capital tax combined with a lump-sum transfer. The room for redistribution under BSz tax policies is small: $\lambda = {}^{c^w}/_{c^c}$ remains close to its autarky value in all 6 scenarios. Finally, in our baseline calibration a lump-sum tax remains optimal even under BSz tax policies, due to the need to finance government spending and debt.

D More on the multiplier μ and optimal waste

SW and BB pointed out that the Chamley-Judd argument may fail because μ may not have a steady state. Indeed, they show examples where optimal policy implies immiseration, i.e., $c_t \to 0$, therefore $\mu_t \to \infty$. However, there is another reason why the standard Chamley-Judd argument might not work, namely, when γ has a steady state $\gamma^{ss} > 0$ and the production function is strictly concave. In this appendix we show that in that case $\mu^{ss} < 0$, and hence it is optimal to waste aggregate consumption, or, allowing for free disposal would improve social welfare.

It follows that the reason BSz obtain $\tau_t^k \nrightarrow 0$ is that in their model μ is negative in some periods, their result is therefore compatible with our Proposition 1 as it requires $\mu_t \geq 0$, $\forall t$. In general, negative μ's are associated with models where $\tau_\infty^k = \widetilde{\tau}$ and $c^{ss} > 0$. Below we also show numerically that negative μ's are optimal in the example of Section IIIA of BSz, and that requiring positive rather than zero government spending improves social welfare. We also show that wasteful government spending improves social welfare.

We modify our baseline model for the purposes of this exercise, and adopt some assumptions from BSz. In particular, we focus on the case with inelastic labour supply, the government can set a (possibly negative) deductible $\mathcal{D}$, as in Section 4.4, all agents are equally productive, i.e., $\phi_1 = \phi_2$, there is no depreciation allowance, the mass of the worker (the wealth-poor agent, agent 1) is zero, and the Pareto weight of the capitalist (agent 2) is zero. $u()$ is CRRA.

Under these assumptions, $c_t = c_{2,t}$, $k_{-1} = k_{2,-1}$, $k_{1,-1} < k_{-1}$ and we can write $\lambda c_{1,t} = c_t$, $\forall t$, for some λ.[2] Due to the deductible $\mathcal{D}$ the optimal allocation sets $\Delta_2 = -\Delta_1$, hence the Lagrangian of the policy-maker's problem is

$$\begin{aligned}\mathcal{L} = \sum_{t=0}^{\infty} \beta^t [& u(c_{1,t}) + \Delta_1 (1-\lambda) u'(c_{1,t}) c_{1,t} \\ & + u'(c_{1,t}) [\gamma_t - \gamma_{t-1} (1 - \delta + r_t (1 - \widetilde{\tau}))] \\ & + \mu_t [F(k_{t-1}) + (1-\delta) k_{t-1} - k_t - c_{1,t}\lambda - g]\} - \mathbf{W},\end{aligned} \tag{1}$$

where $\gamma_{-1} = 0$ and $\mathbf{W} = \Delta_1 u'(c_{1,0}) (k_{1,-1} - k_{-1}) (1 - \delta + r_0 (1 - \widetilde{\tau}))$. The FOCs are easy to derive. They are similar to the ones of our baseline model, see Appendix A.

We assume that c and k have steady states. We are concerned with models where the optimal allocation does not have immiseration, i.e., $c_t \to c^{ss} > 0$, so we ignore the constraint $c_t \geq \widetilde{c}$ for simplicity. Using that $u()$ is CRRA, let $\Omega^c \equiv 1 + \Delta_1 (1 - \lambda)(1 - \sigma_c)$.

[2] BSz denote $1/\lambda$ by α^i.

Proposition 3. *If the tax limit is binding forever in the optimal allocation, then μ and γ have steady states and*

a) either $\mu^{ss}, -\gamma^{ss}, \Omega^l, \Omega^c < 0$,

b) or $\mu^{ss} = \gamma^{ss} = \Omega^l = \Omega^c = 0$.

Proof. Let us rewrite the FOC for capital as

$$\mu_{t+1} = \mu_t \nu_t - \gamma_t \alpha_t, \tag{2}$$

where $\nu_t = \frac{1}{\beta(1-\delta+F_k(k_t,e_{t+1}))}$ and $\alpha_t = -\frac{u'(c_{1,t+1})F_{kk}(k_t,e_{t+1})(1-\widetilde{\tau})}{1-\delta+F_k(k_t,e_{t+1})}$. If the tax limit is binding forever, plugging $\widetilde{\tau} = \tau_t^k$ in (4) of the main text, we have

$$\nu_t = \frac{u'\left(c_{1,t+1}\right)\left(1-\delta+F_k\left(k_t,e_{t+1}\right)\left(1-\widetilde{\tau}\right)\right)}{u'\left(c_{1,t}\right)\left(1-\delta+F_k\left(k_t,e_{t+1}\right)\right)} \to 1 - \frac{F_k\left(k^{ss},e^{ss}\right)\widetilde{\tau}}{1-\delta+F_k\left(k^{ss},e^{ss}\right)} = \nu^{ss},$$

where $0 < \nu^{ss} < 1$.

Let us rewrite (22) of Appendix A as

$$\Omega^l v'\left(l_{1,t}\right) - \gamma_{t-1} B_t = -D_t \mu_t \tag{3}$$

where B_t, D_t are defined by the corresponding terms in (22) of Appendix A. Combining this equation with (2) we have

$$\mu_{t+1} = \mu_t \frac{\nu_t}{1+\frac{D_{t+1}}{B_{t+1}}\alpha_t} - \frac{\Omega^l v'\left(l_{1,t+1}\right)\alpha_t}{B_{t+1}+D_{t+1}\alpha_t}.$$

Given that $(B_t, D_t, \alpha_t) \to (B^{ss}, D^{ss}, \alpha^{ss}) > 0$ we have $0 < \frac{\nu^{ss}}{1+\frac{D^{ss}}{B^{ss}}\alpha^{ss}} < 1$.

Therefore, in the limit this is a stable first-order linear equation in μ, and using a familiar argument we get

$$\mu_t \to \mu^{ss} = -\frac{\Omega^l v'\left(l_1^{ss}\right)\alpha^{ss}}{B^{ss}+D^{ss}\alpha^{ss}}\left(\frac{1}{1-\frac{\nu^{ss}}{1+\frac{D^{ss}}{B^{ss}}\alpha^{ss}}}\right). \tag{4}$$

Further, using (3) we have that

$$\gamma_t \to \gamma^{ss} = \frac{\Omega^l v'\left(l^{ss}\right) + D^{ss}\mu^{ss}}{B^{ss}}. \tag{5}$$

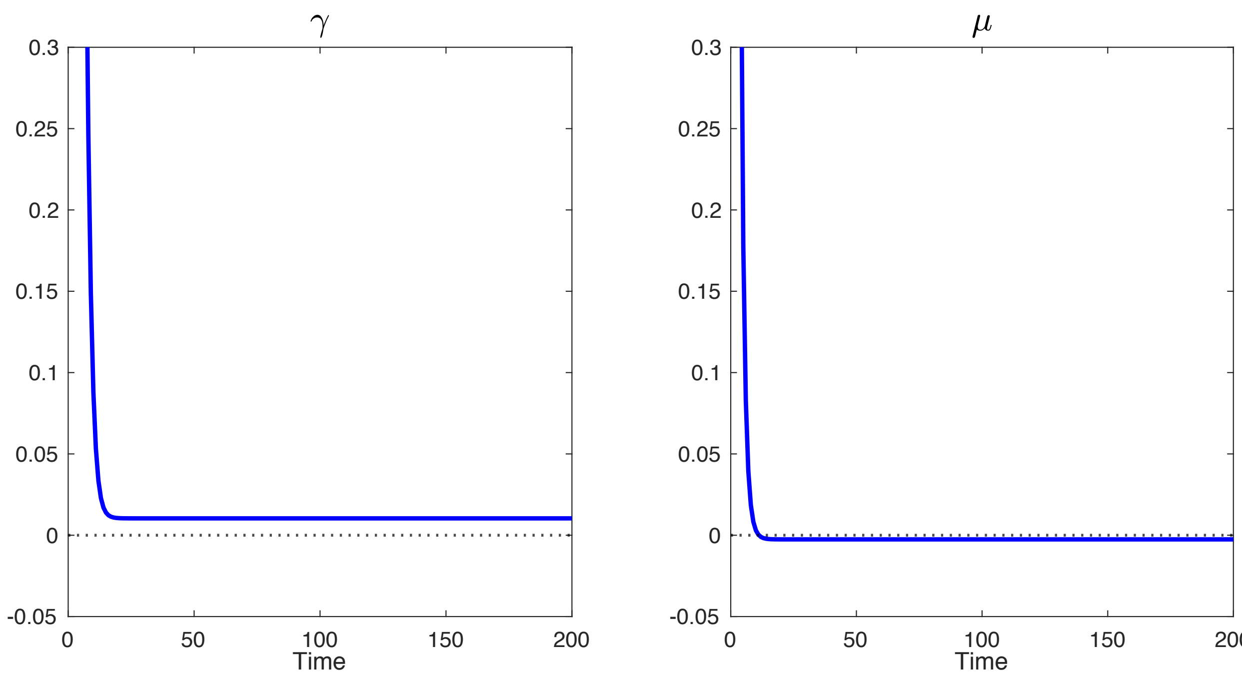
Hence (2) implies

$$\mu^{ss} = -\frac{\alpha^{ss}\gamma^{ss}}{1-\nu^{ss}} \leq 0.$$

Using $(B^{ss}, D^{ss}, \alpha^{ss}) > 0$ and $0 < \nu^{ss} < 1$ in (4), there are only two possibilities: either $\left(\mu^{ss}, \Omega^l, -\gamma^{ss}\right) < 0$ or $\mu^{ss} = \gamma^{ss} = \Omega^l = 0$. Furthermore, taking limits in (21) of Appendix A, with $\xi^{ss} = 0$, it follows that either $\Omega^l, \Omega^c < 0$ or $\Omega^l = \Omega^c = 0$. □

Given these results we expect that in the quantitative example of Section IIIA in BSz, $\mu^{ss} < 0$. To verify this, we compute the dynamic paths of multipliers for that example, where the production function is $F(k,1) = z\left(\rho k^{1-\eta} + (1-\rho)\right)^{\frac{1}{1-\eta}}$ and we use their parameter values $z = 2.5$, $\rho = 0.95$, $\eta = 3$, $\beta = 0.96$, $\sigma_c = 3$, $\delta = 1$, $\widetilde{\tau} = 0.1$, and $g = 0$. The results are shown in Figure 1. The precise values of the multipliers at the steady state are $\gamma^{ss} = 0.0104 > 0$, $\mu^{ss} = -0.0026 < 0$, and $\mu_t < 0$ for $t \geq 12$.[3]

Figure 1: The paths of γ and μ



This demonstrates why our Proposition 1 does not apply here. The question then arises: what does a negative μ in some periods imply for fiscal policy? Indeed, from an economic point of view, this seems like a mistake, as it means that wasting consumption can be optimal even though welfare depends only on consumption.

The explanation is the following: both in this paper and BSz, the government faces the constraint $g_t = g$ for a fixed g. Mathematically the multiplier on such an equality constraint might have either sign at the optimum. Consider now changing this constraint to $g_t \geq g$, giving the government the ability to waste aggregate consumption through fiscal policy. A negative μ_t in this paper or BSz implies that social welfare would improve by setting $g_t > g$.

To demonstrate this we have done the following exercise: since $\mu_t < 0$ for $t \geq 12$, we set

[3] We have computed these multipliers in two ways: imposing the optimal allocation provided by BSz and using our algorithm which jointly solves for allocations and multipliers. The solutions are indistinguishable in the Figure.

$g_t = \overline{g}$ for all periods $t \geq 12$ for some $\overline{g} > 0$, keeping $g_t = 0$ for $t < 12$. We then compute the corresponding allocations when taxes are at the upper bound forever.[4] We find the following.

Table 4: Welfare of the poor agent for various $\overline{g}$

$\overline{g}$	Welfare
0	−3.88675
0.1	−3.88425
0.13	−3.8840711
0.134	−3.8840685
0.14	−3.8840742
0.15	−3.88411

Note that the optimal allocation for $\overline{g} = 0$ is the same as in BSz for $\overline{g} = 0$, and we find the same welfare as using their formula. Social welfare is maximised when $\overline{g} = 0.134$, and this corresponds to a consumption-equivalent welfare gain of 0.109% compared to $\overline{g}$=0.

The reason this happens is that a lower c_t means a higher discount factor $\frac{u'(c_t)}{u'(c_0)}$, therefore increasing g_t could increase the discounted value of tax revenue in the distant future. If capital taxes are already at the upper limit, wasting consumption is the only way to extract more revenue from the capitalist and increase the relative consumption of the worker. This is indeed what happens: relative consumption λ is 0.656 for $\overline{g} = 0$ and 0.666 for $\overline{g} = 0.134$.

E Alternative solution strategies for PO allocations

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

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

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

This is a non-separable utility function, unlike ours, but it is immediate to extend our approach to this case. In addition, Flodén considers a general measure of agents $\widetilde{\lambda}(s)$ ($\lambda(s)$ in Flodén, 2009) of agents of type s. Our two-types-of-agents setup is a special case of his,

[4] This is a feasible but not necessarily optimal policy under free disposal.

therefore this is not an important difference either. Our approach could also be generalized to a general measure of agents.

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

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

and

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

Flodén writes the planner's problem as Atkeson, Chari, and Kehoe (1999), ACK hereafter, by keeping consumption of all agents in the equilibrium conditions, instead of summarizing the allocations of other agents using (6), (7), and λ, as we do. Although this makes computations different, it should not affect the allocations found. We describe this approach in detail below.

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

We use the notation

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

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

Using an ACK Lagrangian

Instead of representing equilibrium conditions with (6) and (7), as we do, Flodén follows ACK and the keeps equilibrium conditions

$$\frac{u_{1c,t}}{u_{1c,t+1}} = \frac{u_{jc,t}}{u_{jc,t+1}} \text{ and } \frac{u_{1l,t}}{u_{1c,t}\phi_1} = \frac{u_{jl,t}}{u_{jc,t}\phi_j},\ \forall j, \tag{8}$$

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

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

The Lagrangian to find the PO allocations using this approach is

$$\begin{aligned}\mathcal{L}=\sum_{t=0}^{\infty}\beta^t\Bigg\{\Bigg(&\sum_{j=1}^{J}\psi_j U_{j,t}+\Delta_j\left[U_{j,t}(1-\mu)+\frac{u_{1c,t}\zeta l_{j,t}^{1/\gamma+1}}{\gamma+1}\right]\\&+\rho_{jt}\left[u_{1c,t}u_{jc,t+1}-u_{jc,t}u_{1c,t+1}\right]+\xi_{jt}\left[u_{1c,t}u_{jl,t}\phi_1-u_{1l,t}u_{jc,t}\phi_j\right]\Big)\\&+\mu_t\left(\sum_{j=1}^{J}\widetilde{\lambda}_j c_{j,t}+g+k_t-(1-\delta)\,k_{t-1}-F\left(k_{t-1},e_t\right)\right)\Bigg\}+\sum_{j=1}^{J}\Delta_j W_{j,-1}.\end{aligned} \tag{9}$$

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

We prefer representing CE in the main text using (6) and (7) to substitute out agent 2's consumption and labor because then the planner's problem can be written as a maximization over $\tau_0^k, \lambda, \{c_t^1, k_t, l_t^1\}_{t=0}^{\infty}$. This reduces enormously the number of variables and multipliers to be computed, and it is much more convenient for computation. More precisely, given the algorithm described in Appendix A, the number of variables to solve for with J agents would be $(2+J)\times T+2+J$ using the ACK approach, while using our approach the number of variables to compute is only $3T+2+J$. But solving the Lagrangian (9) is equally valid, and it should give the same solution as we find. Hence in this appendix we characterize PO solutions to (9), as is done in Flodén (2009).

Using a representative agent

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

$$\sum_{t=0}^{\infty}\beta^t\left[u_{C^{RA},t}C_t^{RA}+u_{l^{RA},t}l_t^{RA}\right]=W_{-1}^{RA}, \tag{10}$$

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

Flodén finds equilibria that arise from the FOCs of the Lagrangian on page 300 in Flodén. The reader can check that one can go from the above Lagrangian (9) to Flodén's with the following three modifications:

1. Equation (10) is introduced in the planner's problem as an additional constraint.

2. The competitive equilibrium conditions (8) are written in terms of ratios of individual marginal utilities to the RA's marginal utilities.

3. Individual consumptions disappear from the feasibility constraint, i.e., $\sum_{j=1}^{J} \widetilde{\lambda}_j c_{j,t}$ is replaced by C_t^{RA} in the feasibility constraint.

Let us comment on the validity of these modifications.

Modification 1 is not needed for an equilibrium, because if all individual implementability constraints are satisfied, constraint (10) is guaranteed to hold. Therefore, modification 1 is redundant. All this means is that the multipliers λ_j and $\wedge$ (in Flodén's notation) are not uniquely defined, but the FOCs obtained from introducing modification 1 should give the same allocations as (9).

Modification 2 is also correct, indeed it implies and is implied by (8).

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

$$\sum_{j=1}^{J} \widetilde{\lambda}_j c_{j,t} = C_t^{RA}, \tag{11}$$

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

As it is written, the Lagrangian on page 300 in Flodén ignores the fact that the aggregate of all individual consumptions and leisure have to satisfy the feasibility constraint. A proper solution would entail incorporating the constraint (11) into the planner's problem, since it is not implied by any combination of the other constraints imposed. Therefore, FOCs (A.6) to (A.14) in Flodén do not provide a PO allocation.

That the FOCs of Flodén's Lagrangian do not give the correct solution can be seen in the following way. Let $\mathcal{L}^2$ represent the expression in the first two lines of (9). The correct FOC with respect to $c_{j,t}$ from (9) is

$$\frac{\partial \mathcal{L}^2}{\partial c_{j,t}} = -\mu_t \widetilde{\lambda}_j. \tag{12}$$

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

$$\frac{\partial \mathcal{L}^2}{\partial c_{j,t}} = 0, \tag{13}$$

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

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

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

If we considered a measure $\widetilde{\lambda}(.)$ with a continuous density $\widetilde{\lambda}'$ (where $\widetilde{\lambda}$ represents the measure of agents denoted λ on page 283 in Flodén (2009), we would have the same problem. Then, to find a PO solution, we would maximize $\sum_{t=0}^{\infty} \beta^t \left(\int_{[0,1]} \psi(j) U_{j,t} dj \right)$ for some density ψ and incorporating in the feasibility constraint that

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

we would find the FOC

$$\frac{\partial \mathcal{L}^2}{\partial c_{j,t}} = -\widetilde{\lambda}'(j)\mu_t, \ \forall j \in [0,1]. \tag{14}$$

This is incompatible with (13). The correct solution would imply $\int_I \frac{\partial \mathcal{L}^2}{\partial c_{s,t}} d\widetilde{\lambda}(s) = -\mu_t \int_I d\widetilde{\lambda}(s)$ for any subset of agents I, but Flodén's FOCs give $\int_I \frac{\partial \mathcal{L}^2}{\partial c_t(s)} d\widetilde{\lambda}(s) = 0$.

The only case where (13) is correct is when an agent has $\widetilde{\lambda}_j = 0$ in the discrete case or $\widetilde{\lambda}'(j) = 0$ in the continuous case. In other words, it seems that the case where the FOCs are valid is where the planner gives full measure in her objective function to agents who have zero measure in the market.

Later on Proposition 5 in Flodén argues that all PO solutions can be traced out by maximizing the utility with respect to one 'optimized' agent, whose initial state is denoted $\overline{s}$. The proof of that proposition shows that the FOCs for this modified problem coincide with the FOCs on page 300 which are as (13). But if (as we think) the latter do not give an PO allocation, then the conclusion of Proposition 5 does not follow. In fact, most PO solutions involve giving weight to all agents in the objective function of the planner, hence (12) has to hold instead of (13). Therefore, it is not true that all PO solutions can be found by selecting an optimized agent even with GHH utility.

In our opinion one can only find all PO allocations by taking properly into account the utilities of all agents in the economy, as we do in the main text, or as a direct application of (9) would do.

A rationale for Flodén's solution

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

Imagine we consider optimizing a weighted sum of utilities of J' agents (where J' is a discrete number) and that these agents have mass zero in the economy. This can either be because $\widetilde{\lambda}_j = 0$ for all $j = 1, ..., J'$ and $J' < J$ in the discrete case or because we consider only a discrete number of agents in the continuous case. For this PO allocation the planner's FOCs are indeed (13). But this is only a very small share of PO solutions. For any welfare function that gives positive weight to all agents, (13) does not work.

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