Optimal Policies with Heterogeneous Agents: Truncation and Transitions*

François Le Grand Xavier Ragot†

April 20, 2022

Abstract

We study the optimal provision of a public good in an heterogeneous-agent economy, with and without aggregate shocks. We rely on a method combining a Lagrangian approach and a truncation procedure that takes advantage of the restrictions imposed by the first-order conditions of the Ramsey problem. We compare these outcomes with those of other solution techniques considering transitions, as usually done in the literature. We have two main results. First, we find that the optimal Ramsey policy faces a time-inconsistency problem specific to incomplete-market economies, which is due to the non-optimality of private savings. This issue affects the solution based on the computation of transitions. Second, we find that the truncation approach provides quantitatively accurate estimates of the value of planner's instruments, both at the steady-state and during the dynamics. We also report a number of quantitative checks.

Keywords: Heterogeneous agents, optimal Ramsey program, truncation method, aggregate shocks.

JEL codes: D31, D52, E21.

*This paper has benefited from the comments of many participants to seminars, where the truncation method has been presented. We thank Diego Rodrigues for outstanding research assistance. We acknowledge financial support from the French National Research Agency (ANR-20-CE26-0018 IRMAC).

†François LeGrand: emlyon business school and ETH Zurich; legrand@em-lyon.com. Xavier Ragot: Sciences Po, CNRS and OFCE: xavier.ragot@sciencespo.fr.

1 Introduction

A frontier in the heterogeneous-agent literature is to compute optimal policies in general equilibrium. In heterogeneous-agent models (or, more precisely, in incomplete insurance-market models for the idiosyncratic risk), redistribution across endogenously heterogeneous agents generates new trade-offs. For instance, any policy affecting agents' income modifies their savings incentives, and consequently the capital stock and future real wages and interest rates, which heterogeneously affects agents' welfare. In this paper, we present a solution technique to solve for optimal Ramsey policy with commitment in a heterogeneous-agent model, for an arbitrary social welfare function. We can thus compute the steady-state optimal value of instruments and their dynamics. As an example, we study the optimal provision of a public good, financed by lump-sum taxes in a model where agents face uninsurable income shocks in an endowment economy à la Huggett (1993) or Aiyagari (1994). This simple environment will allow us to compare our solution techniques with other methods, and to identify some new issues in these environments, such as time-inconsistency.

Our method consists in providing a finite state-space representation of the incomplete market model, by pooling together agents according to their idiosyncratic history. More precisely, we track the consumption and savings of agents sharing the same idiosyncratic history over a given number of consecutive past periods. The model is then expressed in terms of these *truncated* idiosyncratic histories. We can then easy solve for optimal policy in this truncated model, by adapting tools developed in the optimal contract literature (Marcet and Marimon, 2019). We indeed provide an algorithm to compute the for a given number of consecutive past periods steady-state values of instruments in the initial (non-truncated) model. In addition, this finite state-space representation allows us to compute the sequence of Jacobians to simulate the optimal dynamics of economy after aggregate small shocks (**?**), which is both very simple and fast to implement using a standard package like Dynare (Adjemian et al., 2011). Pioneered in LeGrand and Ragot (2021b), we improve here this method by providing a more precise account of the heterogeneity within truncated histories. We compare the accuracy of the truncated method to other solution techniques, both at the steady-state and in the dynamics and show that results are very close to each other.

The truncation method applied to this simple Ramsey problem allows us to identify a time-inconsistency issue that is specific to in incomplete-market models. Whereas the planner is time-consistent in complete market environment, it is in consistent with incomplete markets. As we prove theoretically, this inconsistency is due to the conjunction of two factors. First, the planner is unable to restore the first-best, because of persistent ex-post heterogeneity among agents (in consumption and savings). In our setup, this comes from a combination of limited planner's instruments, incomplete markets and credit constraints. Second, current agent choices imply externalities in future periods. The planner thus has incentives to use its instruments to influence current choices and close the gap with the first-best allocation. In our setup, current savings determine tomorrow's ages and interest rates. As a consequence, the planner of the "next" period does not have to account for past externality because current wage and interest rate are fixed. The "next" planner faces different trade-offs than its predecessor and has thus some incentives to revise the trajectory of its instruments. In the complete-market economy, the Ramsey planner is able to restore the first-best and has therefore no incentive to manipulate agents' choices. This time-inconsistency result is not only of theoretical interest, but of computational interest since it also affects the results of some standard numerical methods used to compute optimal policies in heterogeneous-agent models.

Indeed, we also solve for the optimal long-run lump-sum tax using the transition approach, which determines the constant value of the instrument that maximizes the aggregate welfare while accounting for transitions to the long-run steady-state without aggregate shocks (see Conesa et al., 2009, Dyrda and Pedroni, 2018, Chang et al., 2018, or Ferriere and Navarro, 2020 among many others). Due to the time-inconsistency issue we mentioned, the Ramsey planner does not want to implement a constant policy instrument (even in our simple environment), which biases the solution of the transition approach. This time-inconsistency issue is only taken care if when computing the optimal instruments' path along the transition – but this is computationally very challenging (see Dyrda and Pedroni, 2018). Using the solution of the truncated method, we can quantify the contribution of time-inconsistency to the bias in the constant-instrument method. We also compare the dynamics of the truncated economy to the one simulated in the full-fledged incomplete-market model, using a Reiter (2009) method, which is known to be accurate. To do so, we estimate a rule for the planner's instruments for the truncated economy and we plug this rule into the full-fledged model. We find that the simulation outcomes are very close, confirming that the truncation method is accurate for both the steady-state and the dynamics of Ramsey program.

Our paper contributes to the literature on optimal policies in heterogeneous-agent models. This literature can be divided into three strands. The first branch uses numerical techniques to compute the optimal steady-state value of the planner's instruments. The initial contribution of Aiyagari and McGrattan (1998) finds the optimal value of public debt that maximizes the aggregate steady-state welfare. This is known to have limitations, since for instance it does account for the transition to the steady-state equilibrium. More recent investigations address the latter concern and solve for the optimal value of a constant instrument while taking into consideration the transition to the steady-state equilibrium (see for instance Conesa et al., 2009 for the capital tax). Although addressing the transitions, the analysis still involves that the planner imposes a constant (and not time-varying) value for the instruments. For this reason, its solution differs from the actual steady-state value of the instruments in the Ramsey program due to time-inconsistency (see below in 6.3), and it depends on the choice of the initial distribution (see Açikgöz, 2015 and the results in the current paper). To tackle this difficulty, Dyrda and Pedroni (2018) propose a numerical method to solve for the optimal path over all possible paths

– which is obviously a very computationally intensive technique. Another solution method focuses on the dynamics of the instruments when the optimal steady-state values of the instruments are known (McKay and Wolf, 2022). Numerical techniques can be used to provide a quadratic approximation of the objective of the planner and a linear approximation of the model using the contribution of Auclert et al. (2021). Then with quadratic-linear methods, a first-order approximation of the optimal instrument dynamics can be computed.

To circumvent the previous difficulties and finding the steady-state allocation, a second branch of the literature solves for the Ramsey program by taking advantage of the first-order conditions of the planner when the solution is interior (see Aiyagari, 1995 for an early contribution). A first gain of this second approach is to connect the normative analysis to the public finance literature, which extensively uses marginal valuations (see Heathcote and Tsujiyama, 2017 for a discussion of optimal policies in two-period models and the discussion in Section 3 below). Unfortunately, this generates additional difficulties in intertemporal models, as heterogeneous-agent models typically involve a continuum of Euler equations. Bhandari et al. (2021) provide a numerical procedure assuming that credit constraints are not occasionally-binding. They rely on the so-called primal approach, which implies to substitute interest rate by the ratio of marginal utilities. LeGrand and Ragot (2021b) (studying optimal unemployment insurance) and Açikgöz et al. (2018) (studying fiscal policy) use tools of the Lagrangian approach and follow the whole distribution of Lagrange multipliers on Euler equations, in economies where credit constraint can be occasionally-binding – which is often the quantitatively relevant case in quantitative models. The interest of LeGrand and Ragot (2021b) is to show that the aggregation of heterogeneous agent model along truncated histories can be used to simply solve for optimal Ramsey policies. Here, we improve the former paper to allow for more quantitatively relevant model solutions.

Finally, a third strand of the literature focuses on simplified environment to identify the mechanisms, where the equilibrium distribution is simple enough to yield a tractable setup.[1] Bilbiie (2020) and Bilbiie and Ragot (2020) solve for optimal monetary policies in an environment where a partial insurance structure implies that the equilibrium only features two consumption levels. Acharya et al. (2020) consider a CARA-normal structure without binding credit constraint to easily aggregate consumption. The framework of Lagos and Wright (2005) with a quasi-linear utility function – which simplifies the state-space – is often used to solve for optimal policies in tractable environments (see Angeletos et al., 2020 for a recent investigation concerning public debt). Some recent papers use continuous-time techniques to solve for optimal policies in heterogeneous-agent models (see Nuño and Thomas, 2020 for a recent example) .

The remainder of the paper is organized as follows. Section 2 presents the environment and Section 3 the Ramsey problem. Section 4 details the computation of the Ramsey solution using the truncation approach. Section 5 presents how the transition approach can be used to compute

[1] We do not review here the vast literature on optimal policy with complete insurance-market but ex ante heterogeneous agents. See Bassetto (2014) among others.

an optimal constant value for policy instruments. It also proposes an improvement over current methods to neutralize the impact of the choice of initial distributions. Section 6 contains a numerical exercise that quantifies the differences along several dimensions between the various methods considered in the paper. Section 7 concludes.

2 The environment

We consider an environment similar to the one of Den Haan (2010), which is a heterogeneous-agent economy with aggregate productivity risk and exogenous labor supply. The main twist is the introduction of a public good, whose provision enters into private utility. This public good is financed by a benevolent government through a lump-sum tax raised on all agents. The Ramsey problem we study is the question of the optimal provision of this public good. We consider a discrete-time economy populated by a continuum of agents of size 1. Agents are distributed according to a non-atomic measure ℓ on a set I: $\ell(I) = 1$. We follow Green (1994) and assume that the law of large numbers holds.

2.1 Risk

The economy is affected by two types of risk: an aggregate risk and an individual one. The aggregate risk solely affects TFP, denoted Z_t. It takes values in a possibly continuous set $\mathcal{Z}$ and is assumed to be Markovian. Its precise dynamics will be specified when needed. The history of aggregate risk at period t is denoted $Z^t = \{Z_0, ...Z_t\}$ and is an element of $\mathcal{Z}^{t+1}$.

The other risk is an idiosyncratic labor productivity shock $y \in \mathcal{Y}$ that agents cannot insure away. Every agent provides an inelastic labor supply normalized to 1. Denoting by w_t the date-t hourly wage, the labor income of an agent with productivity y_t amounts to $y_t w_t$. We assume that the individual productivity process follows a first-order Markov chain with constant transition probabilities $(\Pi_{yy'})_{y,y' \in \mathcal{Y}}$. The size of the agents' population with productivity y is constant and denoted S_y. The quantities $(S_y)_{y \in \mathcal{Y}}$ are defined through the recursions, $S_y := \sum_{\tilde{y} \in \mathcal{Y}} \Pi_{\tilde{y}y} S_{\tilde{y}}$, holding for all $y \in \mathcal{Y}$, with $\sum_{y \in \mathcal{Y}} S_y = 1$ since the size of the population is one. Finally, an individual history of productivity shocks up to date t is denoted by $y^t = \{y_{01}, \ldots, y_t\} \in \mathcal{Y}^{t+1}$. The measure over the set of individual histories, denoted by θ_t, can be computed using transition probabilities and the given initial distribution θ_0. The measure θ_t is such that $\theta_t(y^t)$ represents the share of agents with history y^t in period t.

2.2 Preferences

In each period, there are two goods in the economy: a private consumption good and a public consumption good. Households are expected-utility maximizers, who rank streams of private consumption $(c_t)_{t \geq 0}$ and of public consumption $(G_t)_{t \geq 0}$ according to a time-separable intertemporal utility function equal to $\sum_{t=0}^{\infty} \beta^t (u(c_t) + v(G_t))$, where $\beta \in (0,1)$ is a constant discount factor, and $u : \mathbb{R}_+ \to \mathbb{R}$ and $v : \mathbb{R}_+ \to \mathbb{R}$ are instantaneous utility functions reflecting separable preferences over private and public consumption, respectively. As is standard, we assume that u and v are twice continuously differentiable, increasing, and concave, with $u'(0) = v'(0) = \infty$.

2.3 Production

The private consumption good of the economy is produced by a standard profit-maximizing representative firm. At any date t, the firm production function combines labor L_t and capital K_{t-1} – that needs to be installed one period in advance – to produce Y_t units of the consumption good. Since individual labor supply is fixed and normalized to 1, aggregate labor supply is constant and equal to $\overline{L} := \int_i y_i \ell(di)$ efficient units. The production function is assumed to be of the Cobb-Douglas type featuring constant returns-to-scale with parameter $\alpha \in (0,1)$, and capital depreciation at rate $\delta \in (0,1)$. The TFP being denoted Z_t, the production function is formally defined as:

$$Y_t = F(Z_t, K_{t-1}, \overline{L}) = Z_t K_{t-1}^{\alpha} \overline{L}^{1-\alpha} - \delta K_{t-1}, \tag{1}$$

The firm rents labor and capital at respective factor prices w_t and r_t. The profit maximization conditions of the firm implies the following expression for factor prices:

$$w_t = F_L(Z_t, K_{t-1}, \overline{L}) \text{ and } r_t = F_K(Z_t, K_{t-1}, \overline{L}). \tag{2}$$

2.4 Government

In each period t, the government finances an endogenous public good expenditure G_t through a lump-sum transfer T_t. In the absence of public debt, the government budget must be balanced in each period:

$$T_t = G_t. \tag{3}$$

As was specified above, we abstract on purpose from more complex financing schemes to compare numerical methods in a straightforward environment.[2] We will see that the economic trade-offs for this simple scheme are already rich.

2.5 Agents' program, resource constraints and equilibrium definition

Agents can save in capital shares paying off the real interest rate r_t between dates $t-1$ and t. They face credit constraints and their savings must remain greater than an exogenous threshold normalized to 0. In the initial period, each agent i is endowed with an initial wealth a_{-1}^i and an initial productivity status y_0^i that are jointly drawn from an initial distribution Λ_0, defined over $[-\bar{a}; \infty) \times \mathcal{Y}$. Formally, given this initial endowment and given the stream of public spending

[2] See Den Haan (2010) for an early similar strategy for models with aggregate shocks.

$(G_t)_{t\geq 0}$, the agent's program can be expressed as:

$$\max_{(c_t^i, a_t^i)_{t\geq 0}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(u(c_t^i) + v(G_t) \right), \tag{4}$$

$$c_t^i + a_t^i = (1 + r_t) a_{t-1}^i + w_t y_t^i - T_t, \tag{5}$$

$$a_t^i \geq 0,\, c_t^i > 0, a_{-1}^i \text{ given} \tag{6}$$

where $\mathbb{E}_0$ an expectation operator over idiosyncratic and aggregate shocks. In the initial period, the agent chooses her consumption path $(c_t^i)_{t\geq 0}$, and her saving plans $(a_t^i)_{t\geq 0}$ so as to maximize her intertemporal utility (4), subject to the budget constraint (5) and the borrowing limit (6). The budget constraint (5) is straightforward: Agents finance their consumption, savings, and their taxes out of their labor earnings and saving payoffs.

The solution of the previous program is a set of policy rules $c_t : \mathcal{Y}^t \times \mathbb{R} \times \mathcal{Z}^t \to \mathbb{R}^+$ and $a_t : \mathcal{Y}^t \times \mathbb{R} \times \mathcal{Z}^t \to \mathbb{R}^+$ which determine consumption and saving decisions as a function of the idiosyncratic history y_i^t of agent i, her initial wealth a_{-1}^i and the history of aggregate shocks Z^t. However, to simplify the notation, we will simply write c_t^i and a_t^i (instead of $c_t(y_i^t, a_{-1}^i, Z^t)$ and $a_t(y_i^t, a_{-1}^i, Z^t)$). We will use the same notation for all variables, as summarized by the next remark.

Remark 1 (Simplifying Notation) *If an agent has an idiosyncratic history y_i^t, and initial wealth a_{-1}^i at period t, while the aggregate history is Z^t, we will then denote by X_t^i the realization in state (y_i^t, Z^t, a_{-1}^i) of any random variable $X_t : \mathcal{Y}^t \times \mathbb{R} \times \mathcal{S}^t \to \mathbb{R}$*

A consequence of Remark 1 is that the aggregation of the variable X_t in period t over the whole agent population will be written as $\int_i X_t^i \ell(di)$, instead of the more involved explicit notation $\int_{a_{-1}} \sum_{y^t \in \mathcal{Y}^t} \theta_t(y^t) X(y^t, a_{-1}, Z^t) d\Lambda_0(a_{-1}, y_0)$.

Taking advantage of this notation, we denote by $\beta^t \nu_t^i$ the Lagrange multiplier on the agent-i credit constraint. The agent's Euler equation can then be written as:

$$u'(c_t^i) = \beta \mathbb{E}_t \Big[(1 + r_{t+1}) u'(c_{t+1}^i) \Big] + \nu_t^i, \tag{7}$$

Financial market clearing condition and the economy-wide resource constraint can be expressed as follows:

$$\int_i a_t^i \ell(di) = K_t, \tag{8}$$

$$\int_i c_t^i \ell(di) + G_t + K_t = Y_t + K_{t-1}, \tag{9}$$

We can now state our market equilibrium definition.

Definition 1 (Sequential equilibrium) *A sequential competitive equilibrium is a collection of* *individual plans $(c_t^i, a_t^i, \nu_t^i)_{t\geq 0, i\in\mathcal{I}}$, of aggregate quantities $(K_t, Y_t)_{t\geq 0}$, of price processes $(w_t, r_t)_{t\geq 0}$, and of fiscal policy $(G_t, T_t)_{t\geq 0}$, such that, for an initial wealth and productivity distribution $(a_{-1}^i, y_0^i)_{i\in\mathcal{I}}$, and for an initial value of capital stock verifying $K_{-1} = \int_i a_{-1}^i \ell(di)$, we have:*

1. *given prices and fiscal policy, the functions $(c_t^i, l_t^i, \nu_t^i)_{t\geq 0, i\in\mathcal{I}}$ solve the agent's optimization program in equations (4)–(6);*
2. *financial and goods markets clear at all dates: for any $t \geq 0$, equations (8) and (9) hold;*
3. *the government budget is balanced at all dates: equation (3) holds for all $t \geq 0$;*
4. *factor prices $(w_t, r_t)_{t\geq 0}$ are consistent with condition (2).*

3 The Ramsey problem

The Ramsey problem consists for the planner to select an fiscal policy, which corresponds to a competitive equilibrium with the highest aggregate welfare. Regrading the latter, we opt for the standard ex-ante additive criterion – also known as the utilitarian social welfare function – which attributes an identical weight to all agents. Formally, the aggregate welfare criterion can be expressed as follows:[3]

$$W_0 := \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \int_i \Big(u(c_t^i) + v(G_t) \Big) \, \ell(di) \right], \tag{10}$$

which depends on the public spending path $(G_t)_{t\geq 0}$ and on the consumption paths of all agents, $(c_t)_{t\geq 0, i\in I}$. Admittedly, other social welfare functions could be used – and our solution method could indeed be used to solve them – but we restrict our attention to this useful benchmark, which is used in a number of heterogeneous-agent papers since the seminal study of Aiyagari (1995).

We now formalize our definition of Ramsey allocation.

Definition 2 (Ramsey allocation) *An optimal Ramsey allocation is a competitive equilibrium in the sense of Definition 1 that maximizes the aggregate welfare W_0 of equation (10) over the set of competitive equilibria.*

Definition 2 can be formalized as the outcome of an optimization program. Using the governmental budget constraint (3) to substitute G_t by T_t, the Ramsey program can be written as follows:

[3]In the sequential representation, the explicit expression is $W_0 := \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \int_{a_{-1}} \sum_{y^t \in \mathcal{Y}^t} \theta_t(y^t) \left(u(c_t(y^t, a_{-1}, Z^t)) + v(G_t) \right) d\Lambda_0(a_{-1}, y_0.)$

$$\max_{\left(w_t,r_t,T_t,K_t,(a_t^i,c_t^i,\nu_t^i)_{i\in I}\right)_{t\geq 0}} \mathbb{E}_0\left[\sum_{t=0}^{\infty}\beta^t\int_i\Big(u(c_t^i)+v(T_t)\Big)\,\ell(di)\right], \tag{11}$$

$$\forall i\in I,\ a_t^i+c_t^i=(1+r_t)a_{t-1}^i+w_ty_t^i-T_t, \tag{12}$$

$$u'(c_t^i)=\beta\mathbb{E}_t\Big[(1+r_{t+1})u'(c_{t+1}^i)\Big]+\nu_t^i, \tag{13}$$

$$a_t^i\geq 0,\ \nu_t^i(a_t^i+\bar{a})=0,\ \nu_t^i\geq 0,\ c_t^i\geq 0, \tag{14}$$

$$K_t=\int_i a_t^i\ell(di), \tag{15}$$

$$r_t=F_K(K_{t-1},\overline{L}),\ w_t=F_L(K_{t-1},\overline{L}). \tag{16}$$

Equation (11) is the planner's objective (10). Equations (12)–(16) are planner's constraints that guarantee that the chosen allocation is picked up among the competitive equilibria of Definition 1. Equations (12)–(14) are individual constraints: the budget constraint, the Euler equation, and the positivity and credit constraints respectively. In the problem under consideration, the consumption positivity constraint should not be neglected, because the lump-sum tax is the sole source of financing. This means that the consumption of poorer agents becomes negative for large taxes. Equations (15) and (16) are economy-wide constraints, regarding financial market clearing and factor price definitions.

The trade-off faced by the planner in Ramsey program (11)–(16) is rather straightforward. The planner can increase the provision of public good at the cost of a higher tax that reduces the consumption of private goods. This higher tax has an heterogeneous effect on agents, because they have different wealth levels and different income. Thus, the higher tax affects heterogeneously the agents' saving decisions, which in turn modifies the dynamic of the capital stock, the real wage and the real interest rate. These general equilibrium effects, combined with the redistribution motives, are making the Ramsey problem difficult to solve.

We now turn to the expression of first-order conditions of the Ramsey program (11)–(16). We denote by $\beta^t\lambda_t^i$ the Lagrange multiplier on the agent-i Euler equation. To simplify the analysis, following LeGrand and Ragot (2021a), we introduce the concept of *social valuation of liquidity for agent i* denoted by ψ_t^i, and formally defined as:

$$\psi_t^i := \underbrace{u'(c_t^i)}_{\text{direct effet}} - \underbrace{u''(c_t^i)\Big(\lambda_t^i-(1+r_t)\lambda_{t-1}^i\Big)}_{\text{effect on savings' incentives}}, \tag{17}$$

which can be seen as the equivalent for the planner of the marginal utility of consumption.[4] Indeed, it measures, from the planner's perspective, the value of one extra unit of consumption for agent i. Besides the standard marginal utility $u'(c_t^i)$, the term ψ_t^i includes the effect of this

[4]For the sake of clarity, both λ_t^i and ψ_t^i are function of the each idiosyncratic history y_i^t : $\lambda_t(y^t,a_{-1},Z^t)$ and $\psi_t(y^t,a_{-1},Z^t)$.

extra unit of consumption on agent's incentives to save (from yesterday to today, through λ_{t-1}^i, and from today to tomorrow, through λ_t^i).

We derive the first-order conditions in Appendix A. We here discuss the results. The first-order condition (FOC) with respect to savings a_t^i can be written for any agent i as:

$$\psi_t^i = \beta \mathbb{E}_t \left[(1+r_{t+1})\psi_{t+1}^i\right] + \beta \mathbb{E}_t \left[\underbrace{\int_j \psi_{t+1}^j \left(F_{KK,t} a_t^j + F_{KL,t} y_{t+1}^j\right) \ell(dj)}_{\text{indirect effet of the change in saving}} \right] \tag{18}$$

$$+ \beta \mathbb{E}_t \left[\underbrace{\int_j \lambda_t^j F_{KK,t} u'(c_{t+1}^j) \ell(dj)}_{\text{Effect of int. rate on saving.}} \right], \text{ if } i \text{ is not credit-constrained,}$$

$$\lambda_t^i = 0, \text{ if } i \text{ is credit-constrained.} \tag{19}$$

Equation (18) only holds for unconstrained agents, while for constrained agents, equation (19) matters. The first part of equation (18) can be interpreted as a generalized Euler equation for ψ_t^i, which is consistent with the interpretation of ψ_t^i as the generalization of the marginal utility from the planner's perspective. In other words, it reflects that, when setting the savings of agent i, the planner seeks to smooth out her consumption (valued with the social marginal valuation of liquidity, ψ_t^i) through time. However, the planner also internalizes the general-equilibrium impacts of the savings of agent i through wages and interest rates that affect all agents, which is the second term at the right-hand side. When an agent i saves more, this will affect the real interest and the wage rate due to an increase in the capital stock. It affects welfare proportional to the next-period beginning-of-period asset a_t^j (for the real interest rate), and proportional to the next-period productivity y_{t+1}^j (for the real wage). In addition to these redistributive effects, the change in the real interest rate affects the saving incentive and the ability to smooth, which the third term at the right-hand-side. For the planner, the marginal valuation of this effect is λ_t^j, which is the Lagrange multiplier od the Euler equation of agents j.

The second FOC regarding the lump-sum tax T_t can be expressed as follows:

$$v'(T_t) = \int_i \psi_t^i \ell(di). \tag{20}$$

The interpretation is rather straightforward. The marginal benefit of increasing the tax (and hence, the public spending) is $v'(T_t)$ and common to the whole agent population. This marginal benefit is set equal to the marginal cost, which amounts to taxing one unit of private consumption (valued with ψ_t^i for agent i by the planner) to all agents in the population (hence the integral over i).

The relationship (20) can be further clarified by rewriting it using the notation of the literature on the evaluation of public policies (see Hendren and Sprung-Keyser, 2020 for references). In this literature, the willingness-to-pay of agent i at period t (denoted by WTP_t^i) is the ratio of the marginal utility derived from one additional unit of the program (here from one additional unit of public good $G_t = T_t$), which is $v'(T_t)$, to the marginal disutility of reducing the income by one additional unit, which is denoted $\eta_t^i := u'(c_t^i)$. Hence, we have $\mathrm{WTP}_t^i := \frac{v'(T_t)}{u'(c_t^i)}$. With this notation, and expanding (17), we obtain:

$$\underbrace{\bar{\eta}\int_i \mathrm{WTP}_t^i \ell(di)}_{\text{weighted willingness-to-pay}} = \underbrace{\int_i \eta_t^i \ell(di)}_{\text{aggregate private marginal cost}} + \underbrace{\int_i (-u''(c_t^i))\left(\lambda_t^i - (1+r_t)\lambda_{t-1}^i\right)\ell(di)}_{\text{indirect effect of savings}}.$$

where $\bar{\eta} := \int_i \eta_t^i \frac{\mathrm{WTP}_t^i}{\int_i \mathrm{WTP}_t^i \ell(di)} \ell(di)$ is the average social marginal utility of the beneficiaries. The previous equality states that the correctly weighted willingness-to-pay should be equal to private cost plus indirect effects due to the changes in savings.[5] These indirect effects correspond to the general equilibrium effects summarized in equation (18), described above. In the numerical part 6.2, we will show that considering indirect effect of savings is quantitatively not negligible.

A Ramsey allocation is a competitive allocation that additionally verifies the two sets of FOCs (18)–(19) and (20). Computing such a Ramsey allocation is in general a difficult exercise. The main difficulty, even present when computing the steady-state allocation, is that the state space includes the Lagrange multipliers $(\lambda^i)_i$. This can be seen in equation (18), where the past values λ_{t-1}^i enter the definition of ψ_t^i. As a consequence, the relevant state-space is the joint distribution of beginning-of-period wealth and past value of the Lagrange multipliers: $(a_{t-1}^i, \lambda_{t-1}^i)_i$. Before explaining the truncation method to solve for the Ramsey allocation, we present the complete-market case.

The complete-market benchmark: Understanding saving externalities and time-inconsistency

We characterize the complete-market economy. We start with the first-best allocation, which is the one maximizing the aggregate welfare subject to given initial capital K_{-1} and subject to the resource constraint in the economy. If we denote by C_t the consumption of the representative agent, the economy-wide resource constraint can be written as:

$$C_t + G_t + K_t = F(Z_t, K_{t-1}, \bar{L}) + K_{t-1},$$

[5] Hendren and Sprung-Keyser (2020) introduce an additional concept which is the marginal value of public funds (MVPF), which is here equal to the non-weighted average WTP across agents.

which is obviously similar to (9) in the general case. The first-best allocation is determined as the solution of the following program:

$$\max_{(K_t,C_t,G_t)_{t\geq 0}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left(u(C_t) + v(G_t)\right)\right], \tag{21}$$

$$K_t + C_t + G_t = F(Z_t, K_{t-1}, \overline{L}) + K_{t-1}, \tag{22}$$

$$K_{-1} \text{ given}, \tag{23}$$

where we have used the governmental budget constraint $G_t = T_t$. The two first-order conditions are of the first-best program can be written as:

$$u'(C_t) = \beta \mathbb{E}_t \left[(1 + F_K(Z_{t+1}, K_t, \overline{L}))u'(C_{t+1})\right], \tag{24}$$

$$v'(G_t) = u'(C_t). \tag{25}$$

Equations (24) and (25) together with the budget constraint (22) determine a dynamic system in $(C_t, K_t, G_t)_{t\geq 0}$ for a given initial capital K_{-1} and characterize the first-best allocation. The first-best allocation can easily be decentralized by the setting the following prices: $r_t = F_K(Z_t, K_t, \overline{L})$ and $w_t = F_L(Z_t, K_t, \overline{L})$.

In that case, we can verify that the individual budget constraint can be written as:

$$C_t + K_t = K_{t-1} + F_K(Z_t, K_t, \overline{L})K_{t-1} + F_L(Z_t, K_t, \overline{L})\overline{L} - T_t, \tag{26}$$

where we have used the financial market and labor market clearing conditions. Combining (26) with the CES property of the production function and the government budget constraint implies that the individual budget constraint is identical to the resource constraint (22). Therefore, since the representative agent is endowed with the whole amount of capital at the initial date and since the first-best FOC (24) is identical to the Euler equation of the representative agents, a competitive allocation in which the fiscal policy is the same as the first-best, will be identical to the first-best allocation.[6] The Ramsey planner can thus implement the first-best allocation by choosing a fiscal policy according to FOC (25).

To draw the parallel with the Lagrangian approach we used in the general case, it can be observed that in the absence of heterogeneity, equation (18) will simplify into a linear equation in λ_{t-1}^{CM} and λ_t^{CM} with no other terms, with $\lambda_t^{CM} = 0$ as a unique solution.[7] Intuitively, the agents' Euler equation is not a constraint for the planner, since it corresponds to the first-best intertemporal allocation of capital. As a consequence, the Lagrange multiplier on the Euler

[6] Formally, the two dynamic systems will have the same dynamic equations and the same initial conditions. They will therefore coincide at each date.

[7] Indeed, the term in $F_{KK,t}a_t^{CM} + F_{KL,t}y_{t+1}^{CM}$ would be zero and after using the Euler equation to simplify further, equation (18) could be written as $A_t\lambda_t^{CM} + B_{t-1}\lambda_{t-1}^{CM} = 0$ for well-chosen non-zero coefficient A_t and B_{t-1}. With $\lambda_{-1}^{CM} = 0$, this implies by induction $\lambda_t^{CM} = 0$ at all dates.

equation is null and the Lagrangian approach implies $\psi_t^{CM} = u'(C_t) = v'(G_t)$, such that FOCs (20) and (25) are actually identical.

In an incomplete-market economy, the Ramsey planner cannot restore the first-best allocation in general. This is mainly due to the presence of ex-post heterogeneity that the planner does not manage to fully offset. This residual ex-post heterogeneity is incompatible with the first-best – that features perfect equality in the population – and the planner uses its instruments to close the gap to the first-best. This ex-post heterogeneity can result from ex-ante heterogeneity for example and does not necessarily require credit constraint or uninsurable risk. Both can been seen as sufficient markets limitations that contribute to generate ex-post heterogeneity, but they are not necessary. Moreover, credit constraints contribute to widen the gap with the first-best since they imply that for some agents the Euler equation – which is the FOC (24) of the first-best – does not hold. In Appendix E, we have developed a stylized two-period model featuring ex-ante heterogeneity (but no credit constraint and no risk). We show that the first-best cannot be reached by the planner (see Section E.2.1).

Because the Ramsey planner cannot implement the first-best allocation, the capital stock is socially not optimal in the Ramsey economy and the planner will use its instruments to close the gap with the first-best. To set the optimal capital level, the planner has to account for private saving incentives – through individual Euler equations –but also to account for general equilibrium effects of capital on interest rates and wages (of the next period period because of the capital installed one period in advance). More precisely, affecting individual savings today also changes tomorrow's saving incentives and tomorrow's prices. In other words, today's saving choices have externalities on tomorrow's prices. In the Ramsey program featuring full commitment, the full path of instruments is set at the initial date and the planner ties its hands for not modifying it afterwards. This means that setting the instrument value of some future date has to account not only for saving externalities but also for promises regarding past instrument values. This is visible in the fact that the past values of Lagrange multipliers of the Euler equation (i.e., λ_{t-1}^i) affect Ramsey FOCs (see equations (18) and (20)). This combination of commitment and saving externalities creates room for time-inconsistency. Indeed, should the planner be given the opportunity to re-optimize at a given date (and to break its date-0 commitment), it would renege on its past promises and choose a new value for the instrument that does not account for past commitments. In Appendix E, we show with our stylized two-period model the time-inconsistency that clearly appears as the combination of: (i) the incapacity of the planner to reproduce the first-best allocation, and (ii) externality of current saving choices on tomorrow's prices.

The computational application of Section 6 provides a quantification of the magnitude of these effects. Before turning to the numerical exercise, we present the truncation method.

4 Finding the optimal Ramsey policy in an heterogeneous-agent economy using the truncation method

We present the truncation method in several steps.

1. We start with providing the intuition and the basics of the truncation method in Section 4.1.
2. We explain how to implement the truncation method in practice in Section 4.2.
3. We document how the truncation method can solve for steady-state Ramsey policies in Section 4.3.
4. We finally show how to compute optimal policy with aggregate shocks in Section 4.4.

4.1 The truncation method

The truncation method is an aggregation procedure that can be applied to any heterogeneous-agent model. We provide a detailed account of how this method can be used to solve for heterogeneous-agent models in LeGrand and Ragot (2021c).[8] Even though we focus here on how this method can help solve for Ramsey programs, we still provide a concise presentation of the method in this paper's context. The starting point of this method is the sequential solution of the full-fledged incomplete-market model, which can be written as a set of policy rules, mapping histories into choices:[9]

$$a_t(y^t, a_{-1}, Z^t) \text{ and } c_t(y^t, a_{-1}, Z^t), \text{ for } y^t \in \mathcal{Y}^{t+1}, Z^t \in \mathcal{Z}^{t+1},$$

stating that the saving and consumption policy function of any agent depends on her whole idiosyncratic history y^t – and on the history of aggregate shock and her initial wealth. The main idea of the truncation method is to group together agents who have the same productivity history for a given number of consecutive past periods, and then to state the model in terms of this finite number of agents' groups. We call *truncation length* the exogenous parameter setting the length of the shared productivity history and denote it by $N > 0$. The truncation method consists in truncating idiosyncratic histories and in following a finite number of agents' representatives in an adjusted model. A key step of the truncation method is the construction of this adjusted model.

Consider an agent with complete idiosyncratic history $y^\infty = (\ldots, y_{t-N-1}, y_{t-N}, y_{t-N+1}, y_{t-N+2}, \ldots, y_{t-1}, y_t)$ at date t (y_t being the current productivity status). If her history over the last N periods is such that $(y_{t-N+1}, \ldots, y_{t-1}, y_t) = (y^N_{-N+1}, \ldots, y^N_{-1}, y^N_0)$, this agent will be assigned to truncated history $y^N := (y^N_{-N+1}, \ldots, y^N_{-1}, y^N_0)$ at date t, independently of earlier productivity

[8]We notably insist on the matrix notation and the implementation aspects.

[9]We consider the sequential representation to ease exposition, the actual implementation uses the recursive representation, which is the standard input of computational methods, as shown in Section 4.2 below.

levels – i.e, of the sequence $(\ldots, y_{t-N-1}, y_{t-N},)$. Since the number of productivity levels is finite and equal to $n_y := Card(\mathcal{Y})$, the number of truncated histories of length N will also be finite and equal to $N_{tot} = n_y^N$. Because every agent draws a new idiosyncratic status in every period, a given agent is in general assigned to a different truncated history in each period. For instance, if the previous agent with history y^∞ at t is endowed the productivity y_{t+1} at $t+1$, her $t+1$-history will be: $\tilde{y}^\infty := (\ldots, y_{t-N-1}, y_{t-N}, y_{t-N+1}, \ldots, y_{t-1}, y_t, y_{t+1})$ and she will be assigned at date $t+1$ to truncated history $\tilde{y}^N = (y_{-N+2}^N, \ldots, y_1^N, y_0^N, \tilde{y}_0^N)$, where $\tilde{y}_0^N := y_{t+1}$.[10] The probability, denoted by $\Pi_{y^N \tilde{y}^N}$, that an agent transits from history y^N to history $\tilde{y}^N$ is the probability that the agent transits from productivity levels y_0 to $\tilde{y}_0$, or formally:

$$\Pi_{y^N \tilde{y}^N} = 1_{\tilde{y}^N \succeq y^N} \Pi_{y_0^N \tilde{y}_0^N}, \tag{27}$$

where $1_{\tilde{y}^N \succeq y^N} = 1$ if $\tilde{y}^N$ is a possible continuation of y^N (alternatively, if y^N is a possible past history for $\tilde{y}^N$), or formally: $\tilde{y}_{-1}^N = y_0^N$, $\tilde{y}_{-2}^N = y_{-1}^N, \ldots$, $\tilde{y}_{-N+1}^N = y_{-1}^N$); $1_{\tilde{y}^N \succeq y^N} = 0$ otherwise. The population of agents being associated to truncated history y^N can be defined recursively from the previous probabilities as:

$$S_{y^N} = \sum_{\hat{y}^N \in \mathcal{Y}^N} S_{\hat{y}^N} \Pi_{\hat{y}^N y^N}. \tag{28}$$

Since the truncated model aims to express the economy using truncated histories, we need to derive for each truncated history its consumption level and its end-of-period savings, which will be denoted by $c_{t,y^N}(Z^t)$ and $a_{t,y^N}(Z^t)$ respectively, or simply by c_{t,y^N} and a_{t,y^N} when there is no confusion. They defined as the corresponding average value among agents sharing the same truncated history y^N. For instance, for savings:

$$a_{t,y^N} := \frac{1}{S_{y^N}} \int_{a_{-1}} \sum_{\hat{y}^t \in \mathcal{Y}^{t+1} | (\hat{y}_{t-N+1}, \ldots, \hat{y}_t) = y^N} \theta_t(\hat{y}^t) a_t(\hat{y}^t, a_{-1}, Z^t)) d\Lambda_0(a_{-1}, y_0).$$

For computing the beginning-of-period savings, denoted by $\tilde{a}_{t,y^N}$, we have to account for the fact that agents with current truncated history y^N had a possibly different truncated history $\hat{y}^N$ in the previous period. Formally:

$$\tilde{a}_{t,y^N} = \frac{1}{S_{y^N}} \sum_{\hat{y}^N \in \mathcal{Y}^N} S_{\hat{y}^N} \Pi_{\hat{y}^N y^N} a_{t-1,\hat{y}^N}. \tag{29}$$

We can aggregate the individual budget constraint (5) along common truncated history and obtain the following truncated-history budget constraint:

$$c_{t,y^N} + a_{t,y^N} = (1+r_t)\tilde{a}_{t,y^N} + w_t y_0^N - T_t, \tag{30}$$

[10]For the sake of consistency, we will denote with a tilde future truncated histories, with a hat past ones, and without decoration current ones.

where y_0^N is the current productivity level for y^N. The previous aggregation is straightforward since budget constraints are linear. Aggregating utility or its derivatives is less so, because utility levels and marginal utilities are not linear in consumption. To proceed with the the utility aggregation, we define the following history-dependent parameters:

$$\xi^0_{t,y^N} := \frac{1}{u(c_{t,y^N})} \int_{a_{-1}} \sum_{\hat{y}^t \in \mathcal{Y}^{t+1} | (\hat{y}_{t-N+1}, \ldots, \hat{y}_t) = y^N} \theta_t(\hat{y}^t) u(c_t(\hat{y}^t, a_{-1}, Z^t)) d\Lambda_0(a_{-1}, y_0),$$

$$\xi^1_{t,y^N} := \frac{1}{u'(c_{t,y^N})} \int_{a_{-1}} \sum_{\hat{y}^t \in \mathcal{Y}^{t+1} | (\hat{y}_{t-N+1}, \ldots, \hat{y}_t) = y^N} \theta_t(\hat{y}^t) u'(c_t(\hat{y}^t, a_{-1}, Z^t)) d\Lambda_0(a_{-1}, y_0),$$

$$\xi^2_{t,y^N} := \frac{1}{u''(c_{t,y^N})} \int_{a_{-1}} \sum_{\hat{y}^t \in \mathcal{Y}^{t+1} | (\hat{y}_{t-N+1}, \ldots, \hat{y}_t) = y^N} \theta_t(\hat{y}^t) u''(c_t(\hat{y}^t, a_{-1}, Z^t)) d\Lambda_0(a_{-1}, y_0).$$

These parameters enable one to reconcile aggregation of utility (or its derivatives) and utility of aggregate quantities. For instance, the aggregate utility in period t can be expressed as the sum over all histories and all initial asset holdings of individual utility levels:

$$\int_{a_{-1}} \sum_{y^t \in \mathcal{Y}^{t+1}} \theta_t(y^t) u(c_t(y^t, a_{-1}, Z^t)) d\Lambda_0(a_{-1}, y_0).$$

Generally, because the utility function is not linear, it differs from the utility of derived from truncated-history consumption levels. The role of the ξ^0 is precisely to reconcile both and the previous aggregate utility is also: $\sum_{y^N \in \mathcal{Y}^N} S_{y^N} \xi^0_{t,y^N} u(c_{t,y^N})$, which is the sum of truncated-history utility, weighted by ξ^0. A similar mechanism applies for the marginal utility u' with ξ^1, and for u'' with ξ^2. These parameters enable one to capture the residual heterogeneity within each truncated history that is due to the fact that agents experienced different idiosyncratic histories N periods ago and before. Indeed, each truncated history groups together by construction agents sharing the same history over the last N periods, while ignoring the distant past. On the theoretical side, the $\boldsymbol{\xi}$ parameters constructed from the aggregation of a Bewley model can be shown to converge toward 1 when the length of the truncation N increases. However, since N remains small in practice, this asymptotic result has little practical implications. Fortunately, as we check in our quantitative exercise of Section 6, even for short truncation lengths, the ξs allows one to obtain accurate results. We explain in Section 4.2 how to easily compute the ξ^ks for $k = 0, 1, 2$.

We finally denote by $\mathcal{C}_{t,N}$ the set of credit-constrained truncated histories at date t. With this notation and the previous ξs, the Euler equations can be written as follows:

$$\forall y^N \in \mathcal{Y}^N \setminus \mathcal{C}_{t,N},\ \xi^1_{t,y^N} u'(c_{t,y^N}) = \beta \mathbb{E}_t \Big[(1 + r_{t+1}) \sum_{\tilde{y}^N \in \mathcal{Y}^N} \Pi_{y^N \tilde{y}^N} \xi^1_{t+1,\tilde{y}^N} u'(c_{t+1,\tilde{y}^N}) \Big], \quad (31)$$

$$\forall y^N \in \mathcal{C}_{t,N},\ a_{t,y^N} = 0, \quad (32)$$

where the expectation over future idiosyncratic state has been taken care of explicitly and the conditional expectation operator concerns the sole aggregate shock. Equation (31) is the Euler equation at the truncated history level for non-credit-constrained histories, while equation (32) corresponds to credit-constrained truncated histories holding zero asset.

We have characterized the truncated model, whose main advantage is to feature limited heterogeneity, and thereby characterized by a finite number of equations and unknowns.

4.2 Using truncation method for an exogenous tax path

We now explain how the truncated allocation and the ξ^0, ξ^1, ξ^2 can be computed for a steady-state Bewley model for an exogenous tax path $(T_t)_{t\geq 0}$. Since the focus is on the steady-state equilibrium, we assume that the TFP is fixed: $Z_t = Z$ for some $Z > 0$ at all dates. Furthermore, we assume that $\lim_t T_t$ exists and is noted T_∞, which is the steady-state tax. We use a ∞ subscript to highlight that the tax path may be time-varying in the transition; other steady-state variables are denoted without time subscript.

As proved in the literature (for instance in Açikgöz, 2018 for recent results), the solution of the Bewley model is characterized by a steady-state wealth distribution, $\Lambda_\infty : (a, y) \in [0, +\infty) \times \mathcal{Y} \to \mathbb{R}_+$, and a set of policy rules for savings, denoted by $g_a : (a, y) \in [0, +\infty) \times \mathcal{Y} \to \mathbb{R}_+$. Loosely speaking, $\Lambda_\infty(da, y)$ is the measure of agents with a wealth between a and $a+da$ and a productivity level y, while $g_a(a, y)$ is the end-of-period savings of an agent with beginning-of-period wealth a and current productivity level y.

To compute the truncated allocation, we need to obtain the wealth distribution with respect to truncated histories of length N, that will denoted by $\tilde{\Lambda}_N : (a, y^N) \in [0; +\infty) \times \mathcal{Y}^N \to \mathbb{R}_+$ – such that $\tilde{\Lambda}_N(da, y^N)$ is the measure of agents with wealth in $[a, a + da)$ and truncated history $y^N = (y_{-N+1}, \ldots, y_{-1}, y_0)$. The measure $\tilde{\Lambda}_N$ can be computed by starting from the wealth distribution of agents in state y_{-N+1}, which is $\Lambda_\infty(\cdot, y_{-N+1})$ and then applying successively the sequence of savings policy functions corresponding to $y^N = (y_{-N+1}, \ldots, y_{-1}, y_0)$, which is $g_a(\cdot, y_{-N+1})$, $g_a(\cdot, y_{-N+2}), \ldots, g_a(\cdot, y_0)$. From a practical perspective, computing $\tilde{\Lambda}_N$ is both straightforward and very fast, since it consists in multiplying an initial distribution, modeled as a vector, with N different transition matrices.

Using the measure $\tilde{\Lambda}_N$, we deduce that the end-of-period savings, a_{y^N}, for a truncated history y^N can be computed as:

$$a_{y^N} = \int_{a\in[0,\infty)} a\tilde{\Lambda}_N(da, y^N), \tag{33}$$

where in our case the savings are actually bounded from above (see Açikgöz, 2018). We can then deduce from (29), the beginning-of-period savings $(\tilde{a}_{y^N})_{y^N}$, as consumption $(c_{y^N})_{y^N}$ from the truncated-history budget constraint (30).

The ξ^k parameters can be computed following a procedure similar to (33). For instance, for

ξ^1, we have

$$\xi^1_{y^N} = \frac{1}{u'(c_{t,y^N})} \int_{a \in [0,\infty)} u'(c(a, y_0^N)) \tilde{\Lambda}_N(da, y^N), \tag{34}$$

and similar computations can be derived for ξ^0 and ξ^2.

We deduce from equations (28)–(32) that the steady-state economy is then characterized by the following set of equations:

$$\tilde{a}_{y^N} = \frac{1}{S_{y^N}} \sum_{\hat{y}^N \in \mathcal{Y}^N} S_{\hat{y}^N} \Pi_{\hat{y}^N y^N} a_{t-1,\hat{y}^N}, \tag{35}$$

$$c_{y^N} + a_{y^N} = (1+r)\tilde{a}_k + w y_0^N - T_\infty, \tag{36}$$

$$y^N \notin \mathcal{C},\ \xi_{y^N} u'(c_{y^N}) = \beta(1+r) \sum_{\tilde{y}^N=1}^{N_{tot}} \Pi_{y^N \tilde{y}^N} \xi_{\tilde{y}^N} u'(c_{\tilde{y}^N}), \tag{37}$$

$$y^N \in \mathcal{C},\ a_{y^N} = 0, \tag{38}$$

where y_0^N is the current productivity level of history y^N.

4.3 Computing the steady-state Ramsey allocation

We now show how the previous construction can be used to solve for optimal policies at the steady state. This computation proceeds in two steps:

1. For a given steady-state tax level T_∞, we compute the truncated Bewley allocation as explained in Section 4.2.
2. We derive the FOCs of the Ramsey program in the truncated economy, and then compute all Lagrange multipliers for the truncated economy.
3. The Lagrange multipliers allow one to check whether the planner's FOC characterizing the optimal value of T_∞ holds. If the constraint holds, then T_∞ is the optimal steady-state tax. If not, the procedure must be repeated for an updated value of T_∞.

We provides the derivation or all steps and then present the algorithm.

4.3.1 First-order conditions of the Ramsey program in the truncated economy.

The details of the computation can be found in Section B. Before stating the conditions, we need to introduce the quantity $\tilde{\lambda}_{t,y^N}$ defined as follows:

$$\tilde{\lambda}_{t,y^N} = \frac{1}{S_{t,y^N}} \sum_{\tilde{y}^N \in \mathcal{Y}^N} S_{t-1,\tilde{y}^N} \Pi_{t,\tilde{y}^N,y^N} \lambda_{t-1,\tilde{y}^N}, \tag{39}$$

which correspond to the previous period Lagrange multiplier for agents with truncated history y^N at date t. These agents may have different truncated histories in the previous period, which explains the expression (39) – as was the case for beginning-of-period wealth, $\tilde{a}_{t,y^N}$, in equation (29).

We use (39) to express the quantity ψ_{t,y^N}, which is the social valuation of liquidity for truncated history y^N – and is thus the parallel of the individual quantity ψ_t^i of equation (17). The formal definition is:

$$\psi_{t,y_N} = \xi^1_{t,y^N} u'(c_{t,y^N}) - (\lambda_{t,y^N} - \tilde{\lambda}_{t,y^N}(1+r_t))\xi^1_{t,y^N} u''(c_{t,y^N}). \tag{40}$$

With this notation, the first-order conditions for the Ramsey in the truncated economy can be written as follows:

$$\begin{aligned}
\psi_{t,y_N} = {} & \beta \mathbb{E}_t \left[(1+r_{t+1}) \sum_{\tilde{y}^N \in \mathcal{Y}} \Pi_{y^N \tilde{y}^N} \psi_{t+1,\tilde{y}_N} \right] && (41)\\
& + \beta \mathbb{E}_t \left[\sum_{\tilde{y}^N \in \mathcal{Y}} S_{t+1,\tilde{y}^N} \psi_{t+1,\tilde{y}_N} \left(\tilde{a}_{t+1,\tilde{y}^N} F_{KK}(K_t, L) + y_{\tilde{y}^N} F_{LK}(K_t, L) \right) \right] \\
& + \beta F_{KK}(K_t, L) \mathbb{E}_t \left[\sum_{\tilde{y}^N \in \mathcal{Y}} S_{\tilde{y}^N} \tilde{\lambda}_{t+1,\tilde{y}^N} \xi_{\tilde{y}^N} u'(c_{t+1,\tilde{y}^N}) \right], && (42)\\
v'(T_t) = {} & \sum_{y^N \in \mathcal{Y}^N} S_{y^N} \psi_{t,y_N}, && (43)
\end{aligned}$$

which are very similar to the individual conditions (18) and (18). The summary of the dynamic model is provided in Appendix B.2.

4.3.2 Computing the Ramsey allocation at the steady state: using Matrix notation

We now provide the algorithm to solve find the steady-state value of the instrument. A very efficient representation of the truncated model relies on matrix notation, which enables one to derive Lagrange multipliers using simple linear algebra. To implement in practice the numbering of truncated histories, a convenient solution is to use the enumeration in base n_y. A truncated history $y^N = (y_{-N+1}, \dots, y_{-1}, y_0)$ will be assigned the index $1 + \sum_{k=0}^{N-1} n_y^k (n_{y_{-k}} - 1)$ where $n_{y_{-k}} \in \{1, \dots, n_y\}$ is the position of productivity level y_{-k} in the set $\mathcal{Y}$, from 1 for the smallest productivity level to n_y for the largest one. The index belongs by construction to the set $\{1, \dots, N_{tot}\}$.

We then introduce the following matrix notation:

- $\boldsymbol{S} = (S_k)_{k=1,\dots,N_{tot}}$ the N_{tot}-vector of sizes;
- $\boldsymbol{c} = (c_k)_{k=1,\dots,N_{tot}}$, the N_{tot}-vector of consumption levels;
- $u'(\boldsymbol{c}) = (u'(c_k))_{k=1,\dots,N_{tot}}$, the N_{tot}-vector of marginal utilities;

- $\boldsymbol{\xi^j} = (\xi^j_k)_{k=1\ldots N_{tot}}$ the vector of residual-heterogeneity parameters ($j = 0, 1, 2$).
- Finally, we note by "$\circ$" the term-by-term product of two vectors of the same size, which is another vector of the same size: $\boldsymbol{x} \circ \boldsymbol{z} = (x_{y^N}) \circ (z_{y^N}) = (x_{y^N} z_{y^N})$;[11]

We can now state our result regarding the computation of Lagrange multipliers.

Proposition 1 (Steady-state Lagrange multipliers) *Consider a steady-state tax T_∞, for which the truncated model can be computed. Then, there exist two matrices $\boldsymbol{M_1}$ and $\boldsymbol{M_2}$, depending only on the equilibrium allocation, such that:*

$$\boldsymbol{\lambda} = \boldsymbol{M_1}(\boldsymbol{\xi^1} \circ u'(\boldsymbol{c})), \tag{44}$$

$$\boldsymbol{\psi} = \boldsymbol{M_2}(\boldsymbol{\xi^2} \circ u'(\boldsymbol{c})). \tag{45}$$

Proposition 1 states that the steady-state values of Lagrange multipliers and social value of liquidity can be deduced from truncated the truncated allocation, using basic linear algebra. This computation is possible for any value of the instrument for which the steady-state equilibrium can be computed. The proof can be found in Appendix C, where we provide a step-by-step computation of the expressions of matrices $\boldsymbol{M_1}$ and $\boldsymbol{M_2}$. Once, the social value of liquidity vector, $\boldsymbol{\psi}$, has been computed, it is straightforward to check the optimality of the steady-state tax level T_∞ using condition (43), which can be written in a matrix form as

$$v'(T_\infty) = \boldsymbol{S}^\top \boldsymbol{\psi}, \tag{46}$$

where $\boldsymbol{S}^\top \boldsymbol{\psi}$ is a scalar. This result provides the basis for the following algorithm summarizing the successive steps to compute the steady-state Ramsey allocation.

Algorithm 1 (Steady-state Ramsey allocation)

Set a precision criterion $\varepsilon > 0$.

1. *Set an initial value for the steady-state lump-sum tax T_∞.*
2. *Use the method of Section 4.2 to compute the steady-state truncated allocation $(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{\xi^1})$ associated to T_∞.*
3. *Compute matrices $\boldsymbol{M}_1$, and $\boldsymbol{M}_2$ using equations (81) and (82) of Appendix C.*
4. *Compute the vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}$ using equations (44) and (45).*
5. *If $|v'(T_\infty) - \boldsymbol{S}^\top \boldsymbol{\psi}| < \varepsilon$, then the algorithm stops and the steady-state Ramsey allocation is given by $(T_\infty, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{\xi^1})$. Otherwise, update T_∞ and start at Step 2 again.*

[11] This operation is also known as the Hadamard product.

Algorithm 1 explains how to find the steady-state optimal lump-sum tax as a fixed point of an iteration procedure. The algorithm starts with a guess for the steady-state value of the lump-sum tax T_∞. It then involves computing the allocation of the Bewley model corresponding to the lump sum tax T_∞. One can then deduce the steady-state allocation of the truncated model. Taking advantage of the matrix notation (further details in Appendix C) enable sone to compute the social valuation of liquidity $\boldsymbol{\psi}$. Finally, the optimality of the steady-state lump-sum tax T_∞ can be checked with equation (46).

Two remarks are in order. First, Step 2 of Algorithm 1 implies that we compute the Bewley the truncated allocations for each value of the steady-state lump-sum tax. As a consequence, the algorithm by construction converges to a Bewley equilibrium that does exist. Second, the computational implementation of Algorithm 1 is fast. At every step, the computationally intensive task is to simulate the Bewley model for the steady-state lump-sum tax T_∞. The other steps (in particular, 3 to 6) only involve linear algebra and are very fast to perform (less than a second). Besides its speed, this algorithm is also accurate, as we check in Section 6, where we compare its solution to the one of the transition methods.

4.4 Solving the Ramsey model with aggregate shocks

Once the steady-state allocation has been computed using Algorithm 1, the simulation of the Ramsey model in the presence of aggregate shocks is straightforward. Indeed, it relies on perturbation techniques that can be implemented using standard packages, such as Dynare. We summarize the various steps in Algorithm 2.

Algorithm 2 (Simulating the model with aggregate shocks)

We consider as given a truncation length, $N > 0$, a precision criterion $\varepsilon > 0$. We assume that the dynamics of the TFP is given by $Z_t = \exp(z_t)$, where:

$$z_t = \rho_z z_{t-1} + \varepsilon_t, \text{ with } \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_z^2). \tag{47}$$

1. *We use Algorithm 1 to compute the steady-state Ramsey allocation with precision ε.*
2. *The model in the presence of aggregate shocks is thus determined by equation (47) and the equations (58)–(66) of Appendix B.2.*
3. *These equations can be simulated by perturbation method around the steady-state distribution previously computed.*

Algorithm 2 describes a straightforward path to simulate the model in the presence of aggregate shocks. We specify the dynamics of the TFP in equation (47), which states that its log follows a standard AR(1) process. This is a standard specification in the literature and the algorithm could easily be extended to a more complex process. The model equations to be simulated is summarized in Appendix B.2. The core of Algorithm 2 is the perturbation method of Step 3 that can rely on existing and well-tested software, such as Dynare, that are already widely adopted for solving DSGE models.

Algorithm 2 involves two mains assumptions for simulating the truncated model in the presence of aggregate shocks. The first one is that the ξs coefficients remain equal to their steady-state value in the presence of aggregate shocks. This implies that the heterogeneity within truncated histories is constant through time and equal to its steady-state value. Importantly, this assumption does not preclude the existence of heterogeneity within each truncated history. The second assumption is that credit-constrained truncated histories are determined at the steady state – and hence depend on the steady-state optimal fiscal policy. However, in the presence of aggregate shocks, credit-constrained histories remain the same and are unaffected by variations in the optimal policy due to aggregate shocks. In other words, shocks remain small enough for not affecting the set of credit-constrained histories. This second assumption is not a limitation of the truncation method, but of the perturbation method that does not easily allow one to feature time-varying occasionally-binding credit-constraints.

5 Maximizing aggregate welfare with transitions

Before turning to the quantitative exercise, we present two other optimization methods to which we will compare the solution of our Ramsey problem. The first optimization method consists in finding the constant lump-tax that maximizes the aggregate welfare while accounting for transitions, as in Conesa et al. (2009) or Chang et al. (2018) among many others. The initial distribution of agents (over wealth and productivity) is given and exogenous. We will call the resulting tax rate, the *optimal tax rate with transitions and exogenous initial distribution.* We provide a formal definition below.

Definition 3 *Given an initial distribution $\Lambda_0 : [0, \infty) \times \mathcal{Y} \to \mathbb{R}_+$ over wealth and productivity, the optimal tax rate with transitions and exogenous initial distribution is the constant lump-sum tax T such that:*

1. *When the fiscal policy is constant and set to T, the Bewley model with the initial distribution Λ_0 converges to a long-run distribution denoted $\Lambda_\infty(T)$.*[12]
2. *The lump-sum tax T maximizes the period-0 aggregate welfare computed when accounting for the transitions when the agents' distribution evolves from Λ_0 at $t = 0$ to $\Lambda_\infty(T)$ in the long run ($t \to \infty$).*

The algorithm for this solution can be summarized as follows.

[12] We make the dependence in T explicit.

Algorithm 3 (Computing the optimal tax with transitions and exogenous initial distribution)

We consider as given an initial distribution Λ_0. The optimal tax with transitions and exogenous initial distribution can be computed as follows:

1. *Set an initial guess for the lump-sum tax T.*

2. *Solve for the steady-state distribution $\Lambda_\infty(T)$.*

3. *Compute the transition, and the welfare during the transition of the economy from the initial distribution Λ_0 toward $\Lambda_\infty(T)$.*

4. *Update T and start again in Step 2 until the computed welfare is maximal.*

There are two main differences with our truncation method for the Ramsey model – presented in Algorithm 1. First, the tax is assumed to remain constant along the transition. This should be thought as a very-constrained Ramsey problem, where the planner cannot change its instrument that is set once for all in the initial period. In the truncation method, we compute the long-run limit of the tax, but the tax value can change along its path. Second, the result depends on the initial distribution (the initial parameter Λ_0), since the welfare computation accounts for transitions from the initial distribution to the long-run one.[13] With the truncation method, the initial distribution has no influence on the steady state optimal Ramsey tax. We illustrate this point quantitatively in Section 6.

To mitigate the effect of the initial distribution we propose to modify the previous computation by iterating on the initial distribution, such that the initial distribution coincides with the terminal distribution. To our knowledge, this procedure is new. We formalize the definition below of the so-called *optimal tax rate with transitions.*[14]

Definition 4 *The optimal tax rate with transition is the constant fiscal policy T^c such that:*

1. *When the fiscal policy is constant and set to T^c, the Bewley model with the initial distribution Λ^c_∞ converges to the same long-run distribution Λ^c_∞.*

2. *The tax rate T^c maximizes the period-0 aggregate welfare computed when accounting for the transitions when the agents' distribution evolves from Λ^c_∞ at $t = 0$ to Λ^c_∞ in the long run.*

[13] For instance, starting from an economy with a very low initial stock (close to 0) will imply a very low T, as the quantities of goods to tax in the initial periods are close to 0.

[14] To avoid confusion, we will call the "Ramsey optimal tax" or "the optimal tax" (when no ambiguity) to refer to the optimal tax rate computed using Algorithm 1. We will call the "optimal tax rate with transitions" the one of Definition 4 below, where the initial and long-run distributions are identical. We will call "optimal tax rate with transition and exogenous initial distribution" the one of Definition 3, where the initial distribution influences the outcome. Due to this limitation, the latter will barely be used in our applications of Section 6.

The optimal tax rate with transition T^c is such that: (i) the initial and the long-run distributions coincide with each other when the fiscal policy is constant and set to T^c, and (ii) the aggregate welfare with transitions is optimal when fiscal policy is set to T^c. We formalize the computation of T^c in the algorithm below.

Algorithm 4 (Computing the optimal tax with transitions) *The optimal tax with transitions can be computed as follows:*

1. *Choose an initial guess for the tax rate* T

2. *Choose an initial guess for the initial distribution* Λ_0*:*

    (a) *compute the long-run distribution* $\Lambda_\infty(T, \Lambda_0)$ *depending on both* T *and* Λ_0*;*

    (b) *if initial and long-run distributions coincide (i.e.,* $\Lambda_\infty(T, \Lambda_0) = \Lambda_0$*), then stop. Otherwise, update the initial distribution:* $\Lambda_0 \leftarrow \Lambda_\infty(T, \Lambda_0)$ *and start at 2(a).*

3. *Compute the aggregate welfare during the transition of the economy from the initial distribution* Λ_0 *toward* $\Lambda_\infty(T, \Lambda_0) = \Lambda_0$.

4. *Update* T *and start again in Step 2 until the computed welfare is maximal.*

Definition 4 and Algorithm 4 neutralizes the influence – and hence the choice – of the initial distribution in the optimal tax rate with transitions. Step 2 of Algorithm 4 computes the initial distribution Λ_0 such that for the tax rate T, the long-run distribution is identical to the initial one: $\Lambda_0 = \Lambda_\infty(T, \Lambda_0)$. Such an initial distribution Λ_0 will said to be *consistent.* However, even with this consistent initial distribution, the tax rate is kept constant along the transition, which is not the case in the Ramsey program. As quantified in Section 6 with a realistic calibration, we will see that this involves a non-zero – though small – difference with the Ramsey steady-state optimal tax rate.

6 Quantitative exercise

We now turn to the quantitative exercise. The objectives and roadmap of this section are the following:

1. We specify our calibration in Section 6.1.

2. We compute the optimal tax rate using the truncated method in Section 6.2. We also compare the results of the truncated method with other methods to check its accuracy.

3. We compute the optimal tax rate using transitions in Section 6.3.

4. We document that the difference between the two stems from a time-inconsistency issue in Section 6.4.

5. We check that the optimal path obtained with the truncation method actually generates the highest intertemporal welfare in this environment (Section 6.6).

6.1 The calibration

The period is a quarter. The discount factor is set to $\beta = 0.99$. The period utility function for the private good is $\log(c)$. The utility function for the public good is $v(G) = G^\theta$. We set the parameter θ to 24% to target a value of steady-state public consumption over GDP of 8.0%, which corresponds roughly to the US government consumption on final good minus public investment. In the production function of (1), the capital share is set to $\alpha = 36\%$ and the depreciation rate to $\delta = 2.5\%$, as in Krueger et al. (2018) among others.

Idiosyncratic productivity is modeled as an AR(1) productivity process: $\log y_t = \rho_y \log y_{t-1} + \varepsilon_t^y$, with $\varepsilon_t^y \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_y^2)$. We calibrate the parameters ρ_y and σ_y, to use a realistic income process, following estimates of Krueger et al. (2018). We use a quarterly persistence of $\rho_y = 0.996$ and a quarterly standard deviation of $\sigma_y = 4.39\%$, which generate, for the log of earnings, an annual persistence of 0.9849 and an annual standard deviation of 8.71%. The Rouwenhorst (1995) procedure is then used to discretize the productivity process into 5 idiosyncratic states with a constant transition matrix.

Table 1 provides a summary of the model parameters.

Parameter	Description	Value
β	Discount factor	0.98
α	Capital share	0.36
δ	Depreciation rate	0.025
$\tau = T/Y$	Tax rate	8%
θ	Curvature of util. public good	24%
ρ_y	Autocorrelation idio. income	0.996
σ_y	Standard dev. idio. income	4.39%

Table 1: Parameter values in the baseline calibration. See text for descriptions and targets.

We can now compute the steady state equilibrium of the model, using the standard EGM method.[15] The implied capital-output ratio is $K/Y = 2.67$, the consumption-output ratio is $C/Y = 0.65$. Table 2 provides descriptive statistics regarding the wealth distribution in the data and in the model. We use data from the PSID in 2006 and from the SCF in 2007 to abstract

[15]We use the EGM method with 100 points for a exponential grid point for wealth, following Carrol (2006) and Boppart et al. (2018) among others.

from the effects of the 2008 crisis. The model does a relatively good job in reproducing the

	Data		Model
Wealth statistics	PSID, 06	SCF, 07	
Q1 (minimum)	−0.9	−0.2	0.0
Q2	0.8	1.2	0.3
Q3 (median)	4.4	4.6	5.6
Q4	13.0	11.9	21.4
Q5 (maximum)	82.7	82.5	72.7
Gini	0.77	0.78	0.71

Table 2: Steady-state wealth distribution.

wealth distribution. It is known that other mechanisms must be introduced to match the very top of the wealth distribution (such as entrepreneurship or stochastic βs).

6.2 Solving the model with the truncation method

We first compute the optimal long-run value of the tax system using the truncation method of Section 4. We consider a truncation length equal to $N = 5$ – which implies $5^5 = 3125$ histories. We follow Algorithm 1 to compute the steady-state optimal tax T_∞, which we report as a function of GDP and denote $\tau^p_\infty := T_\infty/Y$. The model generates the optimal tax-to-GDP $\tau^p_\infty = T_\infty/Y = 8\%$ – which was the targeted value. The superscript p refers to optimal tax computed with the truncation method. We use the superscript tr for transition. We perform sensitivity test on the choice of the truncation length N in Section 6.7 and the value of N (beyond $N \geq 2$) appears to have a very modest quantitative impact on the optimal provision of public good.

We provide further explanation on the different contributing forces to the optimal tax-to-GDP of 8%. The optimal tax at date t is defined through equation (20), which with the definition (17) of ψ becomes:

$$v'(T_t) = \int_i \Big(\underbrace{u'(c^i_t)}_{\text{direct effet}} \underbrace{- u''(c^i_t)\left(\lambda^i_t - (1+r_t)\lambda^i_{t-1}\right)}_{\text{effect on savings}} \Big) \ell(di). \tag{48}$$

To evaluate the contribution of indirect effects (through savings) to the optimal tax, we can compute the value of the tax $T^{partial}$ that would correspond to a planner valuing only direct effects. Formally, this corresponds to:

$$v'(T^{partial}_t) = \int_i u'(c^i_t)\ell(di).$$

The numerical quantification implies that $T^{partial}$ is found to be 4.3% lower than T^p_∞. The tax-to-GDP ratio decreases from $\tau^p_\infty = 8.0\%$ to $\tau^{partial} = 7.76\%$. A higher tax raises precautionary savings and hence boosts capital and aggregate consumption. When the planner internalizes the savings distortions (through the term in $u''(c^i_t)$ in (48)), it therefore leads to a higher tax – since the benefits of this higher tax are factored in by the planner.

6.3 Results using the transition method

We first solve for the *optimal tax rate with transition and exogenous initial distribution* as presented in Definition 3. We denote by T^{exo} this optimal tax rate, and by τ^{exo} the related tax-to-GDP ratio. To do so, we implement Algorithm 3, which needs an initial distribution Λ_0 as input. As already discussed, the method solution is sensitive to the choice of this initial distribution. To illustrate this, we solve for the optimal tax with two different initial distributions. We first consider the steady-state distribution presented in Table 2 of Section 6.1. We then multiply the wealth level of all agents by 0.9. Doing so we consider a 10% decrease in initial wealth of all agents. We call this distribution the low distribution, which will correspond to the optimal tax-to-GDP τ^{low}. The second initial distribution, called *high*, is a 10% increase in the wealth of all agents starting from the distribution of Table 2. The corresponding optimal tax-to-GDP will be denoted τ^{high}.

The computations of both optimal tax rates with transition and exogenous distribution yields $\tau^{low} = 6.4\%$ and $\tau^{high} = 8.45\%$, respectively. When the initial wealth is low, the tax – that is imposed to remain constant throughout the transition path – affects private consumption at the beginning of the transition when agents have few resources and hence a high marginal utility for private consumption. This contributes to set a low tax. Oppositely, when the initial wealth is high, the tax can also be higher since in the first periods, agents a relatively low marginal utility for private consumption. This illustrates that in the context of Definition 3, the initial distribution has a sizable impact on the optimal outcome.

To neutralize the effect of the initial distribution, we compute the *optimal tax rate with transition*, as presented in Definition 4. We denote the corresponding optimal tax rate-to-GDP by τ^c, where the superscript c means "consistent" as we iterate on the initial distribution until $\Lambda^c = \Lambda_\infty(T^c, \Lambda^c)$. The computation yields $\tau^c = 7.8\%$, which lies between the low and high values, τ^{low} and τ^{high} that we have just computed.

To illustrate the optimality of τ^c, we plot in Figure 1 the aggregate welfare as a function of the tax, when the initial distribution is consistent (plain blue line). More precisely, for each tax value T, we iterate on the initial distribution Λ_0 to compute the fixed point verifying $\Lambda_0 = \Lambda_\infty(T, \Lambda_0)$ and compute the aggregate welfare with transitions when the distribution evolves from Λ_0 to $\Lambda_\infty(T, \Lambda_0) = \Lambda_0$. The tax rate on the x-axis is reported in terms of tax-to-GDP ratio $\tau = T/Y$, while the welfare on the y-axis is reported as the percentage loss in consumption compared to the optimal welfare. In addition to τ^c, we also report in Figure 1 as orange dashed vertical bars, three other tax rates: τ^{low} corresponding to the low-wealth initial distribution, τ^{high} corresponding to the high-wealth initial distribution, and τ^p_∞ corresponding the optimal tax rate computed using the truncation method. Figure 1 confirms that the optimal tax with transitions is highly

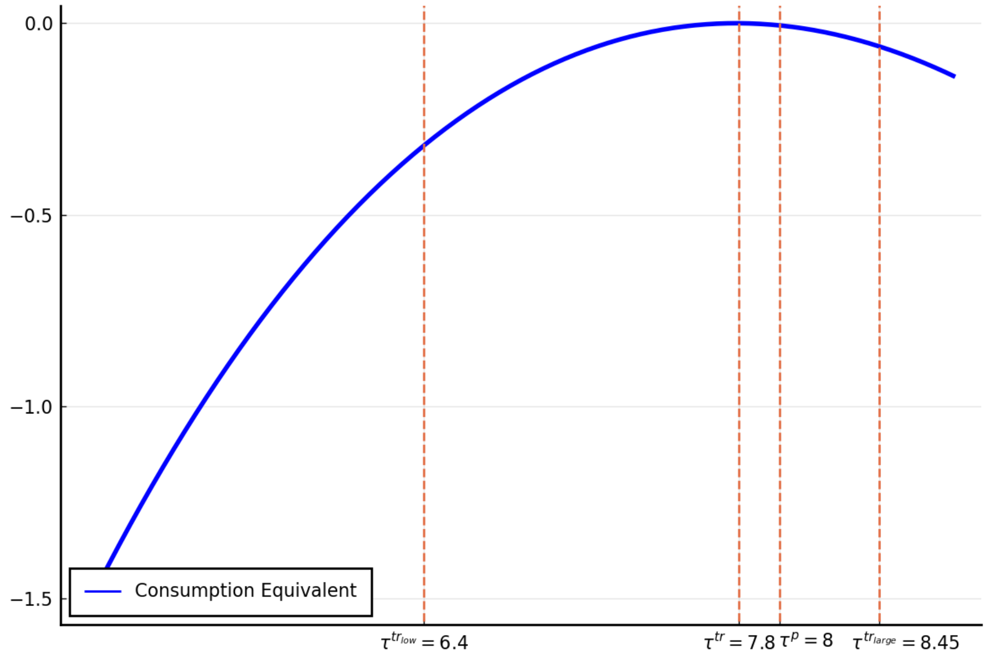

Figure 1: Welfare with transitions and consistent initial distribution computed as a function of the tax rate.

sensitive to the initial distribution, and we can find optimal rates that are greater or smaller than the optimal one with the consistent initial distribution. The optimal capital tax considering the transition with exogenous distribution can be either higher of lower than the optimal tax rate with consistent distribution. Figure 1 also illustrates that the optimal tax rate τ^c computed with transition and consistent initial distribution is close but lower than the optimal tax rate τ^p_∞ computed using the truncation method.

6.4 The time-inconsistency of Ramsey policies

We now take advantage of the tractability of the truncation method to understand the roots of the gap between the optimal tax rates τ^p_∞ and τ^c, computed respectively with truncation and transitions. We perform the following experiment. We set as initial distribution the steady-state distribution, Λ^p_∞, derived in the truncation approach and corresponding to the long-run tax rate T^p_∞. With our former notation, it is denoted by $\Lambda_0(T^p_\infty) = \Lambda^p_\infty$. We then allow the planner to choose an optimal time-varying path for the tax. This is a pure reoptimization shock in period 0, that implies setting to 0 the value of past Lagrange multipliers at date 0: $\lambda^i_{-1} = 0$ for all i. By construction, we know that the long-run steady-state value is also T^p_∞. Hence, the simulation will start from the steady-state distribution, Λ_∞^p, and the tax T_∞^p and will converge back to the same distribution $\Lambda_0(T_\infty^p)$, and the same tax T_∞^p in the long-run. Figure 2 plots the optimal path implied by the reoptimization shock. We denote the this optimal tax path by $(T_t^p)_{t\geq 0}$ and the corresponding optimal tax-to-GDP path by $(\tau_t^p)_{t\geq 0}$. We can verify that we have $\tau_0^p = \lim_{t\to\infty} \tau_t^p = \tau_\infty^p$. Furthermore, the tax rate drops at impact and the tax rate becomes

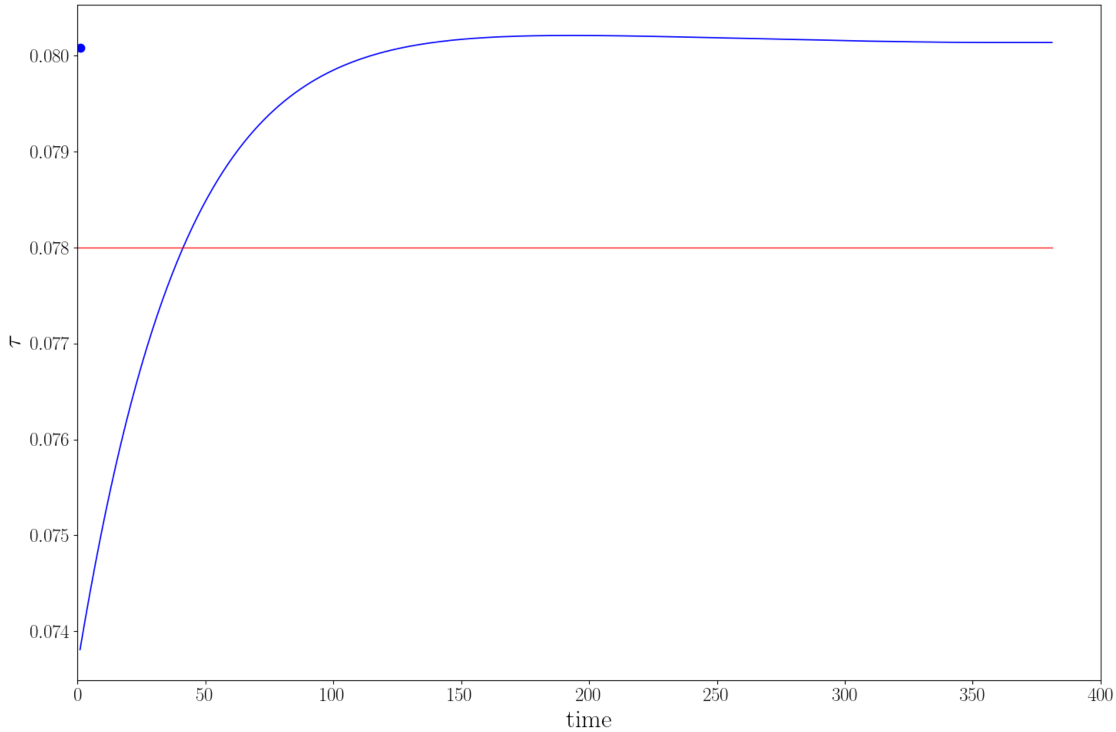

Figure 2: The optimal tax path after a reoptimization shock.

smaller than the optimal tax level with transition τ^c. The intuition as follows. Since $\lambda_{-1}^i = 0$, the planner is not committed to maintain the tax rate equal to τ_∞^p and chooses to decrease the tax, which allows agents, and especially those with a high marginal propensity to consume, to increase private consumption in the short run. The tax path then increases and converges back to the long-run steady-state value τ_∞^p. This raises agents' incentives to build up savings, which increases the capital stock and, in turns, the consumption of private goods. The consumption of public goods also increases with the tax.

Figure 2 allows one to understand the gap between the two optimal tax rates τ_∞^p and τ^c, computed with the truncation method and with transition and consistent initial distribution. This latter method consists in choosing a constant tax level over the whole transition path, whereas the truncation method characterizes the full tax path. The transition method can thus be seen as selecting the "average" tax rate over the whole transition path, so as to balance the benefits and the costs along the transition path. We check this assumption by computing the average discounted tax rate over the transition path $(\tau_t^p)_{t\geq 0}$ where we discount future tax values using the discount factor β, to obtain an approximation of the discounted value of the optimal path of taxes. We denote this weighted discounted tax rate as τ^{weight}, which is formally defined as:

$$\tau^{weight} = \frac{\sum_{t=1}^{400} \beta^{t-1}\tau_t^p}{\sum_{i=1}^{400} \beta^{t-1}}.$$

The computation yields $\tau^{weight} = 7.84\%$, which is very close to the tax rate τ^c computed with transitions. This computation shows that the gap between tax rates computed using truncation and using transition is well explained by the quantification of the reoptimization shock. The reoptimization shock is due to the time-inconsistency of optimal Ramsey policy in heterogeneous-agent models that we discuss in Section 3 and in Appendix E.

6.5 Effect of aggregate shocks

We now present the results for the optimal tax path and for IRFs after a technology shock. We compare the model dynamics implied by the truncation method to those implied by the full-fledged model simulated using the Reiter (2009) method.

The method to solve the model with the truncation method in the presence of aggregate shocks is explained in Section 4.4. Simulating the optimal path with the Reiter method is more involved as it does not able to directly compute the optimal path. We thus need to provide the optimal dynamics of the tax rate as an input. We proceed as follows. First, we simulate the model with TFP aggregate shocks using the truncation method to obtain the path of the optimal tax over the simulation period of 10,000 periods. Second, we approximate the optimal tax path using two observable aggregates of the model, the capital and the GDP.[16] More precisely, we run the following regression:

$$T_t = a + b\,Y_t + c\,I_t + dC_t + \varepsilon_t^T, \text{ where } \varepsilon_t^T \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_T^2). \tag{49}$$

We obtain the following values: $a \approx 0.0$, $b = 12.4867$, $c = -3.3402$, $d = -8.1465$ with a R^2 equal to 1.0. This allows us to capture the dynamics of the tax path using only the GDP and capital. This policy rule is then plugged into a full-fledged heterogeneous-agent with aggregate shock that can be simulated using the Reiter's method – since there is no optimization to perform. [17]

We report the Impulse Response Functions (IRFs) for the main variables in Figure 3. The two methods are labeled "Truncation" and "Reiter". We also plot the aggregate welfare in the two cases economies, using the utilitarian Social Welfare Function, as percentage change in consumption equivalent. It can be observed that the two simulation methods generate very similar results, along the tax path (by construction), aggregate quantities (GDP, consumption

[16]We have considered more involved regressions, but with actual improvement on the fit.

[17]We implement the Reiter method using 100 wealth bins and idiosyncratic states and perform a first-order perturbation of the policy rules as a function of the aggregate shock.

and capital), prices (interest rate and wages), and aggregate welfare.[18] We complete the findings of Figure 3 by reporting in the two first columns of Table 3 the second-order moments in the two simulations (Reiter and the truncation method). As was the case for IRFs, the second-order moments are very close in both cases. Note that the columns (2)–(5) are a robustness check, which is discussed in Section D.

Methods		Reiter	Trunc. $(N = 2)$	Trunc. $(N = 3)$	Trunc. $(N = 4)$	Trunc. $(N = 5)$
Economies		(1)	(2)	(3)	(4)	(5)
GDP	Mean	3.793	3.793	3.793	3.793	3.793
	Std/mean (%)	1.288	1.280	1.280	1.281	1.281
C	Mean	2.475	2.475	2.475	2.475	2.475
	Std/mean (%)	0.978	0.980	0.980	0.979	0.979
K	Mean	40.590	40.590	40.590	40.590	40.590
	Std/mean (%)	1.225	1.196	1.198	1.199	1.201
		.				.
		.				.
$\text{Corr}(C, C_{-1})$		0.9945	0.9941	0.9941	0.9941	0.9942
$\text{Corr}(GDP, GDP_{-1})$		0.9695	0.9691	0.9691	0.9692	0.9692
$\text{Corr}(GDP, C)$		0.9242	0.9300	0.9298	0.9296	0.9294

Table 3: Moments of the simulated model for different computational techniques.

6.6 Checking the optimality of the transition path

We now check the optimality of the transition path, by checking that any perturbation of the optimal tax path of Figure 2 implies a decrease in aggregate welfare, when the full-fledged Aiyagari model is simulated (without relying on the truncation method). Given the optimal tax path of Figure 2, we construct for any real value κ the tax path $(T_t^\kappa)_t$ as:

$$T_t^\kappa = (1 + \kappa)(T_t^p - T_\infty^p) + T_\infty^p,\ t \geq 0,$$

where $T_\infty^p = 30.23$ is the steady-value of the optimal tax path. Independently of κ, any path T_t^κ converges at the steady state toward T_∞^p – which means that there is no steady-state deviation. The parameter κ modifies the initial drop in the tax rate and the speed of convergence to the steady-state value. When $\kappa = 0$, we implement the optimal path, when $\kappa > 0$ ($\kappa < 0$), we implement a higher (lower) path. We then simulate the model for a tax path $(T_t^\kappa)_t$ for different values of κ and set the initial distribution equal to the steady-state distribution $\Lambda(T_\infty^p)$. We use the Reiter method (thus not relying on transition) to compute the model dynamics and the

[18] When computing the aggregate welfare, we use the coefficients ξ^0 to account for within-history heterogeneity. The aggregate welfare at date t is thus: $\sum_{k=1}^{N_{tot}} S_k(\xi_k^0 u(c_{k,t}) + v(G_t))$.

	Trunc. ($N = 7$)	Trunc ($\theta = 65\%$)
τ^p	8.0%	15%
τ^c	7.8%	14.45%
τ^{low}	6.4%	13.92%
τ^{high}	8.45%	15.40%

Table 4: Robustness checks.

associated aggregate welfare. We report the results in Figure 4. The welfare is expressed as the consumption equivalent drop in welfare from optimum. Figure 4 shows that the welfare is maximal for a path that is extremely close to the optimal path (T_t^p) ($\kappa = 0$). We can thus be confident that the truncation method is well-suited to compute optimal policies in heterogeneous-agent models.

6.7 Additional robustness checks

We provide a series of robustness checks, in which we modify some key model parameters. The results are reported in Table 3 for the economy second-order moments, as well as in Table 4 for tax rates. We report the following optimal tax rates: τ_∞^p computed using the truncation method, τ^c computed using transition method and consistent initial distribution, τ^{low} computed with the transition method and the low-wealth initial distribution, and τ^{high} computed with the transition method and the high-wealth initial distribution (same as in Section 6.3).

We consider two robustness checks, one on the truncation length N, and another one on the concavity parameter θ of the public good utility function v. The columns (2)–(5) of Table 4 report the results for $N = 2, \ldots 4$ instead of $N = 5$. Table 4 reports the value of τ_∞^p computed with the truncation method for $N = 7$. We find that the optimal tax rate is basically unchanged when we move from $N = 5$ to $N = 7$. This two sets of results shows that the ξs efficiently capture the overall heterogeneity, even when the truncation length is not too long. The second column of Table 4 considers a change in the concavity parameter θ, and an increase from 24% to 65%. This raises optimal public good provision and thus yields a higher tax rate $\tau_\infty^p(\theta = 65\%) = 15\%$. The tax rate considering transition τ^c is also higher and remains close to $\tau_\infty^p(\theta = 24\%)$ of the benchmark case. Finally, the tax rates τ^{low} and $\tau^{high}$are also higher and the ranking $\tau^{low} < \tau^c < \tau^p < \tau^{high}$ is preserved. The relevant figures are provided in Appendix D.

7 Conclusion

We solve for optimal Ramsey policy in an heterogeneous-agent model with aggregate shocks, where the planner finances a public good by lump-sum taxes. The optimal policy is computed both using the standard transition method and using the truncation method. As the latter method is new, we explain in details the algorithm and the implementation strategy. Considering transitions, we improve on current technique to avoid the effect of initial distributions on the optimal value of the instrument.

This investigation allows us to derive two results. We first identify a time-inconsistency issue that is specific to heterogeneous-agent model and is absent from complete-market models. The time-inconsistency issue comes the combination of two factors: (i) the planner is unable to unwind ex-post heterogeneity, and (ii) individual saving decisions in heterogeneous-agent models create externalities on other agents in the next period. The planner thus needs to internalize these externalities via present and future Euler equations. This holds because the planner is committed to respect its current decisions in the future, but this creates a time-inconsistency since the planner of the future would like to renege on the promises made in the past. The second result is methodological. We show that the truncation method provides a simple and accurate long-run value of the instruments, that is immune to the time-inconsistency problem.

References

Açıkgöz, O. (2015): "Transitional Dynamics and Long-Run Optimal Taxation Under Incomplete Markets," Working Paper, Yeshiva University.

Açıkgöz, O., M. Hagedorn, H. Holter, and Y. Wang (2018): "The Optimum Quantity of Capital and Debt," Working Paper, University of Oslo.

Açıkgöz, O. T. (2018): "On the Existence and Uniqueness of Stationary Equilibrium in Bewley Economies with Production," *Journal of Economic Theory*, 173, 18–55.

Acharya, S., E. Challe, and K. Dogra (2020): "Optimal Monetary Policy According to HANK," Staff Report, Federal Reserve Bank of New York.

Adjemian, S., H. Bastani, M. Juillard, F. Karamé, J. Maih, F. Mihoubi, G. Perendia, J. Pfeifer, M. Ratto, and S. Villemot (2011): "Dynare: Reference Manual, Version 4.6," Dynare Working Paper 1, CEPREMAP.

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109, 659–684.

——— (1995): "Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting," *Journal of Political Economy*, 103, 1158–1175.

Aiyagari, S. R. and E. R. McGrattan (1998): "The Optimum Quantity of Debt," *Journal of Monetary Economics*, 42, 447–469.

Angeletos, G.-M., F. Collard, and H. Dellas (2020): "Public Debt as Private Liquidity: Optimal Policy," *CEPR Discussion Paper.*

Auclert, A., B. Bardóczy, M. Rognlie, and L. Straub (2021): "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," *Econometrica*, 89, 2375–2408.

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

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2021): "Inequality, Business Cycles, and Monetary-Fiscal Policy," *Econometrica*, Forthcoming.

Bilbiie, F. (2020): "Monetary Policy and Heterogeneity: An Analytical Framework," Working Paper, Lausanne University.

Bilbiie, F. and X. Ragot (2020): "Optimal Monetary Policy and Liquidity with Heterogeneous Households," *Review of Economic Dynamics*, forthcoming.

Boppart, T., P. Krusell, and K. Mitman (2018): "Exploiting MIT Shocks in Heterogeneous-Agent Economies: The Impulse Response as a Numerical Derivative," *Journal of Economic Dynamics and Control*, 89, 68–92.

Carrol, C. D. (2006): "The Method of Endogenous Grid Points for Solving Dynamic Stochastic Optimization Problems," *Economic Letters*, 91, 312–320.

Chang, B., Y. Chang, and S.-B. Kim (2018): "Pareto Weights in Practice: Income Inequality and Tax Reform," *Review of Economic Dynamics*, 28, 181–204.

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

Den Haan, W. J. (2010): "Assessing the Accuracy of the Aggregate Law of Motion in Models with Heterogeneous Agents," *Journal of Economic Dynamics and Control*, 1, 79–99.

Dyrda, S. and M. Pedroni (2018): "Optimal Fiscal Policy in a Model with Uninsurable Idiosyncratic Shocks," Working Paper, University of Toronto.

Ferriere, A. and G. Navarro (2020): "The Heterogeneous Effects of Government Spending: It's All About Taxes," Working Paper, Paris School of Economics.

Green, E. (1994): "Individual-Level Randomness in a Nonatomic Population," Working Paper, University of Minnesota.

Heathcote, J. and H. Tsujiyama (2017): "Inverse Optimal Taxation in Closed Form," Working Paper, Federal Reserve Bank of Minneapolis.

Hendren, N. and B. Sprung-Keyser (2020): "A unified welfare analysis of government policies," *Quarterly Journal of Economics*, 3, 1209–1319.

Huggett, M. (1993): "The Risk Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies," *Journal of Economic Dynamics and Control*, 17, 953–969.

Krueger, D., K. Mitman, and F. Perri (2018): "On the Distribution of the Welfare Losses of Large Recessions," in *Advances in Economics and Econometrics: Volume 2, Eleventh World Congress of the Econometric Society*, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuleson, Cambridge University Press, 143–184.

Lagos, R. and R. Wright (2005): "A Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy*, 113, 463–484.

LeGrand, F. and X. Ragot (2021a): "Managing Inequality over the Business Cycle: Optimal Policies with Heterogeneous Agents and Aggregate Shocks," *International Economic Review*, Forthcoming.

——— (2021b): "Sovereign Default and Liquidity: The Case for a World Safe Asset," *Journal of International Economics*, 131, 103462.

——— (2021c): "The Truncation Method to Solve for Heterogeneous-Agent Models," Working Paper, SciencesPo.

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

McKay, A. and C. Wolf (2022): "Optimal Policy Rules in HANK," Working Paper, FRB Minneapolis.

Nuño, G. and C. Thomas (2020): "Optimal Monetary Policy with Heterogeneous Agents," Working Paper, Banco de España.

Reiter, M. (2009): "Solving Heterogeneous-Agent Models by Projection and Perturbation," *Journal of Economic Dynamics and Control*, 33, 649–665.

Rouwenhorst, G. K. (1995): "Asset Pricing Implications of Equilibrium Business Cycle Models," in *Structural Models of Wage and Employment Dynamics*, ed. by T. Cooley, Princeton: Princeton University Press, 201–213.

Appendix

A Computing the FOCs of the Ramsey program

The Ramsey program (11)–(16) can be written using two instruments only: savings $(a_t^i)_i$ and lump-sum tax T_t. We also include the Euler equation into the planner's objective. Recalling that the Euler equation Lagrange multiplier is $\beta^t \lambda_t^i$, we obtain that the program Ramsey program (11)–(16) can equivalently be expressed as follows:

$$\begin{aligned} \max_{(T_t,(a_t^i)_{i\in I})_{t\geq 0}} \quad & \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\int_i u(c_t^i)\ell(di) + v(T_t) \right) \\ & - \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \int_i (\lambda_t^i - \lambda_{t-1}^i (1 + F_K(\int_i a_{t-1}^i \ell(di), \overline{L})))u'(c_t^i)\ell(di), \\ \text{where: } c_t^i = {} & (1 + F_K(\int_i a_{t-1}^i \ell(di), \overline{L}))a_{t-1}^i - a_t^i + F_L(\int_i a_{t-1}^i \ell(di), \overline{L}) y_t^i - T_t. \end{aligned} \tag{50}$$

FOC with respect to a_t^i. Computing the derivative of (50) with respect to a_t^i yields:

$$\begin{aligned} 0 = {} & \left(u'(c_t^i) - (\lambda_t^i - \lambda_{t-1}^i (1 + r_t)) u''(c_t^i) \right) \frac{\partial c_t^i}{\partial a_t^i} \\ & + \beta \mathbb{E}_t \left[\int_j \left(u'(c_{t+1}^j) - (\lambda_{t+1}^j - \lambda_t^j (1 + r_{t+1})) u''(c_{t+1}^j) \right) \frac{\partial c_{t+1}^j}{\partial a_t^i} \ell(dj) \right] \\ & + \beta \mathbb{E}_t \left[\int_j \lambda_t^j F_{KK,t} u'(c_{t+1}^j) \ell(dj) \right], \end{aligned}$$

where $F_{KK,t} = F_{KK}(K_t, \overline{L})$, and similarly for $F_{KL,t}$. Observe that we have:

$$\frac{\partial c_t^i}{\partial a_t^i} \frac{\partial c_{t+1}^j}{\partial a_t^i} = -1, \text{ and } = (1 + r_{t+1}) 1_{i=j} + a_t^j F_{KK,t} + F_{KL,t} y_{t+1}^j,$$

which yields with the definition (17) of ψ_t^j:

$$\begin{aligned} \psi_t^i = {} & \beta \mathbb{E}_t \left[(1 + r_{t+1}) \psi_{t+1}^i \right] + \beta \mathbb{E}_t \left[\int_j \psi_{t+1}^j \left(F_{KK,t} a_t^j + F_{KL,t} y_{t+1}^j \right) \ell(dj) \right] \\ & + \beta \mathbb{E}_t \left[\int_j \lambda_t^j F_{KK,t} u'(c_{t+1}^j) \ell(dj) \right], \end{aligned}$$

for any unconstrained agent i. For a constrained agent, we have $\lambda_t^i = 0$.

FOC withe respect to T_t. Computing the derivative of (50) with respect to T_t yields:

$$0 = v'(T_t) + \int_i \left(u'(c_t^i) - (\lambda_t^i - \lambda_{t-1}^i(1+r_t))u''(c_t^i)\right) \frac{\partial c_t^i}{\partial T_t} \ell(di).$$

Using the definition (17) of ψ_t^i and $\frac{\partial c_t^i}{\partial T_t} = -1$, we obtain:

$$v'(T_t) = \int_i \psi_t^i \ell(di).$$

B Projected model

B.1 Projected program and FOCs

The projected program can be written as:

$$\max_{((a_{t,y^N}, c_{t,y^N},)_{y^N \in \mathcal{Y}^N}, w_t, r_t, T_t)_{t \geq 0}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \sum_{y^N \in \mathcal{Y}} \Big[S_{y^N} \Big(U(c_{t,y^N}) - \left(\lambda_{t,y^N} - \tilde{\lambda}_{t,y^N}(1+r_t)\right) \xi_{y^N}^1 U_c(c_{t,y^N}) \Big) \Big] + v(T_t),$$

subject to:

$$\tilde{\lambda}_{t,y^N} = \frac{\sum_{\tilde{y}^N \in \mathcal{Y}^N} S_{t-1,\tilde{y}^N} \lambda_{t-1,\tilde{y}^N} \Pi_{t,\tilde{y}^N,y^N}}{S_{t,y^N}}, \tag{51}$$

$$c_{t,y^N} + a_{t,y^N} = w_t y_0^N + (1+r_t)\,\tilde{a}_{t,y^N} - T_t, \tag{52}$$

$$a_{t,y^N} \geq 0 \text{ and } \tilde{a}_{t,y^N} = \sum_{\tilde{y}^N \in \mathcal{Y}^N} \Pi_{\tilde{y}^N y^N,t} \frac{S_{\tilde{y}^N}}{S_{y^N}} a_{t-1,\tilde{y}^N}, \tag{53}$$

$$K_t = \sum_{y^N \in \mathcal{Y}} S_{y^N} a_{t,y^N}, \tag{54}$$

$$r_t = F_K(K_{t-1}, \overline{L}),\ \ w_t = F_L(K_{t-1}, \overline{L}). \tag{55}$$

Equivalently, it can also be written as:

$$\max_{((a_{t,y^N}, c_{t,y^N},)_{y^N \in \mathcal{Y}^N}, T_t)_{t \geq 0}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \sum_{y^N \in \mathcal{Y}} \Big[S_{y^N} \Big(U(c_{t,y^N}) - \left(\lambda_{t,y^N} - \tilde{\lambda}_{t,y^N}(1+F_K(K_{t-1}, \overline{L}))\right) \xi_{y^N}^1 U_c(c_{t,y^N}) \Big) \Big] + v\left(T_t\right),$$

with:

$$c_{t,y^N} = F_L(K_{t-1}, L)y_{y^N} + (1 + F_K(K_{t-1}, L)) \sum_{\tilde{y}^N \in \mathcal{Y}^N} \Pi_{\tilde{y}^N y^N,t} \frac{S_{\tilde{y}^N}}{S_{y^N}} a_{t-1,\tilde{y}^N} - a_{t,y^N} - T_t, \quad (56)$$

$$K_t = \sum_{y^N \in \mathcal{Y}} S_{y^N} a_{t,y^N}. \quad (57)$$

The FOC with respect to the saving a_{t,y^N} is, for an unconstrained history:

$$\psi_{t,y_N} = \beta \mathbb{E} \sum_{\tilde{y}^N \in \mathcal{Y}} \psi_{t+1,\tilde{y}_N} \left[(1 + r_{t+1})\Pi_{y^N \tilde{y}^N} + S_{\tilde{y}^N} \left(\tilde{a}_{t+1,\tilde{y}^N} F_{KK}(K_t, \overline{L}) + F_{LK}(K_t, \overline{L}\grave{a} y_{\tilde{y}^N}\right)\right],$$
$$+ \beta F_{KK}(K_t, \overline{L}) \mathbb{E} \sum_{\tilde{y}^N \in \mathcal{Y}} S_{\tilde{y}^N} \tilde{\lambda}_{t+1,\tilde{y}^N} \xi_{\tilde{y}^N} U_c(c_{t+1,\tilde{y}^N}),$$

where ψ_{t,y_N} is defined in (40).

The FOC with respect to the tax T_t is:

$$\sum_{y^N \in \mathcal{Y}} S_{y^N} \psi_{t,y_N} = v'(T_t).$$

B.2 Summary of the dynamic model

The full dynamics of the truncated model can be written as follows:

$$v'(T_t) = \sum_{k=1}^{N_{tot}} S_k \psi_{k,t}, \tag{58}$$

$$K_t = \sum_{k=1}^{N_{tot}} S_k a_{k,t}, \tag{59}$$

$$k = 1...N_{tot}: \ \tilde{a}_{k,t} = \frac{1}{S_k} \sum_{k'=1}^{N_{tot}} S_{k'} \Pi_{k',k} a_{k',t-1}, \tag{60}$$

$$c_{k,t} + a_{k,t} = (1 + r_t)\tilde{a}_{k,t} + w_t y_0, \tag{61}$$

$$\tilde{\lambda}_{k,t} = \frac{1}{S_{k,t}} \sum_{k'=1}^{N_{tot}} S_{k'} \Pi_{k',k} \lambda_{k',t-1}, \tag{62}$$

$$\psi_{k,t} = \xi_k^1 u'(c_{k,t}) - (\lambda_{t,k} - \tilde{\lambda}_{t,k}(1 + r_t))\xi_k^1 u''(c_{k,t}). \tag{63}$$

$$k \notin \mathcal{C}: \ \xi_k u'(c_{k,t}) = \beta \mathbb{E}_t \Big[(1 + r_{t+1}) \sum_{k'=1}^{N_{tot}} \Pi_{k,k'} \xi_{k'} u'(c_{k',t+1}) \Big], \tag{64}$$

$$\begin{aligned} \psi_{k,t} = {} & \beta \mathbb{E}_t \left[(1 + r_{t+1}) \sum_{k'=1}^{N_{tot}} \Pi_{k,k'} \psi_{k',t+1,\tilde{y}^N} \right] \\ & + \beta \mathbb{E}_t \left[\sum_{k'=1}^{N_{tot}} S_{k'} \psi_{k',t+1} \left(\tilde{a}_{k',t+1} F_{KK}(K_t, L) + y_{k'} F_{LK}(K_t, L) \right) \right], \\ & + \beta F_{KK}(K_t, L) \mathbb{E}_t \left[\sum_{k'=1}^{N_{tot}} S_{k'} \tilde{\lambda}_{k',t+1} \xi_{k'} u'(c_{k',t+1}) \right], \end{aligned} \tag{65}$$

$$k \in \mathcal{C}, \ a_{k,t} = \lambda_{k,t} = 0. \tag{66}$$

C Matrix representation

A very convenient way to express the truncated model allocation involves using matrix notation. This notation is very powerful to compute optimal policies at the steady state, as shown below. We define the following elements:

- $\boldsymbol{S} = (S_k)_{k=1...N_{tot}}$ the N_{tot}-vector of sizes;
- $\boldsymbol{\Pi} = (\Pi_{kk'})_{k,k'=1...N_{tot}}$ the transpose of the $N_{tot} \times N_{tot}$ matrix of transition probabilities across histories;
- $\boldsymbol{c} = (c_k)_{k=1...N_{tot}}$, $\boldsymbol{a} = (a_k)_{k=1...N_{tot}}$, $\boldsymbol{\tilde{a}} = (\tilde{a}_k)_{k=1...N_{tot}}$, the N_{tot}-vectors of allocations (consumption, end-of-period and beginning-of-period savings);
- $\boldsymbol{y_0} = (y_{k,0})_{k=1...N_{tot}}$ is the vector of current productivity levels across histories;

- $\boldsymbol{\xi} = (\xi_k)_{k=1\ldots N_{tot}}$ the vector of residual-heterogeneity parameters; their derivation is provided below;

- $\mathcal{C}$ is the set of indices of credit-constrained history;

- $\boldsymbol{P} = \text{diag}((p_k)_{k=1\ldots N_{tot}})$, with $p_k = 1$ if $k \notin \mathcal{C}$ and $p_k = 0$ if $k \in \mathcal{C}$, is the diagonal $N_{tot} \times N_{tot}$-matrix; The matrix $\boldsymbol{P}$ selects the non-constrained histories;

- $\boldsymbol{D}_{\boldsymbol{u'(c)}} = \text{diag}((u'(c_k))_{k=1\ldots N_{tot}})$ is the diagonal $N_{tot} \times N_{tot}$-matrix with $u'(c_k)$ on the diagonal for history k, and 0 elsewhere;

- $\boldsymbol{I}$ is the identity matrix;

- $\mathbf{1}_{N_{tot}}$ is the N_{tot}-vector of 1.

We also introduce the following operations:

- $\circ$ the term-by-term product of two vectors of the same size, which is another vector of the same size: $\boldsymbol{x} \circ \boldsymbol{z} = (x_{y^N}) \circ (z_{y^N}) = (x_{y^N} z_{y^N})$;[19]

- $\times$ the usual matrix product: e.g., for a matrix $\boldsymbol{M}$ and a vector $\boldsymbol{x}$ (of length equal to the number of columns of $\boldsymbol{M}$), $\boldsymbol{M} \times \boldsymbol{x}$ is the vector $(\sum_{k'} M_{kk'} x_{k'})_k$.

We still denote without a sign the usual scalar multiplication – that is assumed to apply to matrix and vectors (e.g., $\lambda \boldsymbol{M} = (\lambda M_{kk'})_{k,k'}$) and with $+$ the addition – that is extended to matrices and vectors of same size (e.g., $\boldsymbol{x} + \boldsymbol{z} = (x_k + z_k)_k$). We also keep the same notation for functions that apply element-wise to vectors: $f(\boldsymbol{x}) = (f(x_k))_k$.

We can rewrite equations characterizing the steady-state of the truncated economy using this notation and explain how to construct the vector $\boldsymbol{\xi}$. We start with equation (28):

$$\boldsymbol{S} = \boldsymbol{\Pi}^\top \times \boldsymbol{S}, \tag{67}$$

which makes it clear that the vector of sizes, $\boldsymbol{S}$, is the eigenvector of matrix $\boldsymbol{\Pi}^\top$ associated to the eigenvalue 1, where the sum of the eigenvector coordinates is normalized to 1.[20] The vector $\boldsymbol{S}$ is thus straightforward to compute.

Second, equation (35) for per-capita beginning-of-period wealth $\tilde{\boldsymbol{a}}$, which yields:

$$\tilde{\boldsymbol{a}} = (1/\boldsymbol{S}) \circ (\boldsymbol{\Pi}^\top \times (\boldsymbol{S} \circ \boldsymbol{a})), \tag{68}$$

where $1/\boldsymbol{S} = (1/S_k)_k$ is the vector of size inverses and $\boldsymbol{a}$ is the given vector of end-of-period wealth. Note that if the size of truncated history is $S_k = 0$, we can set $1/S_k = 0$, which is with a null-size history getting a null wealth.

[19] This operation is also known as the Hadamard product.

[20] The existence of a positive eigenvector vector is guaranteed by the Perron-Frobenius theorem for the positive matrix whose rows sum to 1.

Third, the budget constraint (36) becomes at the steady state, where the tax is T_∞:

$$\boldsymbol{c} + \boldsymbol{a} = (1+r)\tilde{\boldsymbol{a}} + w\boldsymbol{y_0} - T_\infty \mathbf{1}_{N_{tot}}, \tag{69}$$

which allows one to obtain consumption levels using the given vector $\boldsymbol{a}$ of end-of-period wealth, and the vector of beginning-of-period wealth of equation (68).

We define the matrix $\boldsymbol{\Pi}^\lambda$:

$$\Pi^\lambda_{k,k'} = S_{k'}\Pi_{k',k}\frac{1}{S_k}, \tag{70}$$

or equivalently, $\boldsymbol{S} \circ (\boldsymbol{\Pi}^\lambda \boldsymbol{x}) = \boldsymbol{\Pi}^\top(\boldsymbol{S} \circ \boldsymbol{x})$ for any vector $\boldsymbol{x} \in \mathbb{R}^{N_{tot}}$. We denote by $\boldsymbol{\lambda}$, $\tilde{\boldsymbol{\lambda}}$, and $\boldsymbol{\psi}$ the vectors corresponding to $(\lambda_k)_k$, $(\tilde{\lambda}_k)_k$, and $(\psi_k)_k$, respectively.

Using (70), the definitions (39) and (40) of $\tilde{\lambda}_k$ and ψ_k imply in matrix form:

$$\tilde{\boldsymbol{\lambda}} = \boldsymbol{\Pi}^\lambda \boldsymbol{\lambda}, \tag{71}$$

$$\boldsymbol{\psi} = \boldsymbol{\xi} \circ u'(\boldsymbol{c}) - \boldsymbol{D}_{\boldsymbol{\xi}\circ u'(\boldsymbol{c})}(\boldsymbol{I} - (1+r)\boldsymbol{\Pi}^\lambda)\boldsymbol{\lambda}. \tag{72}$$

We start with expressing the first-order condition (18) with respect to saving choices, which only holds for unconstrained truncated histories:

$$\begin{aligned} \boldsymbol{P}\boldsymbol{\psi} = {} & \beta(1+r)\boldsymbol{P}\boldsymbol{\Pi}\boldsymbol{\psi} + \beta\boldsymbol{P}1_{N_{tot}}(\boldsymbol{S} \circ (F_{KK}(K,L)\tilde{\boldsymbol{a}} + F_{LK}(K,L)\boldsymbol{y})^\top\boldsymbol{\psi} \\ & + \beta F_{KK}(K,L)\boldsymbol{P}1_{N_{tot}}\left(\boldsymbol{S} \circ \boldsymbol{\xi} \circ u'(\boldsymbol{c})\right)^\top \boldsymbol{\Pi}^\lambda\boldsymbol{\lambda}, \end{aligned} \tag{73}$$

where it should be observed that $1_{N_{tot}}(\boldsymbol{S}\circ(F_{KK}(K,L)\tilde{\boldsymbol{a}}+F_{LK}(K,L)\boldsymbol{y})^\top$ and $1_{N_{tot}}\left(\boldsymbol{S} \circ \boldsymbol{\xi} \circ u'(\boldsymbol{c})\right)^\top$ are $N_{tot}\times N_{tot}$-matrices (as product of $N_{tot}\times 1$ and $1\times N_{tot}$vectors, and where all rows are identical to row vectors $(\boldsymbol{S} \circ (F_{KK}(K,L)\tilde{\boldsymbol{a}} + F_{LK}(K,L)\boldsymbol{y})^\top$ and $(\boldsymbol{S} \circ \boldsymbol{\xi} \circ u'(\boldsymbol{c}))^\top$, respectively). The pre-multiplication in (73) by the matrix $\boldsymbol{P}$ comes from FOC (18) holding only for unconstrained histories. We define the two following matrices:

$$\boldsymbol{L}_0 = \boldsymbol{I} - \beta(1+r)\boldsymbol{\Pi} - \beta 1_{N_{tot}}(\boldsymbol{S} \circ (F_{KK}(K,L)\tilde{\boldsymbol{a}} + F_{LK}(K,L)\boldsymbol{y}), \tag{74}$$

$$\boldsymbol{L}_1 = \beta F_{KK}(K,L)1_{N_{tot}}\left(\boldsymbol{S} \circ \boldsymbol{\xi} \circ u'(\boldsymbol{c})\right)^\top \boldsymbol{\Pi}^\lambda, \tag{75}$$

such that first-order condition (18) becomes:

$$\boldsymbol{L}_0\boldsymbol{\psi} = \boldsymbol{L}_1\boldsymbol{\lambda}. \tag{76}$$

Using (72) to express $\boldsymbol{\psi}$ using $\boldsymbol{\lambda}$, we deduce:

$$\boldsymbol{P}(\boldsymbol{L}_1 + \boldsymbol{D}_{\boldsymbol{\xi}\circ u'(\boldsymbol{c})}(\boldsymbol{I} - (1+r)\boldsymbol{\Pi}^\lambda))\boldsymbol{\lambda} = \boldsymbol{P}\boldsymbol{L}_0(\boldsymbol{\xi} \circ u'(\boldsymbol{c})). \tag{77}$$

For constrained histories, we simply have $\lambda_{y^N} = 0$, or equivalently using matrix notation:

$$(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\lambda} = 0. \tag{78}$$

Adding equations (77) and (78) yields the important result:

$$\boldsymbol{\lambda} = \boldsymbol{L}_2^{-1}\boldsymbol{P}\boldsymbol{L}_0(\boldsymbol{\xi} \circ u'(\boldsymbol{c})), \tag{79}$$

$$\text{with: } \boldsymbol{L}_2 = \boldsymbol{I} - \boldsymbol{P} + \boldsymbol{P}(\boldsymbol{L}_1 + \boldsymbol{D}_{\boldsymbol{\xi}\circ u'(\boldsymbol{c})}(\boldsymbol{I} - (1+r)\boldsymbol{\Pi}^{\lambda})), \tag{80}$$

Equation (79) provides a closed-form expression for the vector $\boldsymbol{\lambda}$ as a function of steady-state allocations – through matrices $\boldsymbol{L}_0$, $\boldsymbol{L}_1$, and $\boldsymbol{L}_2$ (which only depends on the allocation) of equations (74), (75), and (80). The matrix $\boldsymbol{L}_2$ is invertible when $r > 0$.

Finally, we deduce from (76) and (78):

$$\boldsymbol{\lambda} = \boldsymbol{M_1}(\boldsymbol{\xi}^1 \circ u'(\boldsymbol{c})),$$

$$\boldsymbol{\psi} = \boldsymbol{M_2}(\boldsymbol{\xi}^2 \circ u'(\boldsymbol{c})),$$

$$\text{where: } \boldsymbol{M_1} := \boldsymbol{L}_2^{-1}\boldsymbol{P}\boldsymbol{L}_0, \tag{81}$$

$$\boldsymbol{M_2} := \boldsymbol{I} - \boldsymbol{D}_{\boldsymbol{\xi}\circ u'(\boldsymbol{c})}(\boldsymbol{I} - (1+r)\boldsymbol{\Pi}^{\lambda})\boldsymbol{M_1}. \tag{82}$$

D Robustness check

We summarize below the results for an economy with a different value for the curvature of the public good ($\theta = 65\%$). In this economy, the optimal value of the tax as a share of GDP is 15%, as computed by the truncated method. The optimal value of the tax-to-GDP computed with the transition method is 14.45%. Again, if we consider different initial distributions, we will end up with different values for the optimal tax computed with the transition method. We plot in Figure 5 the welfare as a function of the tax-to-GDP.

We also plot in Figure 6 the tax path in the projected economy (after a pure reoptimization shock), and as in the baseline calibration, the difference in the two tax rates is due to time-inconsistency.

E A stylized model to illustrate time-inconsistency

In this section, we provide a toy model to illustrate that only two ingredients are needed to generate time-inconsistency: (i) agents' heterogeneity that cannot be reduced by planner's instruments and (ii) agents choices have externalities.

We consider a two-period production economy populated by two agents, denoted by A and B. In the first period, both agents have zero initial wealth.[21] They supply one unit of labor with the hourly wage w_0. Their productivity levels, denoted by θ_A and θ_B, are heterogeneous. In the second period, agents still supply one unit of labor, but they have the same productivity θ. The

[21] This model is independent of the main text and mostly serves an illustration purpose for time-inconsistency. As such, notation is completely independent of the one in the main text.

first-period production function is specified such that the first period hourly wage is $w_0 = 1$. The second-period production function is of the Cobb-Douglas-type requiring labor and capital. Capital needs to be installed one period in advance. The output Y_1 in period 1 produced from capital K_0 and labor supply L is:

$$Y_1 = K_0^{\alpha} L_1^{1-\alpha} - \delta K_0,$$

where $\alpha \in (0,1)$ is the capital share and $\delta > 0$ the capital depreciation rate. Firm's profit maximization implies the following factor prices:

$$w_1 = (1-\alpha) K_0^{\alpha} L_1^{-\alpha}, \tag{83}$$
$$r_1 = \alpha K_0^{\alpha-1} L_1^{1-\alpha} - \delta, \tag{84}$$
$$w_0 = 1.$$

Agents derive utility from consumption of a good in both periods, as well as from using a public good that is valued once in the second period. From the consumption of a bundle (c_0, c_1, G), representing respectively consumption in the first and second period and public good supply, the agent is assumed to enjoy utility $u(c_0) + u(c_1) + \frac{1}{2}u(G)$, where $u : \mathbb{R}_+ \to \mathbb{R}$ is strictly increasing, strictly concave with $u'(0) = \infty$. Therefore, the agent's utility is separable in goods and in time, features a discount factor equal to 1, and the period utility function is the same. The one-half weight for the public good utility is only a simplification trick, but does not drive our result. Agents can save through capital shares and do not face any credit constraint.[22]

Finally, the public good is financed out of lump-sum tax raised in the second period by a benevolent government. There is no public debt and the government budget must remain balanced:

$$G = 2T, \tag{85}$$

where T is the lump-sum tax raised on each agent.

Finally, the program of agent $i = A, B$ can be written as:

$$\max_{a_i \in \mathbb{R}} u(c_{0,i}) + u() + \frac{1}{2}u(G),$$
$$\text{s.t. } c_{0,i} = w_0 \theta_i - a_i, \tag{86}$$
$$c_{1,i} = w_1 \theta + (1 + r_1) a_i - T, \tag{87}$$

where a_i is the period-0 saving choice of agent i and $c_{t,i}$ its consumption in period $t = 0, 1$.

[22] Adding credit constraints would not change our results. They are a financial imperfection that generates ex-post heterogeneity, which is not needed here, since the model already features ex-ante heterogeneity.

Solving agent's program implies the following Euler equation:

$$u'(\theta_i - a_i) = (1 + r_1)u'(w_1\theta + (1 + r_1)a_i - T) \tag{88}$$

Finally market clearing conditions imply:

$$a_A + a_B = K_0, \tag{89}$$

$$2\theta = L_1, \tag{90}$$

$$c_{0,A} + c_{0,B} = L_1 - K_0, \tag{91}$$

$$c_{1,A} + c_{1,B} = F(K_0, L_1) + K_0 - G, \tag{92}$$

Equation (89) and (90) are the clearing conditions for capital market (at $t = 0$) and labor market (at $t = 1$), respectively. Given the structure of our economy there is no capital market clearing condition at $t = 1$ since agents do not save and the labor market clearing condition of $t = 0$ is actually embedded within the normalization $w_0 = 1$. Finally, equations (91) and (92) are date-0 and date-1 resource constraints.

The planner will be assumed to adopt a utilitarian welfare criterion, where the aggregate welfare W_t at date t is the sum of individual welfare. Formally, W_0 and W_1 can be expressed as:

$$W_0 = u(c_{0,A}) + u(c_{1,A}) + u(c_{0,B}) + u(c_{1,B}) + u(G), \tag{93}$$

$$W_1 = u(c_{1,A}) + u(c_{1,B}) + u(G), \tag{94}$$

since the public good is enjoyed in the second period only.

E.1 First best

We start with characterizing the first-best, where the planner allocates resources among agents, subject to the sole period resource constraints. We check the absence of time-inconsistency by solving the planner's program both at dates 0 and 1.

Date-0 program. The planner decides at date 0 the allocation of private consumption and public consumption si as to maximize aggregate welfare subject to the two period resource constraints. The aggregate welfare criterion is given in equation (93) and the first-best program can be written as follows:

$$\begin{aligned}
\max_{c_{0,A},c_{1,A},c_{0,B},c_{1,B},G} \; & u(c_{0,A}) + u(c_{1,A}) + u(c_{0,B}) + u(c_{1,B}) + u(G), \\
\text{s.t.} \; & c_{0,A} + c_{0,B} = L_1 - K_0, \\
& c_{1,A} + c_{1,B} = F(K_0, L_1) + K_0 - G.
\end{aligned}$$

Denoting by μ_0 and $\mu_1$the Lagrange multipliers on date-0 and date-1 resource constraints and with a FB superscript first-best allocation, we obtain the following FOCs:

$$\begin{aligned}
u'(c_{0,A}^{FB}) &= u'(c_{0,B}^{Fb}) = \mu_0,\\
u'(c_{1,A}^{FB}) &= u'(c_{1,B}^{FB}) = u'(G^{FB}) = \mu_1,\\
\mu_0 &= \mu_1(1 + r_1^{FB}),
\end{aligned}$$

which implies:

$$u'(c_{0,A}^{FB}) = u'(c_{0,B}^{FB}) = (1 + r_1^{FB})u'(c_{1,A}^{FB}) = (1 + r_1^{FB})u'(c_{1,B}^{FB}) = (1 + r_1^{FB})u'(G^{FB}). \tag{95}$$

We deduce that the first-best allocation is characterized as a function of the capital level K_0:

$$c_{0,A}^{FB} = c_{0,B}^{FB} = \frac{1}{2}(L_1 - K_0^{FB}), \tag{96}$$

$$c_{1,A}^{FB} = c_{1,B}^{FB} = G^{FB} = \frac{1}{3}(F(K_0^{FB}, L_1) + K_0^{FB}), \tag{97}$$

while the capital level K_0^{FB} is determined from $u'(c_{0,A}^{FB}) = (1 + r_1^{FB})u'(c_{1,A}^{FB})$, or as the solution of:

$$u'(\frac{1}{2}(L_1 - K_0^{FB})) = (1 + F_K(K_0^{FB}, L_1))u'(\frac{1}{3}(F(K_0^{FB}, L_1) + K_0^{FB})). \tag{98}$$

In particular, the first-best allocation features perfect equality between the two agents in the two periods. This perfect smoothing is independent of the presence of agent's heterogeneity in productivity.

Date-1 program. We check that there is no time-inconsistency issue in the first-best. Let us assume that the planner can (unexpectedly) break its commitments at date 1. The planner then re-optimizes the date-1 aggregate welfare of equation (94) subject to the date-1 resource constraint. Denoting with a tilde date-1 choices, the formal program can be written as follows:

$$\begin{aligned}
\max_{\tilde{c}_{1,A}, \tilde{c}_{1,B}, \widetilde{G}} \; & u(\tilde{c}_{1,A}) + u(\tilde{c}_{1,B}) + u(\widetilde{G})\\
\text{s.t. } & \tilde{c}_{1,A} + \tilde{c}_{1,B} = F(K_0, \overline{L}) + K_0 - \widetilde{G},
\end{aligned}$$

where K_0^{FB} is given by equation (98). Note that the capital level K_0^{FB} results from date-0 saving choices and thus cannot be modified at date 1. We obtain the following FOCs: $u'(\tilde{c}_{1,A}^{FB}) = u'(\tilde{c}_{1,B}^{FB}) = u'(\widetilde{G}^{FB})$, which together with the resource constraint implies the allocation

$$\tilde{c}_{1,A}^{FB} = \tilde{c}_{1,B}^{FB} = \widetilde{G}^{FB,1} = \frac{1}{3}(F(K_0^{FB}, L_1) + K_0^{FB}),$$

which is exactly the same as equation (97) resulting from date-0 program. Choices of date 1 coincide with those of date 0: there is therefore no time-inconsistency in the first-best allocation.

E.2 Ramsey program

E.2.1 Date-0 program

Formulating the Ramsey program. We now investigate the Ramsey program at date 0. The planner maximizes the aggregate welfare among possible competitive equilibria. In other words, the planner chooses the fiscal policy (consisting of public spending G and lump-sum tax T) maximizing aggregate welfare subject to: (i) individual budget constraints (86) and (87), (ii) individual Euler conditions (88), (iii) factor price definitions (83) and (84), (iv) market clearing conditions (89) and (90), and (v) government budget constraint (85). After the proper substitution, the date-0 Ramsey program can be written as follows:

$$
\begin{aligned}
\max_{a_A,a_B,T}\ & u(2T) + \sum_{i=A,B} u(\theta_i - a_i) + u(w_1\theta + (1+r_1)a_i - T) \\
\text{s.t. }\ & u'(\theta_i - a_i) = (1+r_1)u'(w_1\theta + (1+r_1)a_i - T),\ i = A, B, \\
& w_1 = (1-\alpha)(a_A + a_B)^{\alpha} L_1^{-\alpha}, \\
& r_1 = \alpha(a_A + a_B)^{\alpha-1} L_1^{1-\alpha} - \delta,
\end{aligned}
\tag{99}
$$

where we recall that $L_1 = \theta_A + \theta_B$ is independent of planner's choices.

The Ramsey allocation will differ from the first-best. Before solving for the Ramsey allocation, Euler conditions (99) make it clear that since u' is strictly decreasing, $a_A > a_B$ iff $\theta_A > \theta_B$. This simply reflects the fact that more productive agents earn a larger wage and can thus save more. This directly implies that $c_{1,A} > c_{1,B}$ iff $\theta_A > \theta_B$, and from Euler equations that $c_{0,A} > c_{0,B}$ iff $\theta_A > \theta_B$. Unsurprisingly, the more productive agent saves more and consumes more in both period than the less productive agent.

Another consequence of this remark is that, when $\theta_A > \theta_B$, the Ramsey allocation cannot feature an equal consumption between agents in each period, as it is the case in the first-best. In other words, the Ramsey allocation cannot restore the first-best and offsetting the consequences of agents' heterogeneity is not possible. The presence of ex-post heterogeneity (here due to the combination of ex-ante heterogeneity and a limited set of planner's instruments) presents the planner from reproducing the first-best through the Ramsey program.

Solving for the Ramsey allocation. Denoting by λ_i the Lagrange multiplier on (99), we obtain the following FOCs for the planner's program. First, with respect to lump sum tax T, we have:

$$
2u'(2T) = \sum_{i=A,B} (u'(c_{1,i}) - \lambda_i(1+r_1)u''(c_{1,i})). \tag{100}
$$

Second, with respect to saving choice a_i:

$$u'(c_{0,i}) = (1+r_1)u'(c_{1,i}) + \sum_{j=A,B} \left(\frac{\partial w_1}{\partial a_j}\theta + \frac{\partial r_1}{\partial a_j} a_j \right) \left(u'(c_{1,j}) - (1+r_1)\lambda_j u''(c_{1,j})\right) \qquad (101)$$
$$- \lambda_i \left((1+r_1)^2 u''(c_{1,i}) + u''(c_{0,i}) \right) - \sum_{j=A,B} \lambda_j \frac{\partial r_1}{\partial a_j} u'(c_{1,j})$$

Observe that from factor price definitions (83) and (84), we deduce:

$$\frac{\partial w_1}{\partial a_j} = \alpha(1-\alpha)K_0^{\alpha-1}L_1^{-\alpha},$$
$$\frac{\partial r_1}{\partial a_j} = -\alpha(1-\alpha)K_0^{\alpha-2}L_1^{1-\alpha},$$

which implies:

$$\frac{\partial w_1}{\partial a_j}\theta + \frac{\partial r_1}{\partial a_j} a_j = (1-\alpha)\alpha K_0^{\alpha-2}L_1^{-\alpha} \left(a_A + a_B - 2a_j\right)\theta$$

Using Euler equation $u'(c_{0,i}) = (1+r_1)u'(c_{1,i})$, we obtain:

$$- \frac{(1+r_1)^2 u''(c_{1,i}) + u''(c_{0,i})}{(1-\alpha)\alpha K_0^{\alpha-2}L_1^{-\alpha}} \lambda_i \qquad (102)$$
$$= -\theta(a_A - a_B)(u'(c_{1,B}) - u'(c_{1,A}))$$
$$+ \sum_{j=A,B} \lambda_j \left((1+r_1)u''(c_{1,j}) \left(a_A + a_B - 2a_j\right)\theta - u'(c_{1,j}) \right),$$

which is a linear system that characterizes λ_A and λ_B, as a function of saving choices a_A and a_B.

We can observe that if $\theta_A > \theta_B$, we have $(a_A - a_B)(u'(c_{1,B}) - u'(c_{1,A})) > 0$ and the solution of (102) cannot feature $\lambda_A = \lambda_B = 0$. This means that the FOC (100) characterizing the fiscal policy differs from the one in the first-best case.

No heterogeneity case. In the absence of ex-ante heterogeneity (i.e., when $\theta_A = \theta_B$), Euler conditions (99) imply the absence of ex-post heterogeneity: $a_A = a_B$ and $c_{t,A} = c_{t,B}$ for $t = 0, 1$. The Ramsey savings first-order condition (102) in turn imply that $\lambda_A = \lambda_B = 0$, while the lump-sum tax FOC simplifies into: $2u'(2T) = \sum_{i=A,B} u'(c_{1,i})$, which is identical to the first-best condition. In other words, in the absence of ex-ante heterogeneity (meaning in our context the absence of ex-post heterogeneity), the Ramsey allocation exactly replicates the first-best allocation.

E.2.2 Date-1 program

We assume that the planner can (unexpectedly) break its commitments at date 1 and that it re-optimizes at date 1. As in the first-best case, savings, interest rate, wages and capital all result from date-0 choices and cannot be modified. The only instrument the planner can still influence at date 1 is the fiscal policy. Denoting with $\widetilde{T}$ the date-1 lump-sum tax, the planner's program can be written as:

$$\begin{aligned}
&\max_{\widetilde{T}} \sum_{i=A,B} u(w_1\theta + (1+r_1)a_i - \widetilde{T}) + u(2\widetilde{T}) \\
&\text{s.t. } a_A, a_B, K_0, w_1, r_1 \text{ given,}
\end{aligned}$$

which yields to the FOC:

$$2u'(2\widetilde{T}) = u'(\tilde{c}_{1,A}) + u'(\tilde{c}_{1,B}),$$

which differs from the date-0 FOC of equation (100) because λ_A and λ_B were proved to be different from 0.

In other words, the Ramsey program feature time-inconsistency. The planner would like to change the lump-sum tax in the first period. The reason is that when setting the lump-sum tax at date 0, the planner accounts for the date-1 consequences of its choice. More precisely, the choice of T at date 0 affects agents savings, and in turn the level of capital, the wage and the interest rate in the first period. However, in date 1, prices are fixed and the externality of T on savings does not matter any more. The planner would thus like to "update" its choice of T because it doe snot need to influence agents' savings anymore.

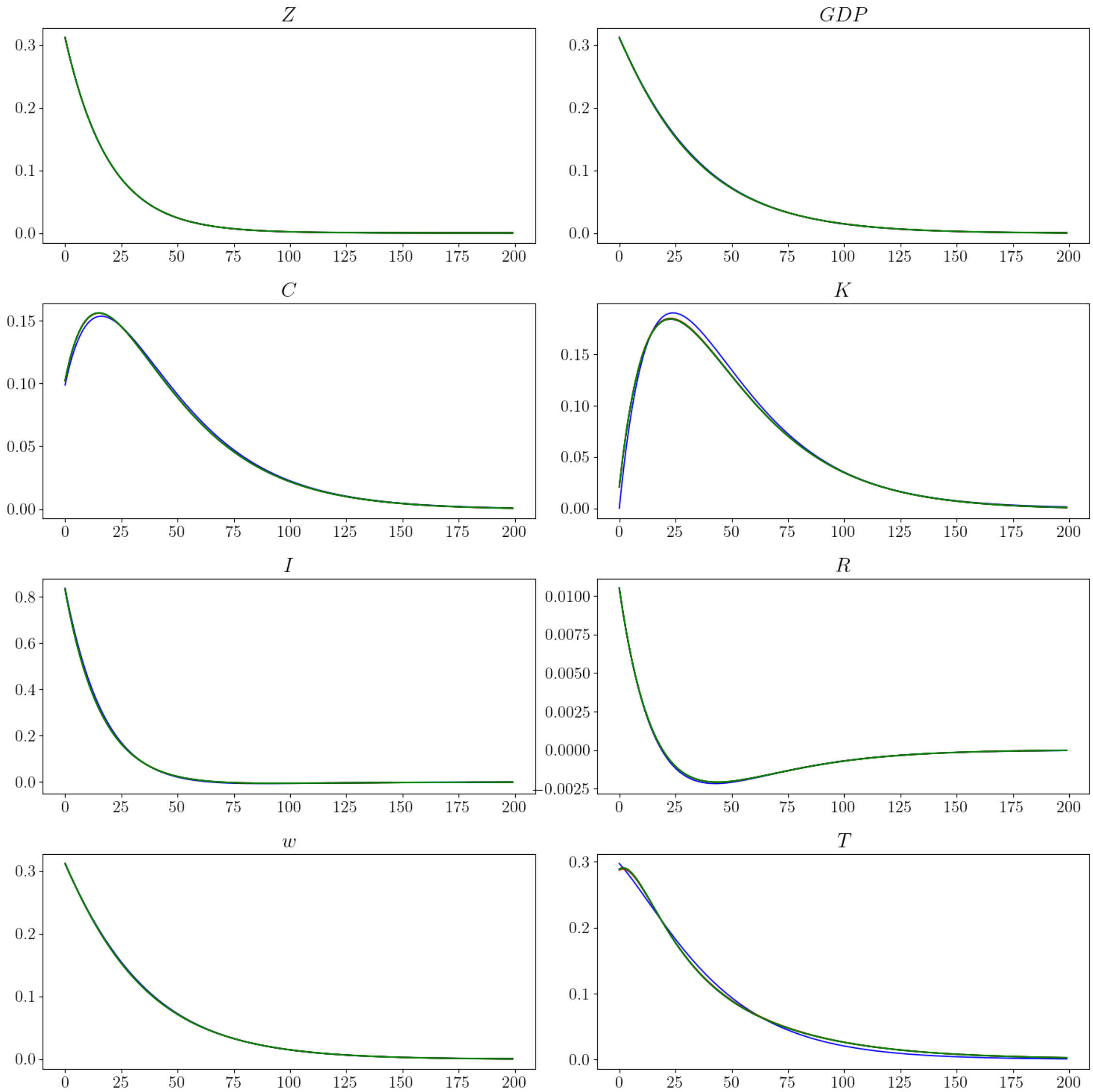

Figure 3: Simulated IRFs after a TFP shock simulated using the truncation and the Reiter methods. See the text for the details of the implementations.

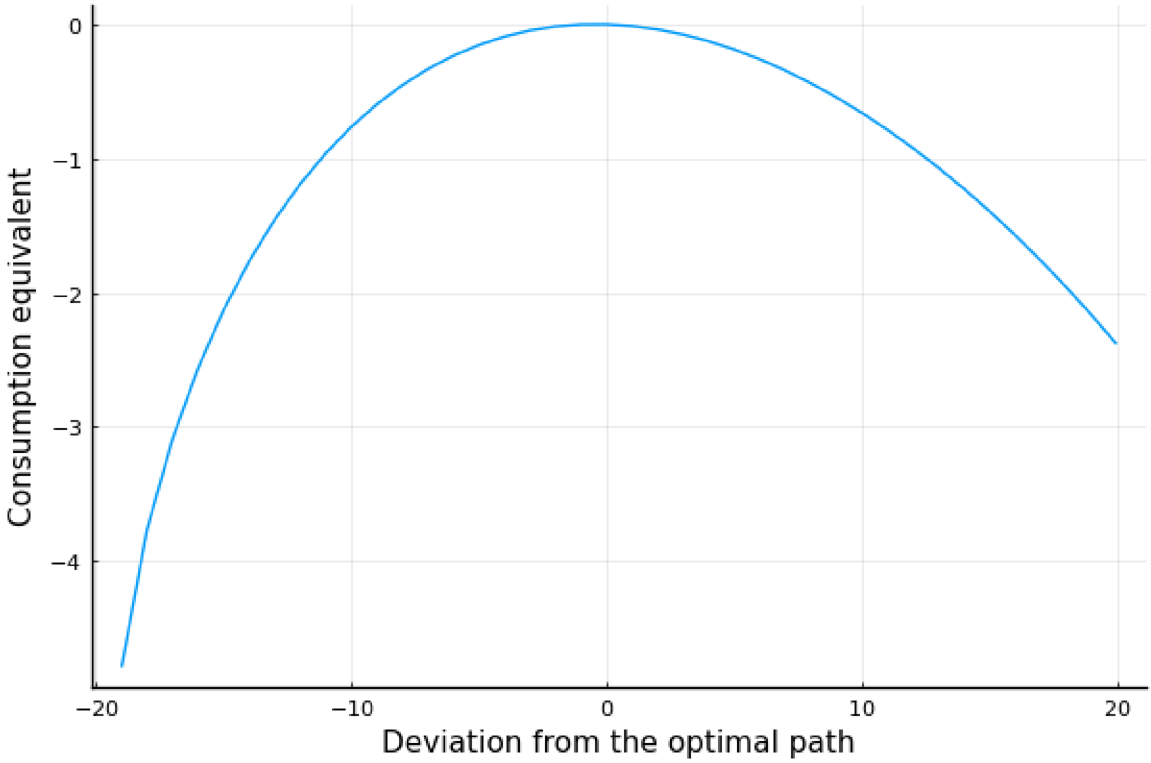

Figure 4: Dynamic of the tax after a reoptimization shock

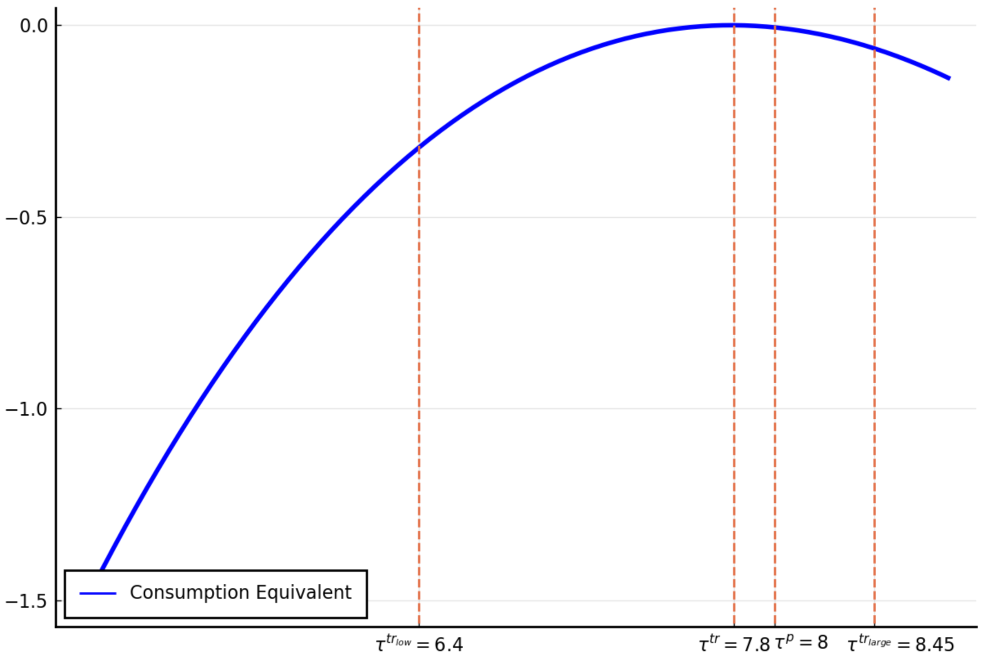

Figure 5: Welfare with transitions and consistent initial distribution computed as a function of the tax rate.

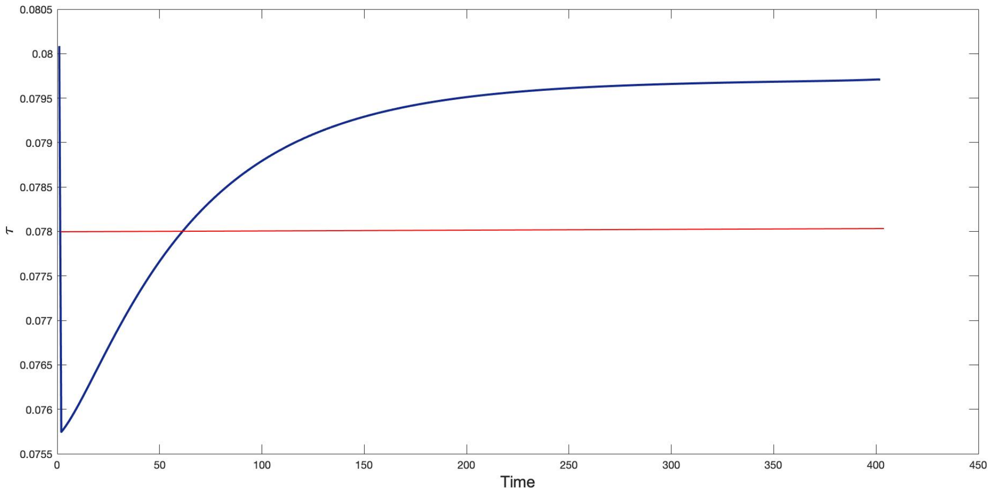

Figure 6: Dynamic of the tax after a reoptimization shock when the curvature of v is $\theta = 65\%$.