Optimal Policies with Heterogeneous Agents: Truncation and Transitions*

François Le Grand† Xavier Ragot‡

July 24, 2023

Abstract

We compare methods to solve for optimal policies in heterogeneous-agent models with and without aggregate shocks, considering the optimal provision of a public good. We first use a method based on transitions that we modify to neutralize the effect of the initial distribution. Second, we consider a truncation method combining a Lagrangian approach and an improved truncation procedure that takes advantage of the restrictions imposed by the first-order conditions of the Ramsey problem. The truncation method provides quantitatively accurate estimates of the value of the planner's instruments, whereas a time-inconsistency issue affects the transition method. We also report a number of quantitative checks, both in time-0 and timeless perspectives, and both with and without aggregate risk.

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

JEL codes: D31, D52, E21.

*This paper has benefited from the comments of many participants in seminars where the truncation method has been presented. We thank Diego Rodrigues for outstanding research assistance.

This work was supported by the French National Research Agency (grant number ANR-20-CE26-0018 IRMAC).

Codes and details about the algorithm can be found at https://github.com/RagotXavier/Truncation_Method_Het.

†François LeGrand: Rennes School of Business; 2 Rue Robert d'Arbrissel, 35065 Rennes; francois.le-grand@rennes-sb.com.

‡Xavier Ragot: Sciences Po, CNRS and OFCE; 27 Rue Saint-Guillaume, 75007 Paris; xavier.ragot@sciencespo.fr.

1 Introduction

A frontier in the heterogeneous-agent literature is the computation of optimal policies in general equilibrium. In heterogeneous-agent models (or, more precisely, in incomplete insurance market models for idiosyncratic risk), redistribution generates new trade-offs. For instance, any policy affecting agents' income modifies the incentives to save for self-insurance motives, and consequently future real wages and interest rates, which heterogeneously affects agents' welfare. A number of papers compute the optimal long-run policy as the constant values of instruments that maximize the aggregate welfare computed considering transitions (see Conesa et al., 2009; Chang et al., 2018; or Ferriere and Navarro, 2023, among many others). In a recent paper, LeGrand and Ragot (2022a) propose a method based on the first-order conditions (FOCs) of the Ramsey planner and a truncation method to compute the solution. In this paper, we compare the two methods in their ability to solve for optimal policy in a heterogeneous-agent model. We consider the provision of a public good, financed by lump-sum taxes, in a model where agents face uninsurable income shocks in a production economy à la Huggett (1993) or Aiyagari (1994). In this problem, the planner's only instrument is the tax level in each period.

The first method we consider is the so-called transition method, which improves on the method of Aiyagari and McGrattan (1998), which maximizes steady-state welfare without accounting for transitions. The truncation method does not optimize over the entire instrument path but is restricted to a constant path. Given an initial distribution, one can indeed simulate the transition to the long-run equilibrium distribution for any constant value of the instrument, and then compute the aggregate welfare of the economy along the transition. Then, by iterating over the value of that instrument, one can find the value that maximizes the aggregate welfare. Thus, the transition method computes the answer to the following question: What is the *constant* value of the instrument that maximizes aggregate welfare given the transition from the initial distribution to a long-run distribution? We will refer later to this question as *Q1*.

The second method we consider is the truncation method introduced by LeGrand and Ragot (2022a) to solve for the Ramsey problem in a heterogeneous-agent model. The truncation method consists of providing a finite state-space representation of the incomplete market model, by pooling together agents according to their idiosyncratic histories, which are truncated at a given truncation length. We track the consumption and savings of agents with the same idiosyncratic history over a given number of consecutive past periods. The model is then expressed in terms of these *truncated* idiosyncratic histories and some parameters (labeled ξs) are introduced to account for the heterogeneity within each truncated history. Indeed, at date t, agents pooled in the same truncated history of length N have the same history from dates $t-N+1$ to t, but differ in their history before period $t-N$. The ξs parameters aim to account for this

Abbreviations: FOC—first-order condition; IRF—impulse response function; PSID—Panel Study of Income Dynamics; SCF—Survey of Consumer Finances; TFP—total factor productivity.

within-history heterogeneity.[1] This construction allows one to derive the first-order conditions of the planner in this finite state-space representation and simulate the optimal time-varying path of the instrument. The transition method answers the question: What is the time-varying path of the instrument that maximizes aggregate welfare (still taking into account the transition from the initial distribution to a long-run distribution)? We will refer later to this question as *Q2*.

Our comparison exercise of the two methods has two main objectives: (i) to identify the differences between the results of the two methods, and (ii) to use the transition method to assess the accuracy of the truncation in computing the optimal instrument path. Conceptually, the two methods involve a Ramsey planner, who maximizes aggregate welfare (i.e., takes into account transitions from an initial distribution to a long-run distribution) and by committing to either a constant instrument value (transition method) or an entire instrument path (truncation method). Therefore, the constant transition tax rate (answer to Question Q1) should be compared to the average tax rate of the optimal time-varying path (answer to Q2).[2] In fact, we verify that the *average* tax rate over the truncation path and the transition tax rate are actually very close to each other. As a consequence, the transition solution does not provide an approximation to the truncation long-run (or steady-state) tax rate, because the planner generally wants the instrument to be time-varying. The initial distribution of agents affects the truncation path (especially in the early periods), and thus also the transition solution. However, the long-run truncation tax rate is independent of the initial distribution – see Açikgöz (2018); Açikgöz et al. (2022) for a formal proof, but we also verify it in our computational exercises. Thus, the transition method is useful when the constant instrument is a constraint of the environment (from an economic or a politic point of view), but it does not approximate the Ramsey steady state, which is the long-run value of the Ramsey problem with time-varying instruments.

We then check the accuracy of the truncation method itself, using the transition method. First, we neutralize the dependence of the transition method on the initial distribution. To do so, we jointly find a constant instrument value and an initial distribution, such that the aggregate welfare is maximized and the final distribution is identical to the initial distribution. Our algorithm works as follows. We start with an initial guess for the initial distribution. We then compute the transition value that maximizes aggregate welfare, as well as the corresponding long-run distribution. We use the latter to update the initial distribution, and apply the transition method again. We keep iterating on the initial distribution, until it coincides with the steady-state distribution. However, and interestingly, the transition tax rate still differs from the steady-state truncation tax rate. This comes from the time-inconsistency of Ramsey

[1] In the current paper we extend LeGrand and Ragot (2022a) and we show how this within-truncated-history heterogeneity can be accounted for, to enable welfare comparisons. See Section 4.1 below.

[2] For the sake of clarity, we will call the *transition tax rate* the constant value of the tax path computed by the transition method. We will call the *truncation (tax) path* the time-varying optimal tax path computed by the truncation method. We will also refer to the *long-run truncation tax rate* or *steady-state truncation tax rate* as the long-run value of the time-varying optimal tax path computed by the truncation method.

problems in heterogeneous-agent models.[3] At time 0, although the distribution is the same as in the steady state, the planner does not have the same commitments as in the steady state and therefore wants to deviate from the steady-state tax rate and to reoptimize the instrument path. This implies a deviation from the steady-state tax rate at impact and a convergence to the steady-state value thereafter (again, even though the initial and steady-state distributions are identical). In the transition method, since the path is constrained to remain constant, the constant value partly reflects the planner's willingness at time-0 to deviate from the steady-state value. Using the truncation method to compute the optimal path, we find that the transition tax rate is almost equal to the "average" discounted value of the instrument along the truncation path. This is a first check of the validity of the truncation method.

We then verify that the truncation path actually maximizes the aggregate welfare. To do this, we use a generalization of the transition method. We set the truncation path as an exogenous input and compute the aggregate welfare as in the transition method (with a time-varying deterministic instrument path instead of a constant one). We then verify that changing the optimal path reduces aggregate welfare. We view this exercise as a second validity check of the truncation method.

Our final comparitive exercise involves the introduction of aggregate shocks. We simulate the optimal Ramsey policy after a TFP using the truncation method. We then regress the response of the instrument on lagged values of aggregate quantities and of the aggregate shock. We then include this estimated instrument as an exogenous rule in the model that we simulate using the Reiter (2009) method. We compare these results with those of the model simulated using the truncation method and find that the truncation method and Reiter methods yield very similar dynamics. Our conclusion from this set of exercises is that the truncation method accurately computes optimal policies in a heterogeneous-agent model. Since this is a methodological paper, we present the truncation method and related algorithms in some detail.

Our paper belongs to the literature on optimal policies in heterogeneous-agent models. This literature can be divided into three strands. A first strand focuses on simplified environments to identify mechanisms where the equilibrium distribution is simple enough to yield a tractable setup.[4] Bilbiie (2021) and Bilbiie and Ragot (2021) solve for optimal monetary policy in an environment where a partial insurance structure implies that the equilibrium features only two consumption levels. Acharya et al. (2023) consider a CARA-normal structure without binding credit constraints to easily aggregate consumption.

A second branch uses numerical techniques to compute the optimal dynamics of the planner's instruments. Yang (2022) solves for the optimal monetary policy by optimizing on the coefficients

[3]Note that the same problem (financing a public good through lump-sum taxation) is not subject to time-inconsistency in the representative-agent economy. We have chosen this economic setting to highlight this difference between RA and HA models.

[4]We do not review here the large literature on optimal policy with a complete insurance-market but ex ante heterogeneous agents; see, among others, Bassetto (2014).

of a Taylor rule. McKay and Wolf (2023) focus on optimal rules when the optimal steady-state values of the instruments are known, building on Auclert et al. (2021) with a quadratic approximation of the objective. These strategies compute the optimal dynamics of the instruments around a known steady state. How to find the optimal steady-state value of the instrument in Ramsey problems is still an open question. A solution to this question is to numerically optimize over all possible paths to find the optimal long-run value of the instruments, which is a difficult task due to the need to discount the long-run welfare effect of the instrument (see Dyrda and Pedroni, 2022 for an example and discussion).

To circumvent the difficulties related to numerical techniques and to find the steady-state allocation, a recent branch of the literature solves for the Ramsey program by exploiting the planner's first-order conditions when the solution is interior (see Aiyagari, 1995 for an early contribution). A first contribution of this approach is to link the normative analysis to the public finance literature, which makes extensive use of marginal valuations (see Heathcote and Tsujiyama, 2021 for a discussion of optimal policies in two-period models and the discussion in Section 2.6 below). Unfortunately, this leads to additional difficulties in intertemporal models, since heterogeneous-agent models typically involve a continuum of Euler equations. Some papers use continuous-time techniques to solve for optimal policies in heterogeneous-agent models (see Dávila and Schaab, 2022, Nuño and Thomas, 2022, and Smirnov, 2022 for recent examples). Other papers develop new techniques to address the issue of aggregate risk. Bhandari et al. (2021) provide a numerical procedure that assumes that credit constraints are not occasionally binding. They rely on the so-called primal approach, which implies substituting the ratio of marginal utilities for the interest rate. Açikgöz et al. (2022) (who study fiscal policy) use tools from the Lagrangian approach and follow the entire distribution of Lagrange multipliers on Euler equations, in economies where the credit constraints are occasionally binding—which is often the quantitatively relevant case in quantitative models.

LeGrand and Ragot (2022a) (studying optimal unemployment insurance) use the Lagrangian approach but in the context of a finite state-space model, which is derived thanks to the truncation method. LeGrand and Ragot (2022b) propose a refinement of the truncation method that allows the consideration of truncated histories of different lengths, but focuses solely on simulating the model in the presence of aggregate shocks, not on computing optimal policies. LeGrand et al. (2022) study the joint optimal monetary-fiscal policy and the truncation method is used only as a tool to solve the model. Our current contribution is to verify the accuracy of the truncation method, using insights from the transition method. To improve the accuracy of the truncation method, we enhance on our aggregation procedure by introducing additional history-dependent parameters. This allows for a precise aggregation of model non-linear equations (see Section 4.1 below).

The remainder of this paper is organized as follows. Section 2 presents the environment and the Ramsey problem. Section 3 describes how the transition method can be used to compute 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 4 details the computation of the Ramsey solution using the truncation method. Section 5 contains a numerical exercise quantifying the differences along several dimensions between the two methods considered in the paper. Section 6 concludes.

2 The environment

We consider an environment similar to that 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 concerns 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. We focus most of the paper on the no-aggregate shock case and introduce aggregate shocks as an extension in Section 5.7.

2.1 Risk

Agents are affected by an idiosyncratic labor productivity shock $y \in \mathcal{Y}$. Where w_t denotes the hourly wage, the labor income of an agent with productivity y_t amounts to $y_t w_t$, because agents provide a unitary labor supply. 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. It is 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 $\sum_{y \in \mathcal{Y}} S_y = 1$ because the size of the population is one. Finally, an individual history of productivity shocks up to date t is denoted $y^t = \{y_0, \ldots, y_t\} \in \mathcal{Y}^{t+1}$, and θ_t is the measure of such histories.

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}, which needs to be installed one period in advance, to produce Y_t units of the consumption good. Individual labor supply is fixed and normalized to 1, and so aggregate labor supply is constant and equal to $\overline{L} := \int_i y_{i,t}\ell(di)$ 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 $\delta \in (0,1)$. TFP is normalized to 1 and the production function is:

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

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

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

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:

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

We choose to abstract from more complex financing schemes in order to compare numerical methods in a straightforward environment.[5] 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 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.[6] In the initial period, each agent i is endowed with initial wealth a_{-1}^i and initial productivity status y_0^i, 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 $(G_t)_{t\geq 0}$, the

[5] See Den Haan (2010) for an early similar strategy for models with aggregate shocks. We have also simulated the model with a proportional tax on labor, instead of a lump-sum tax. The comparisons of the two methods generates the same conclusions.

[6] The credit constraint can be arbitrarily set to 0 without loss of generality up to some monotonic transformation of variables (including the wage process). See Açikgöz (2018, Section 4.1). For a given wage process, changing the credit constraint however affects the economy (through accumulated capital and interest rate).

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 $\mathbb{E}_0$ is an expectation operator over idiosyncratic shock. In the initial period, the agent chooses their consumption path $(c_t^i)_{t\geq 0}$ and their saving plan $(a_t^i)_{t\geq 0}$ to maximize their intertemporal utility (4), subject to the budget constraint (5) and the borrowing limit (6).

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

Remark 1 (Simplifying Notation) *An agent has an idiosyncratic history y_i^t and initial wealth a_{-1}^i; X_t^i denotes the realization in state (y_i^t, a_{-1}^i) of any random variable $X_t : \mathcal{Y}^t \times \mathbb{R} \to \mathbb{R}$.*

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

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 conditions and the economy-wide resource constraint are:

$$\int_i a_t^i \ell(di) = K_t, \text{ and } \int_i c_t^i \ell(di) + G_t + K_t = Y_t + K_{t-1}. \tag{8}$$

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 Λ_0, and for an initial value of capital stock verifying $K_{-1} = \int_{a_{-1}} \sum_{y_0 \in \mathcal{Y}} d\Lambda_0(a_{-1}, y_0)$, 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) 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).*

2.6 The Ramsey problem

The Ramsey problem consists of selecting a fiscal policy that corresponds to a competitive equilibrium with the highest aggregate welfare. Regarding 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:[7]

$$W_0 := \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \int_i \Big(u(c_t^i) + v(G_t) \Big) \ell(di) \right], \tag{9}$$

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

Ramsey allocation, reoptimization shock, and time-inconsistency. We now formalize our definition of the Ramsey allocation.

Definition 2 (Ramsey allocation) *Given an initial distribution Λ_0 over initial wealth and productivity levels:*

1. *A Ramsey allocation is a competitive equilibrium in the sense of Definition 1 that maximizes the aggregate welfare W_0 of equation (9) over the set of competitive equilibria;*

2. *The steady-state Ramsey allocation is characterized by the long-run values, when they exist, of the distribution over savings and productivity levels, denoted Λ_∞^{opt}, and of the instrument, denoted T_∞^{opt}.*

Definition 2 defines the steady-state Ramsey allocation as the limit of the Ramsey allocation. Indeed, the Ramsey allocation characterizes a path for the distributions (over savings and productivity levels) and for the fiscal instrument, which both depend on the initial distribution Λ_0. To highlight this relationship, the paths of the distribution and of the fiscal instrument will be denoted $(\Lambda_t^{opt}(\Lambda_0))_t$ and $(T_t^{opt}(\Lambda_0))_t$, respectively. Conversely, the steady-state Ramsey allocation does not depend on initial conditions, as we will verify numerically, and as discussed by Açikgöz (2018).

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

Characterizing the path of the instruments and time-inconsistency in heterogeneous-agent models. Definition 2 can be formalized as the outcome of an optimization program. Using the governmental budget constraint (3) to substitute T_t for G_t, the Ramsey program can be written as follows:

$$\max_{\left(w_t,r_t,T_t,K_t,(a_t^i,c_t^i,\nu_t^i)_{i\in I}\right)_{t\geq 0}} \mathbb{E}_0\left[\sum_{t=0}^{\infty}\beta^t\int_i\Big(u(c_t^i)+v(T_t)\Big)\,\ell(di)\right], \tag{10}$$

$$\forall i\in I,\ a_t^i+c_t^i =(1+r_t)a_{t-1}^i+w_t y_t^i-T_t, \tag{11}$$

$$u'(c_t^i) =\beta\mathbb{E}_t\left[(1+r_{t+1})u'(c_{t+1}^i)\right]+\nu_t^i, \tag{12}$$

$$a_t^i \geq 0,\ \nu_t^i a_t^i=0,\ \nu_t^i\geq 0,\ c_t^i\geq 0, \tag{13}$$

$$K_t =\int_i a_t^i\ell(di), \tag{14}$$

$$r_t =F_K(K_{t-1},\overline{L}),\ w_t=F_L(K_{t-1},\overline{L}), \tag{15}$$

where $(a_{-1}^i, y_0^i)_i$ is given by the initial distribution Λ_0. Equation (10) is the planner's objective (9). Equations (11)–(15) are the planner's constraints that guarantee that the chosen allocation is picked up among the competitive equilibria of Definition 1. Equations (11)–(13) 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 (14) and (15) are economy-wide constraints, regarding financial market clearing and factor price definitions.

The trade-off faced by the planner in the Ramsey program (10)–(15) is rather straightforward. The planner can increase the provision of the public good at the cost of a higher tax that reduces the consumption of private goods. This higher tax has a heterogeneous effect on agents, because they have different wealth levels and different incomes. Thus, the higher tax heterogeneously affects agents' saving decisions, which in turn modify the dynamic of the capital stock, the real wage, and the real interest rate. These general equilibrium effects, combined with the redistribution motives, make the Ramsey problem difficult to solve. To understand the time-inconsistency of the Ramsey problem, we now use the first-order conditions of the Ramsey program (10)–(15). The quantity $\beta^t\lambda_t^i$ denotes the Lagrange multiplier on the agent-i Euler equation.[8] To simplify the analysis, we introduce the concept of *social valuation of liquidity for agent* i, denoted ψ_t^i and formally defined as:

$$\psi_t^i := \underbrace{u'(c_t^i)}_{\text{direct effect}} - \underbrace{u''(c_t^i)\left(\lambda_t^i-(1+r_t)\lambda_{t-1}^i\right)}_{\text{effect on savings incentives}}, \tag{16}$$

[8] For brevity, we refer to LeGrand and Ragot (2022a), who prove that the Lagrange multipliers can be used with occasionally-binding credit constraints, by taking limits of penalty functions.

which can be seen as the equivalent for the planner of the marginal utility of consumption.[9] Indeed, it measures, from the planner's perspective, the value of one extra unit of consumption for agent i at date t. The term ψ_t^i includes three terms. The first one, $u'(c_t^i)$, is the private value, for agent i, of consuming one extra unit at date t. The two extra terms reflect the planner's valuation of the savings distortions induced by this extra consumption. The extra consumption at date t affects agent's savings incentives between dates $t-1$ and t, as well as between dates t and $t+1$. Since the planner internalizes savings constraints via individual Euler equations, the impact related to savings between $t-1$ and t is proportional to the shadow cost of the Euler equation between these two dates, which is λ_{t-1}^i. Similarly, the impact related to savings between t and $t+1$ is proportional to λ_t^i, which is the shadow cost of the Euler equation between t and $t+1$. Finally, the signs in front of λ_{t-1}^i and λ_t^i are opposite as in the first case the extra consumption means extra future consumption, while in the other it means extra current consumption.[10]

We derive the first-order conditions in Appendix A. We here discuss the results. The first-order condition regarding the lump-sum tax T_t can be expressed as follows:

$$v'(T_t) = \int_i \psi_t^i \ell(di), \tag{17}$$

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

Condition (17) can also be rewritten as:

$$v'(T_t^{opt}) = \int_i u'(c_t^i)\ell(di) - \int_i u''(c_t^i)\left(\lambda_t^i - (1+r_t)\lambda_{t-1}^i\right)\ell(di), \tag{18}$$

which shows a time-inconsistency problem due to the presence of current and past values of the Lagrange multiplier on the right-hand side. In an incomplete-market economy, the Ramsey planner generally cannot restore the first-best allocation and must thus use their instruments to close the gap with the first-best allocation. This involves accounting for private saving incentives that are distorted by the lump-sum tax in a heterogeneous way (because of pervasive heterogeneity among agents). In particular, yesterday's savings choices were affected by the tax amount at that time, hence the presence of λ_{t-1}^i in (18). Should the planner be given the opportunity to re-optimize at some date t_0 (i.e., to set $\lambda_{t_0-1}^i = 0$ for all i, thereby corresponding to the absence of commitment to past promises), they would modify the tax path from date τ onwards and be time-inconsistent.[11]

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

[10]In general we cannot predict the sign of λ_t^i, which depends on whether the planner perceives that agent i is over or under saving (see LeGrand and Ragot, 2022a for an illustration).

[11]Setting the past value of Lagrange multipliers to 0 is sometimes called a reoptimization shock in the literature

Because of this time-inconsistency, there is no reason for the optimal Ramsey tax path to be constant, even when the initial distribution is set to its steady-state value. Indeed, at date-0, there are no past commitments and all $(\lambda_{-1}^i)_i$ are set to zero. However, in the long run $(t \to \infty)$, past promises still matter and past Lagrange multipliers in condition (18) differ from zero. The planner thus faces a different trade-off at date 0 than in the long-run, even though wealth distributions are assumed to be identical. The planner is thus likely to reoptimize at date 0 such that the tax level at impact deviates from the steady-state tax.

In the complete-market environment, there is only one Euler equation for the representative agent, for which the Lagrange multiplier is $\lambda_t^{CM} = 0$, as we formally show in Appendix B. This is because in the absence of heterogeneity, the savings of the representative agent are equal to the aggregate capital stock and hence not influenced by the lump-sum tax. The first-order condition of the Ramsey planner is:

$$v'(T_t^{CM}) = u'(c_t^{CM}),$$

In this case, the Ramsey problem is time-consistent, because no past value of Lagrange multiplier affects the planner's decision. By the same token, it can be expected that in a complete market economy, should the initial distribution be set to the steady-state distribution, the optimal tax path will remain constant in the absence of aggregate shocks.

3 Maximizing aggregate welfare with transitions

A standard optimization method in the literature consists of finding the constant tax that maximizes the aggregate welfare while accounting for transitions, as in Conesa et al. (2009) or Chang et al. (2018), among many others. We call this method the transition method and the resulting tax rate the *transition tax rate*. Indeed, because the method involves simulating the model with transitions, it takes as inputs a constant tax rate T and an initial distribution Λ_0 (over wealth and productivity) and generates a long-run distribution through simulation, denoted $\Lambda_\infty(T, \Lambda_0)$. This long-run distribution depends on the initial distribution and the tax path, hence the notation. We now formalize the definition of the transition method.

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

1. *When the tax path is constant and set to T^{exo}, the Bewley model with the initial distribution Λ_0 converges to a long-run distribution denoted $\Lambda_\infty(T^{exo}, \Lambda_0)$;*

2. *The lump-sum tax T^{exo} maximizes the computed period-0 aggregate welfare, accounting for the transitions when the agents' distribution evolves from Λ_0 at $t = 0$ to $\Lambda_\infty(T^{exo}, \Lambda_0)$ in the long run $(t \to \infty)$.*

(Debortoli et al., 2014).

Formally, the transition method consists of finding the constant tax path with level T that maximizes aggregate welfare for a given Λ_0. The algorithm for this solution can be summarized as follows.

Algorithm 1 (Computing the transition tax rate) *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 constant lump-sum tax path T;*
2. *Solve for the steady-state distribution $\Lambda_\infty(T, \Lambda_0)$;*
3. *Compute the transition and the welfare during the transition of the economy from the initial distribution Λ_0 toward $\Lambda_\infty(T, \Lambda_0)$;*
4. *Update T and start again at Step 2 until the computed welfare is maximal.*

How does the transition method compare to the Ramsey program (10)–(15)? The transition and the truncation methods select a competitive equilibrium (in the sense of Definition 1) that maximize the aggregate welfare (9). The Ramsey program (10)–(15) of the truncation method allows for a time-varying tax path, while the transition method selects a constant tax path.[12] Formally, the transition method solves a problem that could be written as a Ramsey program with one additional condition, stating that the tax path is constant: $T_t = T$ for all T. In particular, the transition method also features time-inconsistency. Indeed, should the truncation planner be given the opportunity to chose a new (constant) tax path at $t = 1$, then it would choose a different tax path than at $t = 0$ – for the same reason as in the Ramsey program.

However, this is a non-trivial restriction. Loosely speaking – and as we will see later in our numerical applications – this means that the transition method has to trade-off short-term aspects with long-term ones. In other words, the initial distribution is affecting the outcome of the transitions method. For instance, starting from an economy with a very low initial capital stock (close to 0) implies a very low T, because the tax base in the initial periods is close to 0. Alternatively, if the initial capital stock is very large, then the planner can set a higher value for its instrument and speed the convergence toward a lower optimal capital stock. Consequently, we denote this optimal tax as $T^{exo}(\Lambda_0)$ to highlight the dependence on the initial distribution.

Addressing the dependence in the initial condition. To neutralize the effect of the initial distribution, we modify the previous computation by iterating on the initial distribution, until initial and long-run distributions coincide for the chosen tax level. Formally, we iterate on the initial distribution Λ_0 and compute a new optimal instrument value $T^{exo}(\Lambda_0)$ until the long-run

[12]Note that in both methods, the initial distribution can be freely calibrated.

distribution equals the initial distribution: $\Lambda_\infty(T^{exo}(\Lambda_0), \Lambda_0) = \Lambda_0$. When convergence has been reached, we obtain a tax and a long-run distribution, denoted (Λ^c_∞, T^c), which verify: $T^c = T^{exo}(\Lambda^c_\infty)$ and $\Lambda^c_\infty = \Lambda_\infty(T^{exo}(\Lambda^c_\infty), \Lambda^c_\infty)$.

To our knowledge, this study is the first to propose this procedure. We formalize the definition below of the *optimal fixed-point transition tax rate.*[13]

Definition 4 *The fixed-point transition tax rate 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 $\Lambda_0 = \Lambda^c_\infty$ converges to the same long-run distribution $\Lambda_\infty(T^c, \Lambda^c_\infty) = \Lambda^c_\infty$;*

2. *The tax rate T^c maximizes the period-0 aggregate welfare computed when accounting for the transitions when the agents' distribution evolves from $\Lambda_0 = \Lambda^c_\infty$ at $t = 0$ to Λ^c_∞ in the long run.*

The fixed-point transition tax rate T^c is such that: (i) the initial and the long-run distributions coincide when the fiscal policy is constant and set to T^c; and (ii) the aggregate welfare with transitions is maximal when the fiscal policy is set to T^c.

We formalize the computation of T^c in the algorithm below.

Algorithm 2 (Computing the fixed-point transition tax rate) *The fixed-point transition tax rate 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 $\Lambda_\infty(T, \Lambda_0)$;*

   (b) *If the initial and long-run distributions coincide (i.e., $\Lambda_\infty(T, \Lambda_0) = \Lambda_0$), then proceed to Step 3. 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 transitions of the economy from the initial distribution Λ_0 toward $\Lambda_\infty(T, \Lambda_0) = \Lambda_0$;*

4. *Update T and start again at Step 2 until the computed welfare at Step 3 is maximal.*

Definition 4 and Algorithm 2 neutralize the influence of the initial distribution on the optimal transition tax rate. However, this still does not mean that the fixed-point transition instrument

[13]To avoid confusion, the "optimal truncation tax rate" refers to the optimal tax rate computed using the truncation method of Algorithm 3. The "fixed-point transition tax rate" refers to Definition 4, where the initial and long-run distributions are identical. The "transition tax rate" refers to Definition 3, where the initial distribution influences the outcome. Due to this limitation, the latter is barely used in our applications in Section 5.

T^c is equal to the Ramsey steady-state value T_∞^{opt}. Indeed, let us assume that the initial and long-term distributions are identical in the transition and truncation methods. In the case of the truncation method, the time-inconsistency implies that in the short-run the Ramsey planner will deviate from the long-term value. Indeed, and as explained before, even though the initial and long-term distributions coincide with each other, the planner's commitment differ in the short-run (past Lagrange multipliers are null at $t = 0$: $\lambda_{-1}^i = 0$) and in the long-run (past Lagrange multipliers λ_{t-1}^i typically differ from 0 for large t). The planner will take advantage of this difference in past commitments to deviate from the long-term tax rate. Oppositely, in the case of the fixed-point transition method, the planner must implement a constant tax path by construction. Because of the absence of short-run deviations, the transition tax must balance short-run and long-run aspects. The fixed-point transition tax rate will again be close to the average truncation tax path and thus differ from the truncation steady-state value. The gap between the two tax rates will reflect the severity of the planner's time-inconsistency. It will be large when the time-0 deviation due to time-inconsistency is large.

Finally, it should be observed that Definition 4 and Algorithm 2 provide a solution that differs from the optimal tax that maximizes the steady state welfare (as in Aiyagari and McGrattan, 1998 for instance). Indeed, the objective solved by the planner is not the same in the two cases: in one case its the intertemporal aggregate welfare, while in the other it is the steady-state aggregate welfare. We illustrate this point in our quantitative exercise.

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

We now present the truncation method of LeGrand and Ragot (2022a). Because the present study improves on this new method, we present it in several steps:

1. We first provide the intuition and the basic details of the improved truncation method in Section 4.1;

2. We explain how to implement the truncation method to simulate the model with an exogenous tax path in Section 4.2;

3. We document how the truncation method can solve for steady-state Ramsey policies in Section 4.3;

4. Finally, we show how to compute the optimal Ramsey path for the instrument and the allocation 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 start with an intuitive presentation before formalizing it.

An intuitive description. The truncation method consists in grouping agents together according to their recent idiosyncratic history and to express the model in terms of these groups of agents. These groups are called *truncated histories*. For the sake of example, we assume that there are 2 productivity levels: $y_l < y_h$ and that we group agents according to their histories over the 3 last periods. Agents with productivity history (y_l, y_l, y_l) over the three last periods will thus be assigned to the same truncated history. Same for history (y_l, y_h, y_l), and so on. Agents move from one truncated history to another when their productivity history is updated in the beginning of the next period. For example, an agent with productivity history (y_l, y_l, y_l) drawing the productivity y_h will have history (y_l, y_l, y_h) – and hence will move from truncated history (y_l, y_l, y_l) to (y_l, y_l, y_h).

All agents of a given truncated history are assigned the same allocation, equal to the average allocation of all agents belonging to the group. *Each* truncated history is thus akin to a "representative agent". A key aspect of the method is that the truncated allocation has to be consistent with the representative agents of truncated history being expected-utility maximizers. This does not raise any difficulty for linear equations, such as budget constraints. Indeed, the average budget constraint is equal to the budget constraint with average quantity (at least as long as there is non-linearity such as progressive taxation). However, this is not the case in general for non-linear equations. For example, the average of marginal utilities of individual consumption is not equal to the marginal utility of average consumption. The discrepancy is due to the heterogeneity within each truncated history (and the non-linearity of the marginal utility function). Hence, we introduce parameters, called ξs, that are necessary to correctly aggregate nonlinear equations. In our case, the non-linearities come the utility function and its derivatives: u, u' and u'', which explains that we introduce three sets of correcting coefficients. They are denoted by ξ^0, ξ^1, and ξ^2 and correspond to u, u' and u'', respectively.[14] This extends LeGrand and Ragot (2022a), who only considered for the sake of simplicity a unique set of ξs.

The truncation method for solving the Ramsey model then works as follows. We set a guess value for the optimal instrument and compute the corresponding Aiyagari allocation (using standard method such as EGM for instance). We then compute the coefficients ξ^j $(j = 0, 1, 2)$ to properly aggregate non-linear equations. Finally, we verify if the Ramsey FOCs hold in the truncated economy. If they do, then the guessed value is the optimal instrument value. Otherwise, we iterate on the instrument value and redo the whole process (including solving the Aiyagari model). See Section 4.3 for further details.

[14] Other ξs could be introduced for other non-linearities – such as progressive taxation for instance.

Two remarks are noteworthy. First, the truncation method converges to an instrument value that corresponds to an existing Aiyagari equilibrium (by construction). Second, the length (equal to 3 here) of the productivity history for building the set of histories is an exogenous model parameters. We have shown that when this length increases, the truncated economy converges to the Aiyagari economy, but in practice small truncation lengths offer very good accuracy.

A more formal description. We provide a detailed account of how this method can be used to solve for heterogeneous-agent models in LeGrand and Ragot (2022b).[15] Although the present paper focuses on how this method can be used to solve for Ramsey programs, we also provide a concise description of the method in this paper's context. This method's starting point 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:[16] $a_t(y^t, a_{-1}, Z^t)$ and $c_t(y^t, a_{-1}, Z^t)$, for $y^t \in \mathcal{Y}^{t+1}, Z^t \in \mathcal{Z}^{t+1}$ using previous notation. 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 use the term *truncation length* for the exogenous parameter setting the length of the shared productivity history, denoted $N > 0$. The truncation method consists of truncating idiosyncratic histories and following a finite number of representative agents in an adjusted model. A key step in 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 their 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, independent of earlier productivity levels (i.e., of the sequence $(\ldots, y_{t-N-1}, y_{t-N})$). Because the number of productivity levels is finite and equal to $n_y := Card(\mathcal{Y})$, the number of truncated histories of length N is also 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 with productivity y_{t+1} at $t+1$, their $t+1$-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 they will be assigned at date $t+1$ to truncated history $\tilde{y}^N = (y^N_{-N+2}, \ldots, y^N_1, y^N_0, \tilde{y}^N_0)$, where $\tilde{y}^N_0 := y_{t+1}$.[17] The probability, denoted $\Pi_{y^N \tilde{y}^N}$, that an agent transitions from history y^N to history $\tilde{y}^N$ is the probability that the agent transitions from productivity level y_0 to $\tilde{y}_0$, or formally:

$$\Pi_{y^N \tilde{y}^N} = 1_{\tilde{y}^N \succeq y^N} \Pi_{y^N_0 \tilde{y}^N_0}, \tag{19}$$

[15] Notably, we insist in the present study on the matrix notation and the implementation aspects.

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

[17] For consistency, we denote future truncated histories with a tilde, past ones with a hat, and current ones without decoration.

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: $\tilde{y}^N_{-1} = y^N_0$, $\tilde{y}^N_{-2} = y^N_{-1}, \ldots$, $\tilde{y}^N_{-N+1} = y^N_{-1}$; $1_{\tilde{y}^N \succeq y^N} = 0$ otherwise. The population of agents with truncated history y^N can be defined recursively from the previous probabilities as:

$$S_{y^N} = \sum_{\hat{y}^N \in \mathcal{Y}^N} S_{\hat{y}^N} \Pi_{\hat{y}^N y^N}. \tag{20}$$

Because the truncated model aims to express the economy using truncated histories, we must derive for each truncated history its consumption level and its end-of-period savings, denoted $c_{t,y^N}(Z^t)$ and $a_{t,y^N}(Z^t)$, respectively, or simply c_{t,y^N} and a_{t,y^N} when there is no ambiguity. These quantities are defined as the corresponding average values among agents sharing the same truncated history y^N. For instance, for savings:

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

To compute beginning-of-period savings, denoted $\tilde{a}_{t,y^N}$, we must account for the possibility that agents with current truncated history y^N had different truncated histories $\hat{y}^N$ in the previous period. Formally:

$$\tilde{a}_{t,y^N} = \frac{1}{S_{y^N}} \sum_{\hat{y}^N \in \mathcal{Y}^N} S_{\hat{y}^N} \Pi_{\hat{y}^N y^N} a_{t-1,\hat{y}^N}. \tag{21}$$

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

$$c_{t,y^N} + a_{t,y^N} = (1 + r_t)\tilde{a}_{t,y^N} + w_t y^N_0 - T_t, \tag{22}$$

where y^N_0 is the current productivity level for y^N. This aggregation is straightforward because budget constraints are linear. Aggregating utility or its derivatives is less straightforward, because utility levels and marginal utilities are not linear in consumption. To proceed with the aggregation of non-linear equations (and in particular of the utility function and its derivatives), we define the following history-dependent parameters:

$$\xi^0_{t,y^N} := \frac{1}{u(c_{t,y^N})} \int_{a_{-1}} \sum_{\hat{y}^t \in \mathcal{Y}^{t+1} | (\hat{y}_{t-N+1}, \ldots, \hat{y}_t) = y^N} \theta_t(\hat{y}^t) u(c_t(\hat{y}^t, a_{-1}, Z^t)) d\Lambda_0(a_{-1}, y_0), \tag{23}$$

$$\xi^1_{t,y^N} := \frac{1}{u'(c_{t,y^N})} \int_{a_{-1}} \sum_{\hat{y}^t \in \mathcal{Y}^{t+1} | (\hat{y}_{t-N+1}, \ldots, \hat{y}_t) = y^N} \theta_t(\hat{y}^t) u'(c_t(\hat{y}^t, a_{-1}, Z^t)) d\Lambda_0(a_{-1}, y_0), \tag{24}$$

$$\xi^2_{t,y^N} := \frac{1}{u''(c_{t,y^N})} \int_{a_{-1}} \sum_{\hat{y}^t \in \mathcal{Y}^{t+1} | (\hat{y}_{t-N+1}, \ldots, \hat{y}_t) = y^N} \theta_t(\hat{y}^t) u''(c_t(\hat{y}^t, a_{-1}, Z^t)) d\Lambda_0(a_{-1}, y_0). \tag{25}$$

These parameters enable us to reconcile the aggregation of the transformation of a given quantity with the transformation of the aggregation of the quantity. 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 derived from truncated-history consumption levels. The role of 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 utilities, weighted by ξ^0.

These parameters capture the residual heterogeneity within each truncated history attributable to the fact that agents experienced different idiosyncratic histories N periods ago and before. Indeed, each truncated history groups together construction agents sharing the same history over the last N periods, while ignoring the distant past. In the absence of within-truncated-history heterogeneity, the ξ^j would be all equal to 1.

On the theoretical side, the ξ^j $(j = 0, 1, 2)$ parameters can be shown to converge toward 1 as the length of the truncation N increases. However, because N remains small in practice, this asymptotic result has little practical implication. Fortunately, as we check in our quantitative exercise in Section 5, even for short truncation lengths, the ξ^j allow us to obtain accurate results. We explain in Section 4.2 how to easily compute the ξ^j for $j = 0, 1, 2$.

Finally, $\mathcal{C}_{t,N}$ denotes the set of credit-constrained truncated histories at date t. With this notation and the previous ξ^j the Euler equations can be written as follows:

$$\forall y^N \in \mathcal{Y}^N, \ \xi^1_{t,y^N} u'(c_{t,y^N}) = \beta(1 + r_{t+1}) \sum_{\tilde{y}^N \in \mathcal{Y}^N} \Pi_{y^N \tilde{y}^N} \xi^1_{t+1,\tilde{y}^N} u'(c_{t+1,\tilde{y}^N}) + \nu_{t,y^N}, \tag{26}$$

$$\forall y^N \in \mathcal{C}_{t,N}, \ a_{t,y^N} = 0, \tag{27}$$

where the expectation over the future idiosyncratic state has been addressed explicitly. Equation (26) is the Euler equation at the truncated history level for non-credit-constrained histories, and equation (27) corresponds to credit-constrained truncated histories holding zero assets.

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

4.2 Using the truncation method for an exogenous tax path

We now explain how the truncated allocation and (ξ^0, ξ^1, ξ^2) can be computed for a steady-state Bewley model for an exogenous tax path $(T_t)_{t \geq 0}$, which converges toward a steady-state denoted T_∞. We use subscript ∞ to highlight that the tax path may be time-varying in the transition; other steady-state variables are denoted without a time subscript.

As proved in the literature (for example by 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 $g_a : (a, y) \in [0, +\infty) \times \mathcal{Y} \to \mathbb{R}_+$. To compute the truncated allocation, we must obtain the wealth distribution with respect to truncated histories of length N, which will be denoted $\tilde{\Lambda}_N : (a, y^N) \in [0; +\infty) \times \mathcal{Y}^N \to \mathbb{R}_+$, 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 $\tilde{\Lambda}_N$ is then computed by starting from the wealth distribution of agents in state y_{-N+1}, which is $\Lambda_\infty(\cdot, y_{-N+1})$, and then successively applying 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, because the process consists of 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:

$$a_{y^N} = \int_{a \in [0,\infty)} a \tilde{\Lambda}_N(da, y^N), \tag{28}$$

where in our case the savings are actually bounded from above (see Açikgöz, 2018). We can then deduce from (21) the beginning-of-period savings $(\tilde{a}_{y^N})_{y^N}$, as consumption $(c_{y^N})_{y^N}$ from the truncated-history budget constraint (22), as well as the Lagrange multipliers $(\nu_{y^N})_{y^N}$ on the credit constraints.

The set $\mathcal{C}$ of credit constrained histories is determined based on these Lagrange multipliers $(\nu_{y^N})_{y^N}$. We take the $n_{\mathcal{C}}$ histories, which have the largest multiplier values (ν_{y^N}) and such that the total size of the $n_{\mathcal{C}}$ histories is as close as possible to the share of credit-constrained agents in the Bewley model.

Finally, the ξ^j parameters need to be computed for properly aggregating non-linear equations.

We deduce from equations (20)–(27) that the steady-state economy is then characterized by the following set of equations:

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

$$c_{y^N} + a_{y^N} = (1+r)\tilde{a}_k + w y_0^N - T_\infty, \tag{30}$$

$$\xi^1_{y^N} u'(c_{y^N}) = \beta(1+r) \sum_{\tilde{y}^N=1}^{N_{tot}} \Pi_{y^N \tilde{y}^N} \xi^1_{\tilde{y}^N} u'(c_{\tilde{y}^N}) + \nu_{y^N}, \tag{31}$$

$$y^N \in \mathcal{C},\ a_{y^N} = 0, \tag{32}$$

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 in the steady state. This computation proceeds in three 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 us 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_0.

We provide the derivation of all steps and then present the algorithm.

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

Details of the computation are provided in Appendix C. Before stating the conditions, we introduce the quantity $\tilde{\lambda}_{t,y^N}$, defined as follows:

$$\tilde{\lambda}_{t,y^N} = \frac{1}{S_{t,y^N}} \sum_{\tilde{y}^N \in \mathcal{Y}^N} S_{t-1,\tilde{y}^N} \Pi_{t,\tilde{y}^N,y^N} \lambda_{t-1,\tilde{y}^N}, \tag{33}$$

which corresponds to the previous period Lagrange multiplier on Euler equation for agents with truncated history y^N at date t. These agents may have different truncated histories in the previous period, which explains equation (33), similar to the beginning-of-period wealth, $\tilde{a}_{t,y^N}$, in equation (21).

We use equation (33) 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 in equation (16). The formal definition features the within-heterogeneity parameters ξ^1 and ξ^2 and is:

$$\psi_{t,y_N} = \xi^1_{t,y^N} u'(c_{t,y^N}) - (\lambda_{t,y^N} - \tilde{\lambda}_{t,y^N}(1+r_t))\xi^2_{t,y^N} u''(c_{t,y^N}). \tag{34}$$

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

$$v'(T_t) = \sum_{y^N \in \mathcal{Y}^N} S_{y^N} \psi_{t,y_N}, \tag{35}$$

which are very similar to condition (17). Equations characterizing the dynamics of the Ramsey model are provided in Appendix C.

4.3.2 Computing the Ramsey allocation at the steady state using matrix notation

We now explain how to compute the steady-state Ramsey allocation, including the planner's instrument value. The truncation model can be represented very efficiently by relying on matrix notation, which enables us to derive Lagrange multipliers using simple linear algebra. To implement the numbering of truncated histories in practice, 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 $k := 1 + \sum_{j=0}^{N-1} n_y^k (n_{y_{-j}} - 1)$, where $n_{y_{-j}} \in \{1, \dots, n_y\}$ is the position of productivity level y_{-j} in the set $\mathcal{Y}$, from 1 for the smallest productivity level to n_y for the largest. 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;
- $\boldsymbol{a} = (a_k)_{k=1,\dots,N_{tot}}$ and $\tilde{\boldsymbol{a}} = (\tilde{a}_k)_{k=1,\dots,N_{tot}}$, the N_{tot}-vector of end-of-period and beginning-of-period asset holdings, respectively;
- $u'(\boldsymbol{c}) = (u'(c_k))_{k=1,\dots,N_{tot}}$, the N_{tot}-vector of marginal utilities;
- $\boldsymbol{\xi^j} = (\xi_k^j)_{k=1,\dots,N_{tot}}$, the vector of residual-heterogeneity parameters ($j = 0, 1, 2$);
- Finally, "$\circ$" denotes the term-by-term product of two vectors of the same size, which is another vector of the same size: $\boldsymbol{x} \circ \boldsymbol{z} = (x_{y^N}) \circ (z_{y^N}) = (x_{y^N} z_{y^N})$.[18]

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

Proposition 1 (Steady-state Lagrange multipliers) *Consider a steady-state tax value T_∞ (not necessarily optimal), for which the truncated model can be computed.*[19] *Then, there exist two matrices, $\boldsymbol{M_1}$ and $\boldsymbol{M_2}$, depending only on the equilibrium allocation, such that:*

$$\boldsymbol{\lambda} = \boldsymbol{M_1}(\boldsymbol{\xi}^1 \circ u'(\boldsymbol{c})) \text{ and } \boldsymbol{\psi} = \boldsymbol{M_2}(\boldsymbol{\xi}^1 \circ u'(\boldsymbol{c})). \tag{36}$$

Proposition 1 states that the steady-state values of Lagrange multipliers and the social value of liquidity can be deduced from 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 D, which provides a step-by-step computation of the expressions of matrices $\boldsymbol{M_1}$ and $\boldsymbol{M_2}$ – the coefficients $\boldsymbol{\xi}^2$s appear in the expression of the matrix $\boldsymbol{M_2}$. After the social value of liquidity vector $\boldsymbol{\psi}$ has been computed, it is straightforward

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

[19] We characterize the full path $(T_t)_{t\geq 0}$ below.

to check the optimality of the steady-state tax level T_∞ using condition (35), which can be written in a matrix form as:

$$v'(T_\infty) = \boldsymbol{S}^\top \boldsymbol{\psi}, \tag{37}$$

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 3 (Steady-state Ramsey allocation) *Set a precision criterion $\varepsilon > 0$ and a truncation length N.*

1. *Set an initial value for the steady-state lump-sum tax T_∞.*
2. *Solve the full-fledged Bewley model for the value of the instrument T_∞.*
3. *Construct the truncated representation for truncation length N and use the method described in Section 4.2 to compute the steady-state truncated allocation $(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{\xi^0}, \boldsymbol{\xi^1}, \boldsymbol{\xi^2})$ and the set of credit-constrained histories $\mathcal{C}$, associated with T_∞.*
4. *Compute matrices $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ using equations (72) and (73) of Appendix D.*
5. *Compute the vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}$ using equation (36).*
6. *If $|v'(T_\infty) - \boldsymbol{S}^\top \boldsymbol{\psi}| < \varepsilon$, then the algorithm stops, and the steady-state Ramsey allocation is given by $(T_\infty, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{\xi^0}, \boldsymbol{\xi^1}, \boldsymbol{\xi^2})$ and T_∞ is the optimal steady-state tax. Otherwise, update T_∞ and start again at Step 2.*

Algorithm 3 shows 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 computes the allocation of the Bewley model corresponding to the lump sum tax T_∞. We can then deduce the steady-state allocation of the truncated model. The matrix notation (further details are provided in Appendix D) enables us to compute the social valuation of liquidity $\boldsymbol{\psi}$. Finally, the optimality of the steady-state lump-sum tax T_∞ is checked using equation (37).

Three remarks are in order. First, Step 2 of Algorithm 3 implies that we compute the Bewley allocation for each value of the steady-state lump-sum tax. Consequently, the algorithm converges by construction to a Bewley equilibrium that does exist. Second, the computational implementation of Algorithm 3 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 (taking less than a second). Third, the computation of the steady-state tax only involves steady-state first-order conditions and is therefore by construction immune to the choice of initial distribution. Compared to Algorithm 2 for the transition method, the computation only requires a fixed-point over the instrument and not a joint fixed-point over the instrument and the initial distribution.

We check in Section 5 that Algorithm 3 yields an accurate solution, and we compare it to that of the transition method.

4.4 Computing the Ramsey tax path

One feature of the truncation method is that it separates the computation of the steady-state allocation (including the planner's instruments) from the computation of the full instrument path. As explained above, this makes the steady-state computation immune to the choice of initial distribution. This also greatly simplifies the computation of the full path of the planner's instruments. Indeed, the steady state being known, standard perturbation methods can be used to compute the path.

The algorithm below summarizes the various steps to compute the instrument path, after the steady-state allocation has been computed using Algorithm 3.

Algorithm 4 (Simulating the full instrument path) *We consider as given a truncation length, $N > 0$, and a precision criterion, $\varepsilon > 0$.*

1. Steady state. *We use Algorithm 3 to compute the steady-state Ramsey truncated allocation with precision ε and truncation length N.*

2. Truncated model. *The dynamics of the truncated model with truncation length N are characterized by equations (49)–(57) of Appendix C.*

3. Initial conditions. *The initial wealth distribution can be set arbitrarily and the initial distribution of Lagrange multipliers $(\lambda_{-1,y^N})_{y^N}$ is set to 0.*

4. Simulation. *Based on the steady-state allocation of Step 1, the model equations of Step 2, and the initial conditions of Step 4, we simulate the path of model variables (including planner's instruments) using perturbation techniques.*

Algorithm 4 describes a straightforward path to simulate the full path of model variables. The core of the algorithm 5 is the perturbation method of Step 4, which can be performed using existing and well-tested software, such as Dynare (Adjemian et al., 2011), which is already widely adopted for solving DSGE models.

Algorithm 4 involves two main assumptions. The first is that the ξ^j coefficients remain equal to their steady-state value along the path. This implies that the heterogeneity within truncated histories is constant through time and is 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 must remain unchanged along the transition. This can be checked especially if the initial distribution markedly differ from the steady-state one.[20]

We use Algorithm 5 to compute a reoptimization shock in Section 5.4, that allows us to assess the severity of time-inconsistency. We explain in Section 5.7 how to adapt Algorithm 4 to handle aggregate shocks.

5 Quantitative exercise

We now turn to the quantitative exercise. This section is structured as follows:

1. We specify our calibration in Section 5.1.
2. We compute the steady-state Ramsey tax using the truncation method in Section 5.2.
3. We compute the optimal fixed-point transition tax rate in Section 5.3.
4. In Section 5.4, we compute the optimal truncation tax path and document the severity of time-inconsistency issue. This also explains the gap between the steady-state truncation tax rate and the fixed-point transition tax rate.
5. In Section 5.5, we check that the time-0 Ramsey tax path obtained with the truncation method actually generates the highest aggregate welfare when the economy is simulated.
6. We perform several robustness checks in Section 5.6.
7. We introduce aggregate shocks in Section 5.7.

5.1 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 $v(G) = G^{\vartheta}$. We set the parameter ϑ to 24% to target a value of steady-state public consumption over GDP of 8.0%, which roughly corresponds to US government consumption of specific final goods minus public investment. In the production function of equation (1), the capital share is set to $\alpha = 36\%$ and the depreciation rate is set 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 the estimates of Krueger et al. (2018). We use a quarterly persistence of $\rho_y = 0.996$ and a quarterly standard deviation of $\sigma_y = 4.39\%$; for the log of earnings, these generate an annual persistence of 0.9849 and an annual standard deviation of 8.71%. The Rouwenhorst

[20] Assuming that the set of credit constrained histories is time-varying would require iterating over the path of this set. We leave that for future research.

(1995) procedure is then used to discretize the productivity process into five idiosyncratic states with a constant transition matrix.

Table 1 summarizes 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 further descriptions and targets.

We can now compute the steady-state equilibrium of the model, using the standard EGM method.[21] The implied capital-output ratio is $K/Y = 2.67$, the consumption-output ratio is $C/Y = 0.65$. Table 2 provides descriptive statistics for the wealth distribution in the data and model. We use data from the Panel Study of Income Dynamics (PSID) in 2006 and from the Survey of Consumer Finances (SCF) in 2007 to abstract from the effects of the 2008 global financial crisis.

	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.

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

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

5.2 Solving the model with the truncation method

We first compute the steady-state truncation tax rate using Algorithm 3. We consider a truncation length $N = 5$, which implies $5^5 = 3,125$ histories. The model by construction generates the steady-state truncation tax-to-GDP $\tau^p_\infty := T^p_\infty / Y = 8\%$, which was the targeted value. The superscript p refers to optimal tax computed with the truncation method. We perform a sensitivity test on the choice of truncation length N in Section 5.6, and the value of N (beyond $N \geq 2$) appears to have a very modest quantitative impact on the optimal provision of the public good.

We provide a further explanation of the different contributing forces to the truncation tax-to-GDP of 8%. 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—thus discarding the role of Lagrange multipliers on Euler equation (18). Formally, this corresponds to:

$$v'(T_t^{partial}) = \int_i u'(c_t^i)\ell(di).$$

The numerical quantification implies that $T^{partial}$ is 4.3% lower than T^p_∞ and the tax-to-GDP ratio decreases from $\tau^p_\infty = 8.00\%$ 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 $u''(c_t^i)$ in equation (18)), the tax is higher because the benefits of this higher tax are factored in by the planner.

5.3 Results using the transition method

We first solve for the *optimal transition tax rate* as presented in Definition 3. To do this, we implement Algorithm 1, which requires an initial distribution Λ_0 as input. We solve for the transition tax rate with two different initial distributions. We first consider the steady-state distribution presented in Table 2 of Section 5.1. We then multiply the wealth level of all agents by 0.9 (implying a 10% decrease in their initial wealth). We call this distribution the low distribution, which corresponds to the transition tax-to-GDP denoted 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 transition tax-to-GDP is denoted τ^{high}.

The computation of the two optimal transition tax rates yields $\tau^{low} = 6.4\%$ and $\tau^{high} = 8.45\%$. 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 setting a low tax. In contrast, when the initial wealth is high, the tax can also be higher because in the first periods agents have 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 fixed-point transition tax rate* T_c, as presented in Definition 4. The tax rate T_c is such that the initial distribution Λ_c equals the long-run distribution: $\Lambda^c = \Lambda_\infty(T^c, \Lambda^c)$. The computation yields a tax-to-GDP ratio $\tau^c = 7.8\%$, which lies between the low and high values, τ^{low} and τ^{high}, that we have just computed.

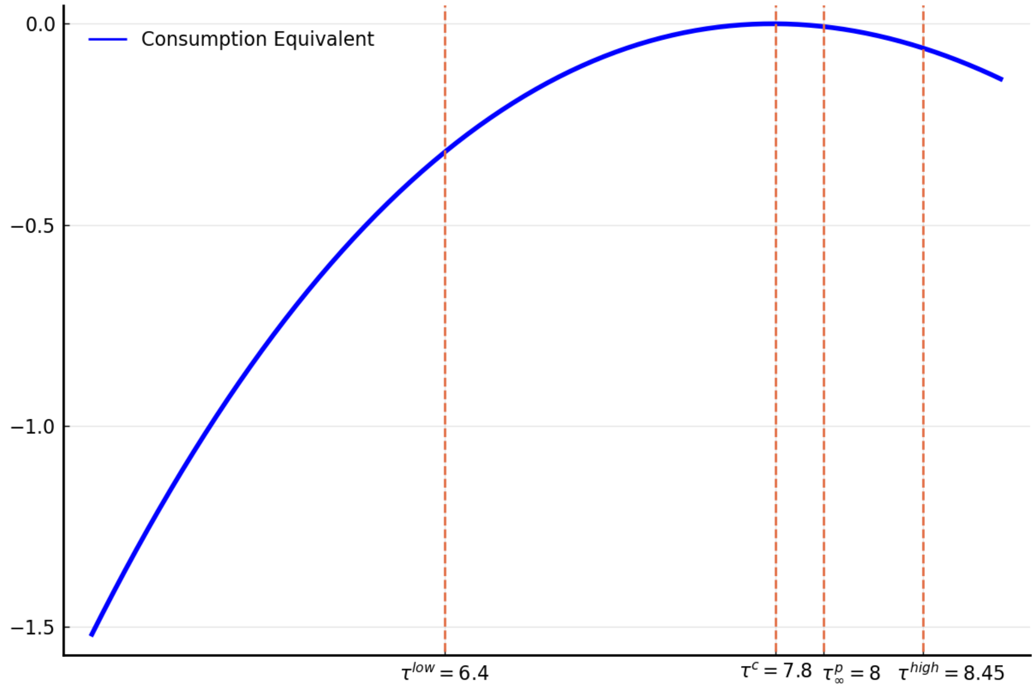

Figure 1: Welfare with transitions and consistent initial distribution computed as a function of the tax rate. See the text for the definitions of other tax rates.

To illustrate the optimality of τ^c, we plot in Figure 1 the aggregate welfare as a function of the tax (plain blue line), and with an initial distribution equal to the long-run one. 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 we then 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 as the tax-to-GDP ratio $\tau = T/Y$, and welfare, on the y-axis, is reported as the percentage loss in consumption compared to the optimal welfare. In addition to τ^c, Figure 1 also shows three other tax rates (as orange dashed vertical bars): τ^{low} corresponding to the low-wealth initial distribution, τ^{high} corresponding to the high-wealth initial distribution, and τ^p_∞ corresponding to the optimal truncation tax rate.

Figure 1 confirms that the optimal transition tax is highly sensitive to the initial distribution, and we can find transition tax rates that are greater or smaller than the fixed-point transition tax rate.

Figure 1 also illustrates that the tax rate τ^c is close to, but lower than, the truncation tax rate τ^p_∞. Although we neutralize the effect of the initial distribution in the computation of the transition tax rate, both tax rates differ. We quantitatively confirm below that this comes from the fact that the Ramsey planner prefers to choose a non-constant tax path.

5.4 The time-inconsistency of Ramsey policies

We further investigate the gap between the truncation and transition tax rates τ^p_∞ and τ^c. We compute the truncation tax path along the lines of Algorithm 4. We neutralize the effect of the initial wealth distribution by setting it equal to its steady-state value. This can be seen as a pure reoptimization shock in period 0, in the sense that the economy starts from the steady state but we turn off the planner's commitment. Figure 2 plots the optimal truncation tax path. We denote this optimal tax path $(T^p_t)_{t\geq 0}$ and the corresponding optimal tax-to-GDP path $(\tau^p_t)_{t\geq 0}$. By construction, the tax path converges to the steady-state tax T^p_∞ in the long-run. Despite starting from the steady-state distribution, the truncation tax path is not constant.

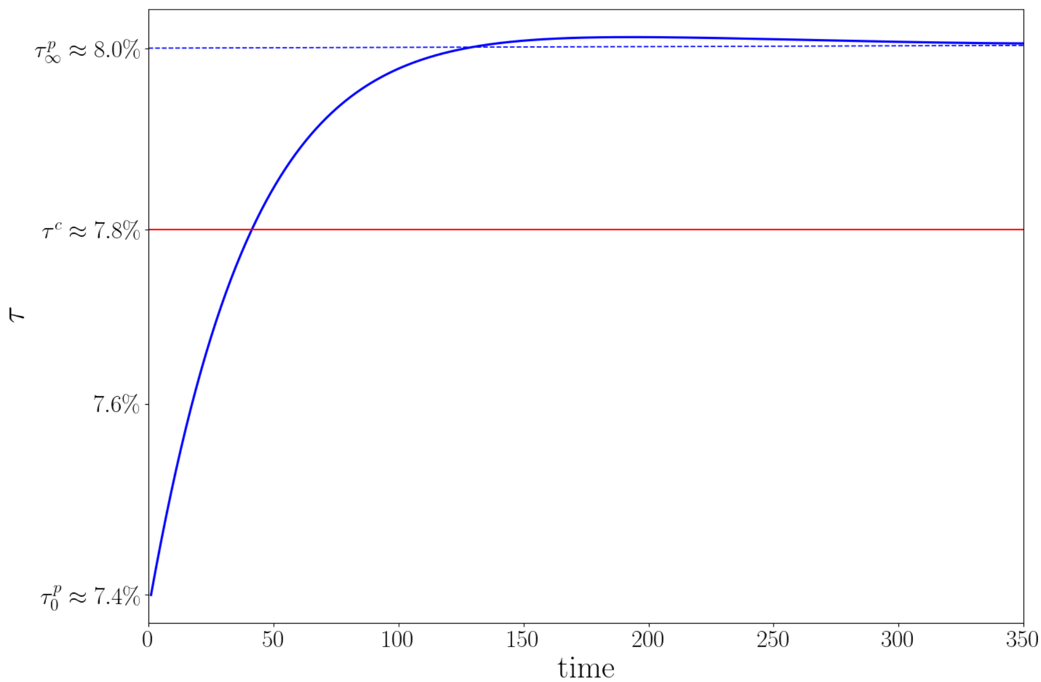

Figure 2: The truncation tax path.

The tax rate drops at impact to a level that is smaller than the fixed-point transition tax rate τ^c. The intuition is as follows. Because $\lambda^i_{-1} = 0$, the planner is not committed to maintain the tax rate equal to τ^p_∞ and chooses to decrease the tax. This 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 with the long-run steady-state value τ^p_∞. This increases agents' incentives to build up savings, which increases the capital stock and, in turn, the consumption of private goods. The consumption of the public good also increases with the tax. The reoptimization that occurs in the truncation tax path stems from the time-inconsistency of the Ramsey program in heterogeneous-agent models, discussed in Section 2.6.

Figure 2 allows us to understand the gap between the steady-state truncation tax τ^p_∞ and the fixed-point transition tax τ^c. As explained above, the transition tax can be seen as a solution of a Ramsey program in which we added the constraint of a constant tax path. With the same view, the transition method thus selects the "average" tax rate over the whole Ramsey path, so as to balance the benefits and costs along the transition path. We check this statement by computing the average discounted tax rate over the transition path $(\tau^p_t)_{t\geq 0}$, 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 τ^{weight}, which is formally defined as:

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

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 the fixed-point transition and the truncation tax rates is well explained by the severity of the time-inconsistency in the Ramsey program.

5.5 Checking the optimality of the transition path

The former exercise was informative to show that the truncation method provides an accurate simulation of the dynamics for a given tax path, but we still need to check the optimality of the truncation tax path. We do this by checking that any perturbation of the truncation tax path of Figure 2 implies a decrease in aggregate welfare. We compute the aggregate welfare using an extended transition method, where the given tax path is not constant but time-varying and deterministic. Given the truncation tax path of Figure 2, we construct for any real value κ the tax path $(T^\kappa_t)_t$ as:

$$T^\kappa_t = T^p_\infty + (1+\kappa)(T^p_t - T^p_\infty),\ t \geq 0, \tag{38}$$

where T^p_∞ is the steady-state value of the truncation tax path. Independent of κ, any path T^κ_t 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 truncation path, and when $\kappa > 0$ ($\kappa < 0$), we implement a higher (lower) path.

We incorporate equation (38) for a given κ as an exogenous rule into the Bewley model. We set the initial distribution equal to the steady-state truncation distribution. The tax path being exogenous, the model simulation does not require any optimization and follows the same lines as the truncation method but with a time-varying deterministic tax path. For each value of κ, we compute the aggregate welfare in this economy, and we report the results in Figure 3. The welfare is expressed as the consumption equivalent drop in welfare from optimum. Figure 3 shows that the welfare is maximal for a path that is extremely close to the optimal path $(T_t^p)_t$ $(\kappa = 0)$. Note that the computation of welfare does not involve the truncation method and is thus an external validity check. We can thus be confident that the truncation method is well-suited to compute optimal policies in heterogeneous-agent models.

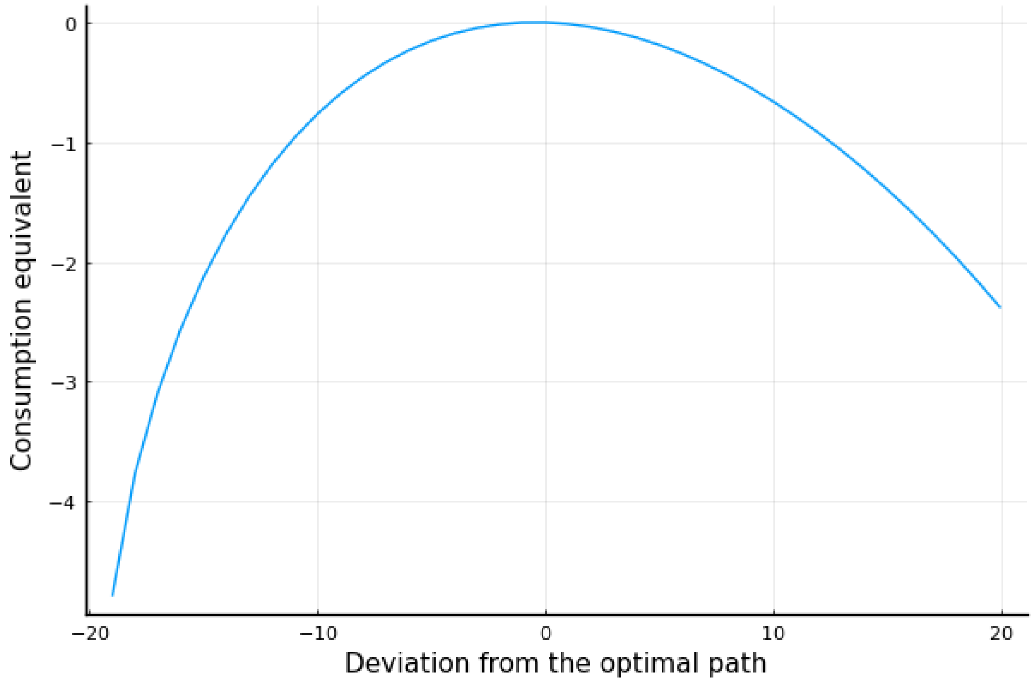

Figure 3: Welfare impact of a deviation from the optimal path. The x-axis plots the value of κ in equation (38) and the y-axis the welfare variation (in percent of consumption equivalent).

5.6 Additional robustness checks

We consider two robustness checks, one on the truncation length N, and another on the concavity parameter ϑ of the public good utility function. The resulting tax rates are reported in Table 3. We report the following optimal tax rates: the long-run truncation tax rate τ_∞^p; the fixed-point transition tax rate τ^c; the transition tax rate with low-wealth initial distribution, τ^{low}; and the transition tax with high-wealth initial distribution, τ^{high}. All these quantities are defined in Section 5.3.

We find that the long-run truncation tax rate is basically unchanged when we move from $N = 5$ to $N = 7$. This shows that the ξ^j efficiently capture the overall heterogeneity, even when the truncation length is not too long. The second column of Table 3 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 $\tau^p_\infty(\vartheta = 65\%) = 15\%$. The optimal transition tax rate, τ^c, is also higher and remains close to $\tau^p_\infty(\vartheta = 24\%)$ of the benchmark case. Finally, the tax rates τ^{low} and τ^{high} are also higher and the ranking $\tau^{low} < \tau^c < \tau^p < \tau^{high}$ is preserved. The corresponding figures are provided in Appendix E.

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

Table 3: Robustness checks.

5.7 Introducing aggregate shocks

As a further check of the truncation method, we compare the dynamics of the same economy with aggregate shocks when simulated using the truncation method (Step 1) and using the Reiter method (Step 3).

We start by introducing an aggregate risk that affects the TFP, denoted by Z_t. The production function (1) becomes: $Y_t = Z_t K^\alpha_{t-1} \overline{L}^{1-\alpha} - \delta K_{t-1}$.

5.7.1 Computing the full instrument path with aggregate shocks

Algorithm 5 can be adapted to handle aggregate shocks.

Algorithm 5 (Simulating the full instrument path with aggregate shocks) *We consider as given a truncation length, $N > 0$, and a precision criterion, $\varepsilon > 0$.*

1. Steady state. *We use Algorithm 3 to compute the steady-state Ramsey allocation with precision ε and truncation length N.*

2. Truncated model. *The dynamics of the truncated model with truncation length N are characterized by equations (49)–(57) of Appendix C.*

3. Aggregate shock. *We specify a functional form for the dynamics of TFP.*

4. Initial conditions. *We have two exclusive cases:*

   (a) Timeless perspective. *We set the initial wealth distribution $(a_{-1,y^N})_{y^N}$ and the initial distribution of Lagrange multipliers $(\lambda_{-1,y^N})_{y^N}$ to their steady-state values.*

   (b) Time-0 perspective. *The initial wealth distribution can be set arbitrarily and the initial distribution of Lagrange multipliers $(\lambda_{-1,y^N})_{y^N}$ is set to 0.*

5. Simulation. *Based on the steady-state allocation of Step 1, the model equations of Step 2, the aggregate shock of Step 3, and the initial conditions of Step 4, we simulate the path of model variables (including planner's instruments) using perturbation techniques.*

Algorithm 5 proposes a versatile approach that covers time-0 and timeless approaches. The timeless approach assumes that the economy has converged to its steady state and is then hit by a shock. The planner optimally reacts to the shock by changing the instrument $(T_t)_t$ and letting the economy adapt to the shock and the new instrument path. As stated by McCallum and Nelson (2000), this timeless perspective is the closest notion to "optimal policy making according to a rule". In the time-0 perspective, the economy is simply assumed to start from date 0 (with or without a shock). The main difference between the two perspectives is the commitment of the planner. In the timeless perspective, the planner is assumed to be constrained by its past commitments (i.e., those of the steady state), whereas in the time-0 perspective there is no past commitment. The truncation method makes it easy to switch between the two perspectives because it mostly reduces to adapting the initial values of Lagrange multipliers on Euler equations that enter into the Ramsey first-order conditions (18). The timeless approach involves setting the initial distribution of Lagrange multipliers, as well as that for wealth, to their steady-state value, thereby mimicking the commitment of the planner and the convergence to the steady state. This is Step 4(a) of Algorithm 5. Conversely, the time-0 approach (Step 4(b)) involves setting past Lagrange multipliers to zero, reflecting the absence of past commitment. The initial wealth distribution can be arbitrary to study the effect of initial inequality on the optimal instrument path. However, in the remainder of this paper, we neutralize this effect and set Lagrange multipliers to their steady-state value, similarly to what is done for the transition method.

Both perspectives can accommodate the presence of an aggregate shock (Step 3).[22] We specify the dynamics of TFP in Step 2. A standard specification assumes that the log of TFP follows a standard AR(1) process – as we do in Section 5.7.2 –, but the algorithm can handle more complex processes. The model equations to be simulated are summarized in Appendix C. As for Algorithm , the perturbation method can be performed using existing software.

Algorithm 5 involves the same two assumptions as Algorithm 4: (i) constant ξ^j along the transition, and (ii) a constant set of credit-constrained truncated histories. The latter constraint means that aggregate shocks must remain small enough not to affect the set of credit-constrained histories.

[22]The timeless perspective combined with no aggregate shock simply involves constant paths for the instrument and allocation, and is therefore of little interest.

5.7.2 Comparison with Reiter in the timeless perspective

As is standard, we assume that the TFP process follows an AR(1) process, with $Z_t = \exp(z_t)$ and $z_t = \rho_z z_{t-1} + \varepsilon_t^z$, where $\varepsilon_t^z \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_z^2)$. We use the common of values $\rho_z = 0.95$ and $\sigma_z = 0.31\%$ to obtain a deviation of the TFP shock z_t equal to 1% at a quarterly frequency (see Den Haan, 2010, for instance). We then follow Algorithm 4 in a timeless perspective to simulate the economy with TFP aggregate shocks over a simulation of 10,000 periods. This allows us to obtain the paths of the optimal truncation tax $(T_t)_t$ and economic aggregates, including aggregate consumption $(C_t)_t$ and GDP $(Y_t)_t$. We use these data to approximate the optimal tax path using two observable aggregates of the model, capital and GDP. More precisely, we run the following regression:[23]

$$T_t = T_\infty + a_Y (Y_t - Y) + a_C (C_t - C) + \varepsilon_t^T, \text{ where } \varepsilon_t^T \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_T^2) \tag{39}$$

and T_∞, C, and Y are the steady-state values computed in the Bewley model. We obtain the following values: $a_Y = 0.0442$, and $a_C = 0.0672$ with an R^2 equal to 0.9982. This allows us to capture the optimal dynamics of the tax path using only GDP and capital.

The estimated equation (39) is then incorporated into a full-fledged heterogeneous-agent model with aggregate shock as an exogenous rule. The model is then simulated using the Reiter method. Indeed, the tax path is then considered as exogenous and the model simulation does not involve any optimization, making simulation via the Reiter method possible.[24]

We compare the outcomes of the truncation and Reiter methods, and report the impulse response functions (IRFs) for the main variables in Figure 4. The two methods are labeled "Truncation" and "Reiter." We also plot the aggregate welfare in the two cases' economies, using the aggregate welfare, as the 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, and capital), and prices (interest rate and wages). A small difference between the two simulations arises from the fact that the rule estimated in equation (39) is close, but not exactly equal, to the actual dynamics of T, which are difficult to capture in the very first periods. We complement the findings of Figure 4 by reporting in Table 4 the second-order moments in the two simulations (the Reiter and truncation methods). As was the case for IRFs, the second-order moments are very similar in the two cases.

5.8 Computational results

We provide here additional statistics about the convergence of the truncation method. The previous section has shown that the Reiter model and the truncation method with $N = 5$ are very

[23] We have considered more involved regressions, but these give little actual improvement on the fit.

[24] 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.

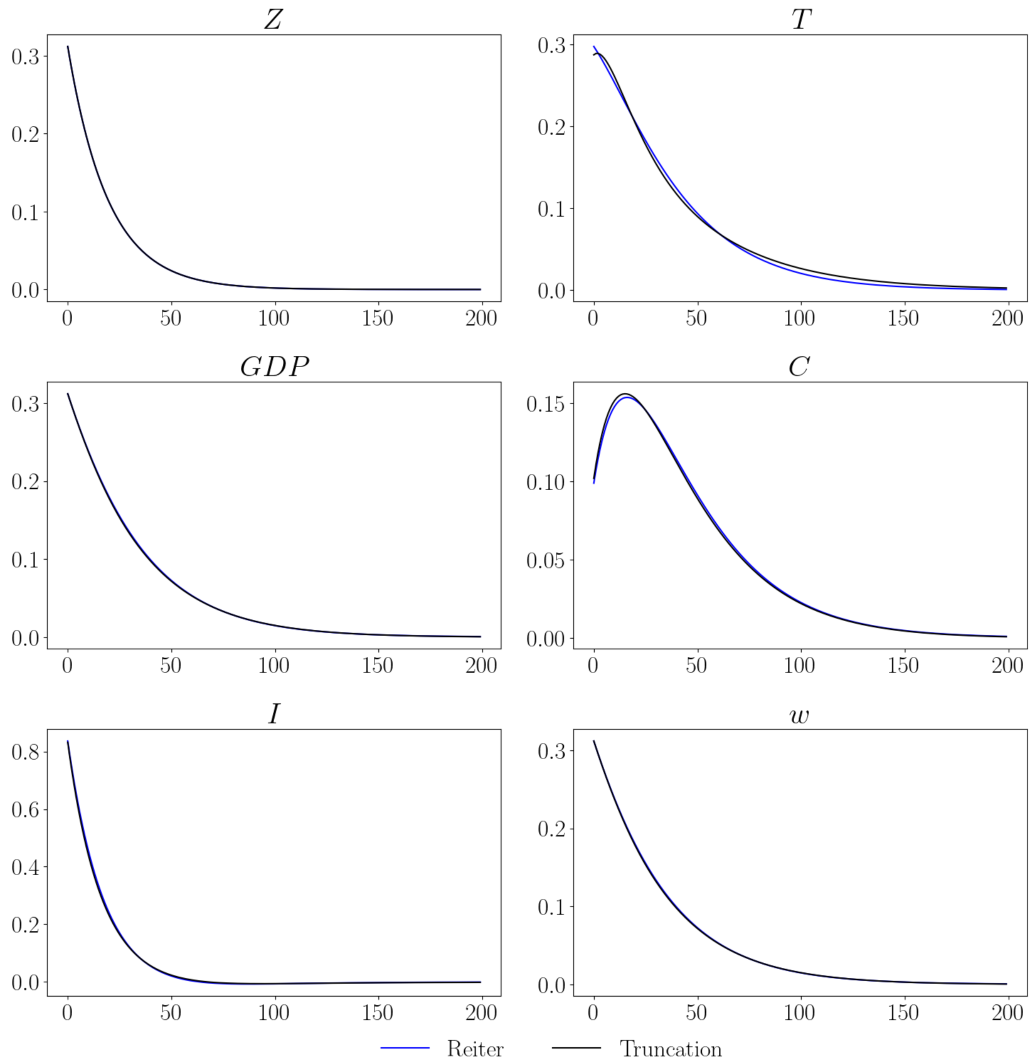

Figure 4: Simulated IRFs after a TFP shock simulated using the truncation and Reiter methods. See the text for the details of the implementations.

Methods		Reiter	Trunc. ($N = 5$)
GDP	Mean	3.793	3.793
	Std/mean (%)	1.288	1.281
C	Mean	2.475	2.475
	Std/mean (%)	0.978	0.979
K	Mean	40.590	40.590
	Std/mean (%)	1.225	1.201
Corr(C, C_{-1})		0.9945	0.9942
Corr(GDP, GDP_{-1})		0.9695	0.9692
Corr(GDP, C)		0.9242	0.9294

Table 4: Moments of the simulated model for different computational techniques.

close (up to the approximation of the optimal tax function 39). We here compare the outcomes of the truncation method with different truncation length N. We also report computational times. Results are provided in Table 5.

mN	2	3	4	5	6
Distance d_N	5.5×10^{-4}	4.1×10^{-4}	2.7×10^{-4}	1.3×10^{-4}	-
Comp. time	3s	5s	20s	2.20min	26min

Table 5: Simulation outcomes of the truncation method for different N. See the text and equation (40) for the definition of the distance d_N.

We apply the truncation method with truncation lengths from $N = 2$ to $N = 6$. In each case, we simulate the economy for 10,000 periods. We then compare the truncation tax paths for the different truncation lengths. We denote by $(T_{N,t})_{t=1,\ldots,10000}$ the simulated truncation tax path for the truncation length $N = 2, \ldots 6$. We then define the distance d_N as:

$$d_N = \max_{t=1,\ldots,10000} \left(\text{abs} \left(\frac{T_{N,t} - T_{6,t}}{T_{6,t}} \right) \right), \tag{40}$$

which is the maximum relative deviation in tax paths of the truncation method between arbitrary N and $N = 6$ (which is the maximum truncation length).

First, one can observe that the maximum relative distance d_N is decreasing with N (line "Distance d_N" of Table 5). An interesting result is that the case $N = 2$ does already a fairly good job and allows one to compute a realistic optimal path for the instrument. Compared to LeGrand and Ragot (2022a), where this result is already discussed, the introduction of multiple ξ^j, instead of a unique set, further improves the accuracy. A proportional deviation of 0.055% (case $N = 2$) implies that the time series for $N = 2$ and $N = 6$ can barely be distinguished.

Second, we report the computing time to simulate the model (Line "Comp. time" of Table 5), once the steady state of the Bewley model (which is the same for all simulations) has been computed.[25] For $N = 2$ the simulation of the model for 10,000 periods take less than 3s. The computation times increases very rapidly with N, as the number of truncated histories increases. For $N = 6$, the computation time is 26 minutes.

6 Conclusion

We solve for an optimal Ramsey policy with commitment in a 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. Because the latter method is recently developed, we explain in detail the algorithm and the implementation strategy. Considering the transition method, we propose a modification of the current technique to avoid the influence of initial distributions on the optimal value of the instrument. However, the method is still affected by a time-inconsistency issue, which may create substantial deviation from the true long-run value of the instrument. Using the truncation method, we first derive the first-order conditions of the Ramsey program for the optimal instrument value. We then show that it provides an accurate outcome even with a relatively short truncation length. However, despite the success of the truncation method in the setting of this paper, it is possible that in some environments, the length of the truncation required for correct accuracy would imply a high-dimensional problem.

We summarize the characteristics of the two methods in Table 6.

Characteristic	**Transitions**	**Truncation**
Instrument path	constant	time-varying
Long-run instrument value	Average path value	indep. of initial distribution
Initial distribution	flexible	flexible but difficult for when initial and long-run distributions differ a lot
Consistency	time-inconsistent	time-inconsistent
Aggregate shocks	not accommodated	accommodated via perturbation method

Table 6: Summary of the two methods.

Declarations of interest: None.

[25]The computation of the steady state takes less than 2s, using a EGM method with 100 grid points for each productivity level, exponentially spaced. In any case, this computational time is mostly independent of N.

Appendix

A Computing the FOCs of the Ramsey program

The Ramsey program (10)–(15) can be written using two instruments only: savings $(a_t^i)_i$ and lump-sum tax T_t. We also incorporate the Euler equation into the planner's objective. Recalling that the Euler equation Lagrange multiplier is $\beta^t \lambda_t^i$, the Ramsey program (10)–(15) can equivalently be expressed as follows:

$$\max_{(T_t,(a_t^i)_{i\in I})_{t\geq 0}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\int_i u(c_t^i)\ell(di) + v(T_t) \right) \tag{41}$$

$$- \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \int_i (\lambda_t^i - \lambda_{t-1}^i (1 + F_K(\int_i a_{t-1}^i \ell(di), \overline{L})))u'(c_t^i)\ell(di),$$

$$\text{where: } c_t^i = (1 + F_K(\int_i a_{t-1}^i \ell(di), \overline{L}))a_{t-1}^i - a_t^i + F_L(\int_i a_{t-1}^i \ell(di), \overline{L})y_t^i - T_t.$$

FOC with respect to a_t^i. Computing the derivative of (41) 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:

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

With definition (16) of ψ_t^j, this yields:

$$\psi_t^i = \beta\mathbb{E}_t \left[(1 + r_{t+1})\psi_{t+1}^i \right] + \beta\mathbb{E}_t \Bigg[\underbrace{\int_j \psi_{t+1}^j \left(F_{KK,t} a_t^j + F_{KL,t} y_{t+1}^j \right) \ell(dj)}_{\text{Indirect effect of the change in saving}} \Bigg] \tag{42}$$

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

$$\lambda_t^i = 0, \text{ if } i \text{ is credit-constrained.} \tag{43}$$

FOC with respect to T_t. Computing the derivative of (41) 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 definition (16) 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 Complete market economy

The complete-market economy is used as a benchmark to identify the effects of heterogeneity. We start with the first-best allocation, which maximizes aggregate welfare subject to a given initial capital K_{-1} and to the economy-wide resource constraint. Let C_t denote the consumption of the representative agent, and the economy-wide resource constraint can be written as:

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

which is obviously similar to equation (8) in the general case. The first-best allocation is determined as the solution of the following program:

$$\max_{(K_t, C_t, G_t)_{t \geq 0}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left(u(C_t) + v(G_t) \right) \right], \tag{44}$$

$$K_t + C_t + G_t = F(Z_t, K_{t-1}, \overline{L}) + K_{t-1}, \tag{45}$$

where we use the governmental budget constraint $G_t = T_t$ and K_{-1} is given. The two first-order conditions of the first-best program can be written as:

$$u'(C_t) = \beta \mathbb{E}_t \left[(1 + F_K(Z_{t+1}, K_t, \overline{L}))u'(C_{t+1}) \right], \tag{46}$$

$$v'(G_t) = u'(C_t). \tag{47}$$

Equations (46) and (47) together with budget constraint (45) 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 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, the individual budget constraint can be written as:

$$C_t + K_t = K_{t-1} + F_K(Z_t, K_t, \overline{L})K_{t-1} + F_L(Z_t, K_t, \overline{L})\overline{L} - T_t, \tag{48}$$

where we use the financial market and labor market clearing conditions. Combining (48) with the constant returns to scale property of the production function and the governmental budget constraint implies that the individual budget constraint is identical to resource constraint (45). Therefore, because the representative agent is endowed with the whole amount of capital at the initial date and because the first-best FOC (46) is identical to the Euler equation of the representative agent, a competitive allocation in which the fiscal policy is the same as the first-best will be identical to the first-best allocation.[26] The Ramsey planner can thus implement the first-best allocation by choosing a fiscal policy according to FOC (47).

To draw the parallel with the Lagrangian approach we used in the general case, observe that in the absence of heterogeneity, equation (42) simplifies 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.[27] Intuitively, the agents' Euler equation is not a constraint for the planner because it corresponds to the first-best intertemporal allocation of capital. Consequently, the Lagrange multiplier on the Euler equation is null and the Lagrangian approach implies $\psi_t^{CM} = u'(C_t) = v'(G_t)$, such that FOCs (17) and (47) are actually identical.

A direct consequence is that in our setup, with an instrument that has only within-period effects, there is no time-inconsistency issue in the complete market setup.

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

[27]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 (42) could be written as $A_t\lambda_t^{CM} + B_{t-1}\lambda_{t-1}^{CM} = 0$ for non-zero coefficients A_t and B_{t-1}. With $\lambda_{-1}^{CM} = 0$, this implies by induction $\lambda_t^{CM} = 0$ at all dates.

C Truncated model

The aggregation of the equations characterizing the individual model implies that the full dynamics of the truncated model can be written as follows:

$$v'(T_t) = \sum_{k=1}^{N_{tot}} S_k \psi_{k,t}, \tag{49}$$

$$K_t = \sum_{k=1}^{N_{tot}} S_k a_{k,t}, \tag{50}$$

$$k = 1...N_{tot}: \; \tilde{a}_{k,t} = \frac{1}{S_k} \sum_{k'=1}^{N_{tot}} S_{k'} \Pi_{k',k} a_{k',t-1}, \tag{51}$$

$$c_{k,t} + a_{k,t} = (1 + r_t)\tilde{a}_{k,t} + w_t y_0, \tag{52}$$

$$\tilde{\lambda}_{k,t} = \frac{1}{S_{k,t}} \sum_{k'=1}^{N_{tot}} S_{k'} \Pi_{k',k} \lambda_{k',t-1}, \tag{53}$$

$$\psi_{k,t} = \xi_k^1 u'(c_{k,t}) - (\lambda_{t,k} - \tilde{\lambda}_{t,k}(1 + r_t))\xi_k^2 u''(c_{k,t}). \tag{54}$$

$$\xi_k^1 u'(c_{k,t}) = \beta \mathbb{E}_t \left[(1 + r_{t+1}) \sum_{k'=1}^{N_{tot}} \Pi_{k,k'} \xi_{k'}^1 u'(c_{k',t+1}) \right] + \nu_{k,t}, \tag{55}$$

$$\begin{aligned} k \notin \mathcal{C}: \; \psi_{k,t} = {} & \beta \mathbb{E}_t \left[(1 + r_{t+1}) \sum_{k'=1}^{N_{tot}} \Pi_{k,k'} \psi_{k',t+1,\tilde{y}^N} \right] \\ & + \beta \mathbb{E}_t \left[\sum_{k'=1}^{N_{tot}} S_{k'} \psi_{k',t+1} \left(\tilde{a}_{k',t+1} F_{KK}(K_t, L) + y_{k'} F_{LK}(K_t, L) \right) \right], \\ & + \beta F_{KK}(K_t, L) \mathbb{E}_t \left[\sum_{k'=1}^{N_{tot}} S_{k'} \tilde{\lambda}_{k',t+1} \xi_{k'}^1 u'(c_{k',t+1}) \right], \end{aligned} \tag{56}$$

$$k \in \mathcal{C}, \; a_{k,t} = \lambda_{k,t} = 0. \tag{57}$$

D Matrix representation

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

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

- $\boldsymbol{y_0} = (y_{k,0})_{k=1\ldots N_{tot}}$ is the vector of current productivity levels across histories;
- $\boldsymbol{\xi}^j = (\xi_k^j)_{k=1\ldots N_{tot}}$ $(j = 0, 1, 2)$ are the vectors of residual-heterogeneity parameters defined in equation (23)–(25) (at the steady state);
- $\mathcal{C}$ 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;
- $\boldsymbol{D_{u'(c)}}= \text{diag}((u'(c_k))_{k=1\ldots N_{tot}})$ 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:

- $\circ$ is the term-by-term product of two vectors of the same size, which is another vector of the same size: $\boldsymbol{x} \circ \boldsymbol{z} = (x_{y^N}) \circ (z_{y^N}) = (x_{y^N} z_{y^N})$;[28]
- $\times$ is 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, which 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 the 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 the equations characterizing the steady-state of the truncated economy using this notation. We start with equation (20):

$$\boldsymbol{S} = \boldsymbol{\Pi}^\top \times \boldsymbol{S}, \tag{58}$$

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.[29] The vector $\boldsymbol{S}$ is thus straightforward to compute.

Second, we consider equation (29) for per-capita beginning-of-period wealth $\tilde{\boldsymbol{a}}$, which yields:

$$\tilde{\boldsymbol{a}} = (1/\boldsymbol{S}) \circ (\boldsymbol{\Pi}^\top \times (\boldsymbol{S} \circ \boldsymbol{a})), \tag{59}$$

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

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

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 with a null-size history yields a null wealth.

Third, the budget constraint (30) becomes at the steady state, where the tax is T_∞:

$$\boldsymbol{c} + \boldsymbol{a} = (1+r)\tilde{\boldsymbol{a}} + w\boldsymbol{y_0} - T_\infty \mathbf{1}_{N_{tot}}, \tag{60}$$

which allows us 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 (59).

Euler equations (7) can be written as follows with matrix notation:

$$\boldsymbol{\xi}^1 \circ u'(\boldsymbol{c}) = \beta(1+r)\boldsymbol{\Pi}\left(\boldsymbol{\xi}^1 \circ u'(\boldsymbol{c})\right) + \boldsymbol{\nu},$$

which defines the vector $\boldsymbol{\nu}$.

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

$$\Pi^\lambda_{k,k'} = S_{k'}\Pi_{k',k}\frac{1}{S_k}, \tag{61}$$

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

Using (61), definitions (33) and (34) for $\tilde{\lambda}_k$ and ψ_k imply in matrix form:

$$\tilde{\boldsymbol{\lambda}} = \boldsymbol{\Pi}^\lambda \boldsymbol{\lambda}, \tag{62}$$

$$\boldsymbol{\psi} = \boldsymbol{\xi}^1 \circ u'(\boldsymbol{c}) - \boldsymbol{D}_{\xi^2 \circ u''(\boldsymbol{c})}(\boldsymbol{I} - (1+r)\boldsymbol{\Pi}^\lambda)\boldsymbol{\lambda}. \tag{63}$$

We start by expressing first-order condition (42) with respect to saving choices, which only holds for unconstrained truncated histories:

$$\begin{aligned}\boldsymbol{P}\boldsymbol{\psi} = {} & \beta(1+r)\boldsymbol{P}\boldsymbol{\Pi}\boldsymbol{\psi} + \beta\boldsymbol{P}1_{N_{tot}}(\boldsymbol{S} \circ (F_{KK}(K,L)\tilde{\boldsymbol{a}} + F_{LK}(K,L)\boldsymbol{y})^\top \boldsymbol{\psi} \\ & + \beta F_{KK}(K,L)\boldsymbol{P}1_{N_{tot}}\left(\boldsymbol{S} \circ \boldsymbol{\xi}^1 \circ u'(\boldsymbol{c})\right)^\top \boldsymbol{\Pi}^\lambda \boldsymbol{\lambda},\end{aligned} \tag{64}$$

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 $1_{N_{tot}}\left(\boldsymbol{S} \circ \boldsymbol{\xi}^1 \circ u'(\boldsymbol{c})\right)^\top$ are $N_{tot} \times N_{tot}$ matrices. The pre-multiplication in (64) by the matrix $\boldsymbol{P}$ is because FOC (42) holds only for unconstrained histories. We define the following two matrices:

$$\boldsymbol{L}_0 = \boldsymbol{I} - \beta(1+r)\boldsymbol{\Pi} - \beta 1_{N_{tot}}(\boldsymbol{S} \circ (F_{KK}(K,L)\tilde{\boldsymbol{a}} + F_{LK}(K,L)\boldsymbol{y}), \tag{65}$$

$$\boldsymbol{L}_1 = \beta F_{KK}(K,L)1_{N_{tot}}\left(\boldsymbol{S} \circ \boldsymbol{\xi}^1 \circ u'(\boldsymbol{c})\right)^\top \boldsymbol{\Pi}^\lambda, \tag{66}$$

such that first-order condition (42) becomes:

$$\boldsymbol{L}_0\boldsymbol{\psi} = \boldsymbol{L}_1\boldsymbol{\lambda}. \tag{67}$$

Using definition (63) to express $\boldsymbol{\psi}$ using $\boldsymbol{\lambda}$, we deduce:

$$\boldsymbol{P}(\boldsymbol{L}_1 + \boldsymbol{D}_{\boldsymbol{\xi}^2\circ u''(\boldsymbol{c})}(\boldsymbol{I} - (1+r)\boldsymbol{\Pi}^{\lambda}))\boldsymbol{\lambda} = \boldsymbol{P}\boldsymbol{L}_0(\boldsymbol{\xi}^1 \circ u'(\boldsymbol{c})). \tag{68}$$

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

$$(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\lambda} = 0. \tag{69}$$

Combining equations (68) and (69) yields the important result:

$$\boldsymbol{\lambda} = \boldsymbol{L}_2^{-1}\boldsymbol{P}\boldsymbol{L}_0(\boldsymbol{\xi}^1 \circ u'(\boldsymbol{c})), \tag{70}$$

$$\text{with: } \boldsymbol{L}_2 = \boldsymbol{I} - \boldsymbol{P} + \boldsymbol{P}(\boldsymbol{L}_1 + \boldsymbol{D}_{\boldsymbol{\xi}^2\circ u''(\boldsymbol{c})}(\boldsymbol{I} - (1+r)\boldsymbol{\Pi}^{\lambda})). \tag{71}$$

Equation (70) 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 depend on the allocation) of equations (65), (66), and (71). Matrix $\boldsymbol{L}_2$ is invertible when $r > 0$.

Finally, we deduce from (67) and (69):

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

$$\boldsymbol{\psi} = \boldsymbol{M_2}(\boldsymbol{\xi}^1 \circ u'(\boldsymbol{c})),$$

$$\text{where: } \boldsymbol{M_1} := \boldsymbol{L}_2^{-1}\boldsymbol{P}\boldsymbol{L}_0, \tag{72}$$

$$\boldsymbol{M_2} := \boldsymbol{I} - \boldsymbol{D}_{\xi^2\circ u''(\boldsymbol{c})}(\boldsymbol{I} - (1+r)\boldsymbol{\Pi}^{\lambda})\boldsymbol{M_1}. \tag{73}$$

E Robustness check

We summarize below the results for an economy with a different value for the curvature of the public good ($\vartheta = 65\%$). In this economy, the optimal value of tax as a share of GDP is 15%, as computed by the truncation method. The optimal value of 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. Figure 5 plots welfare as a function of tax-to-GDP.

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

Finally, we also report a robustness check for the second-order moments in the presence of

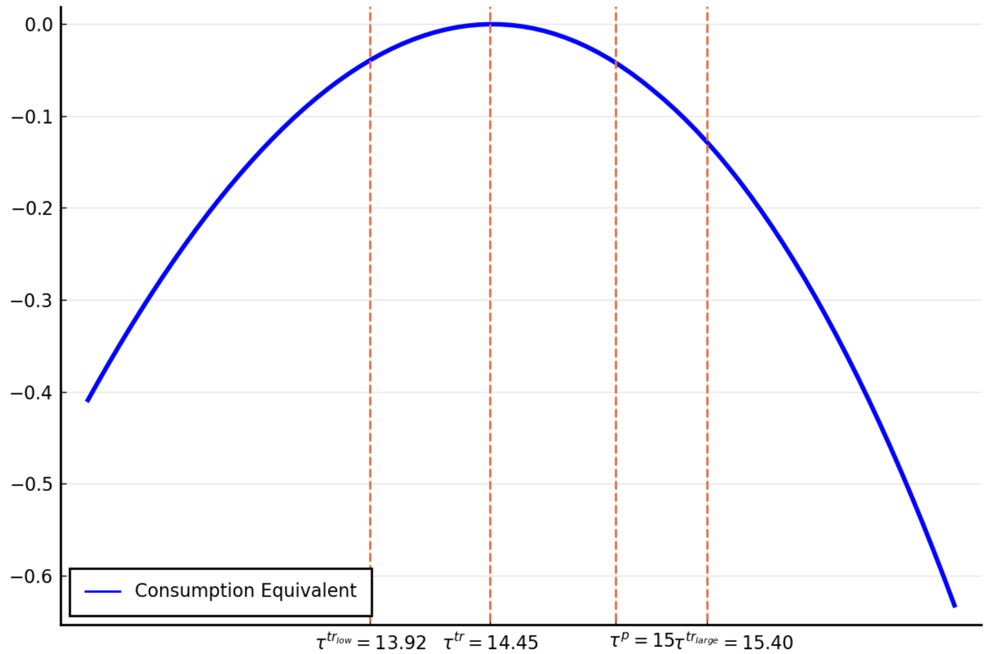

Figure 5: Welfare with transitions and consistent initial distribution computed as a function of the tax rate.

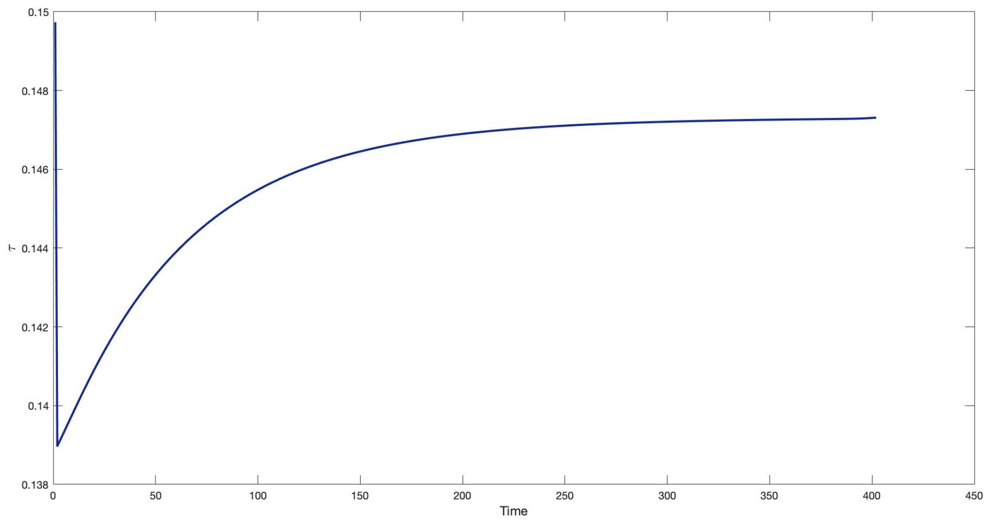

Figure 6: Dynamic of the tax after a reoptimization shock when the curvature of v is $\vartheta = 65\%$.

aggregate risk, as described in Section 5.7.2. Table 7 makes it clear that the truncation length has little impact on second-order moments.

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
Corr(C, C_{-1})		0.9945	0.9941	0.9941	0.9941	0.9942
Corr(GDP, GDP_{-1})		0.9695	0.9691	0.9691	0.9692	0.9692
Corr(GDP, C)		0.9242	0.9300	0.9298	0.9296	0.9294

Table 7: Moments of the simulated model for different computational techniques.

References

Açikgöz, O., M. Hagedorn, H. Holter, and Y. Wang (2022): "The Optimum Quantity of Capital and Debt," Working Paper, University of Oslo.

Açikgöz, 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 (2023): "Optimal Monetary Policy According to HANK," *American Economic Review*, 113, 1741–1782.

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.

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*, 89, 2559–2599.

Bilbiie, F. (2021): "Monetary Policy and Heterogeneity: An Analytical Framework," Working Paper, Lausanne University.

Bilbiie, F. and X. Ragot (2021): "Optimal Monetary Policy and Liquidity with Heterogeneous Households," *Review of Economic Dynamics*, 41, 71–95.

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.

Dávila, E. and A. Schaab (2022): "Welfare Assessments with Heterogeneous Individuals," Working Paper, Yale University.

Debortoli, D., j. Maih, and R. Nunes (2014): "Loose Commitment in medium-scale Macroeconomic Models: Theory and Applications," *Macroeconomic Dynamics*, 18, 175–198.

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 (2022): "Optimal Fiscal Policy in a Model with Uninsurable Idiosyncratic Shocks," *Review of Economic Studies*, 90, 744–780.

Ferriere, A. and G. Navarro (2023): "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 (2021): "Optimal Income Taxation: Mirrlees Meets Ramsey," *Journal of Political Economy*, 129, 3141–3184.

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.

LeGrand, F., A. Martin-Baillon, and X. Ragot (2022): "Should Monetary Policy Care About Redistribution? Optimal Fiscal and Monetary Policy with Heterogeneous Agents," Working Paper, SciencesPo.

LeGrand, F. and X. Ragot (2022a): "Managing Inequality over the Business Cycle: Optimal Policies with Heterogeneous Agents and Aggregate Shocks," *International Economic Review*, 63, 511–540.

——— (2022b): "Refining the Truncation Method to Solve Heterogeneous-Agent Models," *Annals of Economics and Statistics*, 146, 65–92.

McCallum, B. and E. Nelson (2000): "Timeless Perspective vs. Discretionary Monetary Policy in Forward-Looking Models," NBER Working Paper 7915, National Bureau of Economic Research.

McKay, A. and C. Wolf (2023): "Optimal Policy Rules in HANK," Working Paper, FRB Minneapolis.

Nuño, G. and C. Thomas (2022): "Optimal Redistributive Inflation," *Annals of Economics and Statistics*, 146, 3–64.

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.

Smirnov, D. (2022): "Optimal Monetary Policy in HANK," Working paper, Universitat Pompeu Fabra.

Yang, Y. (2022): "Redistributive Inflation and Optimal Monetary Policy," Working paper, Princeton University.