REFINING THE TRUNCATION METHOD TO SOLVE HETEROGENEOUS-AGENT MODELS

Author(s): François Le Grand and Xavier Ragot

Source: *Annals of Economics and Statistics*, June 2022, No. 146 (June 2022), pp. 65-92

Published by: GENES on behalf of ADRES

Stable URL: https://www.jstor.org/stable/10.2307/48674139

Annals of Economics and Statistics, Number 146, June 2022

REFINING THE TRUNCATION METHOD TO SOLVE HETEROGENEOUS-AGENT MODELS

FRANÇOIS LE GRAND[a] AND XAVIER RAGOT[b]

We present a refinement of the uniform truncation method of LeGrand and Ragot (2022) to solve heterogeneous-agent models with aggregate shocks. The method consists in providing a finite state-space representation of such economies by truncating idiosyncratic histories. The innovation compared to the uniform method is to allow for truncated histories of different lengths. This offers a finer representation when needed, while considerably reducing the model dimensionality. The method reproduces the steady-state distribution of any heterogeneous-agent model and solves for its dynamics in the presence of aggregate shocks. As with the uniform method, the refined method can be solved using perturbation methods and hence implemented with standard software, such as Dynare. We show that the refined truncation method provides accurate results that improve on those of the uniform method.

JEL Codes: D31, D52, E21.
Keywords: Heterogeneous Agents, Truncation Method, Aggregate Shocks.

1. INTRODUCTION

Models with incomplete insurance markets for idiosyncratic risks are now a standard workhouse in quantitative macroeconomics. The application of these so-called heterogeneous-agent models is not limited to macroeconomics, since it also covers household finance, corporate finance, and international finance, among other fields.

This paper presents a refinement of the uniform truncation method of LeGrand and Ragot (2022) to solve these models in the presence of aggregate shocks. The *uniform* truncation of LeGrand and Ragot (2022) assumes that the N last periods provide sufficient statistics to capture the relevant heterogeneity (and not the whole idiosyncratic history). This parameter N is called the truncation length. The method then aggregates in each period all agents sharing the same history for the last N periods as if they were the same agent. Since the method truncates histories at the same date, this results in so-called *uniform truncated histories*. Compared to other computational solution methods, a first interest of the truncation is to generate a finite state-space representation. This property considerably eases the simulation of the model dynamics, as the simulations can rely on standard packages, such as Dynare (Adjemian, Bastani, Juillard, Karamé, Maih, Mihoubi, Perendia, Pfeifer, Ratto, and Villemot, 2011). As a consequence, introducing additional frictions – that are sometimes called "bells-and-whistles" in the DSGE literature – is straightforward. A second interest of the truncation method is to allow one to solve optimal

This paper has benefited from the comments of many participants in seminars where the truncation method has been presented. We thank Pablo Winant and two anonymous referees for excellent suggestions on an earlier version of this paper. We also thank Diego Sousa for outstanding research assistance. We acknowledge financial support from the French National Research Agency (ANR-20-CE26-0018 IRMAC). Codes and details about the algorithm can be found at `https://github.com/RagotXavier/Truncation_Method_Het`.

[a]Emlyon business school and ETH Zurich. `legrand@em-lyon.com`
[b]Sciences Po, CNRS and OFCE. `xavier.ragot@sciencespo.fr`

Ramsey programs in incomplete-market environments, which is of independent interest.[1]

However, the main drawback of the uniform truncation method is that the number of truncated histories – and hence the model dimensionality – increases exponentially with the truncation length. If there are k idiosyncratic states, the number of histories is proportional to k^N, which can be large. We present in this paper a new method, called the *refined* truncation method, which differs from the uniform one by allowing for truncated histories of various lengths. Instead of truncating all histories after N periods, the refined method offers to truncate some histories after a longer length, which yields a finer and more precise representation for these histories. This finer representation is obtained in spite of a small number of truncated histories to track. Indeed, the number of truncated histories increases linearly with the refinement length, instead of increasing exponentially with the uniform length. The refined method has thus two main advantages compared to the uniform one: a targeted finer representation for some histories and an overall small number of histories. Furthermore, it preserves the benefits of the uniform method: finite state-space representation and tool to solve Ramsey models.

We present the refined truncation method and its implementation in practice. We illustrate its simplicity and accuracy through an example, in the spirit of Den Haan, Judd, and Juillard (2010). We study a model where households facing idiosyncratic productivity risk have to decide their consumption-saving trade-off. We compare the solution implied by the refined and uniform truncation methods to those of standard alternative methods such as Reiter (2009). The refined truncation method appears to be accurate, even when it involves only a parsimonious finite state-space representation.

Numerical accuracy of the method is obtained thanks to a generalization of the aggregation procedure developed in LeGrand and Ragot (2022) for the uniform truncation method. In the full-fledged incomplete-market model, there are different agents with the same truncated history, but these agents generally have different complete histories (in particular prior to the truncation date). Each truncated history thus features within-heterogeneity that the truncation method accounts for using *residual-heterogeneity parameters*. With the refined truncation, we can control the magnitude of this within-heterogeneity and extend the truncation length for histories in which the within-heterogeneity is too large. The combination of refinement with residual-heterogeneity parameters leads to a very accurate dynamics for the truncated representation, even when the number of truncated histories remains small. In our quantitative exercise, we show that a short uniform truncation length, coupled with long refinement lengths (that thus keeps an overall small number of truncated histories) provides accurate results that are comparable to those of other standard methods (such as Reiter's).

This paper belongs to the literature on solution techniques for heterogeneous-agent models. After the seminal contributions of Krusell and Smith (1998) and Rios-Rull (2001), different solutions have been proposed in the literature. Den Haan (2010) has presented a comparison of these methods. Since then, the method of Reiter (2009) has become the most popular technique to solve for heterogeneous-agent models in discrete times. This method has been refined in posterior developments, in Ahn, Kaplan, Moll, Winberry, and Wolf (2017), Winberry (2018), and Bayer, Luetticke, Pham-Dao, and Tjaden (2019) among

[1] See LeGrand, Martin-Baillon, and Ragot (2021) and LeGrand and Ragot (2022) for the application of the uniform truncation method to solve for optimal policies in these environments. See also Bhandari, Evans, Golosov, and Sargent (2021) for an alternative method in environments without binding credit constraints.

others. Recently, Boppart, Krusell, and Mitman (2018) and Auclert, Bardóczy, Rognlie, and Straub (2021) have presented new methods that are both based on the deterministic non-linear simulations of the economy following small "MIT shocks" that are then used to linearly approximate equilibrium solutions. These methods are known to be quantitatively close to each other and to that of Reiter (2009). Some other methods use continuous-time techniques to facilitate the resolution (see Kaplan, Moll, and Violante, 2018, or Nuño and Moll, 2018, among others). Finally, the current paper provides a refinement on the initial presentation of the truncation method in LeGrand and Ragot (2022). We offer a solution to control for the magnitude of the heterogeneity within truncated histories, while their number remains small.

The paper is organized as follows. Section 2 presents the environment. The truncation method is presented in Section 3. A numerical example is provided in Section 4. Section 5 is the conclusion.

2. THE ENVIRONMENT

The objective of this paper is to present the truncation methodology in a simple setup, so as to make the exposition of the method transparent. We rely on the same environment as does Den Haan (2010), which is a simple heterogeneous-agent economy with aggregate productivity risk. This will allow us to compare the truncation method with other simulation techniques that are known to provide accurate results.[2] The focus of the paper being the quantitative implementation of the truncation method, we do not provide mathematical convergence results (see LeGrand and Ragot, 2022 for a more technical presentation of a similar setup).

We consider a one-good economy populated by a group of ex-ante identical agents. Time is discrete and indexed by $t = 0, 1, \ldots$

2.1. *Risks*

The economy is affected by two types of risk: an aggregate risk and an individual one. The aggregate risk, denoted Z_t, solely affects total factor productivity (TFP). It takes values in a possibly continuous set. Finally, the aggregate risk is assumed to be Markovian. Its precise dynamics will be specified when needed.

The other risk is the idiosyncratic risk. While each agent is assumed to provide an inelastic labor supply normalized to 1, labor productivity is individual and stochastic. The productivity risk is agent-specific, and we assume that agents cannot insure against it. Asset markets are thus incomplete with respect to this productivity risk. The productivity is denoted y and assumed to take value in a finite set $\mathcal{Y}$, where the productivity levels are assumed to be ordered. Higher values of y correspond to higher productivity levels. The productivity process of a given agent follows a first-order Markov chain with constant transition probabilities $(\Pi_{yy'})_{y,y' \in \mathcal{Y}}$, which are in particular assumed to be independent of the aggregate risk. The probability that an agent currently endowed with productivity y will have productivity y' in the following period is equal to $\Pi_{yy'} \in [0, 1]$. Furthermore, transition probabilities verify: $\Pi_{yy'} \geq 0$ and $\sum_{y' \in \mathcal{Y}} \Pi_{yy'} = 1$, reflecting that probabilities

[2] It is interesting to note that since the comparison of Den Haan (2010), the literature has evolved to mostly use the method of Reiter (2009) (with possible improvements) and the recent method of Boppart, Krusell, and Mitman (2018) using the simulation of transitions.

must be nonnegative and that independently of the current productivity y, the agent will be endowed with some productivity level y' in the next period. Finally, an individual history of productivity shocks up to date t is denoted by $y^t = \{y_0, \ldots, y_t\} \in \mathcal{Y}^{t+1}$.

2.2. *Production*

The production sector is standard. A representative profit-maximizing firm produces the sole consumption good of the economy, by combining labor and capital. We consider a Cobb-Douglas production function with constant returns-to-scale, a capital share $\alpha \in (0,1)$, and a capital depreciation rate $\delta \in (0,1)$. Since the individual labor supply is fixed and normalized to 1, the aggregate labor supply is constant and denoted by $\overline{L} > 0$. The TFP Z_t is stochastic. Formally, the production at date t of Y_t units of good is defined as:

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

where the capital K_{t-1} is requested to be installed one period in advance. 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:

$$w_t = (1-\alpha) Z_t K_{t-1}^{\alpha} \overline{L}^{-\alpha} \text{ and } r_t = \alpha Z_t K_{t-1}^{\alpha-1} \overline{L}^{1-\alpha} - \delta. \tag{2}$$

2.3. *Agents' preferences and program*

The instantaneous utility function over consumption is denoted by u. As is standard, $u : \mathbb{R}_+ \to \mathbb{R}$ is assumed to be twice continuously differentiable, strictly increasing, and strictly concave, with $u'(0) = \infty$. Agents are expected-utility maximizers who maximize the expected sum of discounted utility: $\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t)$, where: $0 < \beta < 1$ is a constant discount factor, $(c_t)_{t \geq 0}$ a consumption path, and $\mathbb{E}_0$ an expectation operator over future aggregate and individual shocks.

Agents can save in each period by trading capital shares that pay off the real interest rate r_t. Agents are furthermore prevented from short-selling capital shares and hence face a possible borrowing constraint. The combination of market incompleteness with exogenous borrowing limits is the usual market imperfection in the heterogeneous-agent literature (Bewley, 1983; Imrohoroğlu, 1992; Huggett, 1993; Aiyagari, 1994). Given an initial endowment a_{-1} and an initial productivity level y_0, the agent chooses her consumption path $(c_t)_{t \geq 0}$ and her saving plans $(a_t)_{t \geq 0}$ so as to maximize her intertemporal utility, subject to per-period budget constraints and borrowing limits. Formally, the agent's program can be written as:

$$\max_{(c_t, a_t)_{t \geq 0}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t), \tag{3}$$

$$c_t + a_t = (1 + r_t) a_{t-1} + w_t y_t, \tag{4}$$

$$a_t \geq 0, a_{-1} \text{ given.} \tag{5}$$

The budget constraint (4) states that the agent finances consumption c_t and savings purchases a_t out of saving payoffs $(1 + r_t) a_{t-1}$ and labor earnings $w_t y_t$ – where we recall that labor supply is normalized to 1. Equation (5) is the borrowing constraint.

Denoting by $\beta^t \nu_t$ the Lagrange multiplier on the credit constraint, the agent's Euler equation can be written as:

$$u'(c_t) = \beta \mathbb{E}_t \Big[(1 + r_{t+1})u'(c_{t+1})\Big] + \nu_t, \tag{6}$$

where $\nu_t = 0$ if the agent saves a positive amount, i.e., if $a_t > 0$.

2.4. *Resource constraints and equilibrium definition*

All agents are ex-ante identical but experience different productivity histories. This means that agents are endowed with different streams of income and hence make different choices for the trade-off between consumption and savings. So, though identical ex ante, agents will be heterogeneous ex post. To distinguish between agents, we add a superscript i to refer to an agent i in the population.[3] In the initial period, agents are distributed along a set I with measure $\ell(\cdot)$. Since the population of every agent i is $\ell(di)$, the financial market clearing condition can be written as:

$$\int_i a_t^i \ell(di) = K_t, \tag{7}$$

which formalizes that the sum of all individual savings is equal to aggregate capital. Similarly, the equilibrium on the goods markets states that the sum of all individual consumption plus the current capital should equal the production and past capital. Formally:

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

Finally, the labor market clearing condition implies a relationship between aggregate labor $\overline{L}$ and individual productivity levels:

$$\int_i y^i \ell(di) = \overline{L}, \tag{9}$$

that can also be seen as a definition of $\overline{L}$.

We can now state our market equilibrium definition.

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

1. *given prices, the functions* $(c_t^i, a_t^i)_{t\geq 0, i\in\mathcal{I}}$ *solve the agent's optimization program in equations (3)–(5);*
2. *financial and goods markets clear at all dates: for any* $t \geq 0$*, equations (7) and (8) hold;*
3. *factor prices* $(w_t, r_t)_{t\geq 0}$ *are consistent with condition (2).*

[3] To simplify the exposition, we discuss the recursive representation below.

2.5. *The standard recursive approach*

A difficulty with the equilibrium of Definition 1 is that it features an infinite number of different agents' consumption and saving levels. These agents indeed differ along their productivity history, and therefore along savings and consumption choices. However, the formulation of Definition 1 is unhelpful to tackle the computation of the equilibrium, as the distribution $\ell(\cdot)$ offers very little structure and suggests that we should solve for *all* individual programs and then iterate on prices until the financial market clears. Obviously, given the dimensionality of the problem, this is not an achievable goal.

The standard approach in the literature since Imrohoroğlu (1992) and Huggett (1993) consists in taking advantage of the recursive formulation. More precisely, the heterogeneity in productivity histories is summarized in two idiosyncratic state variables: the beginning-of-period asset holding and the current productivity level. In other words, instead of reasoning on sequences of productivity level realizations, the recursive method focuses on this pair (beginning-of-period asset holding and current productivity level). That the individual sequential program admits an equivalent recursive formulation in an economy without aggregate shocks is proved in Huggett (1993) and generalized in Açikgöz (2018). The recursive formulation lends itself to computing the equilibrium quite easily *without* aggregate shocks (Z being constant). There are several methods to compute the policy function (Endogenous Grid Method of Carrol, 2006, being one of the most efficient ones, see `https://julia.quantecon.org/dynamic_programming/egm_policy_iter.html` for lecture notes). Once policy functions have been calculated, a stationary distribution can be computed either by simulating a population of N agents over a given number of periods T (see Rios-Rull, 2001) or by using the policy rules to construct a transition matrix over the distribution of wealth. The steady-state distribution is the fixed point of this transition matrix (Young, 2010). This distribution can then be aggregated to check whether the financial market clears or not. If the market does not clear, the interest rate is updated and the procedure repeated until market clearing. When the market clears, the equilibrium interest rate and asset distribution are known.[4]

The introduction of aggregate shocks greatly complexifies the resolution of heterogeneous-agent models, since the distribution of wealth becomes time-varying. Seminal contributions for approximating the model solution are Krusell and Smith (1998) and Rios-Rull (2001), who relied on simulation techniques. Den Haan (2010) proposes a comparison of the different methods available at that date. Since this comparison, Reiter (2009) has developed a method that is now widely adopted. More recently, Boppart, Krusell, and Mitman (2018) and Auclert, Bardóczy, Rognlie, and Straub (2021) have developed similar methods that yield results close to the method of Reiter. The truncation method will also generate results close to those obtained with other solution techniques. Its interest relies on its simplicity, allowing for new applications as the derivation of optimal policies mentioned above.

[4]Several lecture notes provide a precise exposition of the resolution of the standard model. See for instance `https://alisdairmckay.com/Notes/HetAgents/` or `https://notes.quantecon.org/submission/5f5f811909e80c001bbd6eaf`.

3. THE TRUNCATION METHOD IN PRACTICE

The truncation method is based on an aggregation procedure in the history representation (or sequential representation). We detail in this section how to use this method and compare it numerically with other solution techniques in Section 4.

3.1. *Constructing the truncated model*

The truncation requires coming back to the sequential model formulation instead of the recursive formulation. Its core idea is to group together agents who share the same productivity history over the recent past and to express the model using these groups of agents, which will act as history-specific representative agents. We present two methods to group agents according to their recent productivity history: (i) the uniform truncation, which involves an identical history length for all agents, and (ii) the refined truncation, which allows one to consider histories of different lengths. The refined truncation model is the specific contribution of this paper.

The uniform truncation.

In the uniform truncation, initially proposed in LeGrand and Ragot (2022), agents, who share a common productivity of a given length, will be grouped together into so-called *truncated histories*. Such agents will be called *companions*. This length, which is identical for all agents, will be called the *truncation length* and is a method parameter. For instance, if the truncation length is set at two periods, and if there are two productivity levels $y_H > y_L$, the truncated representation will feature four truncated histories: $\{(y_H, y_H), (y_L, y_H), (y_H, y_L), (y_L, y_L)\}$. Each agent will then be assigned to one of these four truncated histories, depending on her current and past productivity status. We will use the convention that (y_L, y_H) corresponds to agents who currently have productivity y_H and had productivity y_L in the previous period.

More generally, if we denote by $N > 0$ the truncation length, a truncated history will be a vector $y^N = (y_{-N+1}, \ldots, y_{-1}, y_0)$, where y_0 represents the current productivity status, and y_{-N+1} the productivity N periods ago. The truncation method will then consist in assigning agents with infinite history $y^\infty = (\ldots, y_{-N-1}, y_{-N}, y_{-N+1}, y_{-N+2}, \ldots, y_{-1}, y_0)$ at date t to the truncated history y^N, independently of productivity levels having occurred more than N periods ago (such as the value of y_{-N-1} or y_{-N}). The infinite history of a given agent is updated in every period, and the composition of truncated histories does not remain fixed through time. For instance, if we assume that an agent with history y^∞ at t draws the productivity $\tilde{y}_0$ at date $t+1$, her history at $t+1$ will become: $\tilde{y}^\infty = (\ldots, y_{-N-1}, y_{-N}, y_{-N+1}, \ldots, y_{-1}, y_0, \tilde{y}_0)$.[5] The agent will thus be assigned at date $t+1$ to truncated history $\tilde{y}^N = (y_{-N+2}, \ldots, y_1, y_0, \tilde{y}_0)$. With this construction, if $n_y = Card(\mathcal{Y})$ denotes the number of productivity levels, there will be $N_{tot} = n_y^N$ truncated histories of length N.

It is noteworthy that, in each period, the uniform truncation that assigns an agent with history y^∞ to truncated history y^N is well defined. Indeed, for any agent (more precisely for any history y^∞), there exists one, and exactly one, truncated history $y^N \in \mathcal{Y}^N$ to which

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

the agent will be assigned. The fact that the uniform truncation assignment is well defined makes the model representation with truncated histories $(y^N)_{y^N \in \mathcal{Y}^N}$ feasible. In particular, we can define the quantities $\Pi_{y^N \tilde{y}^N}$ as follows:

$$(10) \qquad \Pi_{y^N \tilde{y}^N} = 1_{\tilde{y}^N \succeq y^N} \Pi_{y_0 \tilde{y}_0},$$

where $\Pi_{y_0 \tilde{y}_0}$ is the (time-independent) probability to transit from productivity y_0 to productivity $\tilde{y}_0$, $1_{\tilde{y}^N \succeq y^N} = 1$ if $\tilde{y}^N$ is a possible continuation of y^N (i.e., if y^N is a possible past history for $\tilde{y}^N$, formally: $\tilde{y}_{-1} = y_0$, $\tilde{y}_{-2} = y_{-1}$,. . ., $\tilde{y}_{-N+1} = y_{-1}$), and 0 otherwise. The quantities $\Pi_{y^N \tilde{y}^N}$ are positive and verify $\sum_{\tilde{y}^N \in \mathcal{Y}^N} \Pi_{y^N \tilde{y}^N} = 1$ for all y^N. The matrix $(\Pi_{y^N \tilde{y}^N})_{y^N, \tilde{y}^N \in \mathcal{Y}^N}$ is thus a well-defined transition matrix for truncated histories – with $\Pi_{y^N \tilde{y}^N}$ being the transition probability from y^N to $\tilde{y}^N$. Truncated histories can thus be thought of as a first-order Markov chain with the state space $\mathcal{Y}^N$ and the transition matrix $(\Pi_{y^N \tilde{y}^N})$.

The refined truncation.

The uniform truncation has the drawback that the number of truncated histories grows exponentially with the length of the truncation. To avoid this exponential growth, this paper proposes a refined truncation method that allows considering truncated histories of unequal length. The intuition is to use longer lengths for some truncated histories that are particularly large or heterogeneous. The constraint is twofold: (i) obtain a well-defined mapping between agents (i.e., infinite histories) and these unequal-length truncated histories, and (ii) obtain a well-defined transition probability matrix between truncated histories. First, an arbitrary set of truncated histories of unequal lengths is unlikely to yield a well-defined mapping. For instance, and to provide intuition, still in the case of two productivity levels, the set $\{(y_H, y_H, y_H), (y_L, y_H), (y_H, y_L), (y_L, y_L)\}$ is not able to assign agents whose terminal history is (y_L, y_H, y_H). Oppositely, with the set $\{(y_L, y_H, y_H), (y_H, y_H), (y_L, y_H), (y_H, y_L), (y_L, y_L)\}$, the same agent with terminal history (y_L, y_H, y_H) can be assigned to two truncated histories. Second, even in the case of a well-defined mapping, we need to check that the transition matrix between truncated histories is also well-defined. Consider for example the set $\{(y_H, y_H), (y_H, y_L, y_H), (y_L, y_L, y_H), \{y_L\}\}$, which corresponds to a well-defined mapping. In that case, the transition matrix is not well-defined (the row associated to $\{y_L\}$ will not sum to 1), reflecting that agents with truncated history $\{y_L\}$ getting the productivity y_H (and hence history $\{y_L, y_H\}$) cannot be assigned to any truncated history.

What we propose here is a construction of an unequal-length truncation that enables us to obtain by construction a well-defined mapping. The starting point of our construction is the observation that only particular truncated histories deserve a longer length. Continuing the 2-state example, quantitative exercises typically involve persistent states (and more precisely the state H being more persistent than the state L). This means that when we focus on the set of uniform truncated histories $\{(y_H, y_H), (y_L, y_H), (y_H, y_L), (y_L, y_L)\}$, a (sizably) larger share of agents will be assigned to truncated histories (y_H, y_H) and (y_L, y_L). A similar observation holds for the uniform truncation of length 3, where truncated histories (y_H, y_H, y_H) and (y_L, y_L, y_L) are also the largest ones. The idea of the refined truncation is to use a longer length for these larger histories. We can illustrate this based on the 2-period uniform truncation, where we perform two refinement rounds. In the first refinement step, the history (y_H, y_H)

will be refined into (y_H, y_H, y_H) and (y_L, y_H, y_H), simply by adding one extra past period. This yields the set $\{(y_H, y_H, y_H), (y_L, y_H, y_H), (y_L, y_H), (y_H, y_L), (y_L, y_L)\}$, which by construction corresponds to a well-defined truncation mapping and a well-defined transition probability matrix.[6] In the second step, the history (y_H, y_H, y_H) can be refined into (y_H, y_H, y_H, y_H) and (y_L, y_H, y_H, y_H). We then obtain the set of truncated histories $\{(y_H, y_H, y_H, y_H), (y_L, y_H, y_H, y_H), (y_L, y_H, y_H), (y_L, y_H), (y_H, y_L), (y_L, y_L)\}$. Starting from a common length of 2 periods and a set of 4 truncated histories, a 2-round refinement leads to a longest history of 4 periods and a set of 6 truncated histories. A similar refinement can be conducted for history (y_L, y_L).

We now provide a more formal construction of a set of refined truncated histories, when there are two productivity levels. We denote a set of refined truncated histories by $R(N, N_H, N_l)$, where N is the uniform truncation length (on which the refinement is based), $N_H \geq N$ is the longest refinement history for the state h, and $N_L \geq N$ is the longest refinement history for the state l. The recursion starts from the set $R(N, N, N)$ of uniform truncated histories that are all of length N. The first refinement step consists in substituting for the history $y_H^N = (\underbrace{y_H, \ldots, y_H}_{N})$ the histories y_H^{N+1} and $(y_L, y_H^N) = (y_L, \underbrace{y_H, \ldots, y_H}_{N})$. This yields the set $R(N, N+1, N)$ that contains $2^N + 1$ histories. Going from $R(N, N+1, N)$ to $R(N, N+2, N)$ involves substituting y_H^{N+2} and (y_L, y_H^{N+1}) for y_H^{N+1}. These steps can be repeated until we obtain $R(N, N_H, N)$. Overall, this refinement has consisted to substitute for y_H^N the set $\{y_H^{N_h}\} \cup \{(y_L, y_H^k) : k = N, \ldots, N_H - 1\}$, while preserving by construction a well-defined mapping between infinite and truncated histories, as well as a well-defined transition probability matrix. The longest history is of length N_H, and the set of refined histories includes $2^N + N_H - N$ histories: increasing the maximal history length implies a linear increase in the number of truncated histories. The construction of $R(N, N_H, N_L)$ is analogous and involves refining the history y_L^N instead of y_H^N.[7] The number of truncated histories in the set $R(N, N_H, N_L)$ is $2^N + N_H + N_L - 2N$. A key implication is that it is possible in practice to combine a small value for N with large values for N_L and N_H to obtain an accurate solution. The increase in the number of histories is now linear instead of exponential.

Compared to the uniform truncation, the refined truncation enables us to target the truncated histories that deserve a more precise representation, while preserving a small total number of truncated histories. We will see in the quantitative section that the refined truncation improves the precision of the uniform truncation. We will also explain in this section how the lengths N, N_H, and N_L can be chosen – where N_L can be chosen independently of N_H as long as $N_H, N_L \geq N$.

[6] Extending an arbitrary truncated history (different from (y_H, y_H) or (y_L, y_L)) is unlikely to yield a well-defined transition matrix. See for instance the above counter-example $\{(y_H, y_H), (y_H, y_L, y_H), (y_L, y_L, y_H), \{y_L\}\}$ that corresponds to the refinement of history (y_L, y_H) in the set $\{(y_H, y_H), (y_L, y_H), (y_L, y_H), \{y_L\}\}$.

[7] Obviously, we can also similarly refine histories when the number of productive states is arbitrary (though finite).

Similarly to equation (10), we can define for any two histories h and $\tilde{h}$ of $R(N, N_H, N_L)$ the quantity $\Pi_{h\tilde{h}}$ as follows:

$$\Pi_{h\tilde{h}} = 1_{\tilde{h}\succeq h}\Pi_{y_0^h y_0^{\tilde{h}}}, \tag{11}$$

where y_0^h is the current productivity for truncated history h, and as previously, $\Pi_{y_0^h y_0^{\tilde{h}}}$ is the (time-independent) probability to transit from productivity y_0^h to productivity $y_0^{\tilde{h}}$, and $1_{\tilde{h}\succeq h} = 1$ if $\tilde{h}$ is a possible continuation of h, and 0 otherwise. Since histories h and $\tilde{h}$ can be of unequal lengths, checking that $\tilde{h}$ is a possible continuation of h is more involved than in the continuous case.[8] Importantly, with our refined truncation, it can be checked that the matrix $(\Pi_{h\tilde{h}})_{h,\tilde{h}\in R(N,N_H,N_L)}$ is a well-defined transition matrix for refined truncated histories, with $\Pi_{h\tilde{h}}$ being the transition probability from h to $\tilde{h}$.

Truncated history representation.

We now present the construction of the truncated model based on the truncated histories. The construction is identical whether we consider a uniform or a refined truncation. We consider as given a set of truncated histories constructed based on an arbitrary number of productivity levels. The cardinal of R is denoted N_{tot} and its elements will be indexed h_k, with $k = 1, \ldots, N_{tot}$.

Truncated histories can thus be thought of as representative agents, whose inside composition evolves through time – but this will be mostly hidden in model equations by the law of large numbers. In particular, the size of a truncated history h_k, denoted by S_{h_k}, can be defined recursively as:

$$S_{h_k} = \sum_{\hat{k}=1}^{N_{tot}} S_{h_{\hat{k}}}\Pi_{h_{\hat{k}}h_k}, \tag{12}$$

which reflects that agents with truncated history h had truncated history $\hat{h}$ in the previous period with probability $\Pi_{h_{\hat{k}}h_k}$.

We can also assign to each truncated history its own consumption level and its own end-of-period savings, which will be denoted c_{t,h_k} and a_{t,h_k} respectively. There is a link between individual consumption and truncated history consumption, since the latter is defined as the average of consumption levels of individuals whose infinite history maps onto the truncated history under consideration. The beginning-of-period savings, denoted $\tilde{a}_{t,h_k}$, reflect the fact that these savings are actually savings at the end of the previous period of agents having a possibly different truncated history. Similarly to the recursion (12) for sizes, end- and beginning-of-period savings are connected through the following relationship:

$$\tilde{a}_{t,h_k} = \frac{1}{S_{h_k}}\sum_{\hat{k}=1}^{N_{tot}} S_{h_{\hat{k}}}\Pi_{h_{\hat{k}}h_k}a_{t-1,h_{\hat{k}}}. \tag{13}$$

[8] If we denote $h = (y^h_{-n_h+1}, \ldots, y^h_{-1}, y^h_0)$ and $\tilde{h} = (y^{\tilde{h}}_{-n_{\tilde{h}}+1}, \ldots, y^{\tilde{h}}_{-1}, y^{\tilde{h}}_0)$, $\tilde{h} \succeq h$ iff: (i) either $n_{\tilde{h}} > n_h$ and $y^{\tilde{h}}_{-k} = y^h_{-k+1}$ for all $k = 1, \ldots, n_h$, (ii) or $n_h \geq n_{\tilde{h}}$ and $y^{\tilde{h}}_{-k} = y^h_{-k+1}$ for all $k = 1, \ldots, n_{\tilde{h}} - 1$. The latter case supersedes the uniform case.

We then deduce that the budget constraint at the truncated history level can be written as:

$$c_{t,h_k} + a_{t,h_k} = (1 + r_t)\tilde{a}_{t,h_k} + w_t y_0^{h_k}. \tag{14}$$

The interpretation is similar to the individual budget constraint (4), and the underlying movement of individual agents between history is not visible once the expression of $\tilde{a}_{t,h}$ is understood.

The second equation characterizing the truncated-history economy is the Euler equation. If we denote by $\mathcal{C}_t$ the set of credit-constrained truncated histories, we have:

$$\forall h_k \in R \setminus \mathcal{C}_t,\ \xi_{t,h_k} u'(c_{t,h_k}) = \beta \mathbb{E}_t \left[(1 + r_{t+1}) \sum_{\tilde{k}=1}^{N_{tot}} \Pi_{h_k h_{\tilde{k}}} \xi_{t+1,h_{\tilde{k}}} u'(c_{t+1,h_{\tilde{k}}}) \right], \tag{15}$$

$$\forall h_k \in \mathcal{C}_t,\ a_{t,h_k} = 0, \tag{16}$$

where the expectation is over aggregate shocks only – the expectation over future truncated histories having been written explicitly. The parameters $(\xi_{t,h})_h$ that appear in (15) can be understood as *residual heterogeneity* parameters, which capture the fact that the truncated history features within-heterogeneity. Indeed, although companions having the same truncated history share by construction the same history over some periods, they differ along their productivity levels in the distant past (prior to the length of the truncated history). The truncated history therefore gathers partly heterogeneous individuals, and this heterogeneity is taken care off in the Euler equation through the so-called ξs. Finally, equation (16) reveals that truncated histories in $\mathcal{C}_{t,N,N_h,N_l}$ are credit-constrained, and that this corresponds to zero asset holdings.

Role of the ξs.

The ξs capture heterogeneity among companions. In the full model, the ξs are time-varying, while in our method, they will be computed using the steady-state allocation and will be kept constant in the presence of aggregate shocks. With this key assumption, the model simulation simplifies into the simulation with a finite set of agents (i.e., the truncated histories), and standard perturbation techniques can be used. Moreover, when heterogeneity is small among companions of the same truncated history h, the corresponding ξ_h will be close to 1. We explore below the quantitative cost of imposing $\xi = 1$ as an approximation (instead of the steady-state values of ξ). This cost appears to be very small for well-chosen truncations, as shown in Section 4.3.1.

To summarize, the truncated model is characterized by parameters $(\xi_h)_k$, the probabilities $(\Pi_{h_k h_{\tilde{k}}})_{k\tilde{k}}$ are derived from the initial transition matrix and equation (10), and the sizes $(S_{h_k})_k$ of (12) and allocations $(c_{t,h_k}, a_{t,h_k}, \tilde{a}_{t,h_k})_{h_k}$ are determined in equations (13)–(15). Importantly, the number of unknowns and equations is finite (and equal to $5N_{tot} + N_{tot}^2$), which offers a limited-heterogeneity model. The truncated model can be thought of as a model with N_{tot} representative agents with different relative sizes and some within-heterogeneity that is addressed by the ξs.

We now present the practical implementation of the truncation method to solve models with aggregate shocks. The method relies on three steps. First, we explain how to select the ξ parameters. In particular, we show how to derive these ξ-parameters from any given steady-state distribution of wealth across histories $(a_{t,h_k})_k$, for a given set R of truncated histories. This initial distribution can be given by different sources, but it is more consistent to derive it from a well-defined Bewley model. The second step is thus to derive this initial distribution of wealth across histories from the aggregation of a full-fledged Bewley model. Third, knowing the ξs, one can use simple perturbation techniques to compute the dynamics with aggregate shocks, using packages like Dynare. The comparison of the outcome of the truncation model with other solution techniques will be done in Section 4.

3.2. *Implementation: From allocation to the determination of the ξs*

We focus here on the steady state that corresponds to a fixed TFP: $Z_t = Z$ for some $Z > 0$ at all dates. Steady-state variables are denoted without time subscript. We denote by $\mathcal{C}$ the set of credit-constrained truncated histories. We consider as given a steady-state distribution of end-of-period wealth, which we denote by a_{h_k}, $1 \leq k \leq N_{tot}$. Our objective is to obtain the corresponding ξ_{h_k}, $1 \leq k \leq N_{tot}$.

The first step is to use this distribution to determine the set of credit-constrained truncated histories $\mathcal{C}$. A truncated history will be assumed to be credit-constrained when end-of-period saving is null: If $a_{h_k} \simeq 0$, then $h_k \in \mathcal{C}$. We deduce from equations (12)–(16) that the steady-state economy is then characterized by the following set of equations:

$$\tilde{a}_{h_k} = \frac{1}{S_k} \sum_{k'=1}^{N_{tot}} S_{h_{k'}} \Pi_{h_k h_{k'}} a_{h_{k'}}, \tag{17}$$

$$c_{h_k} + a_{h_k} = (1+r)\tilde{a}_{h_k} + w y_0^{h_k}, \tag{18}$$

$$h_k \notin \mathcal{C},\ \xi_{h_k} u'(c_{h_k}) = \beta(1+r) \sum_{k'=1}^{N_{tot}} \Pi_{h_k h_{k'}} \xi_{h_{k'}} u'(c_{h_{k'}}), \tag{19}$$

$$h_k \in \mathcal{C},\ a_{h_k} = 0. \tag{20}$$

A very convenient way to express the truncated model allocation involves using matrix notation. We introduce the following notations:

- $\boldsymbol{S} = (S_{h_k})_{k=1,\ldots,N_{tot}}$ the N_{tot}-vector of sizes;
- $\boldsymbol{\Pi} = (\Pi_{h_k h_{k'}})_{k,k'=1,\ldots,N_{tot}}$ the $N_{tot} \times N_{tot}$ matrix of transition probabilities across histories;
- $\boldsymbol{c} = (c_{h_k})_{k=1,\ldots,N_{tot}}$, $\boldsymbol{a} = (a_{h_k})_{k=1,\ldots,N_{tot}}$, $\tilde{\boldsymbol{a}} = (\tilde{a}_{h_k})_{k=1,\ldots,N_{tot}}$, the N_{tot}-vectors of allocations (consumption, end-of-period and beginning-of-period savings);
- $\boldsymbol{y_0} = (y_0^{h_k})_{k=1,\ldots,N_{tot}}$ the vector of current productivity levels across histories;
- $\boldsymbol{\xi} = (\xi_{h_k})_{k=1,\ldots,N_{tot}}$ the vector of residual-heterogeneity parameters;
- $\boldsymbol{P} = \text{diag}(p_k)_{k=1,\ldots,N_{tot}}$, with $p_k = 1$ if $h_k \notin \mathcal{C}$ and $p_k = 0$ if $h_k \in \mathcal{C}$, is a diagonal $N_{tot} \times N_{tot}$-matrix;
- $\boldsymbol{D_{u'(c)}}$= $\text{diag}(u'(c_{h_k}))$ the diagonal $N_{tot} \times N_{tot}$-matrix with $u'(c_{h_k})$ on the diagonal at rank k, and 0 elsewhere;
- $\boldsymbol{I}$ the identity matrix;
- $\mathbf{1}_{N_{tot}}$ the N_{tot}-vector of 1.

We also introduce the following operations:

- $\odot$ the term-by-term product of two vectors of the same size, which is another vector of the same size: $\boldsymbol{x} \odot \boldsymbol{z} = (x_{h_k}) \odot (z_{h_k}) = (x_{h_k} z_{h_k})$;[9]
- $\times$ the usual matrix product: e.g., for a matrix $\boldsymbol{M}$ and a vector $\boldsymbol{x}$ (of length equal to the number of columns of $\boldsymbol{M}$), $\boldsymbol{M} \times \boldsymbol{x}$ is the vector $(\sum_{k'} M_{h_k h_{k'}} x_{h_{k'}})_k$.

To avoid heavy notations, we still denote without a sign the usual scalar multiplication – that is assumed to apply to matrices and vectors (e.g., $\lambda \boldsymbol{M} = (\lambda M_{h_k h_{k'}})_{k,k'=1,\ldots,N_{tot}}$) and with $+$ the addition – that is extended to matrices and vectors of the same size (e.g., $\boldsymbol{x} + \boldsymbol{z} = (x_k + z_k)$). We also keep the same notation for functions that apply element-wise to vectors: $f(\boldsymbol{x}) = (f(x_{h_k}))_{k=1,\ldots,N_{tot}}$.

We can rewrite equations characterizing the steady state of the truncated economy using this notation and explain how to construct the vector $\boldsymbol{\xi}$. We start with equation (12):

$$\text{(21)} \qquad \boldsymbol{S} = \boldsymbol{\Pi} \times \boldsymbol{S},$$

which makes it clear that the vector of sizes, $\boldsymbol{S}$, is the eigenvector of matrix $\boldsymbol{\Pi}$ associated to the eigenvalue 1, where the sum of the eigenvector coordinates is normalized to 1.[10] The vector $\boldsymbol{S}$ is thus straightforward to compute.

Second, equation (17) for per capita beginning-of-period wealth $\tilde{\boldsymbol{a}}$, which yields:

$$\text{(22)} \qquad \tilde{\boldsymbol{a}} = (1/\boldsymbol{S}) \odot (\boldsymbol{\Pi} \times (\boldsymbol{S} \odot \boldsymbol{a})),$$

where $1/\boldsymbol{S} = (1/S_{h_k})_{k=1,\ldots,N_{tot}}$ is the vector of size inverses and $\boldsymbol{a}$ is the given vector of end-of-period wealth. Note that if the size of the truncated history is $S_{h_k} = 0$, we can set $1/S_{h_k} = 0$, which will be consistent such that a null-size history will get a null wealth.

Third, the budget constraint (18) becomes:

$$\text{(23)} \qquad \boldsymbol{c} + \boldsymbol{a} = (1 + r)\tilde{\boldsymbol{a}} + w\boldsymbol{y_0},$$

which allows one to obtain consumption levels using the given vector $\boldsymbol{a}$ of end-of-period wealth, and the vector of beginning-of-period wealth of equation (22).

The final step is to compute the residual heterogeneity parameters, $\boldsymbol{\xi}$. We proceed in two steps. First, as a normalization, we set to 1 the ξs in constrained histories: $\xi_{h_k} = 1$ if $h_k \in \mathcal{C}$. This can be written as:

$$\text{(24)} \qquad (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\xi} = (\boldsymbol{I} - \boldsymbol{P})\mathbf{1}_{N_{tot}}.$$

To understand (24), observe that the matrix $\boldsymbol{I} - \boldsymbol{P}$ is a diagonal matrix with a 1 coefficient for constrained histories and a zero coefficient for constrained histories. In other words, $\boldsymbol{P}$ is a projection matrix on the set of constrained histories. For a generic vector $\boldsymbol{x}$, the product $(\boldsymbol{I} - \boldsymbol{P}) \times \boldsymbol{x}$ therefore selects the coordinates of $\boldsymbol{x}$ corresponding to constrained histories (completed by zeros for unconstrained histories, such that the whole vector is of length N_{tot}). This explains why equation (24) means $\xi_{h_k} = 1$ if $h_k \in \mathcal{C}$. Second, the Euler

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

[10] The existence of a positive eigenvector vector is guaranteed by the Perron-Frobenius theorem for the positive matrix $\boldsymbol{\Pi}$ whose rows sum to 1.

equation (19) for unconstrained histories can be written as:

$$\boldsymbol{P} \times \boldsymbol{D}_{\boldsymbol{u'(c)}} \times \boldsymbol{\xi} = \boldsymbol{P} \times \beta(1+r)\boldsymbol{\Pi} \times \boldsymbol{D}_{\boldsymbol{u'(c)}} \times \boldsymbol{\xi}, \tag{25}$$

The intuition for equation (25) parallels the one for equation (24). Since $\boldsymbol{P}$ is a projection matrix on the set of unconstrained histories, it means that the relationship (25) holds only for unconstrained histories. Equation (25) can be written as $\boldsymbol{P} \times (\boldsymbol{I} - \beta(1+r)\boldsymbol{\Pi}) \times \boldsymbol{D}_{\boldsymbol{u'(c)}} \times \boldsymbol{\xi} = 0$.Adding it to (24), we obtain:

$$\boldsymbol{\xi} = \left(\boldsymbol{P} \times (\boldsymbol{I} - \beta(1+r)\boldsymbol{\Pi}) \times \boldsymbol{D}_{\boldsymbol{u'(c)}} + \boldsymbol{I} - \boldsymbol{P}\right)^{-1} (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{1}_{N_{tot}}, \tag{26}$$

where it can be shown that the matrix $\boldsymbol{P} \times (\boldsymbol{I} - \beta(1+r)\boldsymbol{\Pi}) \times \boldsymbol{D}_{\boldsymbol{u'(c)}} + \boldsymbol{I} - \boldsymbol{P}$ is invertible.[11] The determination of the vector $\boldsymbol{\xi}$ thus involves solving a linear system, which is fast in our applications.

We summarize these results in the following Algorithm.

ALGORITHM 1 (Calibration of the ξs) *Consider a given initial distribution $\boldsymbol{a}$. The set of credit-constrained histories $\mathcal{C}$ can be determined as:*

$$\mathcal{C} = \{k = 1, \dots, N_{tot} : a_{h_k} = 0\},$$

while the steady-state allocation and the residual heterogeneity parameters ξ can be found through the following set of equations:

$$\begin{aligned}
\tilde{\boldsymbol{a}} &= (1/\boldsymbol{S}) \odot (\boldsymbol{\Pi} \times (\boldsymbol{S} \odot \boldsymbol{a})),\\
\boldsymbol{c} &= (1+r)\tilde{\boldsymbol{a}} - \boldsymbol{a} + w\boldsymbol{y_0},\\
\boldsymbol{\xi} &= \left(\boldsymbol{P} \times (\boldsymbol{I} - \beta(1+r)\boldsymbol{\Pi}) \times \boldsymbol{D}_{\boldsymbol{u'(c)}} + \boldsymbol{I} - \boldsymbol{P}\right)^{-1} (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{1}_{N_{tot}}.
\end{aligned}$$

Algorithm 1 allows one to easily find the vector $\boldsymbol{\xi}$ and the equilibrium allocation, which corresponds to a given equilibrium wealth distribution $\boldsymbol{a}$. Even though Algorithm 1 works for any acceptable distribution $\boldsymbol{a}$ (such as the one coming from some empirical estimation), we can derive further results on the convergence of this algorithm with the truncation length N when the distribution $\boldsymbol{a}$ is derived from the aggregation of the steady-state wealth distribution of a full-fledged Bewley model (LeGrand and Ragot, 2022). We now explain how to obtain such a distribution.

3.3. *Deriving the steady-state wealth distribution for a Bewley model*

The initial wealth distribution $\boldsymbol{a}$ is the vector of end-of-period wealth for all truncated histories $h \in R$. For each h, to obtain the component a_h in $\boldsymbol{a}$ from a Bewley model, the intuition is to apply the proper sequence of policy rules that is consistent with truncated history h.

Let us proceed with a more formal description. The solution of a Bewley model is characterized by a steady-state wealth distribution and a set of policy rules for savings. The

[11] The matrix $\boldsymbol{I} - \beta(1+r)\boldsymbol{\Pi}$ is always invertible. Indeed, the Gershgorin circle theorem implies that all eigenvalues of $\boldsymbol{\Pi}$ are of modulus smaller than 1. Since $\beta(1+r) < 1$ at the steady state (see Açikgöz, 2018, for instance), the eigenvalues of $\beta(1+r)\boldsymbol{\Pi}$ remain strictly below 1. It is then straightforward to deduce that the columns of $\boldsymbol{P} \times (\boldsymbol{I} - \beta(1+r)\boldsymbol{\Pi}) \times \boldsymbol{D}_{\boldsymbol{u'(c)}} + \boldsymbol{I} - \boldsymbol{P}$ are independent and hence the matrix is invertible.

steady-state wealth distribution is a mapping $\mu : (a, y) \in [0; +\infty) \times \mathcal{Y} \mapsto \mu(a, y) \in \mathbb{R}_+$, such that $\mu(da, y)$ corresponds to the (steady-state) measure of agents having a wealth lying in interval $[a, a + da)$ and productivity level y. Similarly, saving policy rules are mappings $g_a : (a, y) \in [0; +\infty) \times \mathcal{Y} \mapsto g_a(a, y) \in \mathbb{R}_+$, where $g_a(a, y)$ corresponds to the end-of-period wealth for an agent having the beginning-of-period wealth a and the current productivity level y.

To obtain the vector $\boldsymbol{a}$ of Algorithm 1, we need to compute the wealth distribution $\tilde{\mu} : (a, h) \in [0; +\infty) \times R \mapsto \tilde{\mu}(a, h) \in \mathbb{R}_+$, where $\tilde{\mu}(da, h)$ is the measure of agents with wealth in interval $[a, a + da)$ and truncated history $h = (y^h_{-n_h+1}, \ldots, y^h_{-1}, y^h_0)$. Such a measure can be computed by starting from the wealth distribution of agents in state $y^h_{-n_h+1}$, which is $\mu(\cdot, y^h_{-n_h+1})$, and then applying successively the sequence of policy rules corresponding to $h = (y^h_{-n_h+1}, \ldots, y^h_{-1}, y^h_0)$. Applying first $g(\cdot, y^h_{-n_h+1})$ yields the end-of-period wealth distribution of agents in state $y^h_{-n_h+1}$. We then apply the policy rule $g(\cdot, y^h_{-n_h+2})$ to obtain the end-of-period wealth distribution for agents with history $(y^h_{-n_h+1}, y^h_{-n_h+2})$. Finally, we apply the remaining sequence of policy rules, $g(\cdot, y^h_{-n_h+3}), \ldots, g(\cdot, y^h_0)$, which ultimately yields the end-of-period wealth distribution of agents with truncated history h. The computational implementation of this procedure amounts to multiplying an initial distribution with n_h different transition matrices, which is very fast.

On the theoretical side, the $\boldsymbol{\xi}$-parameters constructed from the aggregation of a Bewley model can be shown to converge toward 1 when the length of the uniform truncation N increases (see LeGrand and Ragot, 2022). However, this asymptotic result is of little practical use, since N remains small in practice. Fortunately, as we check in Section 4, the $\boldsymbol{\xi}$-parameters allow one to generate accurate results for short truncation lengths. As we will see, the refined truncation further offers accuracy gains.

3.4. *Solving the model with aggregate shocks*

Once the steady-state allocation is found following Algorithm 1, the simulation of the model in the presence of aggregate shocks is simple. Standard packages, such as Dynare, can be used to solve the model using perturbation techniques. We provide the Algorithm 2, but the simulation is direct.

ALGORITHM 2 (Simulating the model with aggregate shocks) *We consider as given a truncation length, $N > 0$ and a target steady-state wealth distribution, $\boldsymbol{a}_{ss}$, and the TFP is $Z_t = \exp(z_t)$ with:*

$$z_t = \rho_z z_{t-1} + \varepsilon_t, \text{ where } \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_z^2). \tag{27}$$

1. *We use Algorithm 1 to obtain the vector $\boldsymbol{\xi}$ and the set of credit-constrained histories $\mathcal{C}$ that correspond to the steady-state wealth distribution $\boldsymbol{a}_{ss}$. The elements $\boldsymbol{\xi}$ and $\mathcal{C}$ are assumed to remain constant in the presence of aggregate shocks.*

2. *The model in the presence of aggregate shocks is thus determined by equation (27) and the following set of equations:*

$$(28) \qquad k = 1, \ldots, N_{tot}: \ \tilde{a}_{h_k,t} = \frac{1}{S_{h_k}} \sum_{k'=1}^{N_{tot}} S_{h_{k'}} \Pi_{h_{k'} h_k} a_{h_{k'},t-1},$$

$$(29) \qquad c_{h_k,t} + a_{h_k,t} = (1 + r_t)\tilde{a}_{h_k,t} + w_t y_0,$$

$$(30) \qquad h_k \notin \mathcal{C}: \ \xi_{h_k} u'(c_{h_k,t}) = \beta \mathbb{E}_t \left[(1 + r_{t+1}) \sum_{k'=1}^{N_{tot}} \Pi_{h_k h_{k'}} \xi_{h_{k'}} u'(c_{h_{k'},t+1}) \right],$$

$$(31) \qquad h_k \in \mathcal{C}, \ a_{h_k,t} = 0.$$

3. *The model in equations (27)–(31) can be simulated by perturbation methods around the steady-state wealth distribution $\boldsymbol{a}_{ss}$.*

Algorithm 2 provides a straightforward path to simulate the model in the presence of aggregate shocks. It is assumed that the log of the TFP follows a standard AR(1) process (actually, the algorithm easily extends to a more complex dynamic). The core of the simulation relies on off-the-shelf software, such as Dynare, which are already popular in macroeconomics. The two main assumptions of Algorithm 2 are that: (i) the ξs remain constant in the presence of aggregate shocks; (ii) the set-of credit-constrained truncated histories is also unaffected by aggregate shocks. The second assumption is not directly related to our method and comes from the fact that the model simulation relies on a perturbation method that does not lend itself to time-variations in credit-constrained truncated histories. The first assumption regarding the ξs means that the within-heterogeneity remains constant over time and equal to its steady-state value. However, it is noteworthy that this assumption does not mean the absence of within-heterogeneity.

From a practical aspect, the ξs and the set of credit-constrained histories are determined using Algorithm 1. The model in equations (27)–(31), featuring constant ξs and a constant set of credit-constrained histories can then be input in Dynare.

4. QUANTITATIVE EXERCISE

4.1. *The calibration*

Preferences.

The period is a quarter. The discount factor is set to $\beta = 0.98$ to obtain a realistic capital-output ratio in our environment. The period utility function is $\log(c)$.

Technology and TFP shock.

In the production function of (1), the capital share is set to $\alpha = 36\%$ and the depreciation rate to $\delta = 2.5\%$, as in Krueger, Mitman, and Perri (2018) among others. The TFP process is standard: $Z_t = \exp(z_t)$, with $z_t = \rho_z z_{t-1} + \varepsilon_t^z$, where $\varepsilon_t^z \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_z^2)$. We use the standard values $\rho_z = 0.95$ and $\sigma_z = 0.31\%$ to obtain a deviation of the TFP shock z_t equal to 1% at the quarterly frequency (see e.g., Den Haan, 2010).

TABLE I
PARAMETER VALUES IN THE BASELINE CALIBRATION. SEE TEXT FOR DESCRIPTIONS AND TARGETS.

Parameter	Description	Value
β	Discount factor	0.98
α	Capital share	0.36
δ	Depreciation rate	0.025
ρ_y	Autocorrelation idio. income	$\begin{cases} 0.99 \\ 0.97 \end{cases}$
σ_y	Standard dev. idio. income	10.1%
ρ_z	Autocorrelation TFP	0.95
σ_z	Standard deviation TFP shock	0.31%

Idiosyncratic risk.

Idiosyncratic productivity risk is the key ingredient for the model to generate a realistic earning and wealth distribution. We opt for 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, such that the persistence and variance of the labor income y match some estimated values.

To discuss the accuracy of the method, we consider two calibrations, with two values for the persistence of the idiosyncratic risk. Indeed, this last parameter appears to be crucial to calibrate the truncation, as discussed below.[12]

In the first calibration, we use a quarterly persistence of $\rho_y = 0.99$ and a quarterly standard deviation of $\sigma_y = 10.1\%$, which generate, for the log of earnings, an annual persistence of 0.9617 and an annual standard deviation of 3.96%. This first calibration, which is in line with Krueger, Mitman, and Perri (2018), generates a very high persistence of the idiosyncratic risk. In the second calibration, we reduce the persistence of the idiosyncratic risk to $\rho_y = 0.97$, keeping the standard deviation constant. This second calibration is introduced to discuss the impact of some key parameters of the model on the accuracy of the method.

In both cases, we use the Rouwenhorst (1995) procedure to discretize the productivity process into 2 idiosyncratic states with a constant transition matrix, to be consistent with the theoretical part. Table I provides a summary of the model parameters.

Table II provides the details of the two labor processes used in the paper.

The probability to switch from one state to another is 3 times lower, when the persistence increases from 0.97 to 0.99.

4.2. *Steady-state results*

In the first calibration ($\rho_y = 0.99$), we find a steady-state capital-output ratio equal to $K/Y = 2.15$, and the consumption-output ratio is $C/Y = 0.79$, which are standard statistics in the US. In the second calibration ($\rho_y = 0.97$), the capital output ratio is $K/Y = 2.09$ and the consumption-output ratio is unchanged at $C/Y = 0.79$.

[12]We considered various changes in parameters, such as the volatility of the idiosyncratic risk or the risk aversion, without any significant changes in the results.

TABLE II

PARAMETERS FOR THE LABOR PROCESS FOR THE TWO PERSISTENCES OF THE IDIOSYNCRATIC LABOR PROCESS, DISCRETIZED USING THE ROUWENHORST PROCEDURE.

Parameter	$\rho_y = 0.97$	$\rho_y = 0.99$
y	$[0.60, 1.39]$	$[0.38, 1.62]$
Π	$\begin{bmatrix} 0.985 & 0.015 \\ 0.015 & 0.985 \end{bmatrix}$	$\begin{bmatrix} 0.995 & 0.005 \\ 0.005 & 0.995 \end{bmatrix}$

We now report the steady-state values of the parameters ξs for two truncation strategies and for the two parameters in Table III. In the first calibration ($\rho_y = 0.99$), we consider two truncation strategies. The first (entitled *Uniform*) corresponds to the general truncation procedure, where all histories have the same length. This corresponds to the set of uniform truncated histories $R(N, N, N)$ for different truncation lengths N (see Section 3.1 for the definition of R). In the second strategy, we consider a refined truncation, with the set $R(1, N_H, N_H)$, where all truncated histories have a common length of 1 (i.e., two states only), and where the refinement length N_H is identical for both states. To make the two strategies comparable, we report them for a given total number of histories (N_{tot}). We then report for different values of N_{tot} the standard deviation of the ξs (in percents) for each of the persistence values and two truncation strategies. For instance, when the truncations have only four histories, the standard deviation of ξ is 10.2% for the two strategies when $\rho_y = 0.99$ and 7.3% when $\rho_y = 0.97$. When we consider 1024 histories, the standard deviation in the case when $\rho_y = 0.99$ decreases to 3.3% for the refined truncation, but only 9.0% for the uniform truncation. When the persistence is $\rho_y = 0.97$, these two values are 4.1% and 6.5%, respectively.

We observe that the Uniform truncation is dominated by the Refined truncation in all cases. Moreover, the standard deviation of the ξs decreases faster with the number of histories N_{tot} for the refined strategy than for the uniform one. A final remark is that for the refined strategy, the standard deviation does not decrease much beyond $N_{tot} = 256$. Decreasing it further would require increasing the length of common histories (for instance $R(2, N_H, N_H)$).

TABLE III

STANDARD DEVIATIONS OF THE ξs AS A FUNCTION OF THE NUMBER OF HISTORIES N_{tot}, FOR TWO PERSISTENCE VALUES.

	$\rho_y = 0.99$		$\rho_y = 0.97$	
N_{tot}	Refined	Uniform	Refined	Uniform
4	10.2	10.2	7.3	7.3
8	9.9	10.0	7.1	7.2
16	9.3	9.9	6.7	7.1
32	8.2	9.7	6.0	7.0
64	6.4	9.6	4.9	6.9
128	4.1	9.5	4.1	6.8
256	3.3	9.3	4.1	6.7
512	3.3	9.2	4.1	6.6
1024	3.3	9.0	4.1	6.5

4.3. *Results in the presence of aggregate risk with* $\rho_y = 0.99$

We now report model outcomes in the presence of aggregate risk when the idiosyncratic risk persistence is $\rho_y = 0.99$. The case $\rho_y = 0.97$ provides very similar results and is presented in Appendix A to save some space. We compare the outcomes of the refined truncation method to those of the histogram representation and perturbation methods, developed by Rios-Rull (2001), Reiter (2009), and Young (2010) among others, which we call the Reiter method for brevity.[13] The method is known to provide accurate results, when compared to the global method of Krusell and Smith (1998), as shown in Boppart, Krusell, and Mitman (2018) or in Auclert, Bardóczy, Rognlie, and Straub (2021). We also compare these results to those of a complete market economy, labeled RA for Representative Agent. This last economy is a simple RBC model, where the representative agent supplies one unit of labor with the average productivity. This allows us to quantify the effect of market incompleteness and heterogeneity on model outcomes. We compare the three cases (truncation, Reiter and RA) along three dimensions: (i) Impulse Response Functions (IRFs), (ii) Simulation paths of the economy, and (iii) Second-order moments.

Choice of the truncation lengths.

We consider a refined truncation method corresponding to truncated histories $R(1, N_H, N_L)$, with a one-length common history and a length N_L and N_H for the refined histories in the low and the high productivity states, respectively. To choose N_L and N_H, we progressively increase the refinement length for each state to simulate the dynamics, until the dynamics of the main variables barely change. As can be seen in Table III, the standard deviation of the ξs reaches a plateau, but the dynamics of aggregate variables stop being affected before the minimal standard deviation is reached. In Table IV, we report for each of the two truncation persistences the chosen truncation, the total number of histories N_{tot}, and the associated standard deviation of the ξs.

TABLE IV

CHOICE OF THE OPTIMAL TRUNCATION FOR TWO CALIBRATIONS.

Persistence ρ_y	Set of truncated histories	N_{tot}	std(ξ)
$\rho_y = 0.99$	$R(1, 5, 150)$	155	3.46
$\rho_y = 0.97$	$R(1, 5, 90)$	95	4.32

First, the refined truncation features different lengths for the high and low states. For the case $\rho_y = 0.99$, the chosen truncation corresponds to $R(1, 5, 150)$, implying a maximal history length of 5 for the low productivity state and a maximal history length of 150 for the high productivity state. The reason for this asymmetry is that low productivity agents decumulate their assets rapidly, whereas high productivity agents accumulate their assets progressively. As a consequence, the length of the history of high productive agents needs to be high, while a relatively short history for the low productivity is sufficient to capture the relevant heterogeneity. Thanks to the refinement procedure, the total number of histories remains low and equal to $N_{tot} = 155$. The corresponding standard deviation of

[13]The perturbation of Reiter can be used with bases other than histograms, such as in Winberry (2018) or Bayer, Luetticke, Pham-Dao, and Tjaden (2019).

ξs is 3.46%, which is higher than the minimal value of Table III, but decreasing it further has little impact on the economy dynamics.

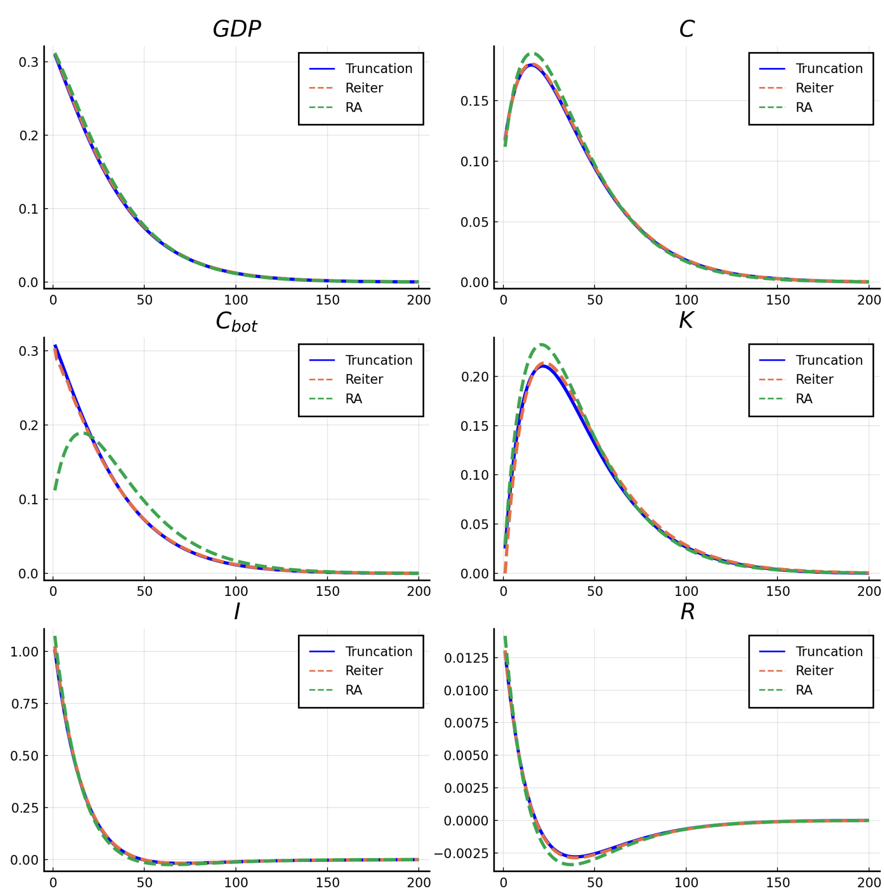
When the persistence is $\rho_y = 0.97$, the chosen truncation is $R(1, 5, 90)$. As the persistence is lower, a shorter history length for high-productive agents is sufficient, as the share of the population remaining in the high productivity state decreases faster than when $\rho_y = 0.99$. The total number of histories is now 95, and the standard deviation of the ξs is 4.32.

We now show that these two truncations deliver very accurate results.

Impulse Response Functions.

IRFs for a selection of key variables are plotted in Figure 1. We focus on the following aggregate variables: GDP (GDP), per-capita aggregate consumption (C), capital (K), and investment (I), as well as the real gross interest rate (R). We also plot the IRF for the per-capita consumption of the bottom 10% in the wealth distribution (C_{bot}).

Figure 1: Comparisons of IRFs for the main variables after a technology shock of 1%. The black line is the Reiter method. The blue dashed line is the truncation method.

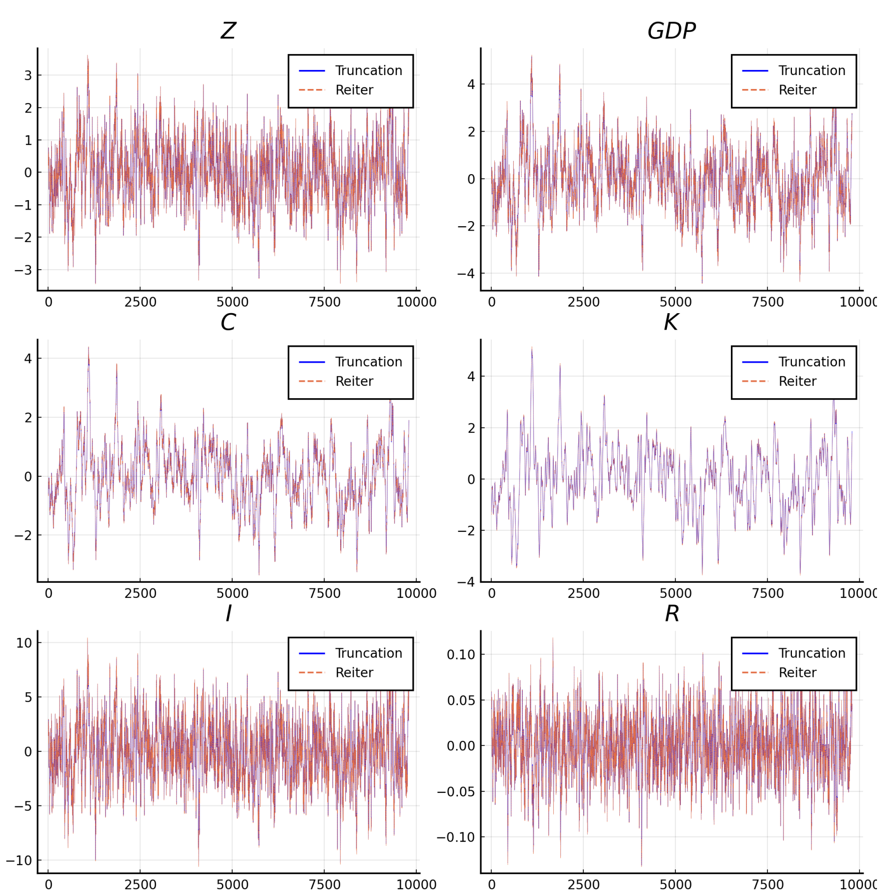


We can draw several lessons from Figure 1. First, the truncation and the Reiter methods yield very close outcomes for all variables of interest – even for those that are very volatile, such as the capital stock, K. Second, the outcomes of the two methods in the incomplete-market economy are sizably different from those of the RA economy. The difference is particularly striking for the per-capita consumption of the bottom 10% in the wealth distribution, C_{bot}.[14] While consumption in the RA economy is by construction the same as the average consumption, the patterns for C_{bot} and C are very different in the incomplete-market model (independently of the solution method). As a conclusion of Figure 1, the assumption of constant ξs along the business cycle does not harm the outcomes of the truncation method, and the time-varying within-history heterogeneity has a second-order effect on the dynamics. This holds in spite of a sizable effect of heterogeneity, as can be seen from the comparison with the RA economy.[15]

Simulation outcomes

Following the Boppart, Krusell, and Mitman (2018) comparison strategy, we plot in Figure 2 the model outcomes corresponding to one simulation of the aggregate shock history over 10,000 periods. The history of aggregate shock is reported in the top-left panel of Figure 2. We also report the associated path for GDP, per-capita aggregate consumption, capital, investment, and interest rate.[16] In each of the panels of Figure 2,

Figure 2: Simulation of aggregate variables for different simulation methods.



[14]With the truncation method, C_{bot} is defined as the sum of consumption of truncated histories, which approximately account for the bottom 10% of the consumption distribution.

[15]In a previous version, we also simulated the model with the uniform truncation method with a similar number of histories. Results were very similar, except for the consumption of the bottom 10%, which was less accurate with the uniform truncation.

[16]We don't plot the outcomes for the RA economy, as levels are very different.

the lines corresponding to the two solution methods are indistinguishable, as they are superimposed. This confirms that the two solution methods thus generate very close model outcomes. Table V provides the median and maximum absolute deviations between the two simulations for the main variables.

TABLE V

COMPARISON OF SIMULATED VARIABLES FOR THE TWO COMPUTATIONAL TECHNIQUES.

	Y	C	K
Median abs. dev. (%)	0.01	0.01	0.04
Max. abs. dev. (%)	0.05	0.03	0.26

4.3.1. *Second-order moments*

The final element of comparison between the three solution methods is the second-order moments of key aggregate variables, as well as the auto-correlations of consumption and GDP. Each economy is simulated over 10,000 periods to compute those moments. The results are reported in Table VI. The results confirm what was said for IRFs and simulation

TABLE VI

SECOND-ORDER MOMENTS FOR DIFFERENT COMPUTATIONAL TECHNIQUES.

Methods		RA	Reiter	Ref.Trunc	$\xi = 1$
GDP	Mean	3.20	3.35	3.35	3.35
	Std/mean (%)	1.36	1.33	1.32	1.32
C	Mean	2.57	2.63	2.63	2.63
	Std/mean (%)	1.12	1.08	1.08	1.08
K	Mean	25.40	28.55	28.55	28.55
	Std/mean (%)	1.38	1.30	1.28	1.28
$corr(C, C_{-1})$ (%)		99.5	99.4	99.4	99.4
$corr(GDP, GDP_{-1})$		97.4	97.2	97.2	97.2

paths. The truncation and the Reiter methods generate very similar results, and only very small differences in the second-order moments generated by the two methods can be observed.

4.3.2. *Are the ξs useful?*

Using the refined truncation method, we set the $\xi s = 1$ (instead of their computed value) and simulate the economy. The result is reported in the last column of Table VI. We find that the simulation outcomes are not quantitatively different when we set $\xi = 1$ when we used the refined truncation method. The result is different from LeGrand and Ragot (2022), who find the ξs are key for the accuracy of the uniform truncation method. The difference in the results comes from the refined truncation. We indeed find that the standard deviation of ξs is as small as 3.46%. As a consequence, setting $\xi = 1$ is a second-order change in the results. We consider this result as an additional gain of the refined truncation.[17]

[17] We don't plot the IRFs with $\xi s = 1$ because they are indistinguishable from those of Figure 1 obtained with the truncation method.

From these experiments, we conclude that the refined truncation method generates very accurate results, with a small number of histories to follow. We provide in Appendix A the simulation results for the case $\rho_y = 0.97$, which confirms what we found for $\rho_y = 0.99$.

5. CONCLUSION

We have presented a refined truncation method to solve heterogeneous-agent models with aggregate shocks. This method elaborates on the uniform truncation method of LeGrand and Ragot (2022), which consists in providing a finite state-space representation of heterogeneous-agent economies by truncating idiosyncratic histories. The core idea involves considering finite-length (truncated) idiosyncratic histories instead of infinite ones. The refinement consists in allowing for truncated histories of different lengths. This allows one to better control the within-heterogeneity of finite-length histories, while keeping a small total number of histories to follow. In the quantitative exercise, the refined method is shown to yield an accurate representation of the model. As with the uniform truncation, the implementation of its refinement can rely on perturbation methods and standard packages (as Dynare). We propose examples in Julia, Matlab and Dynare as supplementary materials.

APPENDIX A: THE CASE $\rho_Y = 0.97$

We report the simulation results for the case $\rho_y = 0.97$. All other parameters are the same as the ones of Table I. We report IRFs, simulation paths, and second-order moments with the same variables as in the main text (with $\rho_y = 0.99$). All conclusions are the same; the Reiter and the refined truncation methods are very similar; and both differ from the RA economy.

TABLE VII
MOMENTS OF THE SIMULATED MODEL ($\rho_y = 0.97$) FOR THE TWO COMPUTATIONAL TECHNIQUES AND THE REPRESENTATIVE AGENT ECONOMY.

Methods		RA	Reiter	Trunc
Economies		(1)	(3)	(4)
GDP	Mean	3.20	3.35	3.35
	Std/mean (%)	1.36	1.33	1.32
C	Mean	2.57	2.63	2.63
	Std/mean (%)	1.12	1.09	1.08
K	Mean	25.40	28.55	28.55
	Std/mean (%)	1.38	1.30	1.27
$corr(C, C_{-1})$ (%)		99.5	99.4	99.4
$corr(GDP, GDP_{-1})$		97.4	97.2	97.2

Figure 3: Comparisons of IRFs for the main variables after a technology shock of 1%. The black line is the Reiter method. The blue dashed line is the truncation method.

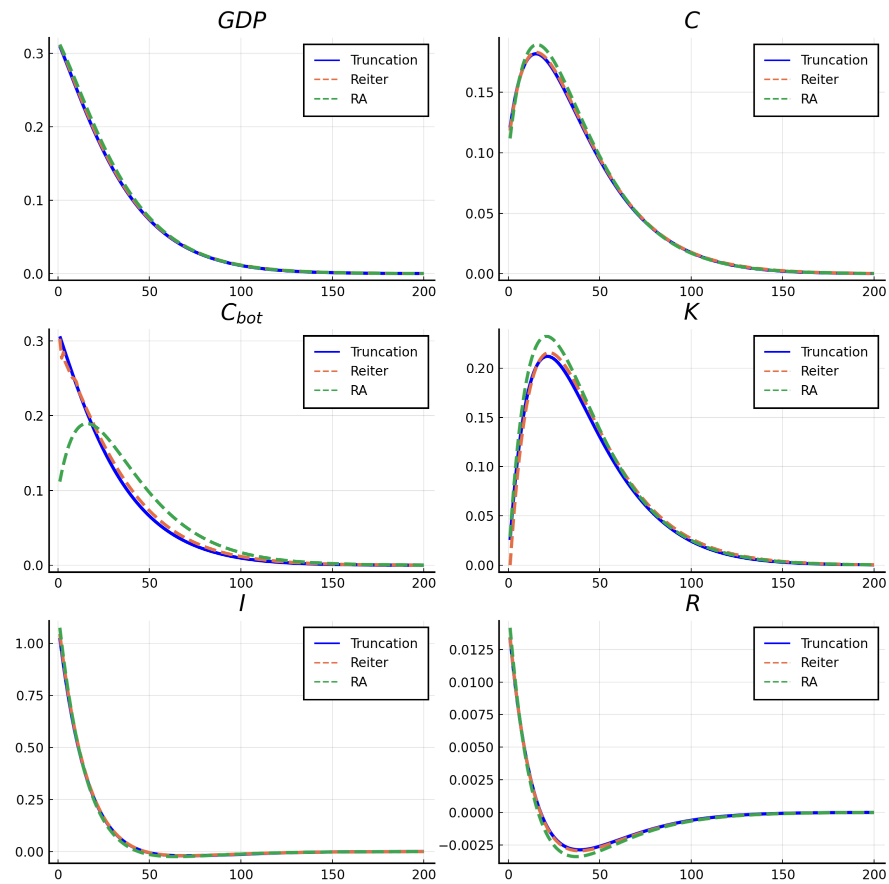
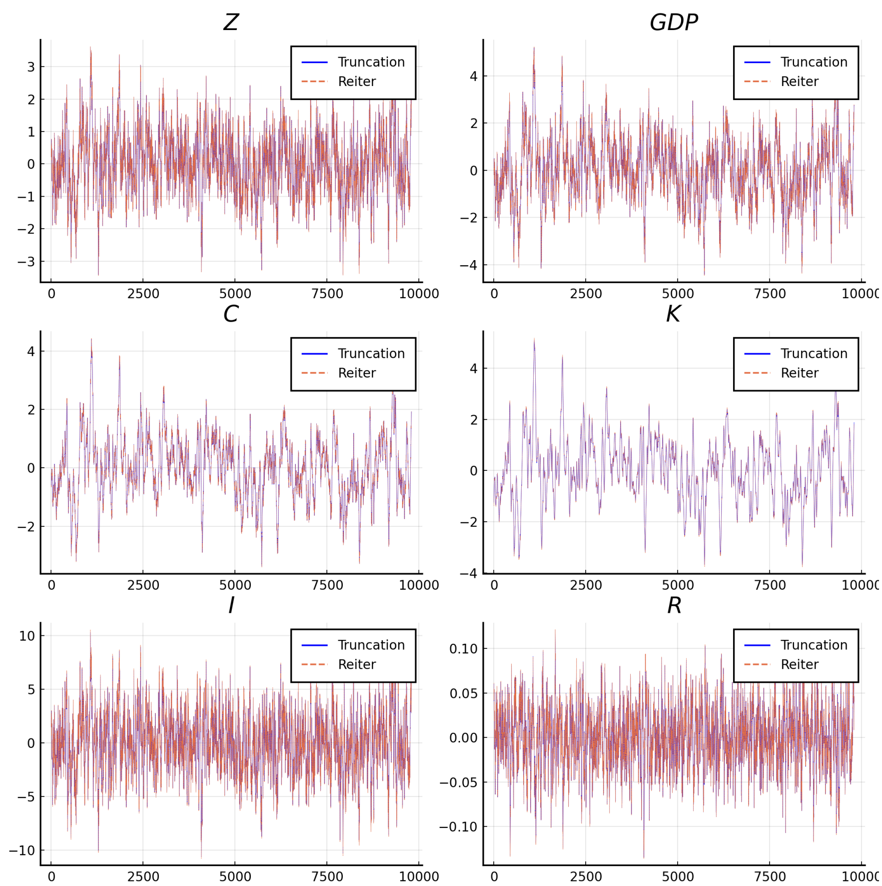


Figure 4: Simulation of aggregate variables for different simulation methods.



REFERENCES

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

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.

Ahn, S., G. Kaplan, B. Moll, T. Winberry, and C. Wolf (2017): "When Inequality Matters for Macro and Macro Matters for Inequality," in *NBER Macroeconomics Annual 2017, volume 32*, ed. by M. Eichenbaum, and J. A. Parker, pp. 1–75. University of Chicago Press.

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109(3), 659–684.

Auclert, A., B. Bardóczy, M. Rognlie, and L. Straub (2021): "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," *Econometrica*, 89(5), 2375–2408.

Bayer, C., R. Luetticke, L. Pham-Dao, and V. Tjaden (2019): "Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk," *Econometrica*, 87, 255–290.

Bewley, T. F. (1983): "A Difficulty with the Optimum Quantity of Money," *Econometrica*, 51(5), 1485–1504.

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2021): "Inequality, Business Cycles, and Monetary-Fiscal Policy," *Econometrica*, 89(6), 2559–2599.

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(3), 312–320.

Den Haan, W. J. (2010): "Comparison of Solutions to the Incomplete Markets Model with Aggregate Uncertainty," *Journal of Economic Dynamics and Control*, 34(1), 4–27.

Den Haan, W. J., K. L. Judd, and M. Juillard (2010): "Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty," *Journal of Economic Dynamics and Control*, 34(1), 1–3.

Huggett, M. (1993): "The Risk Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies," *Journal of Economic Dynamics and Control*, 17(5-6), 953–969.

Imrohoroğlu, A. (1992): "The Welfare Cost of Inflation under Imperfect Insurance," *Journal of Economic Dynamics and Control*, 16(1), 79–91.

Kaplan, G., B. Moll, and G. Violante (2018): "Monetary Policy According to HANK," *American Economic Review*, 3(108), 697–743.

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, pp. 143–184. Cambridge University Press.

Krusell, P., and A. A. J. Smith (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106(5), 867–896.

LeGrand, F., A. Martin-Baillon, and X. Ragot (2021): "Should Monetary Policy Care About Redistribution? Optimal Fiscal and Monetary Policy with Heterogeneous Agents," Working Paper, SciencesPo.

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

Nuño, G., and B. Moll (2018): "Social Optima in Economies with Heterogeneous Agents," *Review of Economic Dynamics*, 48, 150–180.

Reiter, M. (2009): "Solving Heterogeneous-Agent Models by Projection and Perturbation," *Journal of Economic Dynamics and Control*, 33(3), 649–665.

Rios-Rull, V. (2001): "Computation of Equilibria in Heterogenous Agent Models," in *Computational Methods for the Study of Dynamic Economies*, ed. by R. Marimon, and A. Scott, chap. 11. Oxford University Press.

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

Press, Princeton.
WINBERRY, T. (2018): "A Method for Solving and Estimating Heterogeneous Agent Macro Models," *Quantitative Economics*, 9(3), 1123–1151.
YOUNG, E. (2010): "Solving the Incomplete Markets Model with Aggregate Uncertainty using the Krusell-Smith Algorithm and Non-Stochastic Simulations," *Journal of Economic Dynamics and Control*, 34(1), 36–41.