Western University
Scholarship@Western

Department of Economics Research Reports Economics Working Papers Archive

2022

2022-10 Computing Longitudinal Moments for Heterogeneous Agent Models

Sergio Ocampo

Baxter Robinson

Follow this and additional works at: https://ir.lib.uwo.ca/economicsresrpt

Part of the Economics Commons

Citation of this paper:
Ocampo, Sergio, Baxter Robinson. "2022-10 Computing Longitudinal Moments for Heterogeneous Agent Models." Department of Economics Research Reports, 2022-10. London, ON: Department of Economics, University of Western Ontario (2022).

Computing Longitudinal Moments for Heterogeneous Agent Models

by

Sergio Ocampo and Baxter Robinson

Research Report # 2022-10 **September 2022**
Revised March 2023
Revised July 2023
Revised October 2023

Department of Economics

Research Report Series

Department of Economics
Social Science Centre
Western University
London, Ontario, N6A 5C2
Canada

Computing Longitudinal Moments for Heterogeneous Agent Models*†

Sergio Ocampo‡

University of Western Ontario

Baxter Robinson§

University of Western Ontario

This Version: October 31, 2023

Abstract

Computing population moments for heterogeneous agent models is a necessary step for their estimation and evaluation. Computation based on Monte Carlo methods is time- and resource-consuming because it involves simulating a large sample of agents and tracking them over time. We formalize how an alternative non-stochastic method, widely used for computing cross-sectional moments, can be extended to also compute longitudinal moments. The method relies on following the distribution of populations of interest by iterating forward the Markov transition function that defines the evolution of the distribution of agents in the model. Approximations of this function are readily available from standard solution methods of dynamic programming problems. We document the performance of this method *vis-a-vis* standard Monte Carlo simulations when calculating longitudinal moments. The method provides precise estimates of moments like top-wealth shares, auto-correlations, transition rates, age-profiles, or coefficients of population regressions at lower time- and resource-costs compared to Monte Carlo based methods. The method is particularly useful for moments of small groups of agents or involving rare events, but implies increasing memory costs in models with a large state space.

JEL: C6, E2

Keywords: Computational Methods, Heterogeneous Agents, Simulation.

---

*We thank Emmanuel Murray Leclair and Javier Hernández for their excellent work as research assistants on this project. We also thank Émilien Gouin-Bonenfant, Fatih Guvenen, Gueorgui Kambourov, Dirk Krueger, Burhan Kuruscu, Rory McGee, Emily Moschini, Thomas Phelan, Eugene Tan, several anonymous referees, and the editor, Hans Amman, for helpful comments. First Version, September 2022.
We thank the Stephen A. Jarislowsky Chair in Central Banking for generous financial support.

†Replication files for this paper are available at https://github.com/ocamp020/Histogram_Iteration.

‡Email: socampod@uwo.ca; Web: https://sites.google.com/site/sergiocampod/.

§Email: brobin63@uwo.ca; Web: https://sites.google.com/view/baxter-robinson/.

1. Introduction

Computing cross-sectional and longitudinal moments is integral to the estimation and use of heterogeneous agents models that are common in the study of a wide variety of economic phenomena (e.g., Heathcote, Storesletten, and Violante 2009; De Nardi, French, and Jones 2016; De Nardi and Fella 2017). However, calculating these moments frequently poses computational challenges that arise from the repeated simulation of the models. These challenges limit how researchers can use these models and the features they are able to include in them, even as computational power continues to improve.

One key challenge when estimating or evaluating dynamic models is the cost of calculating longitudinal moments or population regressions. Longitudinal moments require following individuals over time, e.g., mobility rates across occupations or the wealth distribution, income persistence, or inter-generational correlations. The standard method for computing these moments uses a Monte Carlo simulation of a panel of agents, however, these simulated panels can fail to be representative of small sub-populations, like the "very rich," or of the effects of rare events, like health shocks. So, in order to obtain accurate moments, these panels must be simulated with a large number of agents, often millions of them, which is computationally costly.

We argue in favor of an alternative non-stochastic method for computing longitudinal moments that works by following the distribution of any sub-population over time, rather than the history of a panel of agents. The method relies on using the Markov kernel that characterizes how agents transition between states to obtain the joint distribution of agents across time.[1] The Markov kernel is already approximated as part of most solution methods (e.g., Young 2010; Heer and Maußner 2005, Ch. 7), so that computing it does not create any additional cost. The method works by iterating forward the Markov kernel and computing

[1] The role of the Markov kernel in characterizing the distribution of agents and the associated theory of Markov processes is presented in Chapter 8 of Stokey, Lucas, and Prescott (1989). See Ríos-Rull (1995) for an early application of these ideas in the context of heterogeneous agent models.

moments based on the full set of potential shocks, rather than a particular realization of them. As a result, it generates moments that can be more precise than Monte Carlo simulations without the additional cost of simulating a large panel of agents.

We formally describe this method and document its performance relative to traditional Monte Carlo simulation.[2] In particular, we formalize how to compute longitudinal moments by exploiting the Markov kernel that defines the distribution of agents in a model. In practice, this entails working with a discrete approximation of the distribution in the form of a histogram,[3] which has already been used to compute cross-sectional moments since at least Conesa and Krueger (1999), making the method recognizable and readily applicable in most heterogeneous agent models used in the literature. In the discrete case, the method computes the future distribution of agents, and its moments, by iterating over an initial histogram. Accordingly, we refer to the method as the *histogram iteration method*.

We show that this method provides fast and precise estimates of moments of interest without involving the computation of new objects, relative to those involved in the model's solution. To evaluate the method's performance, we apply it to two partial equilibrium versions of the standard heterogeneous agent model based on Aiyagari (1994), one with infinitely lived agents and one with overlapping generations. In the infinitely lived agents model, we calculate moments characterizing the right tail of the wealth distribution and the persistence of consumption and wealth. In the overlapping-generations model, we calculate the age-profile of wealth and five- and fifteen-year auto-correlation of wealth. We compare the results with those from a Monte Carlo simulation, which is common in the literature (e.g., Judd 1998, Ch. 8).

We find that the histogram iteration method is at least as precise as using large simulated

[2]Variants of this method can be found in code used in a variety of independent applications to compute longitudinal moments like auto-correlations (e.g., Gourio and Miao 2010; Herreño and Ocampo 2023), but, to the best of our knowledge, no systemic treatment or description of it is available in the literature.

[3]See, e.g., Young (2010), Tan (2020), or Gouin-Bonenfant and Toda (2023). Alternatively, projection methods can also be used to represent the distribution of heterogeneous agents. See Algan, Allais, and Den Haan (2008) for a demonstration for how projection methods can minimize cross-sectional sampling variation in heterogeneous agent models.

panels while significantly reducing computational time. Time savings come from avoiding the simulation of a large enough panel of agents, in favor of computing the histogram that approximates the distribution of agents. In most cases, these time savings more than compensate for the time it takes to compute longitudinal moments by iterating on the histogram, which is somewhat greater than the time it takes to compute them from a simulated panel. There are further gains when computing cross-sectional moments because no iteration or simulation is needed when using the histogram.

The method faces two key limitations. First, as the time horizon of the moment being calculated increases, so does the number of calculations needed to update the approximation of the Markov kernel. Longer time horizons involve more complicated histories for populations of individuals that get reflected in increasingly less sparse approximations of the Markov kernel needed to compute the future distribution of agents. This problem compounds when the moment of interest involves a large sub-population, but we have found that this is not an impediment for moments commonly computed in applied research.

Second, the curse of dimensionality applies to the approximation of the Markov kernel as it grows with the number of states, increasing the memory requirements of the method, while the memory requirements of Monte Carlo methods are (linearly) tied to the sample size of the simulation. This problem, however, is not unique to the histogram iteration method and applies more broadly. The fact that discrete approximations of the Markov kernel are typically sparse helps to alleviate these issues (see, e.g., Tan 2020).

We end the paper with a short discussion of how the method can be extended. The method takes as given the solution to the agents' problem, in the form of policy functions mapping states into actions. This means that the method can be readily applied in non-stationary environments, as is the case with transition paths after the economy faces a change in policy. Finally, the method is also immediately applicable to continuous time settings as the ones described in Achdou, Han, Lasry, Lions, and Moll (2021).

2. Computing moments for heterogeneous agent models

We now describe how to directly compute cross-sectional and longitudinal moments from the solution of heterogeneous agent models. We do this for the general case of continuous state variables. Crucially, the method we describe below does not require any simulations of a sample of agents from the model. In Section 5 below we implement the method for two variants of the baseline Bewley-İmrohoroğlu-Hugget-Aiyagari model.

We take as given the model's solution in the form of policy functions for agents that, together with the stochastic processes of exogenous states, imply the evolution of the distribution of agents in the economy. This evolution is captured by a Markov kernel, $T\left(s^{'}|s\right)$, that maps the transition of a mass of agents from a current state s into a future state s' in the state space $\mathcal{S}$. The stationary distribution, λ, is the solution to

$$\lambda\left(s^{'}\right)=\int_{s\in\mathcal{S}}T\left(s^{'}|s\right)\lambda\left(s\right)ds. \tag{1}$$

We describe how to use λ and T to directly compute cross-sectional and longitudinal moments, rather than using a simulated panel of agents.

Cross-sectional moments. These moments involve taking expectations over some outcome of interest, $x(s)$, for some sub-population characterized by states $s\in S\subseteq\mathcal{S}$,

$$E\left[x|s\in S\right]=\int_{\tilde{s}\in S}x\left(\tilde{s}\right)\lambda\left(\tilde{s}|s\in S\right)d\tilde{s}, \tag{2}$$

where $\lambda\left(\tilde{s}|s\in S\right)\equiv\mathbb{I}_{\tilde{s}\in S}\lambda(\tilde{s})/\int\mathbb{I}_{s\in S}\lambda(s)ds$ is the conditional distribution of the sub-population in S, and where $\mathbb{I}_{s\in S}$ is an indicator variable for whether $s\in S$. Equation (2) applies to a wide range of moments. For example, the skewness or kurtosis of the endogenous wealth distribution for the whole population (when $S=\mathcal{S}$) or for a subgroup (say top income earners).[4] These moments can be computed from the solution of the model's stationary distribution (λ), either by approximating the integral (Judd 1998, Ch. 7) or by calculating the moment from a discrete approximation of the distribution (see, i.a., Young 2010).

Longitudinal moments. Many other moments require knowing either the collective outcomes of a group of agents over time (e.g., for computing transition rates across occupations) or the outcomes of individual agents (e.g., for computing the auto-correlation of their wealth).[5] Calculating these moments is difficult because of the stochastic nature of the individuals' time-paths. However, it is possible to extend the approach described above for cross-sectional moments to the calculation of longitudinal moments at low computational cost. This is accomplished by focusing on the transition of the distribution of agents, taking into account all possible paths an individual can take rather than relying on a sample of realized paths from a Monte Carlo simulation.

A longitudinal moment is the expectation of an outcome $x\left(s_t, s_{t+\Delta}\right)$ over the joint distribution of agents across periods for some horizon $\Delta \in \mathbb{N}$, where the function x describes an outcome that depends on the agents' initial and final states. For example, to calculate the share of individuals in the top 1% of the wealth distribution who remain in the top 1% at some future period, we set $x\left(s_t, s_{t+\Delta}\right) = \mathbb{I}_{s_t, s_{t+\Delta} \in S}$, where $S \subseteq \mathcal{S}$ is the set of states for which an agent is in the top 1% of the wealth distribution, and x is equal to one if both s_t and $s_{t+\Delta}$ place an individual in the top 1% of the wealth distribution. The moment is then calculated as the integral over $x\left(s_t, s_{t+\Delta}\right)$:

$$E\left[x|s_t \in S\right] = \int_{\tilde{s} \in S} \left(\int_{\tilde{s}' \in \mathcal{S}} x\left(\tilde{s}, \tilde{s}'\right) \lambda_{t+\Delta}\left(\tilde{s}'|s_t \in S\right) d\tilde{s}' \right) \lambda_t\left(\tilde{s}|s_t \in S\right) d\tilde{s}, \tag{3}$$

where $\lambda_t\left(\cdot|s_t \in S\right)$ is the distribution of individuals in the top 1% in the initial period (those

[4]Equation (2) can also be used to define percentiles or other descriptors of the distribution. These expectations can also characterize the population value of coefficients in cross-sectional regressions.

[5]Moments that require collective outcomes include mobility rates across the income or wealth distribution, or inter-generational mobility in life cycle models. Moments that require individual outcomes include the distribution of growth rates of income or wealth for individual agents, or the distribution of lifetime earnings.

with $s_t \in S$), and $\lambda_{t+\Delta}(\cdot|s_t \in S)$ is the future distribution of this group of agents.

Crucially, in this example we do not need to know the paths of each individual agent or how many times they transition in or out of the top 1%. Instead, it is sufficient to follow the distribution of agents in the group, as represented by $\lambda_t(\cdot|s_t \in S)$ and $\lambda_{t+\Delta}(\cdot|s_t \in S)$, to calculate the fraction of agents who remain in the top 1% in the final period. This simplifies the construction of the distribution of agents across periods. We compute $\lambda_{t+\Delta}(\cdot|s_t \in S)$ for each $s^{'} \in \mathcal{S}$ starting from the distribution $\lambda_t(\cdot|s_t \in S)$ and updating it according to the Markov kernel T_Δ, computed from the one-period kernel T, obtaining

$$\lambda_{t+\Delta}\left(s^{'}|s_t \in S\right) = \int_{\tilde{s}\in\mathcal{S}} T_\Delta\left(s^{'}|\tilde{s}\right)\lambda_t\left(\tilde{s}|s_t \in S\right)d\tilde{s}. \tag{4}$$

There are other examples of longitudinal moments for which the outcome of each individual agent in the group matters. Computing these moments requires following the distribution of agents who start from the same initial state $s_t = s$, for $s \in S$, as they transition across the state space to their future distribution $\lambda_{t+\Delta}(\cdot|s_t \in \{s\})$. This is the case for the auto-correlation of individual wealth, where we need to compute

$$E\left[x|s_t \in S\right] = \int_{\tilde{s}\in S}\left(\int_{\tilde{s}^{'}\in\mathcal{S}} x\left(\tilde{s},\tilde{s}^{'}\right)\lambda_{t+\Delta}\left(\tilde{s}^{'}|s_t \in \{\tilde{s}\}\right)d\tilde{s}^{'}\right)\lambda_t\left(\tilde{s}|s_t \in S\right)d\tilde{s}, \tag{5}$$

where the outcome of interest is $x(s_t, s_{t+\Delta}) = (a(s_t) - \bar{a}_t)(a(s_{t+\Delta}) - \bar{a}_{t+\Delta})$ with $a(s)$ denoting an agent's wealth when the agent's state is s, and $\bar{a}_t$ and $\bar{a}_{t+\Delta}$ the average wealth across agents in the initial and the final period respectively. The integral in equation (5) requires knowledge of the final distribution of wealth for the agents that start in each initial state $s \in S$, represented by $\lambda_{t+\Delta}(\cdot|s_t \in \{s\})$ and computed using the Markov kernel T_Δ:

$$\lambda_{t+\Delta}\left(s^{'}|s_t \in \{s\}\right) = T_\Delta\left(s^{'}|s\right)\delta_{\{s\}}\left(s\right), \tag{6}$$

where $\delta_{\{s\}}$ is the degenerate distribution concentrated in initial state s.

3. Baseline heterogeneous agent models

We illustrate the histogram iteration method to compute longitudinal moments in the context of the baseline Bewley-İmrohoroğlu-Hugget-Aiyagari model. The economy is populated by a continuum of agents indexed by $i \in [0,1]$ that differ on their age (h), labor productivity (ε), rate of return (ζ), and endogenous asset holdings (a). Labor productivity and rates of return follow discrete Markov processes with transition matrices P^{ε} and P^{ζ}. Agents are price takers. They receive income from the return on their savings, $r(\zeta)$, and from wages, w, paid for their supply of efficiency units of labor, $\ell(h,\varepsilon)$, which depends on their age and labor productivity.

The dynamic programming problem of an agent of age h is

$$V_h(\varepsilon,\zeta,a) = \max_{c}\; u(c) + \beta\phi_h E_{\tilde{\varepsilon},\tilde{\zeta}}\left[V_{h+1}\left(\tilde{\varepsilon},\tilde{\zeta},\tilde{a}\right)|\varepsilon,\zeta\right] \tag{7}$$
$$\text{s.t. } [1+r(\zeta)]\,a + w\ell(h,\varepsilon) = c + \tilde{a};\; \tilde{a} \geq \underline{a},$$

where ϕ_h is the conditional survival probability from age h to age $h+1$ and $\tilde{a}$ is the agent's savings or future assets. The expectation in (7) is taken with respect to the future values of ε and ζ. The solution to (7) is a savings function, that is an optimal rule $a_h^{\star}$ for $\tilde{a}$ such that $a_h^{\star}(\varepsilon,\zeta,a) \geq \underline{a}$ for all (ε,ζ,a) and

$$V_h(\varepsilon,\zeta,a) = u\left([1+r(\zeta)]\,a + w\ell(h,\varepsilon) - a_h^{\star}(\varepsilon,\zeta,a)\right) + \beta E_{\tilde{\varepsilon},\tilde{\zeta}}\left[V_{h+1}\left(\tilde{\varepsilon},\tilde{\zeta},a_h^{\star}(\varepsilon,a)\right)|\varepsilon,\zeta\right]. \tag{8}$$

We will focus on a stationary equilibrium with a time-invariant distribution of agents. $\mathcal{S}$ is the state space with typical element $s = (h,\varepsilon,\zeta,a)$. Given a birth and death process for agents, the transition function of labor productivity and returns, and the savings functions, the stationary distribution is a solution to (1), where the Markov kernel $T\left(s'|s\right)$ is constructed using the policy functions and the evolution of exogenous states.

We solve the model in partial equilibrium taking the wage rate, w, and the average return on savings, $\bar{r}$, as exogenous. We do this to focus on the computation of moments for any given solution of the agents' problem. Our results apply in a general equilibrium setting when computing the moments after finding the market clearing prices.

We solve for two versions of the model that differ in the birth and death process of agents. In both models, we adopt the following functional form for agents' utility:

$$u(c) = \frac{c^{1-\sigma} - 1}{1 - \sigma}. \tag{9}$$

We set σ equal to 2, which is in the range of values used in Aiyagari (1994). We take $\bar{r}$ to be 3.2 percent, in line with historical values for the U.S. and we set w so that labour income in our model matches average labor income for the U.S. in 2019, which is \$53,624.[6] We set $\underline{a} = 0$, preventing borrowing. Below, we outline the differences between the two different versions of the model and their parametrization.

Infinitely lived heterogeneous agents model. We consider a version of the model where agents are infinitely lived and their labor efficiency depends only on their labor productivity. In particular, $\ell(h, \varepsilon) = \exp(\varepsilon)$ and $\phi_h = 1$ for all h. We focus on the age-invariant solutions to (7) and (8), a value function $V(\varepsilon, \zeta, a)$ and a savings function $a^\star(\varepsilon, \zeta, a)$. Accordingly, we drop age from the state vector when referring to the infinitely lived agents model.

Labor productivity follows a discrete Markov process with $n_\varepsilon = 11$ states. We obtain P^ε by discretizing an AR(1) process using Rouwenhorst (1995)'s method and use persistence $\rho_\varepsilon = 0.963$ and innovation variance $\sigma_\varepsilon^2 = 0.162$ from Storesletten, Telmer, and Yaron (2004).

We include heterogeneous returns on savings, a key ingredient for generating high levels of wealth inequality (Benhabib, Bisin, and Zhu 2011; Stachurski and Toda 2019), by setting an agent's returns to be $r(\zeta) = \bar{r}\exp(\zeta)$. The state ζ follows a discrete Markov process with $n_\zeta = 7$ states. We obtain P^ζ by discretizing an AR(1) process with persistence $\rho_\zeta = 0.70$ and

[6] We construct this value from FRED Data (U.S. Bureau of Economic Analysis 2022) as Total Wages and Salaries (BA06RC1A027NBEA) divided by the 12-month average of Civilian Labor Force Level (CLF16OV).

innovation variance $\sigma_\zeta^2 = 1.3$ using Tauchen (1986)'s method.[7]

Overlapping generations heterogeneous agent model. In the second version of the model, agents live for $H > 0$ periods and have a terminal value of $V_{H+1} = 0$. Agents face mortality risk and have a survival probability ϕ_h of surviving into age h, conditional on surviving to age $h-1$. We set the survival probabilities following Bell and Miller (2002) projections for the U.S., with each model period corresponding to a single year. Agents are born at age 20 ($h = 1$) and can live to a maximum age of 100 ($H = 81$), when $\phi_{H+1} = 0$. Upon death, agents are replaced by a newborn who starts life with $a_1^\star = \$1,000$ of assets.

Efficiency units of labor are $\ell(h, \varepsilon) = \exp(\xi(h) + \varepsilon)$, where $\xi(h)$ is a quadratic polynomial that generates a 50 percent rise in average labor income from age 21 to its peak at age 51 as in Guvenen, Kambourov, Kuruscu, Ocampo, and Chen (2023).[8] We use the same process for labor productivity (ε) as in the infinitely lived agent model. Finally, we eliminate rate of return heterogeneity, so that all agents earn $r_i = \bar{r}$. Accordingly, we drop ζ from the state vector when referring to the overlapping generations model.

4. Solving the models

We solve for the policy functions in (7) using readily available solution methods that exploit the optimality conditions of the savings choice (i.e., Carroll 2006). Having computed the policy functions, we approximate the Markov kernel, T, of the distribution of agents by discretizing it over assets on a grid $\vec{a}_{n_a}$ following Young (2010).[9] The result is a transition matrix $\hat{T}$, whose elements $\hat{T}\left(s, s'\right)$ give the probability that an agent with current state

[7] There are alternative methods to discretize Markov processes that outperform those in Rouwenhorst (1995) and Tauchen (1986). Among them, Farmer and Toda (2017) provide a general method to approximate multivariate non-linear processes with non-Gaussian innovations. Their method works by matching relevant moments of the underlying process, generalizing the properties established in Kopecky and Suen (2010) for Rouwenhorst (1995)'s method in the context of AR(1) processes with Gaussian innovations.

[8] $\xi(h) = \left(60(h-1) - (h-1)^2\right)/1800$.

[9] Gouin-Bonenfant and Toda (2023) show that Pareto extrapolation methods can reduce approximation error for models with fat-tailed distributions. The transition probability matrix that these method generate can also be used when calculating moments as we describe below.

s transitions to state s'.[10] This probability depends on the birth and death process (for instance, agents of age $h = H$ transition to age $h = 1$ with certainty), the transition matrix of the labor productivity process (ε), the transition matrix of the return heterogeneity process (ζ), and the approximation of the transition of assets on the fixed grid $\vec{a}_{n_a}$.[11] Finally, we compute the stationary distribution of agents on the discrete grid by iterating over

$$\hat{\lambda}^{n+1}\left(s'\right) = \sum_{s} \hat{T}\left(s, s'\right) \hat{\lambda}^{n}\left(s\right), \tag{10}$$

for some initial $\hat{\lambda}^0$.[12] The stationary distribution, $\hat{\lambda}$, is the limit of $\hat{\lambda}^n$ as n grows large.

In the next section, we use the approximated distribution, $\hat{\lambda}$, and Markov kernel, $\hat{T}$, to compute moments for both models. The Markov kernel plays an important role in computing moments because it describes the evolution of states given any initial distribution. This is crucial for computing longitudinal moments where it is necessary to know how agents transition between states over time. We explore results with grids of different sizes for the approximation of the distribution and the Markov kernel. All grids are curved so that they are denser for low wealth values. In particular, the n^{th} node of an asset grid with N nodes satisfies $\vec{a}_n = \underline{a} + (\overline{a} - \underline{a})\left({}^{n-1}/_{N-1}\right)^{\theta_a}$, where $\theta_a > 1$ measures the curvature. We use a curvature of $\theta_a = 3.5$ and solve for the policy functions on a grid with 250 nodes before approximating the Markov kernel and the stationary distribution.

[10] The transition matrix $\hat{T}$ can be further exploited to speed up the computation of the model's solution as shown by Rendahl (2022).

[11] An agent with state s transitions with certainty to having assets $\tilde{a} = a^{\star}_{h(s)}\left(\varepsilon\left(s\right), \zeta\left(s\right), a\left(s\right)\right) \in \left[\vec{a}_j, \vec{a}_{j+1}\right]$, for some j. In the discrete approximation the agent transitions to either $\vec{a}_j$ with probability ${}^{\vec{a}_{j+1}-\tilde{a}}/_{\vec{a}_{j+1}-\vec{a}_j}$ or $\vec{a}_{j+1}$ with probability ${}^{\tilde{a}-\vec{a}_j}/_{\vec{a}_{j+1}-\vec{a}_j}$.

[12] For the application below, we initialize the infinitely lived agent model with a uniform distribution over states. The computation of the model's solution can be made faster with more tailored choices for the initial distribution of agents, obtained from the solution of smaller-scale models (with fewer states, or coarser grids) or from the simulation of the model. We opt not to optimize over the choice of an initial distribution to maintain the attention on the computation of moments described in Section 5. For the OLG model, the initial distribution of agents is determined by the invariant distribution of the labour productivity with all agents having the same starting level of assets ($a^{\star}_1$).

5. Computing moments

We now compute cross-sectional and longitudinal moments for the entire population and sub-populations of interest, like agents at the top or bottom of the wealth distribution. Cross-sectional moments, like the share of wealth owned by the wealthiest 1% of individuals, can be readily computed from the stationary distribution (λ) or its approximation $\left(\hat{\lambda}\right)$. However, it is often necessary to follow individuals over time when computing longitudinal moments like the auto-correlation of wealth. This is often achieved through costly Monte Carlo based simulations of a sample of individuals. The histogram iteration method relies instead on tracking the distribution of the relevant group of individuals (the sub-population), following its evolution as described by the Markov kernel T. We now describe the method.

The histogram iteration method. Consider a moment describing the expectation over some outcome in some future period $x\left(s_t, s_{t+\Delta}\right)$ for a sub-population with current state $s_t \in S \subseteq \mathcal{S}$, where the subset S contains all states for which agents satisfy some condition of interest, say having a certain level of wealth or income. These moments take the form of the expectations in equations (3) and (5). The objective is to approximate the value of these expectations. We obtain the sub-population's initial distribution, $\hat{\lambda}_t\left(\cdot|s_t \in S\right)$, from the stationary distribution $\hat{\lambda}$ by restricting its domain to S and normalizing. Tracking the distribution of the sub-population over time involves iterating over $\hat{\lambda}_t\left(\cdot|s_t \in S\right)$ with the Markov kernel $\hat{T}$ as in (10).

When the moment requires tracking only the collective outcomes of the group in S, the expectation of interest is as in equation (3),

$$\hat{E}\left[x|s_t \in S\right] = \sum_{\tilde{s}} \sum_{\tilde{s}'} x\left(\tilde{s}, \tilde{s}'\right) \hat{\lambda}_{t+\Delta}\left(\tilde{s}'|s_t \in S\right) \hat{\lambda}_t\left(\tilde{s}|s_t \in S\right), \tag{11}$$

where $\hat{\lambda}_{t+\Delta}\left(\cdot|s_t \in S\right)$ is the distribution of future states for all agents in the initial sub-population, as in equation (4). When the moment requires tracking the outcomes of individuals, the expectation of interest is as in equation (5),

$$\hat{E}\left[x|s_t \in S\right] = \sum_{\tilde{s}} \sum_{\tilde{s}'} x\left(\tilde{s}, \tilde{s}'\right) \hat{\lambda}_{t+\Delta}\left(\tilde{s}'|s_t \in \{\tilde{s}\}\right) \hat{\lambda}_t\left(\tilde{s}|s_t \in S\right), \tag{12}$$

where $\hat{\lambda}_{t+\Delta}\left(\cdot|s_t \in \{s\}\right)$ is the future distribution of agents that started in state $s \in S$, as in equation (6). Below, we apply this method to the models described in Section 3.

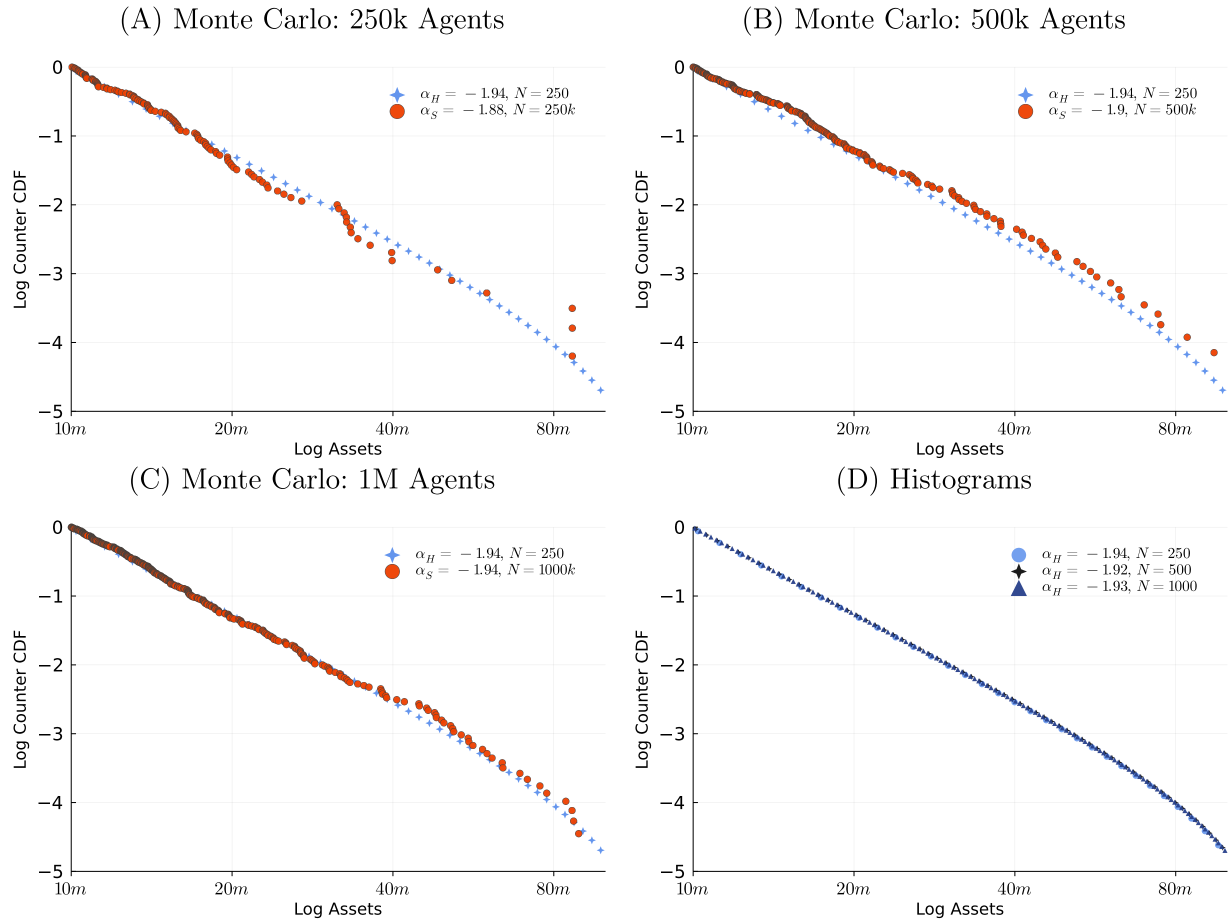
5.1. Moments for the infinitely lived agents model

We now compute several moments for the infinitely lived agents model and compare the performance of the histogram iteration method relative to a traditional Monte Carlo simulation. We focus on moments characterizing the wealth distribution and the behaviour of consumption, which are the endogenous outcomes in our setting. In particular, we present results for the tail of the wealth distribution, top wealth shares, the persistence of consumption and wealth, and the ten-year transition rates across wealth deciles.

In terms of the accuracy, we find that both methods provide similar estimates for the moments, except for those regarding top-wealth holders: the shape of the tail of the wealth distribution and top wealth shares. The challenge for the Monte Carlo simulation method comes from the large number of agents needed in order to obtain a representative sample of top-wealth holders. The histogram iteration method provides more consistent values of these moments when varying the number of grid nodes in the approximation.

In terms of the computational cost of calculating moments, cross-sectional moments come almost for free after solving for the histogram or simulating a panel of agents. Longitudinal moments are significantly more expensive to calculate with the histogram iteration method than from a simulated panel of agents. This is because the histogram iteration method requires iterating forward for all initial states, as all the possible histories of agents are mapped for each initial condition. This makes the computation more expensive than computing the same moment using an already existing panel of agents containing realized histories of consumption

Figure 1. Pareto Tail - Monte Carlo Simulation and Histogram Method



Notes: The figures plot the log counter CDF of the conditional distribution of wealth above \$10 million. Panels 1A to 1C approximate the CDF using samples of agents from a Monte Carlo simulation and differ in the number of agents being simulated. The blue diamonds correspond to the approximation of the counter CDF using the histogram method with 500 grid nodes. The final panel approximates the CDF using the histogram method with 250, 500, and 1000 grid nodes.

and wealth. However, the time required to solve for the histogram is substantially less than the time required to simulate the Monte Carlo panel of agents. As a result, it takes less total time to calculate longitudinal moments when using the histogram iteration method than when using the simulated panel.[13]

Pareto Tail. One characteristic of the cross-sectional distribution of wealth that is often difficult to capture in heterogeneous agent models is the behavior of its right tail and the

[13] All times are for a Mac Mini with an M1 processor running Julia v1.7.

associated level of wealth concentration. These statistics are crucial when studying inequality, particularly because of their implications for taxation. In Figure 1, we report the right tail of the wealth distribution (above ten million dollars) and the corresponding Pareto coefficient for simulations with sample sizes between two hundred and fifty thousand and one million agents, and contrast them with the tail of the stationary distribution of wealth approximated with a histogram with 500 grid points. We find that simulation-based results require a large number of agents to correctly represent the properties of the right tail of the wealth distribution,[14] and that, by contrast, the histogram provides a more stable picture of the distribution at lower computational cost.[15]

The Monte Carlo simulation captures the general shape of the tail, but has issues populating the top end, even with one million agents. This is apparent in the discrepancies between the tail indexes (α) and the wealth shares of the richest agents across simulation samples, as shown in Table 1. Figure 1D shows that the histogram provides more stable outcomes across grid sizes for both the shape of the distribution and the tail index.[16]

The sensitivity of the right tail to the number of agents being simulated becomes an issue in models that aim to capture the extent of wealth inequality in the data. For instance, Guvenen et al. (2023) pose a model capable of reproducing the tail of the wealth distribution in the U.S., including the presence of multi-billionaires. In order to generate these very wealthy agents, they use a Monte Carlo simulation with twenty million agents.

Top Wealth Shares. We compute the share of wealth owned by the top 1% and top 0.1% of individuals in our model and report them in Table 1. Just as with the shape of the right tail, these measures of top wealth concentration are difficult to measure with the Monte Carlo simulation because a small number of "very rich" agents play a large role in determining the

[14] The same logic applies when studying the effects of rare events and tail risks, like extreme health shocks.

[15] This is similar to Gouin-Bonenfant and Toda (2023), who propose replacing the grid at the right end of the distribution with an approximation of the continuous distribution using limit results.

[16] This stability makes moments computed via the histogram iteration method more likely to be continuous in the model parameters than those computed via Monte Carlo simulation. This smoothness can benefit estimation or calibration routines as it smooths the objective function.

Table 1. Cross-sectional and Longitudinal Moments: Infinitely Lived Agents

	Percentage Point Deviations from Reference Value						Ref. Value
	Monte Carlo: Sample Size			Histogram: Grid Size			
	250k	500k	1M	250	500	1000	
Top Wealth Shares							
Top 0.1%	0.12	0.14	−0.52	0.08	−0.04	−0.10	6.29
Top 1%	0.08	0.02	−0.76	0.12	−0.06	0.03	19.00
Pareto Coefficient	−0.04	−0.01	0.03	0.03	0.01	0.02	1.91
Auto-Correlations							
$\rho\left(c_{i,t}, c_{i,t+2}\right)$	0.24	0.00	0.05	0.13	0.01	0.04	82.52
$\rho\left(a_{i,t}, a_{i,t+2}\right)$	0.20	0.97	0.08	1.04	0.07	0.43	49.73
Transition Rates							
$\Pr\left(a_{i,t+10} \in D_1 \| a_{i,t} \in D_1\right)$	0.17	−0.06	0.05	0.59	0.12	0.31	50.04
$\Pr\left(a_{i,t+10} \in D_2 \| a_{i,t} \in D_1\right)$	−0.10	0.24	−0.06	−0.19	0.02	−0.03	34.16
Computational Time							
Simulation	689.3	1386.1	2744.3	—	—	—	—
Distribution $\hat{\lambda}$	—	—	—	478.9	881.4	1827.0	—
Top Inequality	0.01	0.02	0.04	1E-4	4E-4	2E-4	—
Auto-Correlation	0.05	0.08	0.18	9.81	21.47	54.76	—
Transition Rates	0.39	0.83	1.58	13.48	26.04	50.48	—

Notes: The table reports the deviation of calculated moments and computational time in seconds for the infinitely lived agents model. The first block computes the moments approximating the distribution with Monte Carlo simulation on three different samples of 250k, 500k, and 1M agents. The second block computes the moments approximating the stationary distribution with histograms on three different grids with 250, 500, and 1000 nodes. The reference value is obtained from a histogram grid with 5000 nodes.

value of the moments. As a consequence, the top wealth shares are still varying even when the number of simulated agents is increased to one million. The time required to compute the moments is negligible next to the time required to either obtain the stationary distribution of the model or to simulate the agents.

Persistence of Consumption and Wealth. We continue by computing the two-year auto-correlations of consumption and wealth, which are informative about the ability of individuals to insure themselves against temporary income fluctuations. These are longitudinal moments that require comparing the level of consumption and wealth for individuals across time. Both the histogram iteration method and Monte Carlo simulation give very similar results for the moments, but they differ markedly on the time it takes to compute the moments. While it is faster to compute moments from an existing panel of agents, this does not take into account the time it takes to generate the panel. Once this cost is taken into account, it is clear that faster and more accurate estimates of the moment are obtained from the histogram iteration method.

Mobility. Finally, we calculate the ten-year transition rates across deciles of the wealth distribution. These rates are commonly used to study the persistence of wealth inequality and mobility. Unlike the auto-correlation of wealth, computing transition rates does not require following the full path of individuals, rather, it is enough to follow a subset of the population, e.g., those in a given decile. The histogram iteration method takes advantage of this by iterating from the conditional distribution of agents of each decile to obtain their final distribution as in (11). The transition rates are calculated directly as the mass of the final distribution in each decile. As with the auto-correlations, these transition rates take longer to calculate via the histogram method than using an existing panel of agents. However, simulating that panel of agents is costly relative to solving for the histogram.

5.2. Moments for the overlapping generations model

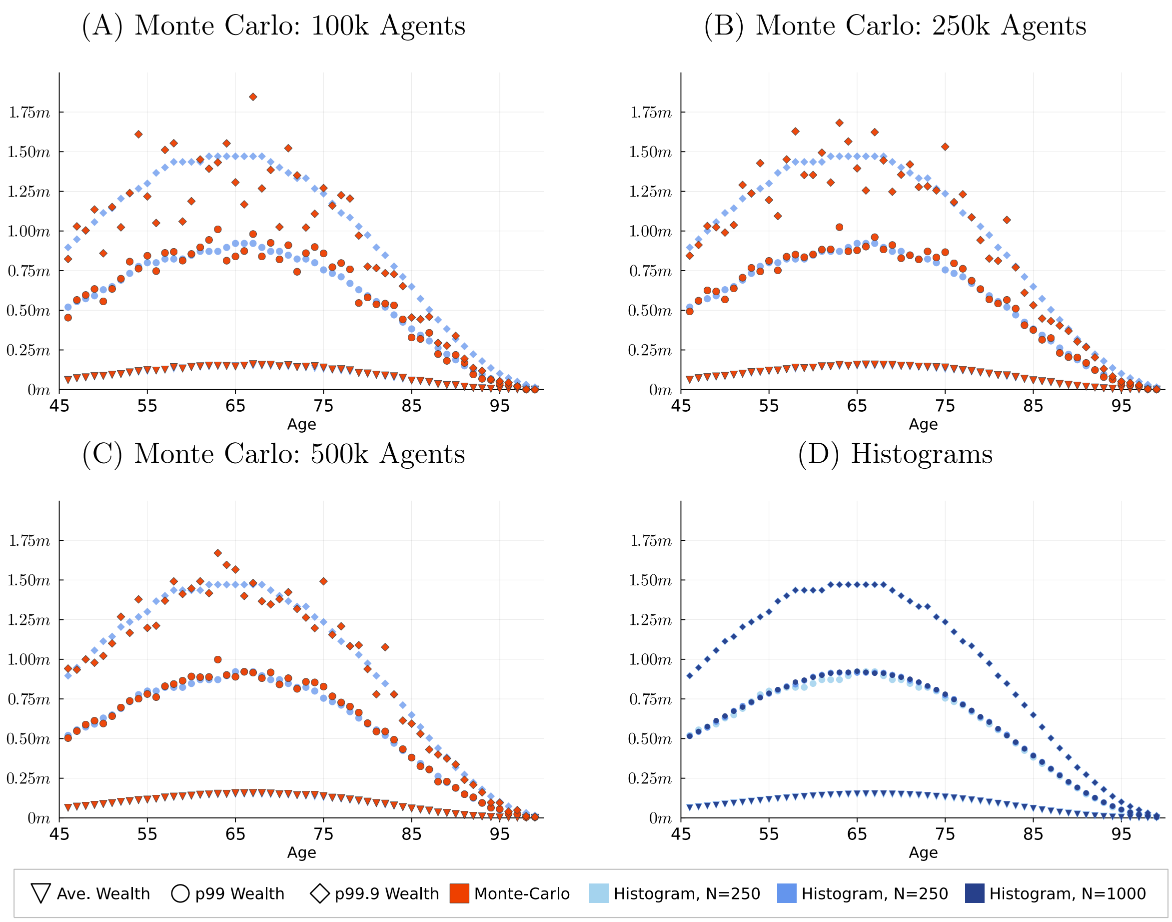
We now conduct similar exercises on the overlapping generations model. We focus on the behavior of agents along their life-cycle. In particular, we present age-profiles of the wealth distribution for agents with above median income at age 45 and the auto-correlation of wealth between the ages of 35 and 40, and the ages of 35 and 50. We compute the moments using the histogram iteration method to iterate over the evolution of a cohort and contrast the results with those of Monte Carlo simulations of up to five hundred thousand agents.[17]

The two methods produce similar moments, with the exception of moments characterizing the top of the wealth distribution. The reason is again that a large number of agents must be

[17] For the cross-sectional moments the sample size refers to the total sample, including agents of all ages. For the auto-correlation we simulate a single cohort of individuals.

simulated in order to have a representative sample of wealthy agents. This is especially the case for life-cycle moments because the sample is also conditioned by age, making it more difficult to ensure large sample sizes. In terms of the computational cost, the simulation time is again the main factor making Monte Carlo simulation-based moments more costly.

FIGURE 2. Wealth-Age Profiles - Monte Carlo Simulation and Histogram Method



Notes: The figures plot the age profile of wealth starting at age 45. Panels 2A to 2C compute the moments using samples of agents from Monte Carlo simulation and differ in the number of agents being simulated. Triangles correspond to the average wealth at every age, circles to the 99^{th} percentile of the wealth distribution, and diamonds to the 99.9^{th} percentile. Markers in blue correspond to the age profiles using the histogram method. The final panel computes the moments from the conditional distribution of wealth by age using the histogram method with 250, 500, and 1000 grid nodes.

Wealth Age Profiles. We report the age profile of wealth for agents with above median income at age 45 in Figure 2. The figures show the average, and the 99th and 99.9th percentile of wealth for every age. It is clear that the results obtained from Monte Carlo simulations

TABLE 2. Cross-sectional and Longitudinal Moments: Overlapping Generations

	Monte Carlo: Sample Size			Histogram: Grid Size		
	100k	250k	500k	250	500	1000
Wealth Auto-Correlation						
Age 35-40	89.43	89.44	89.46	89.39	89.59	89.66
Age 35-50	62.94	62.85	62.69	63.70	63.84	63.96
Computational Time						
Population Simulation	300.5	750.7	1501.1	—	—	—
Cohort Simulation	17.57	43.71	87.33	—	—	—
Distribution $\hat{\lambda}$	—	—	—	113.1	223.9	459.0
Wealth Profiles	0.19	0.71	1.33	0.51	0.99	1.97
Auto-Correlation 35-40	5E-4	4E-3	2E-3	12.45	42.13	145.5
Auto-Correlation 35-50	5E-4	1E-3	2E-3	97.97	355.1	1450.8

Notes: The table reports the auto-correlation of wealth between the ages of 35 and 50. The first block computes the moments approximating the distribution with a Monte Carlo simulation. The second block computes the moments approximating the stationary distribution with histograms on three different grids with 250, 500, and 1000 nodes. The auto-correlation of wealth is computed from the simulation of cohorts between the ages of 35 and 50 of 100k, 250k, and 500k agents, without attrition. The initial distribution is obtained from the histogram with 500 nodes. All times are in seconds.

struggle to capture the top percentiles of the wealth distribution, even though they do successfully capture the average wealth profile. As before, this is because there are only a small number of "very rich" agents in the Monte Carlo simulation, producing volatile age profiles. Even with a simulation of half a million agents, there are only approximately six agents in the top 0.1 percent of the distribution for any given age, making the statistics coming from the simulation unreliable. This is in contrast with the results obtained from the histogram that provides stable results even for relatively coarse grids as shown in Figure 2D.

The time required to compute the distribution or simulate the agents follows the same pattern described above. As we show in Table 2, the bulk of the computational time is accounted for by computing the histogram $\hat{\lambda}$ or performing the Monte Carlo simulation, with the calculation of the wealth profiles taking just a few seconds at most.

Auto-correlation of Wealth. Finally, we compute the five- and fifteen-year auto-correlation of wealth starting at age 35. Both the histogram iteration method and Monte Carlo simulation produce similar results, see Table 2. However, the time required to iterate the histogram increases markedly with the time horizon, making the Monte Carlo simulation faster when computing the fifteen-year auto-correlation.

This result is instructive about the practical limitations of the histogram iteration method. When using Monte Carlo methods, calculating the auto-correlation requires simulating a single cohort of agents, generating a representative sample of paths. This cohort simulation takes less time than a full simulation of the whole population and can take advantage of the histogram by using it to obtain the initial distribution of agents at age 35. By contrast, computing the auto-correlation with the histogram iteration method requires solving for the conditional distribution of agents at age 50 ($\lambda_{t+\Delta}(\cdot|s_t \in \{s\})$) starting from each initial state s at age 35, see (12). $\lambda_{t+\Delta}(\cdot|s_t \in \{s\})$ describes all the possible paths that a 35 year old can take in their next fifteen years. Computing $\lambda_{t+\Delta}(\cdot|s_t \in \{s\})$ requires iterating forward as in equation (6) multiple times. The complexity of this step increases with the time horizon as the initial mass of agents fans out across the state space.

6. Limitations

The histogram iteration method is generally an efficient way for calculating cross-sectional and longitudinal moments. However, longitudinal moments that involve individual outcomes of a large subset of the population, or that involve long periods of time, can be expensive to calculate. As we discussed at the end of Section 5, this is because the full history of individuals' paths must be mapped in order to compare the individuals' initial and final outcomes, unlike for other moments that focus on group outcomes like transition rates. This leads to cases where Monte Carlo methods can be more efficient, as was the case with the computation of the fifteen-year auto-correlation of wealth discussed in Section 5.2.

The same principle applies to moments that involve the outcomes of agents in intervening periods, rather than just the initial and final outcomes. For example, computing the distribution of lifetime earnings in our OLG model with the histogram iteration method proves to be infeasible. Doing so would have required us to compute the time paths of each possible income realization over the 81 year lifespan of agents. With 11 income states, there are $11^{81} \approx 6.8 \times 10^{17}$ possible histories of lifetime income. This distribution, however, can be easily computed from a sufficiently large panel of simulated individuals.

The histogram iteration method also imposes large memory requirements in models with large state spaces. As the state space increases, so does the number of states an individual agent may transition to, and therefore the size and memory requirements of the Markov kernel. By contrast, the memory requirements of Monte Carlo methods scale with the sample size of the panel of agents being simulated. Because of this, the key objects involved in Monte Carlo methods typically demand less memory than those of the histogram iteration method, as we show in Table 3. However, as the state space increases it also becomes necessary to increase the sample size of Monte Carlo simulations to produce an accurate estimate of the moments of interest, thus increasing the computational demands of that method, see Tables 1 and 2 and the discussion in Section 5.[18]

Table 3 makes evident how the curse of dimensionality increases the memory required to store the transition matrix $\hat{T}$, which is the key object in the histogram method. This is particularly evident for the overlapping generations model, where the maximum age is $H = 81$, but the costs are much larger for models that have two or more continuous state variables. Nevertheless, these costs can be substantially reduced in practice because the Markov kernel is typically represented by an extremely sparse matrix, as we show in the table. This sparsity comes from the fact that an agent in a particular state does not face a positive probability of transitioning to every other point on the state space. For instance, an agent with a given optimal level of savings will only transition to one of two asset levels in the discretized Markov kernel $\hat{T}$. Likewise, the age of an agent can only increase by one unit

[18] The fact that Monte Carlo integration requires extremely large samples to address the curse of dimensionality is a major impediment for relying on these methods in high-dimensional models.

Table 3. Memory Costs in Gigabytes - Monte Carlo and the Histogram Method

Infinitely Lived Agents Model

	Monte Carlo: Sample Size				Histogram: Grid Size		
	250k	500k	1M		250	500	1000
Integer Panel	0.003	0.006	0.012	Histogram	2E-04	3E-04	6E-04
Float Panel	0.006	0.012	0.024	Transition Matrix $\hat{T}$	2.965	11.858	47.432
Total Memory	0.018	0.036	0.072	Sparse $\hat{T}$	0.036	0.071	0.142

Overlapping Generations Model

	Monte Carlo: Sample Size				Histogram: Grid Size		
	100k	250k	500k		250	500	1000
Integer Panel	0.0124	0.031	0.062	Histogram	0.002	0.004	0.007
Float Panel	0.0248	0.062	0.124	Transition Matrix $\hat{T}$	396.941	1587.762	6351.048
Total Memory	0.0372	0.093	0.186	Sparse $\hat{T}$	0.059	0.118	0.235

Notes: The table reports the size in memory of key model objects measured in gigabytes. For the Monte Carlo method these objects correspond to the panels of agents storing their states and choices. States like the ε, ζ, and h are discrete and can be saved as (4 byte) integers while assets are saved as (8 byte) floats. The infinitely lived agent model uses three period panels to compute the auto-correlation of wealth and consumption in Table 1, two integer panels for ε and ζ and two float panels for a and c, respectively. The overlapping generation model uses thirty one period panels tracking a single cohort to compute the auto-correlation of wealth in Table 2, one integer panel for ε and one float panel for a. For the histogram method these objects correspond to the histogram itself and the transition matrix. The histogram is a float of the same dimension as the state space, $n_a \cdot n_\varepsilon \cdot n_\zeta$ for the infinitely lived agent model and $n_a \cdot n_\varepsilon \cdot H$ for the overlapping generations model. The transition matrix has the square of the number of states. The sparse matrix for the infinitely lived agent model needs $n_a \cdot n_\varepsilon^2 \cdot n_\zeta^2 \cdot 2$ entries of floats plus an identical matrix of integers for the position in the histogram. The sparse matrix for the overlapping generations model needs $n_a \cdot n_\varepsilon^2 \cdot H \cdot 2$ entries of floats plus a matrix of integers for the position in the histogram.

or be set to 1 upon death.

Once the sparsity of the Markov kernel is taken into account, the memory required by the key objects involved in the Monte Carlo simulation and the histogram method have the same order of magnitude. Further exploiting the sparsity brings reductions in memory and time requirements as in Tan (2020), Rendahl (2022), or the continuous time methods described in Achdou et al. (2021). Besides, the construction and storage of the Markov kernel is often required as part of the solution of the model, as in Young (2010), regardless of how the moments are calculated. In this way the Markov kernel does not add any overhead when computing moments out of the model's solution.

Ultimately, as the moment of interest involves a larger population or a longer span of time, or if the state space increases, it becomes easier to calculate it using Monte Carlo simulation compared to the histogram iteration method, because of both increasing time and memory costs. However, these cost can be effectively reduced in many scenarios through the use of more efficient solution methods that rely on sparse matrices, particularly in models with a single continuous state variable.

7. Extensions

We have demonstrated how to use a histogram approximation of the stationary distribution of agents and its associated Markov kernel to efficiently compute cross-sectional and longitudinal moments without having to simulate large samples of agents through Monte Carlo methods. We illustrated the workings of the method in the context of baseline models that abstract from many of the characteristics of applied work. However, the histogram iteration method can also be used in other scenarios. We outline some of them below.

Solution methods. The histogram iteration method can be easily applied to models that allow for additional endogenous choices (e.g., labor supply). In this case the policy functions can be solved with extensions of the endogenous grid method like those in Barillas and Fernández-Villaverde (2007) and Fella (2014). Once the policy functions are obtained, the construction of the Markov kernel and the histogram that approximates the distribution follow as above.

Non-stationary environments. Similarly, the method applies to non-stationary problems where the distribution of agents changes over time, or the agents' choices change (therefore making the Markov kernel time-varying). This can happen, for example, in the transition path to a new steady state after changes in policy variables. The histogram iteration method is already built to capture changes in the distribution, as the iteration in equation (6) shows.

The only change required is to index the Markov kernel by time when iterating over an initial distribution of agents.

Continuous-time methods. The histogram iteration method can also be applied with only minor changes to continuous-time heterogeneous agent models as in Herreño and Ocampo (2023). The solution of these models by means of the Finite Difference method is constructed from a (sparse) matrix A that characterizes the approximation to the Hamilton-Jacobi-Bellman equation (see, Achdou, Han, Lasry, Lions, and Moll 2021). The adjoint of this matrix plays the same role as the Markov kernel T described above and characterizes the solution to the Kolmogorov Forward equation that describes the evolution of the distribution of agents.

The changes in the implementation of the histogram iteration method come from how the method can take advantage of the sparseness of matrix A, the stipulation of a time step (Δt), and the need to integrate with respect to the distribution of continuous states. Alternative solution methods for continuous time models as those in Phelan and Eslami (2022) also allow for a direct implementation of the histogram iteration method.

8. Conclusion

The histogram iteration method takes advantage of the histogram approximation of the distribution of agents and the associated Markov kernel, which are often already computed as part of solving heterogeneous agent models. Because of this, the histogram iteration method will usually generate time savings even when the computation of specific moments is costlier than the computation from a simulated Monte Carlo panel, as the simulation has to be conducted on top of the model solution. This makes the key computational trade-off for computing moments clear: the complexity of the moment is weighed against the complexity of simulating a representative sample of agents. This trade-off will usually land in favor of using the model's own stationary distribution and Markov kernel, allowing researchers to avoid both coding and running computationally costly Monte Carlo simulations.

References

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2021. "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach." *The Review of Economic Studies* 89 (1):45–86. URL https://doi.org/10.1093/restud/rdab002.

Aiyagari, S. Rao. 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving." *The Quarterly Journal of Economics* 109 (3):659–684. URL http://www.jstor.org/stable/2118417.

Algan, Yann, Olivier Allais, and Wouter J Den Haan. 2008. "Solving Heterogeneous-Agent Models with Parameterized Cross-Sectional Distributions." *Journal of Economic Dynamics and Control* 32 (3):875–908. URL https://doi.org/10.1016/j.jedc.2007.03.007.

Barillas, Francisco and Jesús Fernández-Villaverde. 2007. "A generalization of the endogenous grid method." *Journal of Economic Dynamics and Control* 31 (8):2698–2712. URL https://doi.org/10.1016/j.jedc.2006.08.005.

Bell, Felicite C. and Michael L. Miller. 2002. "Life Tables for the United States Social Security Area: 1900-2100"." Actuarial Study 116, Office of the Actuary, Social Security Administration. URL https://www.ssa.gov/oact/NOTES/as120/LifeTables_Body.html.

Benhabib, Jess, Alberto Bisin, and Shenghao Zhu. 2011. "The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents." *Econometrica* 79 (1):123–157. URL https://doi.org/10.3982/ECTA8416.

Carroll, Christopher D. 2006. "The method of endogenous gridpoints for solving dynamic stochastic optimization problems." *Economics Letters* 91 (3):312–320. URL https://doi.org/10.1016/j.econlet.2005.09.013.

Conesa, Juan C and Dirk Krueger. 1999. "Social Security Reform with Heterogeneous Agents." *Review of Economic Dynamics* 2 (4):757–795. URL https://doi.org/10.1006/redy.1998.0039.

De Nardi, Mariacristina and Giulio Fella. 2017. "Saving and wealth inequality." *Review of Economic Dynamics* 26:280–300. URL https://doi.org/10.1016/j.red.2017.06.002.

De Nardi, Mariacristina, Eric French, and John Bailey Jones. 2016. "Savings After Retirement: A Survey." *Annual Review of Economics* 8 (1):177–204. URL https://doi.org/10.1146/annurev-economics-080315-015127.

Farmer, Leland E. and Alexis Akira Toda. 2017. "Discretizing nonlinear, non-Gaussian Markov processes with exact conditional moments." *Quantitative Economics* 8 (2):651–683. URL https://doi.org/10.3982/QE737.

Fella, Giulio. 2014. "A generalized endogenous grid method for non-smooth and non-concave problems." *Review of Economic Dynamics* 17 (2):329–344. URL https://doi.org/10.1016/j.red.2013.07.001.

Gouin-Bonenfant, Émilien and Alexis Akira Toda. 2023. "Pareto Extrapolation: An Analytical Framework for Studying Tail Inequality." *Quantitative Economics* 14 (1):201–233. URL https://doi.org/10.3982/QE1817.

Gourio, François and Jianjun Miao. 2010. "Firm Heterogeneity and the Long-Run Effects of Dividend Tax Reform." *American Economic Journal: Macroeconomics* 2 (1):131–68. URL https://doi.org/10.1257/mac.2.1.131.

Guvenen, Fatih, Gueorgui Kambourov, Burhan Kuruscu, Sergio Ocampo, and Daphne Chen. 2023. "Use It or Lose It: Efficiency and Redistributional Effects of Wealth Taxation." *The Quarterly*

Journal of Economics 138 (2):835–894. URL https://doi.org/10.1093/qje/qjac047.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2009. "Quantitative Macroeconomics with Heterogeneous Households." *Annual Review of Economics* 1 (1):319–354. URL https://doi.org/10.1146/annurev.economics.050708.142922.

Heer, Burkhard and Alfred Maußner. 2005. *Dynamic General Equilibrium Modelling*. Springer.

Herreño, Juan and Sergio Ocampo. 2023. "The Macroeconomic Consequences of Subsistence Self-Employment." *Journal of Monetary Economics* 136:91–106. URL https://doi.org/10.1016/j.jmoneco.2023.02.002.

Judd, Kenneth L. 1998. *Numerical Methods in Economics*. MIT Press.

Kopecky, Karen A. and Richard M.H. Suen. 2010. "Finite state Markov-chain approximations to highly persistent processes." *Review of Economic Dynamics* 13 (3):701–714. URL https://doi.org/10.1016/j.red.2010.02.002.

Phelan, Thomas and Keyvan Eslami. 2022. "Applications of Markov chain approximation methods to optimal control problems in economics." *Journal of Economic Dynamics and Control* 143:104437. URL https://doi.org/10.1016/j.jedc.2022.104437.

Rendahl, Pontus. 2022. "Continuous vs. Discrete Time: Some Computational Insights." *Journal of Economic Dynamics and Control* 144:1–16. URL https://doi.org/10.1016/j.jedc.2022.104522.

Ríos-Rull, José-Víctor. 1995. "Models with Heterogeneous Agents." In *Frontiers of Business Cycle Research*, edited by Thomas Cooley, chap. 10. Princeton University Press, 394–330. URL https://doi.org/10.2307/j.ctv14163jx.10.

Rouwenhorst, K. Geert. 1995. "Asset Pricing Implications of Equilibrium Business Cycle Models." In *Frontiers of Business Cycle Research*, edited by Thomas Cooley, chap. 10. Princeton University Press, 394–330. URL https://doi.org/10.2307/j.ctv14163jx.16.

Stachurski, John and Alexis Akira Toda. 2019. "An impossibility theorem for wealth in heterogeneous-agent models with limited heterogeneity." *Journal of Economic Theory* 182:1–24. URL https://doi.org/10.1016/j.jet.2019.04.001.

Stokey, Nancy L., Robert E. Lucas, and Edward C. Prescott. 1989. *Recursive Methods in Economic Dynamics*. Harvard University Press. URL https://doi.org/10.2307/j.ctvjnrt76.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron. 2004. "Cyclical Dynamics in Idiosyncratic Labor Market Risk." *Journal of Political Economy* 112 (3):695–717. URL https://doi.org/10.1086/383105.

Tan, Eugene. 2020. "A fast and low computational memory algorithm for non-stochastic simulations in heterogeneous agent models." *Economics Letters* 193:109285. URL https://doi.org/10.1016/j.econlet.2020.109285.

Tauchen, George. 1986. "Finite state markov-chain approximations to univariate and vector autoregressions." *Economics Letters* 20 (2):177–181. URL https://doi.org/10.1016/0165-1765(86)90168-0.

U.S. Bureau of Economic Analysis. 2022. "Economic Data." FRED, Federal Reserve Bank of St. Louis; April 9, 2022., https://fred.stlouisfed.org.

Young, Eric R. 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. URL https://doi.org/10.1016/j.jedc.2008.11.010.