Markov-Chain Approximations for Life-Cycle Models

Giulio Fella[*] Giovanni Gallipoli[†] Jutong Pan[‡]

December 22, 2018

Abstract

Non-stationary income processes are standard in quantitative life-cycle models, prompted by the observation that within-cohort income inequality increases with age. This paper generalizes Tauchen (1986), Adda and Cooper (2003), and Rouwenhorst's (1995) discretization methods to non-stationary AR(1) processes. We evaluate the performance of these methods in the context of a canonical life-cycle, income-fluctuation problem with a non-stationary income process. We also examine the case in which innovations to the persistent component of earnings are modeled as draws from a mixture of Normal distributions. We find that the generalized Rouwenhorst's method performs consistently better than the others even with a relatively small number of states.

JEL classification: C63

Keywords: Numerical methods, finite state approximations

[*]Queen Mary University of London, CFM and IFS. E-mail: g.fella@qmul.ac.uk
[†]UBC, HCEO and CEPR. E-mail: gallipol@mail.ubc.ca
[‡]Analysis Group. E-mail: panjutong@gmail.com

1 Introduction

Life-cycle models featuring idiosyncratic risk are used extensively to quantitatively examine a wide range of issues, such as the determinants of consumption (Storesletten et al., 2004*b*) and wealth (Huggett, 1996; De Nardi, 2004; Cagetti and De Nardi, 2006), optimal tax progressivity (Conesa and Krueger, 2006; Krueger and Ludwig, 2013; Heathcote et al., 2017) and educational choices (Abbott et al., 2018), just to name a few.

Idiosyncratic labor income risk is most often a crucial ingredient in this class of models. The stylized fact, first documented by Deaton and Paxson (1994), that both income and consumption inequality increase with age implies that the non-stationarity of income must, a fortiori, be driven by a non-stationary persistent component. For this reason, most quantitative life-cycle analyses assume a persistent labor income component whose (unconditional) variance increases with age. Typically this is obtained by positing that the persistent component has a stationary *conditional* distribution (i.e. the persistence parameter and the distribution of innovations are age-independent) but either (a) the process has a unit-root (Storesletten et al., 2004*b*); or, if the persistence parameter is less than unit, (b) the variance of its initial conditions is small relative to the variance of subsequent shocks[1] (Huggett, 1996; Storesletten et al., 2004*a*; Kaplan, 2012). More recently, a number of papers (Karahan and Ozkan, 2013; Blundell et al., 2015; Guvenen et al., 2016; De Nardi et al., 2018) have documented that even the *conditional* distribution of the persistent component of labor income is non-stationary; namely, both its persistence and the variance of innovations change with age. As shown by Karahan and Ozkan (2013) and De Nardi et al. (2018) these latter features are important to account for the pass-through of persistent income shocks onto consumption and for the evolution of cross-sectional consumption dispersion in the data, as well as for the welfare costs of labor income risk.

To sum up, non-stationarity of the (persistent component of the) labor income process is an important feature of any life-cycle model that aims to account for the distribution of consumption and wealth and other forms of heterogeneity in individual outcomes.

Introducing such a process into a quantitative model usually involves approximat-

[1] More formally, the variance of the initial conditions η_0 is smaller than that of the (stationary) asymptotic distribution of η_t as t diverges to infinity.

ing the continuous stochastic process through a Markov chain with a finite state space. As one would expect, the accuracy of such an approximation affects quantitative predictions. Different methods are available to perform such approximation for *stationary* AR(1) processes. Among these, Tauchen (1986) and its variant Tauchen and Hussey (1991), Rouwenhorst (1995), and Adda and Cooper (2003) are the most commonly used in economics. Yet, there is currently no standard, off-the-shelf method for discretizing a *non-stationary* AR(1) process. The quantitative implementations in the extensive literature on life-cycle, heterogeneous-agent models use a variety of approaches and we review some well-known examples in Section 2.4. In most cases these methods are only partially documented, hence we know little about their performance.

Our work is meant to provide a more systematic analysis of this approximation problem. We show how to extend standard discretization methods for stationary AR(1) processes—namely Tauchen (1986), Rouwenhorst (1995), and Adda and Cooper (2003)—to non-stationary AR(1) processes, and we evaluate their performance. As in the original methods, our extensions keep the number of states in each time period constant. The main difference is that both the state vector and the transition matrix are allowed to change over time in accordance with changes in the moments of the original process. In all cases, the defining properties of the original stationary method are preserved.

The properties of alternative discretization methods to approximate *stationary* AR(1) processes in the context of stationary infinite horizon problems have been studied in some detail by Kopecky and Suen (2010). They find that: (a) the choice of discretization method may have a significant impact on the model simulated moments; (b) the performance of Rouwenhorst's (1995) method is more robust, particularly for highly persistent processes. Like in the analysis of Kopecky and Suen (2010) for the stationary case, we compare the respective performance of the three methods both in approximating the original continuous process *and* in generating accurate model solutions. The latter is the relevant metric to assess the impact of alternative discretization methods on the variables of interest to the researcher. Our baseline analysis is carried out within a standard life-cycle, income-fluctuation model with a canonical labor income process featuring both an AR(1) persistent component and a transitory white noise component (Abowd and Card, 1989).

In our numerical implementations we consider three sets of assumptions for the AR(1) component. In the first, more standard, case the conditional distribution is stationary, while the unconditional is not; i.e., both the persistence coefficient and the variance of the shocks are age-independent. In the second case, both the persistence coefficient and the shocks' variance are functions of age, as in Karahan and Ozkan (2013). In both these specifications the innovations to the AR(1) component are assumed to be normally distributed. Finally, in a third case, we maintain the assumption that the persistence coefficient and the shocks' variance are age-independent, but we assume that shocks have a non-normal distribution following Guvenen et al. (2016). In both the first and third case, we report results under different degrees of persistence of the AR(1) process, including the unit root case, as it is well known that the performance of standard discretization methods worsens as the degree of persistence increases.

We find that Rouwenhorst's method tends to perform better even with a relatively small number of grid-points.

The remainder of the paper is structured as follows. Sections 2.1-2.3 discuss how to extend Tauchen (1986), Adda and Cooper (2003) and Rouwenhorst's (1995) methods to non-stationary AR(1) processes. Section 2.4 reviews some well-known implementations of the life-cycle model with idiosyncratic labor income risk, with a focus on their respective approaches to discretizing non-stationary AR(1) processes. Section 3 presents the quantitative framework we use to assess the accuracy of the various methods. Section 4 compares the accuracy of our three discretization methods. Section 5 concludes.

2 Discrete approximations of AR(1) processes

Consider an AR(1) process of the following form,

$$\eta_t = \rho_t \eta_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{id}{\sim} N(0, \sigma_{\varepsilon t}), \tag{1}$$

where the standard deviation of the innovation $\sigma_{\varepsilon t}$ and the autoregressive coefficient ρ_t are both allowed to depend on time t. Furthermore, ρ_t is not restricted to lie inside the unit circle. The initial realization η_0 may be deterministic or a random draw from some distribution.[2] Let σ_t denote the unconditional standard deviation of η_t. It follows from equation (1) that

$$\sigma_t^2 = \rho_t^2 \sigma_{t-1}^2 + \sigma_{\varepsilon t}^2 \tag{2}$$

is also potentially time-dependent and therefore is not, in general, covariance-stationary.

Sufficient conditions for stationarity are that the process in equation (1) is restricted to

$$\eta_t = \rho \eta_{t-1} + \varepsilon_t, \quad |\rho| < 1, \; \varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon) \tag{3}$$

with constant persistence ρ, standard deviation σ_ε, and η_0 randomly drawn from the asymptotic distribution of η_t; namely, $N(0, \sigma)$ where $\sigma = \sigma_\varepsilon / \sqrt{1-\rho^2}$. We call this case the stationary case in what follows, to distinguish it from the general, unrestricted process in equation (1). A number of methods have been proposed to discretize stationary AR(1) processes of this kind by means of an N-state Markov chain with time-independent state space Υ^N and transition matrix Π^N.

In what follows we extend three of these methods—Tauchen (1986), Adda and Cooper (2003) and Rouwenhorst's (1995)—to a *non-stationary* AR(1) of the general form in equation (1).[3] Basically, in each case we approximate the non-stationary AR(1) process by means of a Markov-chain with a time-independent *number* of states N, but time-dependent state space Υ_t^N and transition matrix Π_t^N.

2.1 Tauchen's (1986) method

2.1.1 Stationary case

Tauchen (1986) proposes the following method to discretize a stationary AR(1) process by means of an N-state Markov chain. The state space $\Upsilon^N = \{\bar{\eta}^1, \ldots, \bar{\eta}^N\}$ is uniformly-

[2] Without loss of generality, we assume its expectations to be zero.

[3] Unlike Kopecky and Suen (2010), we do not examine the quadrature-based methods of Tauchen and Hussey (1991) and Flodén (2008). As pointed out in Flodén (2008), Tauchen and Hussey (1991) performs poorly when approximating highly persistent processes, as it uses the quadrature nodes implied by the conditional distribution (the distribution of the shock ε_t in eq. 1). This approach places grid points too close to the mean of the process compared to the unconditional distribution. Flodén (2008) improves on that by selecting the nodes implied by a distribution with a variance that is a weighted average of the conditional σ_ε and the variance of the asymptotic distribution. His choice of weights, however, is appropriate for a stationary AR(1) process and, as Kopecky and Suen (2010) show, the performance of Flodén (2008) is never significantly better, and usually worse, than Rouwenhorst (1995). For these reasons, we restrict our attention to methods that require no extra parameter choices and can be more immediately adapted to non-stationary settings.

spaced with

$$\bar{\eta}^N = -\bar{\eta}^1 = \Omega\sigma$$

where Ω is a positive constant.[4] The transition matrix Π^N is determined as follows. Let Φ denote the cumulative distribution function for the standard normal distribution and $h = 2\Omega\sigma/(N-1)$ the step size between grid points. For any $i, j = 1, \ldots, N$, the transition probabilities satisfy

$$\pi^{ij} = \begin{cases} \Phi\left(\frac{\bar{\eta}^j - \rho\bar{\eta}^i + h/2}{\sigma_\varepsilon}\right) & \text{if } j = 1, \\ 1 - \Phi\left(\frac{\bar{\eta}^j - \rho\bar{\eta}^i - h/2}{\sigma_\varepsilon}\right) & \text{if } j = N, \\ \Phi\left(\frac{\bar{\eta}^j - \rho\bar{\eta}^i + h/2}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\bar{\eta}^j - \rho\bar{\eta}^i - h/2}{\sigma_\varepsilon}\right) & \text{otherwise.} \end{cases} \tag{4}$$

Basically, the method constructs the transition probabilities π_{ij} to equal the probability (truncated at the extremes) that η_t falls in the interval $(\bar{\eta}^j - h/2, \bar{\eta}^j + h/2)$ conditionally on $\eta_{t-1} = \bar{\eta}^i$.

2.1.2 Non-stationary case

Our non-stationary extension of Tauchen (1986) constructs a state space $\Upsilon_t^N = \{\bar{\eta}_t^1, \ldots, \bar{\eta}_t^N\}$ with constant size N, but time-varying grid-points with

$$\bar{\eta}_t^N = -\bar{\eta}_t^1 = \Omega\sigma_t \tag{5}$$

and step size $h_t = 2\Omega\sigma_t/(N-1)$. The associated transition probabilities are

$$\pi_t^{ij} = \begin{cases} \Phi\left(\frac{\bar{\eta}_t^j - \rho\bar{\eta}_{t-1}^i + h_t/2}{\sigma_{\varepsilon t}}\right) & \text{if } j = 1, \\ 1 - \Phi\left(\frac{\bar{\eta}_t^j - \rho\bar{\eta}_{t-1}^i - h_t/2}{\sigma_{\varepsilon t}}\right) & \text{if } j = N, \\ \Phi\left(\frac{\bar{\eta}_t^j - \rho\bar{\eta}_{t-1}^i + h_t/2}{\sigma_{\varepsilon t}}\right) - \Phi\left(\frac{\bar{\eta}_t^j - \rho\bar{\eta}_{t-1}^i - h_t/2}{\sigma_{\varepsilon t}}\right) & \text{otherwise.} \end{cases} \tag{6}$$

The main difference between our extension and its stationary counterpart is that the range of the equidistant state space in equation (5) is time varying and, as a result, so are the transition probabilities.

[4]Tauchen (1986) sets $\Omega = 3$. Kopecky and Suen (2010) calibrate it so that the standard deviation of the Markov chain coincides with that of the original AR(1) process.

2.2 Adda and Cooper's (2003) method

2.2.1 Stationary case

The Adda-Cooper discretization method is similar to Tauchen's, except that the support of the unconditional distribution of η is partitioned into N intervals $\{[x^i, x^{i+1}]\}_{i=1}^N$ each having equal probability mass. Under the maintained assumptions that η_t has zero mean and the innovation ε_t is normally distributed, we can define the intervals by solving the following system of equations,

$$\Phi\left(\frac{x^i}{\sigma}\right) = \frac{i-1}{N} \qquad i = 1, ..., N+1,$$

where $\Phi(\cdot)$ is the standard Normal cumulative probability function.[5]

Element $\eta^i,\ i = 1, \ldots, N$ of the state space equals the expected value of the variable η conditional on it assuming values within the interval $[x^i, x^{i+1}]$. The transition probability $\pi^{i,j}$ is defined as the conditional probability of η moving from interval $[x^i, x^{i+1}]$ to interval $[x^j, x^{j+1}]$ from one period to the next and have to be computed numerically.

2.2.2 Non-stationary case

To implement the Adda and Cooper method in a non-stationary setting we first construct N intervals, $\{[x_t^i, x_t^{i+1}]\}_{i=1}^N$, *for each period* t. The cut-off points are the solutions to

$$\Phi\left(\frac{x_t^i}{\sigma_t}\right) = \frac{i-1}{N} \qquad i = 1, ..., N+1, \tag{7}$$

where the unconditional mean is still zero and the variance σ_t of η_t now depend on t.

The transition probability $\pi_t^{i,j}$ is then defined as the probability of η moving from the interval $[x_t^i, x_t^{i+1}]$ to the interval $[x_{t+1}^j, x_{t+1}^{j+1}]$ between t and $t+1$.

2.3 Rouwenhorst's (1995) method

The Rouwenhorst method of is best understood as determining the parameters of a two-state Markov chain, with equidistant state space, in such a way that the conditional first

[5]The formula implies $x^1 = -\infty$ and $x^{N+1} = +\infty$. For more details, see pages 57-58 of Adda and Cooper (2003).

and second moments of the Markov chain coincide with the same moments of the original AR(1) process.[6]

2.3.1 Stationary case

In the case of the stationary AR(1) process in equation (3), the state space for the two-state Markov chain is $\bar{\eta}^2 = -\bar{\eta}^1$ and the transition matrix is written as

$$\Pi^2 = \begin{bmatrix} \pi^{11} & 1-\pi^{11} \\ 1-\pi^{22} & \pi^{22} \end{bmatrix}. \tag{8}$$

The moment condition for the expectation conditional on $\eta_{t-1} = \bar{\eta}^2$ is

$$\mathrm{E}(\eta_t | \eta_{t-1} = \bar{\eta}^2) = -(1-\pi^{22})\bar{\eta}^2 + \pi^{22}\bar{\eta}^2 = \rho\bar{\eta}^2, \tag{9}$$

where the left hand side is the conditional expectation of the Markov chain and the right hand side its counterpart for the AR(1) process with η_{t-1} evaluated at the grid point $\bar{\eta}^2$. It follows that

$$\pi^{22} = \frac{1+\rho}{2} = \pi^{11}, \tag{10}$$

where the second equality follows from imposing the same condition for $\eta_{t-1} = \bar{\eta}^1 = -\bar{\eta}^2$.

If $\eta_{t-1} = \bar{\eta}^2$, the moment condition for the variance is[7]

$$\mathrm{Var}(\eta_t | \eta_{t-1} = \bar{\eta}^2) = (1-\pi^{22})\left(-\bar{\eta}^2 - \rho\bar{\eta}^2\right)^2 + \pi^{22}\left(\bar{\eta}^2 - \rho\bar{\eta}^2\right)^2 = \sigma_\varepsilon^2, \tag{11}$$

which, after replacing for π^{22} from equation (10), implies

$$\bar{\eta}^2 = \sigma. \tag{12}$$

Having determined Π^2 and Υ^2, the method scales to an arbitrary number of grid points N in the following way.[8] The state space $\Upsilon^N = \{\bar{\eta}^1, \ldots, \bar{\eta}^N\}$ is uniformly-spaced

[6] A Markov chain of order N is characterized by N^2 parameters (N states plus $(N^2 - N)$ linearly-independent transition probabilities) and can be uniquely identified by N^2 linearly-independent moment conditions. Rouwenhorst's method is, therefore, a special case of a general moment-matching procedure.

[7] By symmetry the other conditional-variance equation is satisfied whenever equation (11) holds with equality.

[8] We refer the reader to Rouwenhorst (1995) and Kopecky and Suen (2010) for a rigorous derivation.

with

$$\bar{\eta}^N = -\bar{\eta}^1 = \sigma\sqrt{N-1}. \tag{13}$$

For $N \geq 3$, the transition matrix satisfies the recursion

$$\Pi^N = \pi \begin{bmatrix} \Pi^{N-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} + (1-\pi) \begin{bmatrix} \mathbf{0} & \Pi^{N-1} \\ 0 & \mathbf{0}' \end{bmatrix} + \pi \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \Pi^{N-1} \end{bmatrix} + (1-\pi) \begin{bmatrix} \mathbf{0}' & 0 \\ \Pi^{N-1} & \mathbf{0} \end{bmatrix}, \tag{14}$$

where $\pi = \pi^{11} = \pi^{22}$ and $\mathbf{0}$ is an $(N-1)$ column vector of zeros. The main difference between Rouwenhorst and the other two methods is that in the former the transition probabilities do not embody the normality assumption about the distribution of the shocks. Rather, Rouwenhorst matches exactly, by construction, the first and second conditional moments and, by the law of iterated expectations, also the unconditional moments of the continuous process, independently of the shock distribution.

2.3.2 Non-stationary case

Just like in the Tauchen's case, our non-stationary implementation of Rouwenhorst (1995) constructs an equally-spaced, symmetric, state space $\Upsilon_t^N = \{\bar{\eta}_t^1, \ldots, \bar{\eta}_t^N\}$. The grid cardinality N is constant but the grid points and transition matrix Π_t^N change with the index t. If $N = 2$, it follows that $\bar{\eta}_t^2 = -\bar{\eta}_t^1$ and the counterpart of the first-moment condition (9) becomes

$$\mathrm{E}(\eta_t | \eta_{t-1} = \bar{\eta}_{t-1}^2) = -(1 - \pi_t^{22})\bar{\eta}_t^2 + \pi_t^{22}\bar{\eta}_t^2 = \rho_t \bar{\eta}_{t-1}^2,$$

with unique solution

$$\pi_t^{22} = \frac{1}{2}\left(1 + \rho_t \frac{\bar{\eta}_{t-1}^2}{\bar{\eta}_t^2}\right) = \frac{1}{2}\left(1 + \rho_t \frac{\sigma_{t-1}}{\sigma_t}\right) = \pi_t^{11}, \tag{15}$$

where the second equality follows from the counterpart of the second moment condition (11) which implies

$$\bar{\eta}_t^2 = -\bar{\eta}_t^1 = \sigma_t. \tag{16}$$

The third equality in equation (15) follows from the expression for the conditional first moment for $\eta_{t-1} = \bar{\eta}_{t-1}$.

As in the non-stationary version of Tauchen, the points of the state space are a function of the time-dependent unconditional variance of η_t. Comparing equations (10) and (15) reveals that, relative to the stationary case, the probability π_t^{22} of transiting from $\bar{\eta}_{t-1}^2$ to $\bar{\eta}_t^2$ depends on the rate of growth of the unconditional variance of η_t.

Equation (15) implies that the condition for the Markov chain to be well defined, and have no absorbing states, namely $0 < \pi_t^{11} = \pi_t^{22} < 1$, is equivalent to

$$\rho_t^2 \frac{\sigma_{t-1}^2}{\sigma_t^2} < 1. \tag{17}$$

It follows from equation 2(2) that this condition always holds. Therefore Rouwenhorst's approximation can be applied to any process of the type defined in equation (1).[9]

As in the stationary case, the approach scales to an N-dimensional, evenly-spaced state space Υ_t^N by setting

$$\bar{\eta}_t^N = -\bar{\eta}_t^1 = \sigma_t \sqrt{N-1} \tag{18}$$

and Π_t^N to satisfy the recursion (14) with the transition matrices and the probability $\pi_t = \pi_t^{11} = \pi_t^{22}$ indexed by t.

2.4 Alternative discretization approaches

This section reviews some well-known numerical implementations of the life-cycle model, with a focus on their approach to the discretization of non-stationary AR(1) income processes. Rather than providing an exhaustive survey of the wide-ranging literature in this area, we aim to present few illustrative examples and use them to highlight key issues in the approximation of income processes within finite life-cycle models.

Huggett (1996) examines the ability of a life cycle model to reproduce the age-wealth distribution in US data. He assumes an AR(1) labor income process as in equation (3) and normally distributed initial conditions η_0 with standard deviation σ_{η_0} lower than the asymptotic standard deviation $\sigma = \sigma_\varepsilon / \sqrt{1-\rho^2}$. Therefore, the cross-sectional variance of income increases with age. To discretize the income process Huggett (1996) uses a variant of the (stationary) method of Tauchen (1986). Specifically, he posits an age-invariant, uniformly-spaced grid of 18 states with $\bar{\eta}^N = -\hat{\eta}^1 = 4\sigma_{\eta_0} \approx 3\sigma_\eta$—where the last relation follows from the fact that $\sigma_{\eta_0}/\sigma \approx .7$ in his calibration—plus an extra grid point equal

[9] This is also trivially true for Tauchen's method.

to $6\sigma_{\eta_0}$. Basically, this approach constructs the age-invariant state space using Tauchen's approximation of the asymptotic distribution of labor income.[10] Huggett (1996) reconciles the non-stationarity of the earnings process with the stationarity of the state space by imposing that draws of the initial condition η_0 are restricted to points of age-invariant state space.[11] This way, at the simulation stage, the Markov chain has a non-stationary distribution despite a time-invariant state space and transition probabilities. The price to pay for holding the state space constant is that one uses the same range for grid points at all ages, which implies that grid points are placed too far out, relative to Tauchen, for ages at which the unconditional distribution is less disperse than the asymptotic one. De Nardi (2004) studies a similar model, with the addition of inter-generational bequests and transmission of earnings ability, to explain wealth concentration at the top. Along the lines of Huggett (1996), she uses a time-invariant Markov chain with a distribution of η_0 that differs from the asymptotic one. The main difference is that the 4 grid points and the associated transition probabilities are chosen following Tauchen and Hussey (1991).[12] These two variations of Huggett's approach to introduce non-stationary AR(1) processes into life-cycle models have been extensively used in the literature (e.g. Conesa and Krueger, 2006).

Storesletten et al. (2004*b*) explore the extent to which individual-specific earnings risk can account for the fanning-out of cross-sectional income and consumption inequality over the life-cycle observed in data. Their assumed income process features a permanent (random walk) and a transitory component, both with stationary conditional distributions. They approximate the random walk through a binomial tree with innovations taking two possible values $\{\pm 0.127\}$, each with probability 0.5.[13] Given an initial condition $\eta_0 = 0$, the support for the shocks fans out over the life-cycle and traces the income dynamics

[10] The sentence in the paper, "The transition probabilities between states are calculated by integrating the area under the normal distribution conditional on the current value of the state." suggests that also the transition probabilities follow Tauchen.

[11] In other words, η_0 is distributed over the grid-points $\bar{\eta}^i$ with probabilities given by equation (4) with $\rho = 0$ and normalizing by σ_{η_0} instead of σ_ε.

[12] This is likely even more problematic because Tauchen and Hussey (1991) chooses grid points on the basis of the *conditional distribution* of the process (that is, the distribution of the innovations) implying that the fit worsens as the unconditional distribution fans out with age.

[13] The two values are those implied by the Gaussian-Hermite nodes. To be precise, they write that the transition probabilities are "...chosen following Tauchen and Hussey (1991)." This implies that the transition probabilities are 0.5 given symmetry and the requirement that they add up to one.

over an equally-spaced state space whose dimension depends on the length of the working life. Their approximation features a Markov chain with 62[14] states and age-dependent transition probabilities implied by the binomial process.

Kaplan (2012) structurally estimates a life-cycle model with endogenous labor supply to account for the joint distribution of wages, hours and consumption. He estimates a wage process featuring both a persistent AR(1) component (with less-than-unit persistence) and a transitory component. The initial condition for the persistent component is $\eta_0 = 0$ and he approximates the persistent component through an 11-state Markov chain defined over an "...age-varying state. Values and transition probabilities are chosen to match the age-varying unconditional variance and dependence structure...to that implied by the continuous process." Unlike the papers mentioned above, Kaplan's approach is effectively a moment-matching method similar in spirit to that of Rouwenhorst's. In fact his description is reminiscent of, and consistent with, our non-stationary adaptation of Rouwenhorst's method, as described in Section 2.3.2. While the description of this method does not contain enough details to reproduce it, we expect this approach to enjoy some of the advantages of the Rouwenhorst's moment-matching method that we document below.

All the above papers assume that the *conditional* distribution of the persistent income component (namely, the persistence parameter and the shocks' variance) is time-variant. Karahan and Ozkan (2013) were the first to examine the implications of age-varying persistence and shocks' variances for the age-profile of consumption and the pass-through of persistent income shocks onto consumption. In their analysis they use 61 grid points for the normally-distributed, persistent labor income component. While there is no detailed information about the implementation steps, we conjecture that the solution method is based on the same highly accurate quadrature-based approach used in Guvenen et al. (2016), which we describe below.

Guvenen et al. (2016) document how the distribution of labor earnings growth rates displays significant deviations from normality and age-dependence of conditional moments. They estimate a sophisticated parametric labor earnings process by simulated

[14] Strictly speaking, given a working life of $T = 43$ years in their calibration, the process should actually trace 87 (i.e., $2T+1$), not 62, states. We are unsure whether they actually truncate the support of their process. In our implementation of their method in Section 4.1.1 we do not truncate it.

method of moments using W2 Social Security individual earnings data for the US. The process features substantial individual heterogeneity as well as a persistent and a transitory component. To capture the non-normality in the data, the persistent component is modeled as a mixture of two AR(1) processes, with shocks that are, in turn, a mixture of a normal distribution and a mass point. Mixtures probabilities are allowed to depend on age. The estimated income process is then introduced into a life cycle model in order to assess its implications for wealth inequality, self-insurance and welfare. Each of the two AR(1) components is discretized by using a relatively dense, age-invariant grid (with, respectively, 21 and 41 points). At each grid point, expectations with respect to the Gaussian component of innovations are computed using Gaussian-Hermite quadrature with 7 nodes and bi-dimensional interpolation. This approach is highly accurate and, for this reason, we use it to compute our benchmark solutions and to assess the performance of the three alternative discretization methods in Section 4. The downside of this approach is that it is computationally very costly and harder to implement because it relies heavily on interpolation to compute expectations.

Finally, De Nardi et al. (2018) estimate a "persistent plus transitory" process for disposable household earnings using US data from the PSID. They use the semi-parametric method developed by Arellano et al. (2017) to estimate a polynomial approximation to the distributions of both the persistent and the transitory component, allowing for age-dependence, non-normality and non-linearity. A discretized version of the estimated process is then introduced into a life-cycle model to assess its implications for the age profile of cross-sectional consumption dispersion, self-insurance and welfare. In the numerical implementation the different earnings components are discretized as follows. First, for each component, they simulate a large panel of individual histories using the estimated distribution. Second, they discretize the simulated marginal distribution of earnings at each age into a (age-independent) set of bins and replace the (heterogeneous) values of earnings in each bin with their median. The associated, age-specific transition matrices are then obtained by computing the proportion of observations transiting from bins of the earnings distribution at age t to bins at age $t+1$. The result is a non-parametric representation of the process that follows a Markov chain with an age-dependent transition matrix and a fixed number of age-dependent earnings states. This method is conceptually very similar to our non-stationary extension of Adda and Cooper (2003) in Section 2.2.2 with a few differences. In both cases the marginal distributions are discretized into a finite number of bins and transition probabilities are computed by numerically integrating the conditional distributions. Their simulation-based, counting procedure is effectively Monte Carlo integration. The main difference is that contrary to our method in Section 2.2.2 (a) their bins have different mass, with a finer partition in the tails of the distribution;[15] (b) they assign to each interval the median value among observations within that interval, rather than the mean.[16]

In Section 4 we examine the approximation quality of the well-known discretization methods by Tauchen, Adda-Cooper and Rouwenhorst, appropriately adapted to non-stationary income processes. For comparison we also report approximation results for two of the ad-hoc methods described above (Huggett, 1996 and Storesletten et al., 2004) whose implementation procedures are fairly straightforward and easily reproducible. Of course, as we mentioned, we implement the more complex and computationally intensive approach of Guvenen et al. (2016) to recover a highly accurate benchmark solution for all the income models we study below.

3 Evaluation

In order to assess the performance of the different discretization methods described above, we solve a finite-horizon, income-fluctuation problem with a persistent, non-stationary labor income process. Countless variations of this model have been studied motivated by Deaton and Paxson's (1994) finding that within-cohort income inequality increases with the age of a cohort.

[15] They use 18 bins for the persistent labor income component with the top and bottom five bins having each 2% mass and the remaining 8 bins containing 10% of observations.

[16] They choose non-homogeneous bin sizes, heuristically, to obtain a good fit of the age-dependent moments of the original distribution.

The problem has the following recursive representation[17]

$$
\begin{aligned}
V_t(z_t, \eta_t) = \max_{c_t, a_t} \log(c_t) + \beta \mathbb{E}_t V_{t+1}(z_{t+1}, \eta_{t+1}) \qquad (19) \\
\text{s.t. } z_t = (1+r)a_{t-1} + y_t \\
a_t + c_t = z_t \\
\log y_t = \eta_t + u_t \\
\eta_t = \rho_t \eta_{t-1} + \varepsilon_t, \\
\varepsilon_t \sim F(0, \sigma_{\varepsilon t}), \quad u_t \overset{i.i.d.}{\sim} N(0, \sigma_u) \\
a_{t+1} \geq 0, \quad a_0, y_0 \text{ given.}
\end{aligned}
$$

Individuals start life in period 1 and live until period $T = 40$. In each period t, they allocate their total cash at hand z_t between consumption c_t and the stock a_{t+1} of an asset paying a risk-free interest rate r. In each period, agents receive a stochastic flow of labor income y_t whose logarithm is the sum of an Gaussian i.i.d. transitory component ε_t and a persistent AR(1) components η_t. The allow for the possibility that the autoregressive coefficient of η_t and the standard deviation of its innovation ε_t depend on time, as well as for the distribution of ε_t to be non-Gaussian.

Since, as is well known, accurately approximating a transitory stochastic process poses no serious difficulty, we apply our discretization methods to the persistent component η_t. Similarly, we do not consider fixed effects in the labor income process because, by definition, they are drawn only once and, therefore, the only relevant issue in their modeling is the choice of the number and location of grid points.[18]

We evaluate the accuracy of the different discretization methods in the following way. We solve the optimization problem by using the endogenous grid point method of Carroll (2006) and Gaussian-Hermite quadrature and linear interpolation over z to compute the expectation with respect to the transitory component u. As for computing the expectation with respect to the persistent component, we use either of our three discretization methods and compare it to an alternative, highly accurate, benchmark

[17]The zero lower bound for next period's assets is without loss of generality. It is always possible to rewrite the problem so that the lower bound on, the appropriately translated, asset space is zero.

[18]In our model with no retirement stage, fixed effects would simply induce a parallel shift in consumption at each age.

solution using a time-invariant grid for η with a large number of points and Gaussian quadrature to compute the expectation.[19] In the latter case, we interpolate bi-linearly over z and η. We describe our benchmark solution in detail in Appendix A.1.

In all cases we solve for the policy functions using a common, 1,000-point grid, for z and 5 quadrature nodes for the transitory shock u.[20] As for η, we use a method-dependent N-state ($N = 5,\ 10,\ 25$) Markov chain under our three discretization methods against 10,000 grid points and 5 quadrature nodes for the benchmark solution.

After solving for the policy functions, we generate 2,000,000 individual income histories. We do this in two different ways. In the first case, we generate the income histories using the discrete Markov-chain approximation. The simulation involves linearly interpolating the policy functions only with respect to z. In the second case, as in the benchmark solution, we generate income histories using the *continuous* $AR(1)$ process for η. We then interpolate bi-linearly over both z *and* labor income η. The key difference between these two approaches has to do with the sources of the errors that they introduce. Both cases suffer from approximation errors in the policy function, relative to the benchmark, due to the fact that the policy functions solve the Euler equations exactly only at a relatively small number of grid points for labor income. The continuous $AR(1)$ simulation does not suffer from the approximation error of discretization that exists in the Markov-chain simulation, but introduces an additional source of error due to the bi-linear interpolation with respect to η and z.

In both cases, we assess the accuracy of the three discretization methods by comparing simulated moments under each method to moments generated by the benchmark solution.

We set the common parameters to the following standard values. The discount rate is $\beta = 0.96$, the interest rate $r = 0.04$, initial wealth $a_0 = 0$ and $\eta_0 = 0$.

3.1 Accuracy of the benchmark solution

As stated above, our chosen benchmark involves solving the problem by the endogenous grid point method on a finite set of grid points for (z, η), Gaussian-Hermite quadrature to compute expectations and bi-linearly interpolation in between grid points. Although the

[19]The same integration method is used in Guvenen et al. (2016) (see their Appendix D.1.2).

[20]Given n quadrature nodes, Gaussian quadrature approximates exactly the integral of any polynomial function of degree up to $2n - 1$.

accuracy properties of the endogenous gridpoint methods with linear interpolation in one dimension are well known (see Barillas and Fernández-Villaverde, 2007), one may want to be reassured that the method, combined with Gaussian quadrature, remains accurate when interpolating bi-linearly over z *and* η.[21]

In order to assess the accuracy of our method we study the special case in which the persistent component η is a random walk and the variances of the innovations are constant. The advantage of the assumption is that, as first shown in Carroll (2004), the combination in such a case problem (19) can be normalized by the permanent labor income η_t component, thereby reducing the effective state space to the single variable $\hat{z}_t = z_t/\eta_t$.[22]. It follows that, under the assumption that income innovations are log-normally distributed, one can solve the model in (19) using an alternative procedure. This entails implementing the endogenous gridpoint method interpolating just with respect to $\hat{z}_t$ and using Gaussian-Hermite quadrature to integrate with respect to u_t and ε_t. Given the well-known properties of quadrature, the model solution based on the endogenous gridpoint method and quadrature can be considered highly accurate.

Furthermore since, by construction, the non-normalized policy function $a_t(z_t, y_t) = \hat{a}_t(\hat{z}_t)y_t$ is linear in labor income, our benchmark simulation does not require any approximation with respect to labor income. Therefore, the simulated moments generated by this benchmark solution constitute a highly accurate approximation to the true model moments.

As we show in Appendix A.2 this alternative solution of the benchmark unit root model delivers simulated moments that are effectively identical to the ones obtained when we use our baseline solution method. This is reassuring and indicates that our standard benchmark solution is extremely accurate and provides a reliable benchmark when assessing the different discretization methods.

[21] Kopecky and Suen (2010) use a non-finite state space benchmark solution—the Parameterized Expectations Algorithm (PAE) of den Haan and Marcet (1990)—which does not require interpolation when assessing the accuracy of alternative methods to discretize stationary AR(1) processes used in *stationary, infinite-horizon* models. The computational costs of PAE are very large in life-cycle models with realistic lifetimes, given the non-stationarity of the policy functions. This likely explains why we are not aware of any paper using PAE to solve a life-cycle model.

[22] Appendix A.1 reports the derivation.

4 Results

This section compares the performance of the three discretization methods considered under alternative assumptions about the stochastic process of the persistent earnings component. Section 4.1 considers the most common case (e.g. Storesletten et al., 2004*b*) in which both the autoregressive coefficient ρ_t and the variances of the income innovations are constant over time. In this setting the non-stationarity of the permanent component is due to the initial draw η_0 of the persistent component not being drawn from the limiting (stationary) distribution of η and/or to the fact that the autoregressive process has a unit root. Section 4.2 considers the case in which both ρ_t and the variance of the innovations $\sigma^2_{\varepsilon_t}$ are non-stationary, as in Karahan and Ozkan (2013). Finally, in Section 4.3 we allow for non-normality of the innovations.

We evaluate the accuracy of the alternative discretization methods by comparing simulated moments obtained under the benchmark approach to those obtained under either of the three discretization methods. We report percent deviations from benchmark values for the unconditional mean and the standard deviation of income, consumption and assets. Given the growing interest in wealth concentration, we also compute deviations from the benchmark share of aggregate wealth held by households in the top 5% of the wealth distribution.

4.1 Canonical persistent-transitory process

We set the variance of the AR(1) innovations $\sigma^2_\varepsilon = .0161$ and the variance of the transitory component $\sigma^2_u = .063$, as in Storesletten et al. (2004*b*).[23] We consider three possible values for the autoregressive coefficient, namely $\rho \in \{0.95,\ 0.98,\ 1\}$.[24]

Tables 1 and 2 report the size of the approximation error for moments simulated using, respectively, Tauchen, Adda and Cooper, and Rouwenhorst's income discretization methods. In the case of Tauchen we set $\Omega = 3$ in equation (5).[25] Approximation errors are

[23]The parameterization implies an aggregate wealth-income ratio slightly below 0.7 when $\rho = 1$, which is in line with the baseline calibration in Carroll (2009) for a similar model with no retirement and deterministic lifetime.

[24]Storesletten et al. (2004*b*) set $\rho = 1$ in their benchmark parameterization.

[25]This is the value in Tauchen (1986). Kopecky and Suen (2010) shows that Tauchen's method is sensitive to the choice of Ω. For this reason, in Appendix A.3 we calibrate a time-dependent value of Ω to match the unconditional variance of the process at each age. Our key finding is that the relative

Table 1: Percentage deviations from benchmark moments: Markov-chain simulation.

	$N=5$			$N=10$			$N=25$		
	Tau (%)	AC (%)	Rou (%)	Tau (%)	AC (%)	Rou (%)	Tau (%)	AC (%)	Rou (%)
					$\rho=0.95$				
Mean of y	3.73	-0.75	**-0.08**	1.48	-0.33	**-0.06**	0.08	-0.14	**-0.05**
SD of y	32.46	-9.96	**-2.18**	13.17	-4.99	**-0.98**	0.75	-2.06	**-0.37**
Mean of c	2.49	**-0.02**	-0.04	1.91	**-0.01**	-0.03	0.16	-0.08	**-0.04**
SD of c	29.79	-11.59	**-1.89**	12.84	-5.81	**-0.80**	0.71	-2.48	**-0.30**
Mean of a	-30.16	19.24	**0.98**	13.28	8.48	**0.72**	2.09	1.46	**0.21**
SD of a	-21.02	**1.11**	-3.40	11.28	**0.41**	-1.56	0.57	**-0.14**	-0.66
Top 5% wealth share	9.99	-18.09	**-4.60**	**-0.85**	-8.54	-2.22	-1.13	-1.40	**-0.85**
					$\rho=0.98$				
Mean of y	5.72	-1.31	**-0.17**	3.15	-0.61	**-0.11**	0.30	-0.25	**-0.07**
SD of y	34.64	-13.66	**-4.04**	18.87	-7.49	**-1.89**	1.31	-3.43	**-0.67**
Mean of c	4.32	0.19	**-0.14**	3.61	0.15	**-0.09**	0.43	**-0.03**	-0.06
SD of c	27.85	-13.11	**-3.80**	18.84	-7.13	**-1.75**	1.42	-3.40	**-0.61**
Mean of a	-38.11	45.31	**0.79**	17.37	22.91	**0.38**	4.12	6.62	**0.19**
SD of a	-47.14	21.25	**-3.79**	11.36	11.09	**-1.73**	2.18	4.62	**-0.77**
Top 5% wealth share	-22.56	-17.80	**-4.38**	-3.37	-7.55	**-1.90**	-1.16	**0.51**	-0.86
					$\rho=1.00$				
Mean of y	8.82	-2.34	**-0.38**	5.48	-1.12	**-0.18**	0.76	-0.48	**-0.15**
SD of y	40.02	-20.43	**-7.92**	23.35	-12.42	**-3.67**	1.74	-6.48	**-1.44**
Mean of c	8.21	**-0.06**	-0.32	5.68	**0.06**	-0.15	0.98	**-0.07**	-0.13
SD of c	37.00	-16.12	**-7.29**	22.43	-9.57	**-3.33**	2.30	-5.10	**-1.28**
Mean of a	-14.86	85.45	**1.72**	13.12	44.58	**0.83**	9.42	15.61	**0.25**
SD of a	-26.26	69.61	**-0.41**	-2.92	39.39	**-0.16**	4.84	17.43	**-0.11**
Top 5% wealth share	-19.61	-4.17	**-1.11**	-13.26	6.01	**-0.43**	-2.05	8.31	**-0.19**

Note: we report in bold the lowest deviation, for each moment and number of grid points.

Table 2: Percentage deviations from benchmark moments: continuous process simulation.

	$N = 5$			$N = 10$			$N = 25$		
	Tau (%)	AC (%)	Rou (%)	Tau (%)	AC (%)	Rou (%)	Tau (%)	AC (%)	Rou (%)
				$\rho = 0.95$					
Mean of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean of c	-1.95	1.03	**-0.06**	0.18	0.44	**-0.03**	0.06	0.10	**-0.01**
SD of c	-4.97	6.13	**0.00**	0.39	2.79	**-0.12**	0.22	0.96	**-0.05**
Mean of a	-53.22	28.24	**-1.72**	4.80	11.93	**-0.83**	1.54	2.67	**-0.28**
SD of a	-38.71	31.37	**0.45**	2.24	16.47	**-0.63**	1.32	6.87	**-0.30**
Top 5% wealth share	16.59	4.83	**2.30**	-2.38	5.74	**0.26**	-0.11	4.47	**0.03**
				$\rho = 0.98$					
Mean of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean of c	-1.81	1.85	**-0.19**	0.19	0.90	**-0.12**	0.10	0.27	**-0.04**
SD of c	-4.40	10.71	**0.32**	**-0.01**	5.52	-0.23	0.44	2.25	**-0.11**
Mean of a	-56.47	57.77	**-5.96**	5.89	28.08	**-3.74**	3.15	8.52	**-1.30**
SD of a	-54.30	72.31	**3.80**	**0.17**	40.70	-1.23	3.65	18.68	**-0.73**
Top 5% wealth share	-9.93	13.24	**8.94**	-5.74	12.61	**2.11**	**0.15**	9.08	0.53
				$\rho = 1.00$					
Mean of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean of c	-1.00	2.71	**-0.38**	**0.06**	1.38	-0.25	0.21	0.49	**-0.10**
SD of c	**-0.95**	17.99	2.03	-0.66	10.71	**-0.18**	1.01	5.19	**-0.18**
Mean of a	-38.61	104.51	**-14.57**	**2.15**	53.21	-9.72	8.20	18.96	**-3.82**
SD of a	-39.32	176.63	**21.34**	-8.26	108.81	**-0.80**	8.82	54.58	**-1.16**
Top 5% wealth share	**-9.26**	43.60	29.11	-12.61	38.80	**6.58**	**-0.16**	23.67	1.96

Note: we report in bold the lowest deviation, for each moment and number of grid points.

expressed in percentage deviations from the benchmark solution and each table reports results for the three values of ρ we consider.

Table 1 reports results for the case in which the discretized income processes are used both to compute the expectation in the decision problem and to simulate the model (Markov-chain simulation). It is apparent that Rouwenhorst's method provides the best approximation by a wide margin, with Tauchen and Adda and Cooper's methods delivering much less precise approximations even as the grid size increases. While all methods tend to provide better approximations when the number of grid points increases, Rouwenhorst's is the only method that results in highly accurate moments even with just 5 grid points for the income process. In fact, a five points approximation using Rouwenhorst's method is generally more accurate than the 25-point approximation using either of the other methods.

Differences in performance become noticeably larger when income persistence grows: in the unit root case, Tauchen and Adda-Cooper's methods do a rather unremarkable job of approximating the distributions of consumption and assets even when using a dense grid, while Rouwenhorst's approximation is consistently good regardless of grid size.

The largest discrepancies in the quality of the approximation occur for the distribution of assets, where the only discretization method delivering consistently accurate moments is Rouwenhorst, regardless of the number of grid points. This performance gap is especially visible when considering the standard deviation of assets. In the latter case, the Rouwenhorst approximation using the Markov-chain simulation approach has an error of at most 3.79 per cent for $N = 5$ and $\rho = 0.98$, while the lowest error is only 0.11 per cent for $N = 25$ and $\rho = 0.98$. In contrast, the Tauchen approximation is off by a much wider margin relative to the benchmark quadrature method, up to a 47% deviation for the SD of assets. The Adda and Cooper approximation does better than Tauchen when ρ is relatively low but performs poorly with higher persistence, even with a dense grid. Striking differences are also evident in the approximation of the upper tail of assets (top 5% share) where Rouwenhorst's method is roughly one order of magnitude more accurate than either Tauchen or Adda and Cooper.

It is also apparent that, under the Tauchen discretization, the approximation error

performance of the three methods is not significantly affected.

does not necessarily shrink as the number of grid points increases. Intuitively, when comparing the range of the income grid for Tauchen (equation 5) and Rouwenhorst (equation 13) methods, the range of the income grid increases faster with N for the latter method. This implies that, in the case of Tauchen, a larger number of simulated observations get piled onto the bounds relative to the benchmark method, reducing accuracy. This problem appears to be quite important when approximating the standard deviation of wealth holdings.

Table 2 reports the approximation errors for the case in which the discretized income processes are used only to compute the expectation in the decision problem but the true (continuous) income processes are used in the model simulation (continuous process simulation). By construction, there is no approximation error for the income process in this case, which explains the zero deviations for the moments of y. On the other hand, the simulation now involves sampling and interpolation errors, the latter stemming from interpolating the policy functions with respect to η using the grid points as interpolating nodes. The approximation of the other moments is generally better under this simulation approach when $\rho = 0.95$. This is particularly true in the case of Rouwenhorst. For all methods, it is substantially worse than the Markov-chain simulation for ρ greater or equal 0.98. Clearly, for a given number of grid points, the effect of the interpolation errors increases with ρ. As in the previous case, though, Rouwenhorst's approximation errors are roughly one order of magnitude smaller than either Tauchen or Adda and Cooper.

In summary, these exercises indicate that Rouwenhorst's method provides considerable advantages when approximating the life-cycle income processes considered in this section. Specifically, we find that Rouwenhorst's approximation: (i) consistently performs better than the popular alternatives considered, across all moments and grid sizes; (ii) delivers a significantly better approximation of the assets' distribution, especially at the top end of its range; (iii) is more robust to the choice of simulation method, exhibiting smaller discrepancies between discrete and continuous simulations of the income process.

Table 3: Percentage deviations from benchmark moments: comparison of Huggett method and the three methods. $\rho = 0.95$.

	$N=5$				$N=10$				$N=25$			
	Hug (%)	Tau (%)	AC (%)	Rou (%)	Hug (%)	Tau (%)	AC (%)	Rou (%)	Hug (%)	Tau (%)	AC (%)	Rou (%)
						Markov-chain simulation						
Mean of y	3.21	3.69	-0.69	**-0.08**	3.29	1.45	-0.27	**-0.06**	1.43	0.08	-0.08	**-0.05**
SD of y	28.12	32.46	-9.48	**-2.18**	27.36	13.13	-4.49	**-0.97**	12.02	0.75	-1.54	**-0.37**
Mean of c	1.68	2.49	0.08	**-0.04**	4.53	1.88	0.09	**-0.05**	2.17	0.15	0.02	**-0.03**
SD of c	25.21	29.75	-11.16	**-1.88**	26.17	12.80	-5.35	**-0.83**	11.12	0.71	-2.01	**-0.30**
Mean of a	-38.95	-29.41	20.38	**0.97**	37.58	13.14	9.52	**0.13**	21.59	2.05	2.46	**0.29**
SD of a	-16.92	-20.88	2.29	**-3.35**	42.94	11.16	1.58	**-1.46**	26.32	**0.60**	1.02	-0.67
Top 5% wealth share	30.38	9.04	-17.91	**-4.53**	5.55	**-0.83**	-8.34	-1.71	4.98	-1.06	-1.20	**-0.93**
						Continuous simulation						
Mean of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean of c	-2.63	-1.89	1.02	**-0.06**	0.18	0.18	0.43	**-0.05**	0.05	0.05	0.09	**-0.01**
SD of c	-5.46	-4.92	6.08	**0.00**	0.32	0.39	2.75	**-0.15**	0.17	0.22	0.94	**-0.05**
Mean of a	-72.37	-52.17	28.02	**-1.72**	5.09	4.83	11.75	**-1.39**	1.48	1.50	2.54	**-0.21**
SD of a	-42.69	-38.33	31.19	**0.44**	1.42	2.26	16.26	**-0.53**	0.85	1.33	6.71	**-0.31**
Top 5% wealth share	72.61	14.93	4.90	**2.29**	-3.36	-2.39	5.72	**0.77**	-0.53	-0.08	4.44	**-0.03**

Note: we report in bold the lowest deviation, for each moment and number of grid points.

4.1.1 Comparison with two classic 'ad-hoc' approaches

In what follows we implement two of the well-known ad-hoc approaches that we discussed in Section 2.4, and we assess their accuracy under the same type of income process as in the previous subsection. The methods we consider are relatively easy to implement and have well-documented steps that allow us to reproduce them accurately.

Huggett (1996) Table 3 reports percentage deviations from benchmark moments obtained using the ad-hoc approximation in Huggett (1996). We compare this approximation to that obtained with the three methods considered so far. For this comparison exercise we set ρ to 0.95.[26] All parameter values are as in Section 4.1 with the exception

[26]This choice of ρ makes comparison to the previous tables easy and is almost identical to the value of $\rho = 0.96$ used in the original paper by Huggett. Additional results are available from the authors.

Table 4: Percentage deviations from benchmark moments: comparison of STY method and the three methods, $\rho = 1.00$.

	$N = 81$	$N = 25$			$N = 50$		
	STY (%)	Tau (%)	AC (%)	Rou (%)	Tau (%)	AC (%)	Rou (%)
	Markov-chain simulation						
Mean of y	-0.12	0.76	-0.48	-0.15	-0.17	-0.26	**-0.08**
SD of y	-1.17	1.74	-6.48	-1.44	-2.70	-3.89	**-0.61**
Mean of c	-0.12	0.98	**-0.07**	-0.13	-0.10	-0.07	-0.08
SD of c	-1.08	2.30	-5.10	-1.28	-2.34	-3.05	**-0.53**
Mean of a	-0.18	9.42	15.61	0.25	2.37	7.10	**0.10**
SD of a	-0.31	4.84	17.43	-0.11	1.46	9.67	**-0.08**
Top 5% wealth share	-0.17	-2.05	8.31	-0.19	**-0.07**	5.86	-0.12

Note: we report in bold the lowest deviation, for each moment and number of grid points.

of the variance of η_0 which, in line with Huggett's parameterization, we set to 0.7 times the asymptotic variance rather than zero.[27] It is apparent that, under the Markov-chain simulation, Huggett's method does worse than the other three. However, when simulating using the continuous process, Huggett's method does sometimes better than Tauchen. For all simulation approaches (Markov-chain or continuous) and all grid sizes, the Rouwenhorst method delivers a more accurate approximation of the benchmark moments. This is not surprising given that Huggett's method posits an age-invariant state-space and transition probabilities using the same approximation as in Tauchen (1986). Since Rouwenhorst does better than our non-stationary (age-varying) adaptation of Tauchen's method, a fortiori one would expect it to outperform this age-invariant approach.

Storesletten et al. (2004*b*) Table 4 reports approximation results obtained implementing the ad-hoc approach of Storesletten et al. (2004*b*) (henceforth STY) in our canonical model. By construction this approach is only applicable to unit-root processes: specifically, it is designed to approximate a random walk using binomial innovation steps (that is, only two possible shocks are allowed in any period, each with probability 0.5). Given initial conditions, the shocks generate a fanning out over the life-cycle and result in a sequence of overlapping Markov chains which expand incrementally over an age-

[27] In line with intuition, the performance of Huggett's discretization worsens as the difference between the variance of the initial condition and the asymptotic one becomes larger.

invariant state-space. Given a 40-year lifetime, the state-space has 81 grid points, which is significantly more than what we use in our baseline implementations of Tauchen, Adda and Cooper, and Rouwenhorst. As shown in Table 4, the Rouwenhorst discretization with 25 grid points performs just as well as the STY approach does with more than three times as many grid nodes. For comparison, we also show that if one were to use Rouwenhorst with 50 grid points (still well below the 81 points of STY) the approximation performance would become significantly better.[28] As mentioned above, unlike the other three methods, the binomial-tree approximation is not easily generalized to AR(1) processes with root less than unity.

4.2 Time-dependent conditional second moments

A recent literature (see Karahan and Ozkan, 2013; Guvenen et al., 2016; De Nardi et al., 2018) has pointed out that both the persistence and conditional second moments of earnings may not be constant over the life cycle, contrary to what the canonical earnings process of the previous section assumes.

As discussed in Section 2, our non-stationary discretization methods can accommodate processes with age-dependent conditional second moments ρ_t and σ_t. In this section, we gauge the accuracy of the three discretization methods in approximating age-dependent processes of this kind. In particular, we parameterize the income process following Karahan and Ozkan (2013).[29] Namely, we set the variance of the transitory income component to $\sigma_u^2 = 0.0564$, the age-dependent persistence to $\rho_t = 0.7596 + 0.2039(t/10) - 0.0535(t/10)^2 + 0.0028(t/10)^3$ and, finally, the variance of innovations in the persistent component to $\sigma_{\varepsilon t}^2 = 0.0518 - 0.0405(t/10) + 0.0105(t/10)^2 - 0.0002(t/10)^3$.

Table 5 reports the results of this exercise. First, we note that the quality of the approximation using the Rouwenhorst method is just as accurate as in the age-independent parameterization case. The Adda and Cooper method delivers a slightly improved approximation for most grid sizes, although it remains far less accurate than Rouwen-

[28] In fairness, while the suggested alternatives with 50 grid points run faster than our straightforward implementation of STY with 81 points, the latter could be speeded up by exploiting the fact that, at each age t, only t states are reached with positive probability. This would reduce the number of evaluations to approximately the same as the other methods with 25 grid points, although at the price of some extra coding effort.

[29] The only difference is that we set the distribution of η_0 to be degenerate at zero.

Table 5: Percentage deviations from benchmark moments: AR(1) with age-dependent conditional second moments.

	$N = 5$			$N = 10$			$N = 25$		
	Tau (%)	AC (%)	Rou (%)	Tau (%)	AC (%)	Rou (%)	Tau (%)	AC (%)	Rou (%)
					Markov-chain simulation				
Mean of y	4.50	-1.05	**-0.11**	1.85	-0.53	**-0.07**	0.12	-0.28	**-0.04**
SD of y	33.04	-11.80	**-2.81**	13.79	-6.47	**-1.30**	0.72	-3.19	**-0.47**
Mean of c	3.28	-0.35	**-0.08**	2.28	-0.26	**-0.05**	0.19	-0.29	**-0.04**
SD of c	29.35	-13.17	**-2.56**	13.71	-7.08	**-1.17**	0.68	-3.47	**-0.42**
Mean of a	-26.03	16.31	**0.66**	12.55	6.45	**0.38**	1.80	-0.49	**-0.01**
SD of a	-28.98	**1.73**	-3.81	10.15	**-0.38**	-1.79	0.73	-1.60	**-0.71**
Top 5% wealth share	-5.09	-16.05	**-5.08**	**-1.33**	-7.68	-2.31	-0.84	-0.96	**-0.79**
					Continuous simulation				
Mean of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean of c	-1.86	1.12	**-0.10**	0.17	0.50	**-0.06**	0.05	0.13	**-0.03**
SD of c	-4.94	6.85	**0.00**	0.31	3.25	**-0.17**	0.24	1.21	**-0.08**
Mean of a	-46.65	28.16	**-2.57**	4.37	12.59	**-1.47**	1.26	3.22	**-0.64**
SD of a	-42.20	38.24	**0.88**	1.71	20.45	**-0.74**	1.74	8.85	**-0.34**
Top 5% wealth share	**1.09**	10.42	3.56	-2.55	9.08	**0.68**	0.43	6.03	**0.28**

Note: we report in bold the lowest deviation, for each moment and number of grid points.

horst. Finally, the quality of results obtained using the Tauchen's method is generally unchanged. Hence, the key findings of our baseline numerical analysis are confirmed: Rouwenhorst's method performs significantly better for all grid sizes and moments, exhibiting approximation errors that are often an order of magnitude smaller than the alternatives. The largest discrepancies in the quality of approximations occur, again, when looking at the distribution of assets.

4.3 Non-normal innovations

The seminal contributions of Guvenen et al. (2014) and Guvenen et al. (2016) document how the distribution of earnings growth rates displays significant deviations from normality. Given this observation, processes with non-normal innovations have become more common in numerical implementations of heterogeneous agents' models.[30] Non-normality is usually introduced parametrically by assuming stochastic processes that are mixtures of normal processes.

In this section, we examine how the different discretization methods perform when innovations to the persistent component of earnings (η in problem 19) are a mixture of two normals. Namely, we consider the following non-normal shocks:

$$\varepsilon_t \sim \begin{cases} N(\mu^1, \sigma^1) & \text{with probability } p^1, \\ N(\mu^2, \sigma^2) & \text{with probability } p^2. \end{cases} \tag{20}$$

While our parameterization of the ε innovations' process assumes the exact same persistence and variances as in Section 4.1, it also exhibits negative skewness and high kurtosis. The values of skewness and kurtosis are, respectively, -1.36 and 17.95 and are taken from Civale et al. (2017), who target the corresponding moments in Guvenen et al. (2016).[31] Civale et al. (2017) also provide closed-form expressions to calibrate a mixture of two normals so as to match the first four moments of a non-normal distribution. Applying their formulas, and imposing that ε has mean zero, we obtain the following values for the mixing probabilities p^i, the means μ^i and the standard deviations σ^i ($i = 1, 2$) of the two

[30] See, for example, McKay (2017), Kaplan et al. (2018), De Nardi et al. (2018).

[31] Civale et al. (2017) propose a method to discretize a *stationary* AR(1) process with mixture-of-normals innovations. Their method is based on Tauchen but differs from it because the state space of the Markov chain is calibrated to minimize the distance between a set of targeted moments of the undiscretized process and their counterparts computed using the Markov chain.

normals: $p^1 = .9$, $p^2 = .1$, $\mu^1 = .0089$, $\mu^2 = -\mu^1 p^1/p^2$, $\sigma^1 = .0635$, $\sigma^2 = .3430$. Note that, as in Section 4.1, the persistent component is non-stationary because the initial condition $\eta_0 = 0$ is not a random draw from the limiting distribution of η.

Adapting Tauchen to allow for Gaussian-mixture innovations is straightforward. The only change is that the transition probabilities satisfy

$$\pi_t^{ij} = \begin{cases} F_\varepsilon\left(\bar{\eta}_t^j - \rho\bar{\eta}_{t-1}^i + h_t/2\right) & \text{if } j = 1, \\ 1 - F_\varepsilon\left(\bar{\eta}_t^j - \rho\bar{\eta}_{t-1}^i - h_t/2\right) & \text{if } j = N, \\ F_\varepsilon\left(\bar{\eta}_t^j - \rho\bar{\eta}_{t-1}^i + h_t/2\right) - F_\varepsilon\left(\bar{\eta}_t^j - \rho\bar{\eta}_{t-1}^i - h_t/2\right) & \text{otherwise.} \end{cases} \tag{21}$$

where F_ε is the cumulative density of the Gaussian mixture (20).

While adapting Adda and Cooper is also conceptually straightforward, its implementation is more involved. The cutoff points $\{[x_t^i, x_t^{i+1}]\}_{i=1}^N$ of the N intervals in period t must now satisfy

$$\int_{-\infty}^{x_t^i} f_{\eta_t}(u)du = \frac{i-1}{N}, \qquad i = 1, ..., N+1, \tag{22}$$

where f_{η_t} is the (marginal or, equivalently, unconditional) density of η_t, which in turn satisfies the recursive formula

$$f_{\eta_t}(u) = \int_{-\infty}^{\infty} f_\varepsilon(u - \rho v) f_{\eta_{t-1}}(v)dv \tag{23}$$

with initial condition $f_{\eta_1}(\cdot) = f_\varepsilon(\cdot)$. That is the probability of η_t to be in a neighborhood of u is the probability of transiting from $\eta_{t-1} = v$ to $\eta_t = u$ integrating over all possible value of η_{t-1}. Then, the transition probability $\pi_t^{i,j}$ is defined as the probability of η moving from the interval $[x_t^i, x_t^{i+1}]$ to the interval $[x_{t+1}^j, x_{t+1}^{j+1}]$, between t and $t+1$.

Note that, contrary to the Normal case, both the transition probabilities *and* the density used to compute the cutoff points in equation (22) are not available in closed-form and have to be computed numerically. We do this by using Montecarlo integration, drawing a panel of 2,000,000 histories.

Lastly, Rouwenhorst's method requires no adaptation because, by construction, it matches only moments up to the second order and cannot target higher order moments. Nonetheless we still include it in our evaluation for the purpose of comparison.

Tables 6 and 7 report the percentage deviations from the benchmark moments under, respectively, the Markov-chain and continuous simulations. Given the non-normality of the earnings process we also report measures of skewness and kurtosis for each variable. In order to reduce the impact of sampling error, in line with the literature (see, e.g., Guvenen et al., 2016), we report measures—Kelly skewness and Crow-Siddiqui kurtosis—that are robust to extremes. More precisely, Kelly's measure of skewness is defined as

$$\mathcal{S}_{\mathcal{K}} = \frac{(P90 - P50) - (P50 - P10)}{P90 - P10}$$

and takes values between -1 and +1, while Crow-Siddiqui kurtosis is

$$\kappa_{\mathcal{C}-\mathcal{S}} = \frac{P97.5 - P2.5}{P75 - P25}$$

and equals 2.91 for the normal distribution.

In the Markov-chain simulations, Rouwenhorst tends to deliver the most accurate approximation of the first two moments. This is not surprising in the case of income, since it targets them explicitly. Similarly, Tauchen and Adda and Cooper yield a better approximation of income's higher moments, as they do exploit the distributional assumptions.[32] The same pattern applies to the consumption moments. Looking at the other moments, Rouwenhorst tends to perform better than the alternatives at lower grid sizes, especially when approximating the asset distribution. In fact, it provides a better fit of the higher asset moments even with just 5 grid points.

In the continuous simulation case, where the Markov-chain approximation is not used to simulate income histories, the performance of Rouwenhorst's discretization is even better. In fact, it tends to outperform the other two methods for most moments. This suggests that its advantage lies partly in placing grid points in a way that reduces interpolation errors.

One caveat is in order: all discretization methods appear to provide a relatively poor approximation of the higher-order moments of consumption and assets when $\rho = 1$ and grid size is small. This problem is especially severe in the case of discrete (Markov-chain) simulations. In such cases, even the best performing approach (Rouwenhorst) can

[32]Note that, by construction, the Rouwenhorst Markov chain implies a symmetric (i.e., zero-skewness) process up to sampling error. This accounts for an approximation error equal, or close, to -100% (zero as opposed to the true negative value) for income skewness.

Table 6: Percentage deviations from benchmark moments: AR(1) with non-normal innovations, Markov-chain simulation.

	$N=5$			$N=10$			$N=25$		
	Tau (%)	AC (%)	Rou (%)	Tau (%)	AC (%)	Rou (%)	Tau (%)	AC (%)	Rou (%)
					$\rho=0.95$				
Mean of y	-4.36	-0.98	**0.13**	-0.92	-0.51	**0.15**	0.17	-0.25	**0.16**
SD of y	52.77	-12.54	**0.78**	14.91	-7.43	**2.03**	**-2.60**	-3.87	2.66
Skewness of y	212.23	**8.04**	-100.00	**51.00**	-84.13	-220.48	-35.73	**28.04**	-100.00
Kurtosis of y	**9.16**	-22.54	-35.76	**3.41**	-12.56	16.04	**-0.61**	16.35	-18.61
Mean of c	-2.61	-0.48	**0.29**	-0.35	**-0.22**	0.30	0.21	**-0.10**	0.31
SD of c	52.72	-14.46	**1.02**	16.56	-8.14	**2.24**	**-2.53**	-4.02	2.86
Skewness of c	37.41	**-21.89**	35.34	28.14	**2.82**	11.93	5.97	**2.83**	13.28
Kurtosis of c	**13.65**	-20.11	-20.53	**-1.07**	-7.03	-3.35	-2.53	**-1.14**	-8.78
Mean of a	57.78	16.83	**5.73**	19.32	9.69	**5.42**	**1.53**	5.09	5.31
SD of a	-26.11	-6.23	**-1.63**	-6.34	-5.57	**0.88**	-5.49	-3.85	**2.04**
Skewness of a	-89.32	**-3.55**	5.97	-24.66	**-2.81**	5.51	**-2.33**	-4.38	5.28
Kurtosis of a	-28.01	-24.48	**-8.50**	-12.87	-6.58	**-6.13**	-4.04	**1.06**	-4.77
Top 5% wealth share	-49.37	-24.63	**-7.64**	-19.88	-14.54	**-4.77**	-6.02	-6.46	**-3.41**
					$\rho=0.98$				
Mean of y	-2.83	-1.55	**0.21**	-1.89	-0.81	**0.26**	0.33	-0.39	**0.30**
SD of y	54.28	-15.55	**-0.79**	22.25	-9.60	**1.44**	**-1.40**	-5.25	2.69
Skewness of y	**-13.14**	24.13	-100.00	**33.24**	-71.17	-203.91	-86.80	**8.28**	-100.00
Kurtosis of y	**14.34**	-22.08	-26.85	0.18	**0.05**	27.91	-4.66	4.25	**0.39**
Mean of c	-0.74	-0.43	**0.31**	-0.44	**-0.12**	0.37	0.48	**-0.04**	0.40
SD of c	48.50	-15.39	**-0.63**	20.04	-9.33	**1.58**	**-1.20**	-5.06	2.82
Skewness of c	43.17	**-2.73**	25.29	25.94	**6.60**	26.62	7.11	**3.62**	16.27
Kurtosis of c	24.52	-16.07	**-15.32**	-6.37	-5.93	**4.30**	-1.08	**-1.06**	-5.02
Mean of a	79.75	42.66	**4.34**	55.35	26.37	**4.47**	6.53	13.45	**4.50**
SD of a	-32.87	21.65	**-1.80**	-31.69	11.81	**0.98**	-2.88	5.16	**2.34**
Skewness of a	-138.25	**10.00**	10.35	-109.63	**4.54**	11.15	**-3.19**	-4.41	10.16
Kurtosis of a	-45.06	-18.51	**-4.42**	-35.12	8.52	**-3.74**	-4.68	14.57	**-2.83**
Top 5% wealth share	-66.17	-19.27	**-5.36**	-55.39	-8.93	**-2.76**	-7.08	-1.60	**-1.52**
					$\rho=1.00$				
Mean of y	0.42	-2.53	**0.30**	-1.68	-1.36	**0.50**	**0.32**	-0.67	0.53
SD of y	59.99	-22.34	**-4.71**	24.01	-14.96	**-0.32**	**-1.54**	-8.96	1.99
Skewness of y	95.85	**49.84**	-87.93	58.60	**-41.44**	-191.01	-60.76	**9.87**	-91.19
Kurtosis of y	**19.30**	-20.22	-22.27	**5.22**	6.59	16.89	-4.84	**0.80**	-2.66
Mean of c	2.39	-0.85	**0.43**	**0.01**	-0.29	0.61	0.69	**-0.10**	0.62
SD of c	54.89	-18.59	**-4.17**	20.75	-12.19	**-0.01**	**-1.05**	-7.29	2.16
Skewness of c	53.47	**1.24**	18.40	24.65	**7.39**	38.46	6.69	**2.55**	3.78
Kurtosis of c	10.06	**-9.62**	-12.54	**-0.34**	-3.95	16.10	3.57	**0.75**	-4.21
Mean of a	90.75	74.76	**6.64**	76.00	47.97	**5.60**	17.36	25.19	**4.98**
SD of a	95.59	190.83	**12.44**	58.54	132.06	**9.21**	16.34	80.17	**7.18**
Skewness of a	-510.78	1059.18	**394.40**	-484.98	636.37	**232.31**	**-19.55**	-96.30	175.85
Kurtosis of a	-48.85	**7.85**	8.54	-37.59	50.69	**5.57**	-7.85	54.11	**4.78**
Top 5% wealth share	-27.14	50.83	**5.55**	-29.08	61.42	**4.12**	-3.31	54.81	**2.77**

Note: we report in bold the lowest deviation, for each moment and number of grid points.

Table 7: Percentage deviations from benchmark moments: AR(1) with non-normal innovations, continuous simulation.

	$N=5$			$N=10$			$N=25$		
	Tau (%)	AC (%)	Rou (%)	Tau (%)	AC (%)	Rou (%)	Tau (%)	AC (%)	Rou (%)
					$\rho=0.95$				
Mean of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Skewness of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Kurtosis of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean of c	1.06	0.84	**-0.04**	0.42	0.44	**-0.01**	0.08	0.21	**0.03**
SD of c	-0.87	5.86	**0.29**	-0.56	3.01	**0.00**	0.36	1.29	**0.03**
Skewness of c	15.51	13.73	**1.36**	17.65	7.10	**0.61**	2.88	5.08	**-0.98**
Kurtosis of c	-17.78	2.35	**-0.28**	-6.58	-1.21	**-0.17**	**-0.03**	-0.75	-0.09
Mean of a	37.54	29.90	**-1.44**	14.93	15.64	**-0.23**	2.91	7.46	**0.90**
SD of a	-46.92	38.16	**2.94**	-13.58	22.01	**0.33**	2.85	10.13	**0.09**
Skewness of a	-119.12	2.50	**1.17**	-35.56	-2.95	**0.66**	-1.57	-4.40	**-0.22**
Kurtosis of a	-34.39	17.95	**3.68**	-17.27	14.79	**0.39**	-0.49	4.03	**-0.28**
Top 5% wealth share	-60.45	7.31	**3.64**	-24.33	6.34	**0.30**	**-0.39**	2.21	-0.73
					$\rho=0.98$				
Mean of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Skewness of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Kurtosis of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean of c	1.02	1.52	**-0.14**	1.25	0.88	**-0.06**	0.18	0.43	**0.01**
SD of c	-2.70	10.03	**0.77**	-1.33	5.91	**0.00**	0.65	2.91	**-0.01**
Skewness of c	13.25	10.92	**-0.80**	15.44	5.55	**-0.93**	2.91	3.56	**0.46**
Kurtosis of c	-15.45	5.66	**-0.23**	-8.92	**-0.07**	-0.14	0.09	-0.95	**0.02**
Mean of a	40.30	60.11	**-5.65**	49.22	34.75	**-2.55**	7.18	16.90	**0.52**
SD of a	-43.26	100.25	**13.41**	-33.73	65.60	**1.61**	8.52	35.87	**0.13**
Skewness of a	-146.22	17.60	**7.49**	-123.74	**0.51**	3.62	-4.05	-5.88	**0.58**
Kurtosis of a	-45.63	34.30	**8.23**	-39.30	35.71	**0.40**	-2.34	14.62	**-0.43**
Top 5% wealth share	-64.61	23.79	**13.21**	-57.43	20.60	**2.32**	-1.10	11.05	**-0.39**
					$\rho=1.00$				
Mean of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Skewness of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Kurtosis of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean of c	0.85	2.18	**-0.25**	1.48	1.33	**-0.15**	0.36	0.67	**-0.01**
SD of c	**-0.66**	17.78	3.28	-0.67	11.91	**0.42**	1.32	6.85	**-0.04**
Skewness of c	8.83	8.38	**-1.21**	11.89	4.60	**-2.15**	3.34	2.61	**0.38**
Kurtosis of c	-12.61	8.07	**-0.76**	-7.72	1.42	**-0.09**	**0.01**	-0.59	0.09
Mean of a	39.12	100.18	**-11.44**	68.25	61.00	**-6.87**	16.73	30.83	**-0.42**
SD of a	**64.95**	490.95	137.61	53.36	361.55	**28.41**	56.66	236.93	**0.83**
Skewness of a	220.01	1108.27	**-46.60**	-323.92	429.33	**5.20**	-51.04	-148.81	**-0.09**
Kurtosis of a	-35.01	87.15	**12.36**	-35.47	89.04	**-0.45**	-8.45	43.41	**-1.22**
Top 5% wealth share	**-12.50**	158.32	71.73	-28.66	137.87	**9.12**	6.75	90.00	**0.56**

Note: we report in bold the lowest deviation, for each moment and number of grid points.

deliver very imprecise approximations of wealth, and to a smaller extent, consumption higher moments. The poor approximation of higher-order moments is an issue with all of these discretization methods and its severity increases with the persistence of the income process.

5 Conclusion

Non-stationary persistent income processes are commonplace in quantitative studies of life-cycle behavior and inequality. Introducing such processes into numerical models requires approximating their distribution by means of a finite-state approximation, usually a Markov chain. The quality of the approximation is important for the accuracy of implied evolution of the cross-sectional distribution of consumption, income and wealth across age groups. Large approximation errors may result in misleading inference.

This paper derives new generalizations of Tauchen (1986), Adda and Cooper (2003) and Rouwenhorst's (1995) discretization methods to the case of non-stationary processes, like the ones commonly employed in life-cycle economies. It also provides a systematic examination of the performance of these methods when used in a life-cycle income fluctuation problem under a variety of assumptions for the income process.

Throughout our analysis we find large differences in the quality of the approximations relative to a very accurate benchmark. In many cases we find that deviations from benchmark moments are large. Adda and Cooper and Tauchen are generally much less precise than Rouwenhorst. This discrepancy is most severe when considering wealth moments, for which the relative size of approximation errors can differ by an order of magnitude. In many cases we find that Rouwenhorst delivers with just 5 grid points more accurate moments than the other methods do with 10 or 25 grid points.

We also find, that when shocks are non-normally distributed, all methods do a pretty poor job in approximating higher order moments of consumption and wealth, although Rouwenhorst still tends to do somewhat better. This problem is worse the higher the persistence of the income process. Even adding grid points does not always improve the quality of the approximations by much. Arguably, adding more grid points results in consistent and significant improvements in the approximation of higher-order moments of consumption and wealth only when using the Rouwenhorst method. This is surprising insofar Rouwenhorst does not directly target the higher order moments of income, whereas both Tauchen and Adda and Cooper incorporate the distributional assumptions by construction.

A Appendix

A.1 The benchmark solution

Let $c_t(z_t, \eta_t)$ denote the consumption policy function with state variables current cash at hand z_t and the AR(1) component of (log) labor income η_t. Define two time-invariant, exogenous, grids $G_a = \{a^i\}_{i=1}^m$ and $G_\eta = \{\eta^j\}_{j=1}^n$, respectively for assets a and for η,[33] where the lower bound on the asset grid $a^1 = 0$, the value of the borrowing constraint.

For any $t < T$, the optimal saving policy when borrowing is unconstrained satisfies the Euler equation

$$(z_t - a_t)^{-1} = \frac{\partial \tilde{V}_{t+1}(a_t, \eta_t)}{\partial a_t} \tag{24}$$

where, from the envelope theorem, the marginal continuation utility satisfies

$$\frac{\partial \tilde{V}_{t+1}(a_t, \eta_t)}{\partial a_t} = \beta(1+r)\mathbb{E}_t[c_{t+1}((1+r)a_t + e^{(\rho\eta_t + \varepsilon_{t+1} + u_{t+1})}, \rho\eta_t + \varepsilon_{t+1})]^{-1}. \tag{25}$$

Suppose for the moment that the function $\partial \tilde{V}_{t+1}(a_t, \eta_t)/\partial a_t$ is known. For each pair $(a_t, \eta_t) \in G_a \times G_\eta$, we use Carroll's (2006) endogenous gridpoint methods to solve the Euler equations *backward* for the "endogenous" value of cash at hand

$$z_t^i(\eta^j) = a^i + \left(\frac{\partial \tilde{V}_{t+1}(a^i, \eta^j)}{\partial a_t}\right)^{-1} \tag{26}$$

which satisfies (24); namely the value of cash at hand for which a^i is the optimal saving choice given $\eta_t = \eta_j$. From the dynamic budget identity, the value of the consumption function at the gridpoints $(z_t^i(\eta^j), \eta^j),\ i = 1, \ldots, m,\ j = 1, \ldots, n$ is given by

$$c_t(z_t^i(\eta^j), \eta^j) = z_t^i(\eta^j) - a^i. \tag{27}$$

The finite mapping thus obtained can be used to construct an interpolating approximation to the (continuous) consumption function.

Note that, by construction, the endogenous grid method returns the exact lower bound on cash at hand $z_t^1(\eta_j)$ below which the borrowing constraint $a_t \geq a^1 = 0$ for $\eta_t = \eta^j$.

[33] We use 1,000 grid Points for a and 10,000 grid points for η.

Therefore, for all z_t below such bound the consumption function has the linear form $c_t = z_t$.

Given the consumption function at time $t + 1$, the conditional expectation $\partial\tilde{V}_{t+1}(a_t, \eta_t)/\partial a_t$ is computed using equation (25) and involves integration over the innovations ε_{t+1} and u_{t+1}. We approximate their distributions using Gaussian Hermite nodes and weights. Finally, evaluating the consumption function on the right hand side of (25) in between grid points requires interpolating over both (z_t, η_t), for which we use bi-linear interpolation.

The system of difference equations (25)-(27) is solved backward starting from the last period $t = T$ in which the consumption function satisfies

$$c_T(z_T, \eta_T) = z_T.$$

A.2 Normalized problem with unit-root labor income

In the case in which the (log) income process has a unit root and the felicity function has the CRRA form, it is well known from Carroll (2004) that it is possible to normalize problem (19) by the permanent component $\exp(\eta_t)$ of labor income $y_t = \exp(\eta_t + u_t)$, thereby reducing the effective state space to a rescaling of z_t.

To see this, replace for $c_t = z_t - a_t$ in (19) and consider the problem in the second-to-last period

$$\mathbb{V}_{T-1}(z_{T-1}, \eta_{T-1}) = \max_{a_{T-1}} \log(z_{T-1} - a_{T-1}) + \beta \mathbb{E}_{T-1} \log(z_T) \tag{28}$$

If one defines the state variables $\hat{z}_t = z_t / \exp(\eta_t)$ and $\hat{a}_t = a_t / \exp(\eta_t)$, equation (28) can be rewritten as

$$\begin{aligned} \mathbb{V}_{T-1}(z_{T-1}, \eta_{T-1}) &= \max_{\hat{a}_{T-1}} \log(\exp(\eta_{T-1})(\hat{z}_{T-1} - \hat{a}_{T-1})) + \beta \mathbb{E}_{T-1} \log(\exp(\eta_T)\hat{z}_T) \\ &= (1+\beta)\eta_{T-1} + \left\{ \max_{\hat{a}_{T-1}} \log(\hat{z}_{T-1} - \hat{a}_{T-1}) + \beta \mathbb{E}_{T-1} \log(\hat{z}_T) \right\}, \end{aligned} \tag{29}$$

where the second line follows from $\mathbb{E}_{T-1} \log(\exp(\eta_T)) = \eta_{T-1}$.

Note that by definition

$$\hat{z}_t = (1+r)\frac{a_{t-1}}{\exp(\eta_{t-1}+\varepsilon_t)} + \exp(u_t) = (1+r)\frac{\hat{a}_{t-1}}{\exp(\varepsilon_t)} + \exp(u_t), \tag{30}$$

which implies that the curly bracket in (29) is equal to $V_{T-1}(\hat{z}_{T-1})$ where the latter satisfies the Bellman equation

$$V_{T-1}(\hat{z}_{T-1}) = \max_{\hat{a}_{T-1}} \log(\hat{z}_{T-1} - \hat{a}_{T-1}) + \beta \mathbb{E}_{T-1} V_T(\hat{z}_T) \tag{31}$$

with $V_T(\hat{z}_T) = \log(\hat{z}_T)$.

Equations (29) and (31) imply that $\mathbb{V}_{T-1}(z_{T-1}, \eta_{T-1}) = (1+\beta)\eta_{T-1} + V(\hat{z}_{T-1})$. The same logic implies that this holds also for any $t < T-1$.

Therefore the Bellman equation for the problem in normalized form satisfies

$$V_t(\hat{z}_t) = \max_{\hat{a}_t} \log(\hat{z}_t - \hat{a}_t) + \beta \mathbb{E}_t V_{t+1}(\hat{z}_{t+1}), \tag{32}$$

for all t. It follows from (30) and the envelope condition that the associated Euler equation is

$$\frac{1}{\hat{c}_t} = \beta(1+r)\mathbb{E}_t \frac{1}{\hat{c}_{t+1}}. \tag{33}$$

The advantage of the normalized problem (32) is that one can first solve, very accurately (see Barillas and Fernández-Villaverde, 2007), for the saving function $\hat{a}_t(\hat{z}_t)$ using the endogenous gridpoint method with only one-dimensional interpolation with respect to $\hat{z}_t$. Under the assumption that ε_t and u_t are i.i.d. and normally distributed, the expectation in equation (33) can be computed using the, again, very accurate Gaussian Hermite quadrature.

Recovering the non-normalized policy function $a_t(z_t, \eta_t) = \hat{a}_t(\hat{z}_t)\exp(\eta_t)$ then does not require interpolation with respect to η_t.

In Table 8 we compare the results from this alternative accurate solution of the random-walk special case and the results from our baseline solution. As one can see, they are effectively identical.

Table 8: Percentage deviations of moments of the benchmark solution relative to moments obtained for quasi-exact solution of normalized problem (random walk case).

Moment	Deviation (%)
Mean of y	0.0000
SD of y	0.0000
Mean of c	-0.0001
SD of c	-0.0001
Mean of a	-0.0031
SD of a	-0.0032
Top 5% wealth share	-0.0001

A.3 Robustness: calibration of Ω in Tauchen method

In what follows we examine the effect of using alternative values for Ω on the performance of Tauchen's method. In his original paper, Tauchen (1986) simply sets $\Omega = 3$. Kopecky and Suen (2010) argue that the choice of Ω affects the approximation performance of this method. For this reason, they calibrate Ω so that the standard deviation of the Markov chain is equal to the standard deviation of the (stationary) AR(1) process.

Our results in the main text are based on the original Tauchen parameterization (that is, we set $\Omega = 3$). To check whether a different value of Ω would improve the performance of Tauchen's method significantly, we also experiment with a more flexible parameterization. Namely, in our non-stationary setting, we allow Ω to vary with age t and we calibrate it to match the unconditional standard deviation of the AR(1) process σ_t at each age.[34] We then compare the performance of the baseline Tauchen approximation with that of the Tauchen's method with age-dependent Ω's. This exercise is carried out in the context of our canonical AR(1) process, described in Section 4.1, with $\rho = 1$. Results are in Table 9, where we also show the performance of Rouwenhorst's method for comparison. As before, Table 9 displays the percentage deviations of the approximated moments from the benchmark counterparts.

[34] The calibrated values of Ω_t vary with both age t and the number of grid points N (as well as with the parameters of the AR(1) process). Specifically, the calibrated Ω_t decreases with age and increases with N. For $N = 5$, Ω_t ranges from 1.669 at $t = 40$ to 1.934 at $t = 1$; in the cases of $N = 10$ and $N = 25$, Ω_t takes values between $[1.946, 2.439]$ and $[2.430, 2.983]$, respectively. While, prima facie, it may be surprising that Ω_t is decreasing with age, the explanation for this is simple: as shown in equation 5, the highest and lowest grid points are defined as the product of Ω_t and σ_t, with σ_t growing significantly with age as described in equation 2. For this reason the range of variation of earnings increases with age even if Ω_t gets smaller. All Ω_t values are available upon request.

When $N = 5$, the calibrated Tauchen method performs better than the baseline Tauchen; however, these performance gains are reverted when $N = 10$ and $N = 25$: in these cases Tauchen's method with calibrated Ω does a much poorer job than its counterpart with $\Omega = 3$ for most moments. Notably, neither of the two Tauchen implementations performs better than Rouwenhorst's method, for all grid sizes. We thus conclude that, while the choice of Ω does have an effect on the performance of Tauchen's method when using few grid points, even a carefully calibrated choice of this parameter does not guarantee a performance boost sufficient to make the approximation quality comparable to that obtained using the Rouwenhorst's method.

Table 9: Robustness check: baseline Tauchen ($\Omega = 3$) vs Tauchen with calibrated Ω. Rouwenhorst approximation reported for comparison.

	$N = 5$			$N = 10$			$N = 25$		
	(1) Tau (%)	(2) Tau* (%)	(3) Rou (%)	(4) Tau (%)	(5) Tau* (%)	(6) Rou (%)	(7) Tau (%)	(8) Tau* (%)	(9) Rou (%)
					Markov-chain simulation				
Mean of y	8.82	-0.62	**-0.38**	5.48	-0.51	**-0.18**	0.76	-0.30	**-0.15**
SD of y	40.02	-12.46	**-7.92**	23.35	-10.29	**-3.67**	1.74	-5.96	**-1.44**
Mean of c	8.21	-0.80	**-0.32**	5.68	0.15	**-0.15**	0.98	-0.09	**-0.13**
SD of c	37.00	-12.66	**-7.29**	22.43	-8.39	**-3.33**	2.30	-5.10	**-1.28**
Mean of a	-14.86	-7.48	**1.72**	13.12	24.89	**0.83**	9.42	7.69	**0.25**
SD of a	-26.26	-16.29	**-0.41**	-2.92	17.24	**-0.16**	4.84	5.74	**-0.11**
Top 5% wealth share	-19.61	-11.14	**-1.11**	-13.26	**-0.18**	-0.43	-2.05	0.80	**-0.19**
					Continuous simulation				
Mean of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD of y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean of c	-1.00	**-0.25**	-0.38	**0.06**	0.71	-0.25	0.21	0.24	**-0.10**
SD of c	**-0.95**	2.34	2.03	-0.66	5.60	**-0.18**	1.01	2.55	**-0.18**
Mean of a	-38.61	**-9.72**	-14.57	**2.15**	27.52	-9.72	8.20	9.21	**-3.82**
SD of a	-39.32	**16.54**	21.34	-8.26	56.99	**-0.80**	8.82	24.61	**-1.16**
Top 5% wealth share	**-9.26**	16.82	29.11	-12.61	20.89	**6.58**	**-0.16**	9.10	1.96

Notes: (1) We report in bold the lowest deviation, for each moment and number of grid points.
(2) *Tau* (columns 1, 4 and 7) denotes results from the baseline Tauchen method ($\Omega = 3$ and constant over time); *Tau** (columns 2, 5 and 8) denotes the Tauchen method with calibrated, age-varying Ω_t. Columns 3, 6 and 9 reports the percentage deviations for the Rouwenhorst approximation.

References

Abbott, Brant, Gallipoli, Giovanni, Meghir, Costas and Violante, Giovanni L. (2018), 'Education Policy and Intergenerational Transfers in Equilibrium', *Journal of Political Economy* (forthcoming).

Abowd, John and Card, David (1989), 'On the covariance structure of earnings and hours changes', *Econometrica* **57**(2), 411–45.

Adda, Jerome and Cooper, Russell W. (2003), *Dynamic Economics: Quantitative Methods and Applications*, MIT Press.

Arellano, Manuel, Blundell, Richard and Bonhomme, Stéphane (2017), 'Earnings and consumption dynamics: A non-linear panel data framework', *Econometrica* **85**(3), 693–734.

Barillas, Francisco and Fernández-Villaverde, Jésus (2007), 'A generalization of the endogenous grid method', *Journal of Economic Dynamics and Control* **31**, 2698–2712.

Blundell, Richard, Graber, Michael and Mogstad, Magne (2015), 'Labor income dynamics and the insurance from taxes, transfers, and the family', *Journal of Public Economics* **127**, 58–73.

Cagetti, Marco and De Nardi, Mariacristina (2006), 'Entrepreneurship, frictions and wealth', *Journal of Political Economy* **114**(5), 835–870.

Carroll, Christopher (2004), Theoretical foundations of buffer stock saving, Working Paper 10867, National Bureau of Economic Research.

Carroll, Christopher D. (2006), 'The method of endogenous gridpoints for solving dynamic stochastic optimization problems', *Economics Letters* **91**(3), 312–320.

Carroll, Christopher D. (2009), 'Precautionary saving and the marginal propensity to consume out of permanent income', *Journal of Monetary Economics* **56**, 780–790.

Civale, Simone, Díez-Catalán, Luis and Fazilet, Fatih (2017), 'Discretizing a process with non-zero skewness and high kurtosis', Mimeo, University of Minnesota.

Conesa, Juan Carlos and Krueger, Dirk (2006), 'On the optimal progressivity of the income tax code', *Journal of Monetary Economics* **53**(7), 1425 – 1450.

Deaton, Angus and Paxson, Christina (1994), 'Intertemporal choice and inequality', *Journal of Political Economy* **102**(3), 437–67.

den Haan, Wouter J and Marcet, Albert (1990), 'Solving the Stochastic Growth Model by Parameterizing Expectations', *Journal of Business and Economic Statistics* **8**(1), 31–34.

De Nardi, Mariacristina (2004), 'Wealth inequality and intergenerational links', *Review of Economic Studies* **71**, 743–768.

De Nardi, Mariacristina, Fella, Giulio and Paz-Pardo, Gonzalo (2018), 'Nonlinear household earnings dynamics, self-insurance and welfare', *Journal of the European Economics Association* (forthcoming).

Flodén, Martin (2008), 'A note on the accuracy of Markov-chain approximations to highly persistent AR(1) processes', *Economics Letters* **99**(3), 516 – 520.

Guvenen, Fatih, Karahan, Fatih, Ozkan, Serdar and Song, Jae (2016), What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk? Mimeo, University of Minnesota.

Guvenen, Fatih, Ozkan, Serdar and Song, Jae (2014), 'The nature of countercyclical income risk', *Journal of Political Economy* **122**(3), pp. 621–660.

Heathcote, Jonathan, Storesletten, Kjetil and Violante, Giovanni L. (2017), 'Optimal tax progressivity: An analytical framework', *The Quarterly Journal of Economics* **132**(4), 1693–1754.

Huggett, Mark (1996), 'Wealth distribution in life-cycle economies', *Journal of Monetary Economics* **38**(3), 469–494.

Kaplan, Greg (2012), 'Inequality and the life cycle', *Quantitative Economics* **3**(3), 471–525.

Kaplan, Greg, Moll, Benjamin and Violante, Giovanni L. (2018), 'Monetary policy according to HANK', *American Economic Review* **108**(3), 697–743.

Karahan, Fatih and Ozkan, Serdar (2013), 'On the Persistence of Income Shocks over the Life Cycle: Evidence, Theory, and Implications', *Review of Economic Dynamics* **16**(3), 452–476.

Kopecky, Karen and Suen, Richard (2010), 'Finite State Markov-chain Approximations to Highly Persistent Processes', *Review of Economic Dynamics* **13**(3), 701–714.

Krueger, Dirk and Ludwig, Alexander (2013), 'Optimal progressive labor income taxation and education subsidies when education decisions and intergenerational transfers are endogenous', *American Economic Review* **103**(3), 496–501.

McKay, Alisdair (2017), 'Time-varying idiosyncratic risk and aggregate consumption dynamics', *Journal of Monetary Economics* **88**, 1 – 14.

Rouwenhorst, Geert (1995), Asset pricing implications of equilibrium business cycle models, *in* Thomas F.Cooley, ed., 'Frontiers of business cycle research', Princeton University Press, chapter 10.

Storesletten, Kjetil, Telmer, Chris I. and Yaron, Amir (2004*a*), 'Cyclical dynamics in idiosyncratic labor-market risk', *Journal of Political Economy* **112**, 695–717.

Storesletten, Kjetil, Telmer, Christopher I. and Yaron, Amir (2004*b*), 'Consumption and risk sharing over the life cycle', *Journal of Monetary Economics* **51**(3), 609–633.

Tauchen, George (1986), 'Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions', *Economics Letters* **20**(2), 177–181.

Tauchen, George and Hussey, Robert (1991), 'Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models', *Econometrica* **59**(2), 371–96.