Finite-State Markov-Chain Approximations: A Hidden Markov Approach

Eva F. Janssens† and Sean McCrary‡

July 10, 2022

PLEASE DO NOT QUOTE OR REDISTRIBUTE WITHOUT PERMISSION.

Abstract

This paper proposes a novel finite-state Markov chain approximation method for Markov processes with continuous support. The method can be used for both uni- and multivariate processes, as well as non-stationary processes such as those with a life-cycle component. The method is based on minimizing the information loss between a misspecified approximating model and the true data generating process. In contrast to existing methods, we provide both an optimal grid and transition probability matrix. We provide guidance on how to select the optimal number of grid points. The method outperforms existing methods in several dimensions, including parsimoniousness. We compare the performance of our method to existing methods through the lens of an asset-pricing model, and a life-cycle consumption-savings model. We find the choice of the discretization method matters for the accuracy of the model solutions, the welfare costs of risk, and the amount of wealth inequality a life-cycle model can generate.

Keywords: Numerical methods, Kullback–Leibler divergence, life-cycle dynamics, earnings process

JEL classification codes: C63, C68, D15, E21

* Acknowledgements: Both authors are grateful to the invaluable comments from José Víctor Ríos Rull, Frank Kleibergen, Christian Stoltenberg, Robin Lumsdaine, as well as seminar participants at the University of Zürich and University of Amsterdam. Janssens is grateful to the Dutch Research Council for the NWO Research Talent Grant, project number 406.18.514 and to Erasmus Trustfonds for the Professor Bruins Prize 2018, funding the research visit to University of Pennsylvania during which this paper was written, as well as to Frank Schorfheide for hosting this visit. We thank the Society of Computational Economics for the CEF 2022 Student Prize. All errors are our own.

Contact information:

† Eva F. Janssens: University of Amsterdam, Postbus 15867, 1001 NJ Amsterdam, The Netherlands, e-mail: e.f.janssens@uva.nl

‡ Sean McCrary: University of Pennsylvania, The Ronald O. Perelman Center for Political Science and Economics 133 South 36th Street, Suite 150, Philadelphia, PA 19104, United States, e-mail: smccrary@sas.upenn.edu

1 Introduction

Numerical methods to solve nonlinear dynamic stochastic models often rely on finite-state Markov chain approximations of continuous stochastic processes. These models are used to answer many policy-relevant questions and to study business cycles, asset pricing, intra-household insurance and more. The stochastic process is an important input for these models and its finite-state Markov chain approximation should therefore resemble the original process as closely as possible. This paper proposes a novel method that can be used for the discretization of continuous Markov processes while providing both an optimal grid and transition probability matrix, as well as a way to select the optimal number of grid points for the discretization. This is an improvement to the existing literature that typically assumes (a given) equal-distant or equal-quantile grid and only provides the corresponding transition probability matrix. For multivariate processes, our discretization method does not rely on tensor products, avoiding the curse-of-dimensionality issue other discretization methods face. We also extend our method so that it can be used for non-stationary processes with life-cycle dynamics, providing age-dependent grids and transition probability matrices.

Approximating a continuous Markov process by a discrete Markov process inherently comes down to picking a misspecified model according to a certain objective function. Existing discretization methods such as those in Rouwenhorst (1995) and Farmer and Toda (2017) focus on choosing the transition probability matrix such that the discretized process matches a set of low order moments of the underlying continuous-support process. For a small class of optimization problems, this procedure may be optimal.[1] However, for the more general class of these decision problems, all moments – both conditional and unconditional – of the stochastic process may matter to the agent, and to what extent will depend on the nature of the decision problem and the characteristics of the stochastic process.

Ideally, a discretization procedure would therefore minimize the welfare loss of an agent that uses the discretized process instead of the continuous-support process when solving their optimization problem. However, given that the solution to the optimization problem under the continuous process is typically unknown, which is why a discretization is needed for numerical computation, this welfare loss cannot be evaluated. Instead, we propose to minimize the information loss that the agent faces when using the misspecified discrete process instead of the continuous process. This objective has the feature that it is not problem specific, and if the information loss between the misspecified process and true process can

[1] For example, in a consumption-saving or portfolio-selection problem with quadratic utility, such a discretization should match the conditional mean and variance.

be made arbitrarily small asymptotically, an implication of Portmanteau's lemma is that the welfare loss of using the misspecified process can also be made arbitrarily small.

To minimize the information loss, we use the Kullback-Leibler (KL) divergence between the continuous-support process and the discretized process. To link the continuous and discrete distributions, we assume that in the misspecified process, each observation is equal to the sum of a state-dependent level and an error term. This state is unobserved, and the evolution of the unobserved state is governed by a discrete first-order Markov process. This effectively embeds a discrete Markov chain into a continuous support process via a continuous measurement error, i.e., a Hidden Markov Model (HMM). By Douc and Moulines (2012), the maximum likelihood estimator of a misspecified HMM minimizes the KL-divergence between the model and the true distribution. The objective of minimizing the KL-divergence is asymptotically equivalent to simulating data from the continuous-support process, and estimating the HMM parameters via maximum likelihood on this simulated data. Our method can be seen as a full-information discretization method, compared to moment-matching methods like Rouwenhorst (1995), Gospodinov and Lkhagvasuren (2014), and Farmer and Toda (2017).

For computational reasons, a discrete process ideally is low dimensional. Our HMM method can be seen as a probabilistic clustering method, where each realization of the continuous-support stochastic process has a certain probability to fall into a certain cluster. We can rely on methods from the clustering literature to select the optimal number of grid points. Therefore, we propose using a scree-plot in the log likelihood of the HMM, such that the number of grid points is chosen such that the information gain from including an additional grid point is diminishing. The interpretation of an HMM as a dimension reduction method for dependent data is common in the statistics literature (McLachlan, Lee, and Rathnayake, 2019).

We apply our discretization method to a large number of stochastic processes, namely an autoregressive process with Gaussian errors and with errors from a normal mixture distribution (colloquially, "fat tails"), an AR(1) with stochastic volatility (SV), a Vector AutoRegression (VAR) process, the earnings process in Guvenen, Karahan, Ozkan, and Song (2021) that features life-cycle dynamics, non-employment shocks and fat tails, and the non-parametric and highly non-linear earnings process of Arellano, Blundell, and Bonhomme (2017). We compare how our method performs compared to the methods by Rouwenhorst (1995), Tauchen (1986), Farmer and Toda (2017), and the binning method of Adda and Cooper (2003) (and the adaption of Adda and Cooper (2003) by De Nardi, Fella, and Paz-Pardo (2020) to discretize Arellano et al. (2017)). We find that our method can outperform existing methods in various dimensions, where the performance is most pronounced in the AR(1) with fat tails, the AR(1)-SV, the Guvenen et al. (2021) and the Arellano et al. (2017) process. However, even for simple

AR(1) processes, we find that our method produces the lowest mean-squared forecast error, implying that a decision maker makes smaller forecasting errors when using our discretized process than the discretized process that follows from existing methods. In addition, we show that our method can generate a more parsimonious discretization than other methods without sacrificing in terms of information loss to the continuous-support process, which is attractive for computational efficiency.

We evaluate the performance of our method in two economic applications. First, an asset pricing model where dividend growth follows an AR(1) process with stochastic volatility. As shown by De Groot (2015), this model has a closed-form solution. We use this solution as a benchmark to compare the performance of our method against the standards in the literature. We find our method is more accurate than the method of Farmer and Toda (2017) or a binning method as in Adda and Cooper (2003) with a small number of gridpoints. In particular, we analyze the accuracy of the three discretization methods for estimates of the certainty equivalent level of consumption (CEC) and find that our method deviates 0.8-1.9% from the closed-form solution of De Groot (2015), while the method of Farmer and Toda (2017) results in deviations ranging from 8.3-12.2%, and the Adda and Cooper (2003) method has deviations ranging from 4%-5.4%. These results highlight the importance of considering a full-information approach, as for a highly non-linear object as the CEC, all information of the stochastic process is important and should be incorporated in the discretization.

Second, we analyze the performance of our method through the lens of a life-cycle consumption-saving model. In this application, we consider two discretized processes that both feature life-cycle dependence; the process proposed in Guvenen et al. (2021), henceforth GKOS, and the non-parametric process in Arellano et al. (2017), henceforth ABB. We find that the choice of the discretization method matters greatly for the mean and variance of asset holdings and consumption over the life-cycle, and also matters for other relevant statistics that are often reported in a life-cycle context, such as the covariance between earnings changes and consumption changes, the partial insurance to permanent earnings shocks as measured by Blundell, Pistaferri, and Preston (2008), and the welfare cost of risk.

For the GKOS process, binning-based methods can underestimate the welfare cost of risk by as much as 17-24 percentage points relative to our method, because they fail to capture the rich dynamics of non-employment and the higher-order moments of fat-tailed processes. Our discretization of ABB results in a welfare cost of risk that is 6.7 percentage points larger than the one obtained by applying the binning method used in De Nardi et al. (2020). This discrepancy comes from the fact that our discretization better captures the excess kurtosis and

skewness of the ABB process over the life-cycle. Welfare cost estimates are important tools for policy analysis and their sensitivity to the choice of the discretization method highlights the importance of having an accurate approximation that captures all moments of the stochastic process well.

To our knowledge, this paper is the first to discretize the GKOS process, which, together with the ABB process, is considered to be at the frontier of the earnings dynamics literature (Altonji, Hynsjö, and Vidangos, 2022). Our parsimonious discretization allows for an easy interpretation of and comparison between both processes. We find the largest source of risk in GKOS comes from the probability of non-employment, which is a highly persistent state whose persistence rises over the life-cycle. In contrast, most risk in ABB comes from the highest earnings state, which features a considerable probability of earnings loss next period, especially at younger ages. In contrast to GKOS, the low-earnings states in ABB are fairly constant in their dynamics over the life-cycle while the transition probabilities related to the highest earnings states do change with age. In our life-cycle model, we find that our discretization of ABB can generate wealth inequality similar to that observed in the United States. This is not the case for other discretization methods, nor for the GKOS process.

The paper proceeds as follows. The next subsection discusses the related literature. Section 2 discusses our discretization method and how it can be applied to an AR(1) process. Section 3 presents the asset pricing model with stochastic volatility and Section 4 discusses the life-cycle model and the discretization of the GKOS and ABB processes. Section 5 concludes. Appendix Section B presents the discretization of the AR(1) process in more detail, as well as an AR(1) process with fat tails, the VAR process and a different specification of the AR(1)-SV process. In Appendix Section F, these processes are also incorporated in the life-cycle model of Section 4, to demonstrate that even for linear Gaussian processes, the discretization method matters and can lead to different conclusions.

Related literature Several papers have proposed methods to discretize stochastic processes. Most of these, such as Tauchen (1986), Rouwenhorst (1995), Tauchen and Hussey (1991), Terry and Knotek II (2011), and Gospodinov and Lkhagvasuren (2014) are designed for specific linear and Gaussian processes, such as AR(1) or VAR processes. Fella, Gallipoli, and Pan (2019) adapt the methods of Rouwenhorst (1995), Tauchen and Hussey (1991) and Adda and Cooper (2003) to processes with a life-cycle component, and analyze how it performs under settings where the innovations are drawn from a mixture of normals. Kopecky and Suen (2010) assess the performance of various methods for AR(1) processes close to unit-root, and finds that Rouwenhorst (1995) is more robust to these highly persistent environments. Galindev and

Lkhagvasuren (2010) adapt Rouwenhorst (1995) to a setting with highly-persistent correlated AR(1) shocks. Civale, Díez-Catalán, and Fazilet (2016) adapt the Tauchen (1986) method to accommodate autoregressive processes with innovations drawn from a normal mixture[2]. An important difference of these methods to ours is that our method is generally applicable to any process, and provides both an optimal grid and transition probability matrix, while these methods typically take a grid as input, and/or assume equal-distant or equal-quantile grids.

Some discretization methods are applicable to a larger class of stochastic processes. Binning methods as in Adda and Cooper (2003), which discretize via a partition in the quantile space, are applicable to any stochastic process. However, binning methods take the grid spacing as an input. Our discretization method provides an optimal grid as well. Another method that is applicable to any Markov process is Farmer and Toda (2017), who propose a method to refine discrete approximations by moment matching. Their method takes as inputs a grid, an initial transition probability matrix, and a set of moments to match, where the goal is to match these moments exactly - if possible - with a transition matrix that is close to the initial approximation measured through relative entropy. Our method, in contrast, can be seen as a full-information discretization method that does not rely on prior information to obtain identification.

For multivariate processes, most existing methods rely on tensor grids, which leads to a curse of dimensionality and is computationally unattractive. As stated by Gordon (2021), tensor grids are inefficient, because many of the grid points will rarely be visited. Gordon (2021) proposes the use of pruning and sparse grids for VAR models. Our method results in optimal grids that do not suffer from the issue that Gordon (2021) aims to solve, and is applicable to any type of process.

2 Discretization using Hidden Markov Models

Let $y_{it} \in \mathbb{R}^k$, $i = 1, ..., N$, $t = 1, ..., T$, denote a random variable for which the data generating process is any discrete-time continuous-support Markov process. Denote its probability distribution by $f(\mathbf{y})$. The objective is to approximate the distribution of $\mathbf{y}$ by a misspecified model, with probability distribution $p(\mathbf{y}; \theta)$, by choosing parameter vector θ such that the relative entropy from the misspecified distribution P to the distribution F of the misspecified model is minimized.

The relative entropy is defined as the logarithmic difference between the distributions F and P, where the expectation is taken using the distribution F, also known as the Kullback–Leibler

[2]Normal mixtures can generate non-zero skewness and excess kurtosis

(KL) divergence:

$$\mathrm{D}_{\mathrm{KL}} = \int \log\left(\frac{f(\mathbf{y})}{p(\mathbf{y};\theta)}\right) f(\mathbf{y}) d\mathbf{y}. \tag{1}$$

Minimizing the KL divergence with respect to parameter vector θ requires taking the derivative of Equation 1 with respect to θ:

$$\begin{aligned} &\int \nabla\theta \log p(\mathbf{y};\theta) f(\mathbf{y}) dy = 0 \\ \Leftrightarrow\ &\mathbb{E}_f\left[\nabla\theta \log\left(p(\mathbf{y};\theta)\right)\right] = 0. \end{aligned}$$

Typically, $\mathbb{E}_f(\cdot)$ is hard to evaluate, and can be replaced by an estimate, by simulating data $\mathbf{y}_{\text{sim}} = \{y_t\}_{t=1}^T$ from $f(\mathbf{y})$, and evaluating $\nabla\theta \log\left(p(\cdot;\theta)\right)$ in the simulated data.

2.1 Hidden Markov Model

As our misspecified model, we propose using the following Hidden Markov Model (HMM). By Douc and Moulines (2012), the maximum likelihood estimator of a misspecified HMM minimizes the KL divergence between the misspecified process and the true underlying process, making an HMM suitable for our above-stated objective. Denote the data by $y_{i,t} \in \mathbb{R}^k$, and denote unobserved states $x_{i,t} \in \{1, ..., m\}$. The latent state $x_{i,t}$ lies in a finite discrete set $\{1, 2, \ldots, m\}$ which evolves according to a time-homogeneous first-order Markov process.

$$y_{i,t}|x_{i,t} = \mu_t(x_{i,t}) + \mathrm{diag}(\sigma_t)\varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim N(0, I_k) \tag{2}$$

$$x_{i,t+1}|x_t \sim \Pi_{ij,t}. \tag{3}$$

Denote bold variables $\boldsymbol{y} = \{y_{i,1}, y_{i,2}, \ldots, y_{i,T}\}_{i=1}^N$ and $\boldsymbol{x} = \{x_{i,1}, x_{i,2}, \ldots, x_{i,T}\}_{i=1}^N$ as realizations of this random process. The transition matrix Π_t has stationary distribution $\boldsymbol{\delta}_t = (\delta_{1,t}, \delta_{2,t}, \ldots, \delta_{m,t})$. Note, $\boldsymbol{\delta}_t$ is a $1 \times m$ row vector. Parameter vector θ in Equation (1) thus consists of:

(i) the parameters in transition probability matrix Π_t, denoted by $\Pi_{ij,t}$. This matrix is allowed to be time-varying. In the case that there is no time dependence, that is, $\Pi_t = \Pi$ for all $t = 1, ..., T$, the number of parameters in Π is $m \times m$, of which $m \times (m-1)$ are linearly independent, given that each row sums to one;

(ii) μ_t is the grid, and is allowed to be time-varying. When there is no time dependence, $\mu_t = \mu$ is a $m \times k$ matrix;

(iii) σ_t^2 is the variance of the error term, which is allowed to be varying. If $y_{i,t} \in \mathbb{R}^k$ has $k > 1$, we assume the variance is a diagonal matrix $\text{diag}(\sigma_1, ..., \sigma_k)$. It is also possible to allow σ to be state-dependent, that is, to let σ vary with each realization of $\boldsymbol{x}$.

The HMM in Equation (2) can be estimated using the Expectation-Maximization (EM) algorithm that we will describe below. As follows from Douc and Moulines (2012), the maximum likelihood estimator of parameter vector $\theta = (\text{vec}(\Pi), \text{vec}(\mu), \sigma)$ in the model in Equation (2) is the set of parameters that minimizes the information loss as specified in Equation (1). These parameters $\theta = (\Pi, \mu, \sigma)$ gives a discretization of the process $f(\mathbf{y})$, where μ is the grid of the discretized process, and Π governs the transitions between the m states.

We will consider time series settings where $N = 1$, as well as panel $N \geq 2$. The inclusion of a panel dimension allows for the estimation of parameters that vary with t (e.g., over the life-cycle).

2.2 Estimation of HMM

We first discuss the general procedure we use for the estimation of the HMM. Let $\phi_j^t(y_{i,t}) = P(y_{i,t}|x_{i,t} = j)$ denote the density of $y_{i,t}$ conditional on $x_{i,t}$ being in state j. That is,

$$\phi_j^t(y_{i,t}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_{i,t}-\mu_t(j))^2}, \tag{4}$$

if $k = 1$, or $\det(2\pi\Sigma_t)^{-\frac{1}{2}} e^{-\frac{1}{2}(y_{i,t}-\mu_t(j))'(\Sigma_t)^{-1}(y_{i,t}-\mu_j)}$ for $k > 1$, where $\Sigma_t = \text{diag}(\sigma_t^2)$. It will be useful to think of the following matrix form for the observation densities:

$$\mathbf{\Phi}^t(y_{i,t}) = \begin{pmatrix} \phi_1^t(y_{i,t}) & & 0 \\ & \ddots & \\ 0 & & \phi_m^t(y_{i,t}) \end{pmatrix}, \tag{5}$$

that is, $\mathbf{\Phi}^t$ is an $m \times m$ diagonal matrix with the observation densities as diagonal elements.

The complete data likelihood (CDL) of the model in Equation (2) is given by

$$\mathcal{L}(\theta|\boldsymbol{y}, \boldsymbol{x}) = p(\boldsymbol{y}, \boldsymbol{x}|\theta) = p(\boldsymbol{y}|\boldsymbol{x}, \theta)p(\boldsymbol{x}|\theta), \tag{6}$$

and the maximum likelihood estimator is given by

$$\theta^* = \underset{\theta}{\text{argmax}}\, \mathcal{L}(\theta|\boldsymbol{y}, \boldsymbol{x}). \tag{7}$$

If the latent states $\boldsymbol{x}$ were observed, the log-likelihood would be straightforward to maximize. This is because the log-likelihood is given by

$$\log\left(\mathcal{L}(\theta|\boldsymbol{y},\boldsymbol{x})\right) = \log\left(p(\boldsymbol{y}|\boldsymbol{x},\theta)\right) + \log\left(p(\boldsymbol{x}|\theta)\right), \tag{8}$$

and, conditional on $\boldsymbol{x}$, the parameters Π do not influence $\boldsymbol{y}$ and, similarly, the parameters $(\boldsymbol{\mu}, \sigma)$ do not matter for $\boldsymbol{x}$. Together this implies the log-likelihood is given by

$$\log\left(\mathcal{L}(\theta|\boldsymbol{y},\boldsymbol{x})\right) = \log\left(p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\mu},\sigma)\right) + \log\left(p(\boldsymbol{x}|\Pi)\right) \tag{9}$$

That is, the parameters governing the observation equation and state transition equation could be solved for separately, given $\boldsymbol{x}$. Intuitively, if the states $\boldsymbol{x}$ are observed, one could estimate Π using only data on transitions from $\boldsymbol{x}$, estimate $\mu(x_{i,t} = j)$ by averaging the $y_{i,t}$ that are observed when $x_{i,t}$ is in state j, and then estimate Σ using the sample variance of the observations $\boldsymbol{y}$ demeaned by the estimates of $\boldsymbol{\mu}$.

In practice, the latent states $\boldsymbol{x}$ are unobservable, but we can use the EM algorithm to maximize the likelihood. The EM algorithm (or Baum-Welch algorithm in the case of HMMs) iterates between updating the posterior distribution over the latent states $p_{\boldsymbol{x}} = p(\boldsymbol{x}|\boldsymbol{y},\theta)$ taking the parameters and observations $(\boldsymbol{y},\theta)$ as fixed in the E step, and updating the parameters $\theta^{(i)} \rightarrow \theta^{(i+1)}$ taking the latent states and observations $(p_{\boldsymbol{x}},\boldsymbol{y})$ as fixed in the M step.

We now describe the E-step. Let $y_i^t = (y_{i,1}, y_{i,2}, \ldots, y_{i,t})$, i.e., the observed values up to time t for individual i. The forward probabilities $\boldsymbol{\alpha}_{i,t}(j)$ are given by

$$\boldsymbol{\alpha}_{i,t}(j) = p\left(y_i^t, x_{i,t} = j|\theta\right) \tag{10}$$

We can define $\boldsymbol{\alpha}_{i,t}$ recursively as

$$\begin{aligned}
\boldsymbol{\alpha}_{i,1}(j) &= \delta_{1,j}\phi_j^t(y_{i,1}) \\
\boldsymbol{\alpha}_{i,t+1}(j) &= \left(\sum_{k=1}^{m} \boldsymbol{\alpha}_{i,t}(k)\Pi_{kj,t}\right)\phi_j^t(y_{i,t+1}),
\end{aligned} \tag{11}$$

or in matrix form

$$\boldsymbol{\alpha}_{i,t} = \boldsymbol{\alpha}_{i,t-1}\Pi_t\boldsymbol{\Phi}^t(y_{i,t}). \tag{12}$$

Similarly, let $y_{i,t+1}^{T} = (y_{i,t+1}, y_{i,t+2}, \ldots, y_{i,T})$, i.e., the observed values from time $t+1$ to T for individual i. The backward probabilities $\boldsymbol{\beta}_{i,t}(k)$ are given by

$$\boldsymbol{\beta}_{i,t}(k) = p\left(y_{i,t+1}^{T} | x_{i,t} = k, \theta\right) \tag{13}$$

We can define $\boldsymbol{\beta}_{i,t}$ recursively as

$$\begin{aligned} \boldsymbol{\beta}_{i,T}(k) &= 1 \\ \boldsymbol{\beta}_{i,t}(k) &= \sum_{j=1}^{m} \Pi_{kj,t} \phi_j^t(y_{i,t+1}) \boldsymbol{\beta}_{i,t+1}(j), \end{aligned} \tag{14}$$

or, in matrix form,

$$\boldsymbol{\beta}_{i,t}' = \Pi \boldsymbol{\Phi}^t(y_{i,t+1}) \boldsymbol{\beta}_{i,t+1}'. \tag{15}$$

Using these probabilities, we can define the probability of being in state k at time t, and observing $\boldsymbol{y}_{i,t}$ as

$$p(\boldsymbol{y}_{i,t}, x_{i,t} = k|\theta) = \boldsymbol{\alpha}_{i,t}(k) \boldsymbol{\beta}_{i,t}(k). \tag{16}$$

This leads to a posterior probability of being in state k, given by

$$\gamma_{i,t}(k) = p(x_{i,t} = k|\boldsymbol{y}_{it}, \theta) = \frac{p(\boldsymbol{y}_{i,t}, x_{i,t} = k|\theta)}{p(\boldsymbol{y}_{i,t}|\theta)} = \frac{p(\boldsymbol{y}_{i,t}, x_{i,t} = k|\theta)}{\sum_{j=1}^{m} p(\boldsymbol{y}_{i,t}, x_{i,t} = j|\theta)} = \frac{\boldsymbol{\alpha}_{i,t}(k) \boldsymbol{\beta}_{i,t}(k)}{\sum_{j=1}^{m} \boldsymbol{\alpha}_{i,t}(j) \boldsymbol{\beta}_{i,t}(j)}. \tag{17}$$

We can also define the posterior transition probability between state i at time t and state j at time $t+1$ as

$$\begin{aligned} \xi_{i,t}(k,j) &= p(x_{i,t+1} = j, x_{i,t} = k|\boldsymbol{y}_{i,t}, \theta) \\ &\propto \boldsymbol{\beta}_{i,t+1}(j) \phi_j^t(y_{i,t+1}) \Pi_{kj,t} \boldsymbol{\alpha}_{i,t}(k), \end{aligned} \tag{18}$$

where the last line follows from the definition of $\boldsymbol{\gamma}_{i,t}(k)$ from above.

At last, the M step is given by

$$\mu_t^l(j) = \frac{\sum_{i=1}^N y_t^l p(x_{i,t} = j|\boldsymbol{y}_{i,t}, \theta)}{\sum_{i=1}^N p(x_{i,t} = j|\boldsymbol{y}_{i,t}, \theta)} = \frac{\sum_{i=1}^N y_{i,t}^l \gamma_{i,t}(j)}{\sum_{i=1}^N \gamma_{i,t}(j)} \tag{19}$$

$$(\sigma_t^l)^2 = \frac{\sum_{i=1}^N \sum_{j=1}^m (y_{i,t}^l - \mu_t^l(j))^2 p(x_{i,t} = j|\boldsymbol{y}_{i,t}, \theta)}{\sum_{i=1}^N \sum_{j=1}^m p(x_{i,t} = j|\boldsymbol{y}_{i,t}, \theta)} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^m (y_{i,t}^l - \mu_t^l(j))^2 \gamma_{i,t}(j) \tag{20}$$

$$\Pi_{qj,t} = \frac{\sum_{i=1}^N p(x_{i,t} = j, x_{i,t-1} = q|\boldsymbol{y}_{i,t}, \theta)}{\sum_{i=1}^N p(x_{i,t-1} = q|\boldsymbol{y}_{i,t}, \theta)} = \frac{\sum_{i=1}^N \xi_{i,t}(q, j)}{\sum_{i=1}^N \gamma_{i,t}(q)}, \tag{21}$$

for $l = 1, ..., k$.

When omitting time-dependence, the M-step becomes

$$\mu_l^i = \frac{\sum_{t=1}^T \sum_{i=1}^N y_{i,t}^l \gamma_{i,t}(j)}{\sum_{t=1}^T \sum_{i=1}^N \gamma_{i,t}(j)} \tag{22}$$

$$(\sigma^l)^2 = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^m (y_{i,t}^l - \mu_j^l)^2 \gamma_{i,t}(j) \tag{23}$$

$$\Pi_{qj} = \frac{\sum_{t=2}^T \sum_{i=1}^N \xi_{i,t}(q, j)}{\sum_{t=2}^T \sum_{i=1}^N \gamma_{i,t}(q)}, \tag{24}$$

for $l = 1, ..., k$.

Given the updated transition matrix Π_t we can update the stationary probabilities as

$$\boldsymbol{\delta}_t = \mathbf{1}' \left(I_m - \Pi_t + U\right)^{-1}. \tag{25}$$

Here U is a $m \times m$ matrix of ones.

For the estimation of models with rich life-cycle dynamics, we will use an iterative adaption of this algorithm. This is described in Appendix Section A.

2.3 Imposing structure through restrictions

One can impose additional structure by estimating the process under a set of restrictions. In the estimation, this happens through modifying the M step. For example, for symmetric processes, a symmetry restriction can be imposed on μ. In case of a process that is symmetric around zero and an odd number of grid points m, this means that:

$$\mu(\lceil m/2 \rceil) = 0, \text{ and } \mu(\lceil m/2 \rceil - r) = -\mu(\lceil m/2 \rceil + r), \quad \text{for } r = 1, ..., \lfloor m/2 \rfloor \tag{26}$$

Similarly, a process can also be symmetric in its dynamics, as reflected by the transition probability matrix. In that case, the restriction takes the form

$$\Pi_{i,j} = \Pi_{(m+1-i),(m+1-j)}. \tag{27}$$

2.4 Imposing structure through a penalty term

For the specific restrictions in Equations (26)-(27), a closed-form solution is available for the M-step. In other cases, one may want to introduce restrictions through penalty terms rather than hard restrictions. For example, one wants the discretized process to target certain moments. Denote a certain set of moments functions of the discretized process by $\mathcal{M}(p(\mathbf{y}; \theta))$ and the moments of the continuous process by $\mathcal{M}(f(\mathbf{y}))$. In that case, instead of maximizing the log-likelihood of the simulated data $\mathbf{y}_{\text{sim}}$, maximize:

$$\log(\mathcal{L}(\theta|\mathbf{y}, \mathbf{x})) - \lambda \mathcal{D}\left(\mathcal{M}(f(\mathbf{y})), \mathcal{M}(p(\mathbf{y}; \theta))\right) \tag{28}$$

where $\lambda \in \mathbb{R}^+$ is a scalar parameter and $\mathcal{D}(\cdot, \cdot)$ a distance measure of choice. λ is chosen by the researcher. A higher λ should be chosen if the researcher considers it more important that the discretization matches the moments $\mathcal{M}$. Typically the M-step will no longer be analytically tractable and numerical optimization will be necessary.

Another example is when one wants to encourage sparsity in one or more dimensions of the grid in the case that $k > 1$, where k is the dimension of y_{it}. This can limit the number of distinct entries in the grid, and can force the grid to be more tensor-like.

2.5 Allowing for life-cycle dynamics

For age-dependent life-cycle processes, like GKOS and ABB, we want to allow for age-dependent transition probabilities and grid placement. The EM algorithm above already

allows for this, and, using Equations (19)-(21), the transition probability matrix Π_t and the grid μ_t are fully time-varying (where time is corresponding to age). In this case, the asymptotics depend on N and a different transition probability matrix and grid are estimated for every age group.

2.6 Selecting the number of grid points

To select the number of grid points m, we face a trade-off between minimizing the extent of misspecification from the discretization on the one hand, and wanting a low number of grid points as is desirable for computational reasons. The intuition of using a HMM is that this provides a so-called "soft-clustering". Therefore, we propose to use an elbow plot based on the log-likelihood of the misspecified HMM on the y-axis, and the number of grid points on the x-axis. Recall that the log-likelihood is proportional to the information loss of the misspecified model relative to the true continuous process. An elbow plot is a heuristic commonly used for the selection of the number of factors or clusters when reducing the dimensionality of a dataset. This will ensure that m is chosen such that that the decrease in the information loss from adding an additional grid point is diminishing.

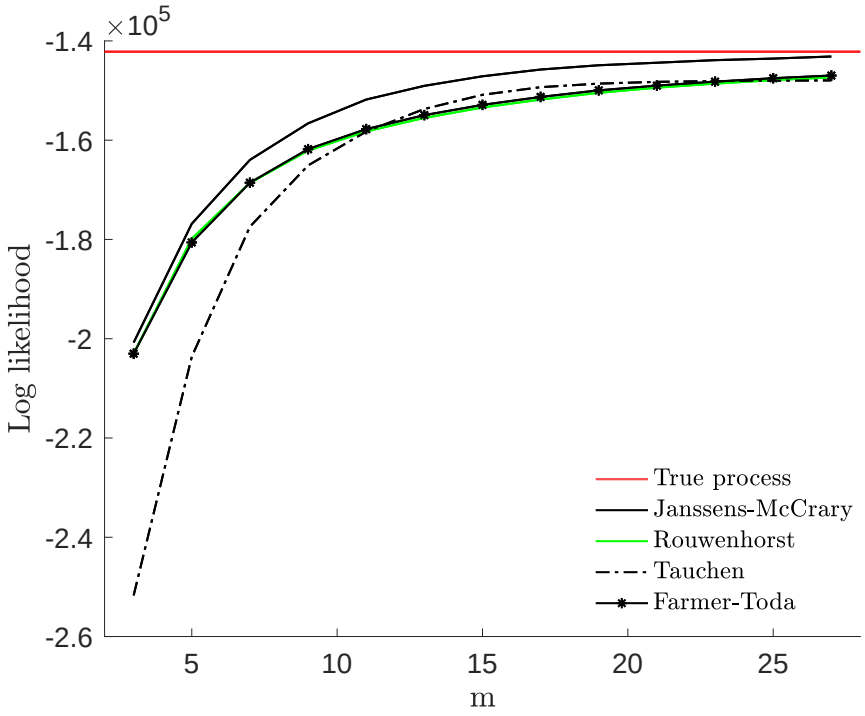
Example: AR(1) process To demonstrate the use of the elbow plot, we use an AR(1) process with Gaussian errors:

$$y_t = \rho y_{t-1} + \sigma_\varepsilon \varepsilon_t, \quad \varepsilon_t \sim N(0,1), \quad t = 1, ..., T. \tag{29}$$

We simulate a time series of length $T = 100,000$ from the process in Equation (29), and estimate the HMM of Equation (2) on the simulated data for different choices of the grid size m. We impose both restriction (26) and (27) because of the symmetry of the process. The black sold line in Figure 1 then displays the maximum log likelhood obtained as m varies from 3 to 27. As can be seen from this figure, the black solid line is elbow shaped, and the elbow lies at $m = 7$. Alternatively, if one wants to choose m such that the elbow graph is almost flat, we would in this case recommend $m = 15$.

Figure 1 also displays the true log likelihood of the continuous-support AR(1) process in red. As can be seen, as m becomes larger, the HMM log likelihood converges to the true log likelihood. In addition, the figure visualizes the implied log likelihood of three competing methods that are often used for discretizing AR(1) processes. To obtain these log likelihoods, we use the obtained discretization from the existing method in our HMM model, and interpret the given grid and transition probability matrix Π as a restriction, but estimate the correspond-

Figure 1: Elbow plot for an AR(1) process with Gaussian innovations. $\rho = 0.95$, $\sigma_\varepsilon = 1$ *and* $T = 100,000$. *The red line displays the true log likelihood for the AR(1) process, the other lines visualizes existing methods.*



ing variances Σ. We can use this statistic to demonstrate how much more parsimonious our method is than existing methods, and that we can achieve the same relative information loss as existing methods with fewer grid points. As can be seen from Figure 1, while we already capture almost all of the information of the true process at 25 grid points, this does not hold for the existing methods. We can achieve the same information loss they achieve at 25 grid points with only 15.

2.7 Evaluation criteria

We compute a number of statistics in addition to the relative information loss described in the previous subsection. We will focus on the standard set of statistics, such as unconditional and unconditional moments, as well as accuracy of one-step ahead predictions. The mean squared forecast error (MSFE) of the misspecified model measures the one-step ahead forecasting error that the agent makes. For this statistic, we assume that an agent assigns the grid point closest

to the current realization of y_t for forecasting y_{t+1}. Formally, the MSFE is defined as

$$\begin{aligned} &\text{MSFE} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2, \\ &\text{where } \hat{y}_t = \sum_j \Pi_{ij} \cdot \mu(x_t = j), \text{ and } i = \underset{i\in\{1,\dots,m\}}{\operatorname{argmin}} |y_{t-1} - \mu(x_{t-1} = i)| \end{aligned} \tag{30}$$

Note, the accuracy of the MSFE captures both the effect of an accurate transition matrix and grid point placement.

2.8 Application to linear Gaussian processes

We apply our method to a large number of stochastic processes and compare its performance to existing methods. For the AR(1) process with Gaussian innovations, the AR(1) process with fat tails and the VAR process, these results are summarized in Appendix B. In the applications that follow below, we will apply our method to an AR(1)-SV process, and the earnings processes of Guvenen et al. (2021) and Arellano et al. (2017).

3 Application I: asset pricing model with stochastic volatility

In this section, we evaluate the performance of our method in an asset pricing model with stochastic volatility. Most models that involve solving a dynamic stochastic optimization problem with a continuous-support process do not have a closed-form solution. This makes analyzing the implications of the choice of the discretization method hard due to the lack of a benchmark solution. However, as shown by De Groot (2015), the model we present below does have a closed-form solution for the price-dividend ratio and the conditional expected return on equity. This allows us to compare the true value of those statistics with the solutions from a model with a discretized process.

The first subsection will present the analytically tractable asset pricing model of De Groot (2015). Next, we demonstrate how to discretize the AR(1)-SV process in the De Groot (2015) model using our and two other methods, and analyze their respective performances at capturing various moments of the stochastic process. Finally, we assess how the numerical solutions with each methods perform relative to the analytical benchmark solution of De Groot (2015).

3.1 Analytical benchmark

We use the Lucas tree asset pricing model of De Groot (2015), who derives closed-form solutions for the price-dividend ratio and conditional expected return on equity when dividend growth is assumed to follow an autoregressive process with stochastic volatility.

As in De Groot (2015), there is a representative agent maximizing her discounted stream of utility:

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\sigma}}{1-\sigma} \tag{31}$$

$$\text{s.t. } c_t + s_{t+1} p_t \leq (d_t + p_t) s_t, \tag{32}$$

where c_t is consumption, s_t an asset with price p_t and dividends d_t. Parameter $\beta \in (0, 1)$ denotes the discount factor and σ is the coefficient of relative risk aversion.

The growth rate of dividends $y_t = \ln(d_t/d_{t-1})$ is assumed to follow an AR(1) process with stochastic volatility:[3]

$$y_t = \bar{y} + \rho(y_{t-1} - \bar{y}) + \sqrt{\eta_t}\varepsilon_t \tag{33}$$

$$\eta_t = \bar{\eta} + \rho_\eta(\eta_{t-1} - \bar{\eta}) + \omega\varepsilon_{\eta,t}. \tag{34}$$

with persistence in levels $\rho \in (-1, 1)$, and ε_t is i.i.d. $N(0, 1)$. The random variable η_t is the time-varying conditional variance of dividend growth. Parameter $\rho_\eta \in (-1, 1)$ is the persistence of the stochastic volatility process, and $\varepsilon_{\eta,t}$ is also i.i.d. $N(0, 1)$. Market clearing, $s_t = 1$, implies that $c_t = d_t$. Defining the price-dividend ratio as $v_t := p_t/d_t$, the first order condition of the representative agent's maximization problem is given by:

$$v_t = \mathbb{E}_t \beta \left(\frac{d_{t+1}}{d_t}\right)^{1-\sigma} (v_{t+1} + 1). \tag{35}$$

[3]Note that this is not a desirable way to define an AR(1)-SV process, given that η_t can become negative, in which case $\sqrt{\eta_t}$ is imaginary. In the parametrization we use, taken from Bansal and Yaron (2004), the probability of a negative value for η is very small. However, to demonstrate our performance in discretizing the more-commonly encountered specification of an AR(1)-SV process, we also discretize a different specification of the AR(1)-SV process in Appendix B.

De Groot (2015) derives a closed-form solution for the price-dividend ratio v_t and the conditional expected return on equity defined as:

$$\mathbb{E}_t R^e_{t+1} = \mathbb{E}_t \left(\frac{d_{t+1} + p_{t+1}}{p_t} \right), \tag{36}$$

which we provide in more detail in Appendix Section C.

Instead of using the continuous-support process in Equations (33)-(34), one can assume y_t follows a discrete Markov process, and obtain exact solutions for the price-dividend ratio, the conditional expected return on equity, and other objects of interest. The expressions for the solution to the discrete model are also provided in Appendix Section C.

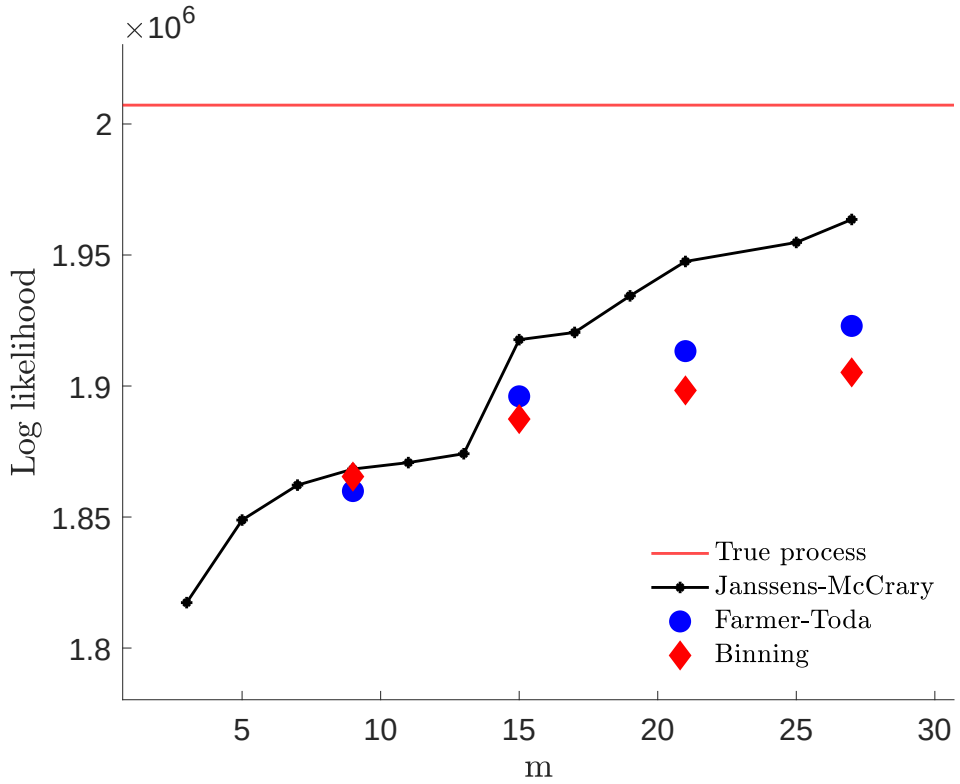
3.2 Discretizing the AR(1)-SV process of De Groot (2015)

The process of Equations (33)-(34) is multivariate, so when applying our method we discretize over y_t and η_t jointly. First, we demonstrate how our method selects the optimal grid and the number of grid points. Figure 2 visualizes the log likelihood of the misspecified HMM for different choices of grid size m.

Figure 2 looks different from the elbow plot of an AR(1) in Figure 1. In particular, we see a jump in the log likelihood when going from 13 to 15 grid points. The increase of the log likelihood can be understood from looking at the optimal grids that our method selects, which are visualized in Figure 3. Figure 3(a) shows the optimal grid for $m = 11$ grid points, and shows how our method assigns grid points in the tails with higher variances than in the center. This is consistent with the intuition behind an AR(1)-SV process, as it is more likely to end up at a higher absolute value of y_t with a higher realization of the variance η_t. As m becomes 15 or higher, our optimal grid adds what we call 'double' or 'triple' states. These are grid points with very similar levels for y, but different values for the variance η. These grid points will have different dynamics to next period's states, as will be reflected by different rows in the transition probability matrix for these states.

Figure 2 also visualizes the log likelihood of the existing methods when we use the grid and transition probability matrix we obtain from them as constraints for our maximum likelihood estimation of the HMMs, and compute the log-likelihood that corresponds to their discretization. As can be seen, once $m = 15$, the information loss of existing methods relative to ours increases, and when we give their methods 27 grid points ($m_y = 7,\ m_h = 3$), we can achieve the same information loss with 15 only grid points.

Figure 2: Elbow plot for the AR(1)-SV process in Equations (33)-(34).



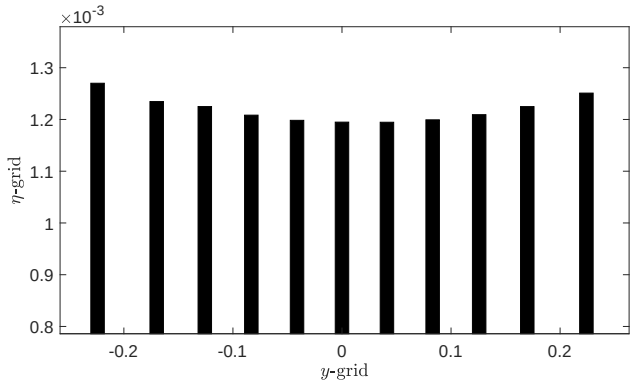
Parameters taken from Bansal and Yaron (2004): $\sigma = 1.5$, $\rho_\eta = 0.855$, $\omega = 7.4000 \times 10^{-5}$, $\bar{\eta} = 0.0012$, $\beta = 0.95$, $\rho = 0.868$, $\bar{y} = 0.0179$. Based on $T = 200,000$. We only visualize a selected number of grid points for the binning method and the Farmer-Toda method, because their methods rely on a tensor grid, and cannot be computed for any choice of m. For their methods, we keep the dimension of η, denoted m_η, fixed to three, and increase the dimension of y, denoted m_y.

Table 1: Comparison for an AR(1) process with stochastic volatility as in Equation (33)-(34) *($T = 200,000$) parametrized as in Bansal and Yaron (2004).*

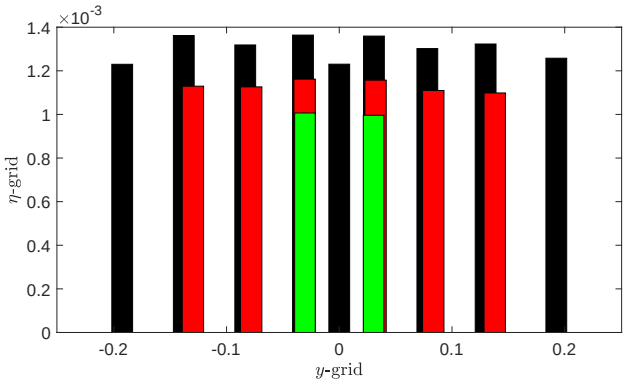
Method	Janssens-McCrary	Farmer-Toda	Binning
m = 15 ($m_y = 5$, $m_\eta = 3$ for Farmer-Toda and binning)			
Dev. uncond. mean y	0.018	**0.018**	0.018
% dev. uncond. variance y	−15.4	**0.525**	−23.7
% dev. autocorrelation y	3.35	**−0.053**	−5.66
Abs. dev. uncond. skewness y	**<0.001**	0.044	0.023
% dev. uncond. kurtosis y	**−9.42**	-19.2	−37.5
% abs. dev. cond. mean y	0.003	**0.002**	0.004
% abs. dev. cond. variance y	33.3	26.3	**26.0**
Abs. dev. cond. skewness y	**0.543**	1.16	0.580
% abs. dev. cond. kurtosis y	51.3	81.1	**18.5**
MSFE y	**0.0013**	0.0018	0.0017

Figure 3: Visualisation of optimal grid for two optimal grids ($m = 11, 17$), and the tensor grid following from the Farmer and Toda method, where the data generating process is an AR(1) process with stochastic volatility as in Equation (33)-(34).

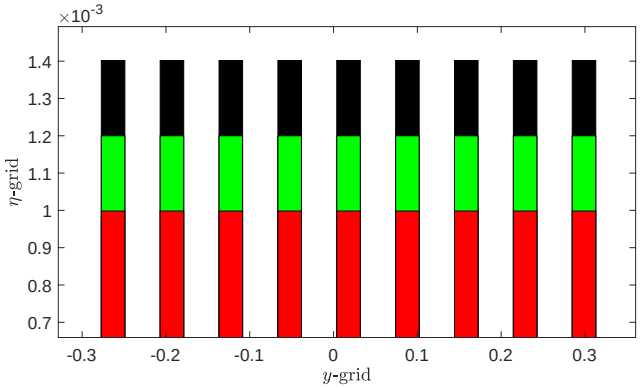
(a) $m = 11$



(b) $m = 17$



(c) Farmer-Toda method: $m = 27$ ($m_y = 9$, $m_\eta = 3$)



Parameters taken from Bansal and Yaron (2004): $\sigma = 1.5$, $\rho_\eta = 0.855$, $\omega = 7.4000 \times 10^{-5}$, $\bar{\eta} = 0.0012$, $\beta = 0.95$, $\rho = 0.868$, $\bar{y} = 0.0179$. Height of the bars depicts the variance level at the grid points, positioning on the x-axis of the bars depicts the level of y. The distinction between the red, green and black bars is to indicate that for m large, pairs of grid points arise where y has very similar values, but the value of the variance differs.

We compare our discretization with two existing methods, one being the method of Farmer and Toda (2017), and the other a simulation-based binning method based on the method by Adda and Cooper (2003).[4] Table 1 computes several statistics to compare the performance of our method and the existing methods at capturing moments of y. As can be seen, the Farmer and Toda (2017) method does well at the mean and variances, as these are the moments they match, while we do well at higher order moments such as the skewness and kurtosis. The MSFE of the other methods is 30-40% larger than ours, supporting that we do give an agent a better process to make forecasts with.

[4]We use the codes provided on the personal website of A.A. Toda, available at https://alexisakira.github.io/discretization/.

3.3 Accuracy of the models solutions

To compare the relative performance of our method versus existing methods at solving the asset pricing model, we compare moments of the discrete solutions to the analytical benchmark. To compute moments, we simulate a long time series of dividend growth y_t and variances η_t and compute, using the closed-form solutions of De Groot (2015), the implied price-dividend ratios v_t and expected returns on equity $\mathbb{E}_t R^e_{t+1}$ at each point in time. We compute several moments of these time series. Next, we also simulate data from the discretized processes, and compute the corresponding expected return on equity and price-dividend ratio using Equations (C.2) and (C.1).

To assess the accuracy of the different solutions, we compute the summary statistic

$$\log_{10}(|\hat{M}/M - 1|),$$

for different moments $\hat{M}$ of the time series simulated from the discretization and moments M computed from the simulation from the benchmark solution. Lower values of $\log_{10}(|\hat{M}/M-1|)$ indicate the moments of the discrete model are closer to those of the benchmark.

These results are summarized in Table 2. The parametrization used for the results in the table are based on the estimates of the stochastic volatility process in Bansal and Yaron (2004), annualized as in De Groot (2015). We choose risk aversion σ and the discount factor β such that the price-dividend ratio is finite.[5] Overall, our method always perform best at the mean, and often the variance of both statistics.

An object economists often care about is the welfare cost of risk, measured in terms of consumption. In the endowment economy considered in this application, we can measure the welfare cost of risk using the certainty equivalent consumption (CEC). This object is defined as follows:

$$V(d) = u(d) + \beta\mathbb{E}[V(d')|d],$$

where $V(d)$ is the value to the household of being in state d where d is the level of aggregate dividends. This value reflects the present discounted value of the risky dividend (i.e., consumption) stream. One could ask what the certainty equivalent level of consumption is that would make the household indifferent between the risky consumption stream and a certain (constant) level of consumption. We denote that constant value by $x(d)$, which is the solution

[5]De Groot (2015) provides an expression the parameters have to satisfy such that the price-dividend ratio is finite, we provided it in Appendix C.

Table 2: Accuracy of asset pricing model solutions for the price-dividend ratio v_t and the conditional expected return on equity $\mathbb{E}R^e_{t+1}$.

	M	$\log_{10}(\lvert\hat{M}/M - 1\rvert)$		
		Janssens-McCrary	Farmer-Toda	Adda-Cooper
		$m = 9$	$m = 3$ x 3	$m = 3$ x 3
Mean v_t	18.10	**−1.67**	−1.51	−1.13
Variance v_t	9.61	**−1.33**	−0.29	−0.07
Skewness v_t	0.59	−0.23	**−0.36**	−0.03
Kurtosis v_t	3.79	−0.32	**−1.01**	−0.22
Mean $\mathbb{E}_t(R^e_{t+1})$	1.08	**−3.10**	−2.37	−2.67
Variance $\mathbb{E}_t(R^e_{t+1})$	0.01	**−0.64**	−0.49	−0.28
Skewness $\mathbb{E}_t(R^e_{t+1})$	0.29	−0.28	**−1.02**	−0.09
Kurtosis $\mathbb{E}_t(R^e_{t+1})$	3.20	−0.42	**−1.55**	−0.28
		$m = 15$	$m = 5$ x 3	$m = 5$ x 3
Mean v_t	18.10	**−2.77**	−2.23	−1.29
Variance v_t	9.61	−0.65	**−2.33**	−0.19
Skewness v_t	0.59	−0.60	**−0.63**	−0.09
Kurtosis v_t	3.79	−0.54	**−0.58**	−0.27
Mean $\mathbb{E}_t(R^e_{t+1})$	1.08	**−4.51**	−2.44	−2.73
Variance $\mathbb{E}_t(R^e_{t+1})$	0.01	**−0.64**	−0.49	−0.28
Skewness $\mathbb{E}_t(R^e_{t+1})$	0.29	**−0.57**	−0.39	−0.16
Kurtosis $\mathbb{E}_t(R^e_{t+1})$	3.20	**−0.74**	−0.63	−0.36

Comparison of moments of simulated time-series from the discretized model solutions (denoted by $\hat{M}$) and the analytical closed-form model solution (denoted by M), such that the relative accuracy of the solution for moment M is measured by $\log_{10}(\lvert\hat{M}/M - 1\rvert)$. The lower (more negative) this value is, the closer this moment of the simulated time series of the discrete model solution is to the moment of time series from the exact model solution. Lowest values are marked in bold.

Parameters taken from Bansal and Yaron (2004): $\sigma = 1.5$, $\rho_\eta = 0.855$, $\omega = 7.4000 \times 10^{-5}$, $\bar{\eta} = 0.0012$, $\beta = 0.95$, $\rho = 0.868$, $\bar{y} = 0.0179$.

to:

$$V(d) = \frac{u(x(d))}{1 - \beta}.$$

We solve for $x(1)$ numerically by simulation using the true stochastic process for dividend growth and the discretized processes. Certainty equivalent consumption x measures the willingness to pay to remove all risk. Lower values of x indicate a higher willingness to pay, so to the extent the discretizations fail to capture risk, they will overstate x relative to the true value.

Table 3: Accuracy of asset pricing model solutions for the certainty equivalent of consumption (CEC): true value of CEC compared to those following from three different methods. The lower the percentage deviation, the closer the solution of the discretized model is to the truth. Different grid sizes are presented.

	Janssens-McCrary	Farmer-Toda	Adda-Cooper
CEC (true)	% dev	% dev	% dev
	$m = 9$	$m = 3$ x 3	$m = 3$ x 3
1.65	**0.76%**	8.28%	5.41%
	$m = 15$	$m = 5$ x 3	$m = 5$ x 3
1.65	**1.93%**	12.22%	3.95%

Lowest values are marked in bold. Parameters taken from Bansal and Yaron (2004): $\sigma = 1.5$, $\rho_\eta = 0.855$, $\omega = 7.4000 \times 10^{-5}$, $\bar{\eta} = 0.0012$, $\beta = 0.95$, $\rho = 0.868$, $\bar{y} = 0.0179$. The notation $m = a \times b$ indicates that y is discretized with a grid points and η is discretized with b grid points. This is because Farmer-Toda and Adda-Cooper make use of tensor grids. Janssens-McCrary does not. Average is taken over 50 simulations of the CEC.

In Table 3, we analyze the accuracy of the different methods when it comes to the computation of the CEC for two different grid sizes. As follows from Table 3, our method produces the most accurate estimates of the CEC, with deviations in percentage points 0.8-2% from the truth. The other two methods are at best 4% away from the truth, and at worst 12%, underestimating the amount of consumption the household is willing to give up to remove risk.

4 Application II: life-cycle model

In this section, we evaluate the quantitative implications of using different methods for consumption, wealth and welfare through a life-cycle consumption-savings model. While simple, our model forms the basis for most of the heterogeneous agent quantitative macro literature, so we would expect our results on the importance of accurate discretization to hold in richer models. In addition, we use this section to demonstrate how our method can be applied to non-linear non-Gaussian processes with life-cycle dynamics where the parameters of the discrete process are allowed to vary over time.

We consider the discretized version of the GKOS and ABB earnings processes. We first discuss the life-cycle model we will use in our analysis. Next, we discuss the two stochastic processes, our performance at discretizing these processes, and what the implications are for the model solutions, using our and existing methods.

4.1 Model and calibration

We begin by discussing the model environment, followed by the household optimization problem, and the details of the calibration strategy.

Environment. We consider a partial equilibrium life-cycle version of the canonical incomplete-markets model without aggregate uncertainty. Households live up to T periods, where the first $t < T_r$ are spent working and the remaining periods are spent in retirement. Working households supply one unit of labor inelastically with pre-tax earnings $\mathfrak{e}_t$ that evolve stochastically as described in more detail below. Retired households receive pension b and die with probability $1 - s_t$ each period. Asset markets are incomplete. Agents can borrow and save via an uncontingent bond, at risk-free interest rate r, up to an exogenous borrowing limit $\underline{a}$.

Household problem. At every age, agents choose consumption c and saving a' subject to the budget constraint which depends on the current state of assets a and earnings $\mathfrak{e}$. During their working age $t < T_r$, households solve the following optimization problem:

$$V_t(a, \mathfrak{e}) = \max_{c, a'} \left\{ u(c) + \beta \mathbb{E}_t V_{t+1}(a', \mathfrak{e}') \right\},$$
$$\text{s.t. } c + a' = \tau(\mathfrak{e}) + (1 + r)a$$
$$a' \geq \underline{a},$$

where earnings satisfy

$$\mathfrak{e}_t = g_t z_t.$$

That is, earnings in levels $\mathfrak{e}_t$ are the product of a common deterministic age component g_t and an idiosyncratic stochastic component z_t that evolves according to a (possibly age-dependent) Markov transition matrix Π_t. The specification for g_t is taken from Guvenen et al. (2021).

Retired households solve the following problem:

$$V_t(a) = \max_{c, a'} \left\{ u(c) + \beta s_t V_{t+1}(a') \right\},$$
$$\text{s.t. } c + a' = b + (1 + r)a$$
$$a' \geq \underline{a},$$

where b is a pension benefit paid to all retired households, and s_t is the probability of surviving from period t to $t + 1$.

Calibration. Agents enter the model at age 25 and work until age $T_r = 65$, after which they can be retired up to 25 years, so $T = 80$. If agents reach $T = 80$, they die with certainty. The exact year of death after retirement is stochastic, and the survival probabilities are from the Social Security Administration actuarial life table. Retirement benefit b is chosen to match the 45% replacement rate of average earnings, which is a good approximation of the system in the United States (Mitchell and Phillips, 2006).

Utility has CRRA form:

$$u(c) = c^{1-\sigma}/1 - \sigma.$$

The coefficient of relative risk aversion is set to 2. The risk free rate is 2% and the borrowing limit is 12% of average earnings, which De Nardi et al. (2020) find is roughly the ratio of credit cards limits to income in the Survey of Consumer Finances. The discount factor β is calibrated to match a wealth to income ratio of 3.1 for the working age population, and this will be re-calibrated for each process, and each discretization method we use for these income processes.

Following Benabou (2002) the labor income tax function has the form

$$\tau(y) = (1 - \chi)y^{1-\mu}. \tag{37}$$

The parameters χ and μ govern the level and progressivity of the tax function. Following Krueger and Wu (2021) we set the progressivity parameter to 0.1327, and the level parameter to 0.1575. The calibration is summarized in Table 4.

Table 4: Model parameters

Parameter	Description	Value	Motivation
σ	Risk aversion	2.0	De Nardi et al. (2020)
b	Retirement benefits	0.45	Mitchell and Phillips (2006)
r	Risk-free interest rate	0.04	De Nardi et al. (2020)
$\underline{a}$	Borrowing limit	-0.12	De Nardi et al. (2020)
μ	Income tax progressivity	0.1327	Krueger and Wu (2021)
χ	Income tax level	0.1575	Krueger and Wu (2021)
W/I	Wealth-to-income ratio	3.1	De Nardi et al. (2020)

Model statistics. After solving the model, we will report several statistics, such as covariances and variances of consumption, asset holdings and earnings over the life cycle. In addition, we compute two other statistics. First, we compute the certainty equivalent value (CEV). This is the fraction of lifetime consumption an individual would be willing to give up to live in a world without risk instead. Specifically let c^1 be the sequence of consumption arising in an economy with risk and c^0 be the sequence of consumption without risk. The CEV is defined in term of welfare as

$$W\big((1-CEV)c^0\big) = W\big(c^1\big)$$

that is, the fraction of consumption one would be willing to give up to remain in the economy without risk. Second, we report the partial insurance to persistent income shocks coefficient as in Blundell et al. (2008).

$$\psi^P_{\text{BPP}} = 1 - \frac{\text{cov}(\Delta c_{it}, y_{i,t+1} - y_{i,t-2})}{\text{cov}(\Delta y_{it}, y_{i,t+1} - y_{i,t-2})}.$$

This statistic measures the extent to which consumption adjusts to unpredictable persistent changes in income. This is a statistic that is sometimes used to calibrate life-cycle models, in which case this statistic is computed both for actual data and for the model. It is therefore important to know how sensitive this coefficient is to the discretization method used for the earnings process.

4.2 Discretizing Guvenen, Karahan, Ozkan and Song (2021)

We now consider the earnings process of Guvenen et al. (2021). We first discuss the details of the process, then how each method performs at discretizing this process in terms of capturing the different moments over the life-cycle.

The GKOS earnings process is given by:[6]

$$\begin{aligned}
y^i_t &= (1-\nu^i_t)e^{(z^i_t+\varepsilon^i_t)} \\
z^i_t &= \rho z^i_{t-1} + \eta^i_t \\
z^i_0 &\sim N(0, \sigma_{z_0}) \\
\eta^i_t &\sim \begin{cases} N(\mu_{\eta,1}, \sigma_{\eta,1}) & \text{with prob. } p_z \\ N(\mu_{\eta,2}, \sigma_{\eta,2}) & \text{with prob. } 1-p_z \end{cases}
\end{aligned} \tag{38}$$

[6]Note that we leave out the non-stochastic elements of the income-level, such as the fixed-effect. For the estimated values of the parameters, we refer to their paper.

$$\varepsilon_t^i \sim \begin{cases} N(\mu_{\varepsilon,1}, \sigma_{\varepsilon,1}) & \text{with prob. } p_\varepsilon \\ N(\mu_{\varepsilon,1}, \sigma_{\varepsilon,2}) & \text{with prob. } 1 - p_\varepsilon \end{cases}$$

$$v_t^i \sim \begin{cases} 0 & \text{with prob. } 1 - p_v(t, z_t^i), \\ \min\{1, \exp(\lambda)\} & \text{with prob. } p_v(t, z_t^i) \end{cases}$$

$$p_v^i(t, z_t) = \frac{e^{\xi_t^i}}{1 + e^{\xi_t^i}}, \quad \text{where } \xi_t^i \equiv a + bt + cz_t^i + dz_t^i t.$$

Here y_t^i is the income level of individual i at time t, z_t^i is the persistent component of income, ε_t^i is the transitory component and v_t^i is a non-employment indicator. The process can be interpreted as an extended persistent-transitory earnings process, where the main novelties are (i) the fat-tailed innovations to the persistent and transitory component, and (ii) the non-employment shocks v_t.

We use a multivariate discretization on $\log(y_t^i + 1)$ and z_t^i jointly. We allow the grid and transition probabilities of our discretization to be age-dependent without any restrictions. We use nine grid points, because, as one can see below in Section 4.2.1, nine grid points is sufficient to capture the process well.[7]

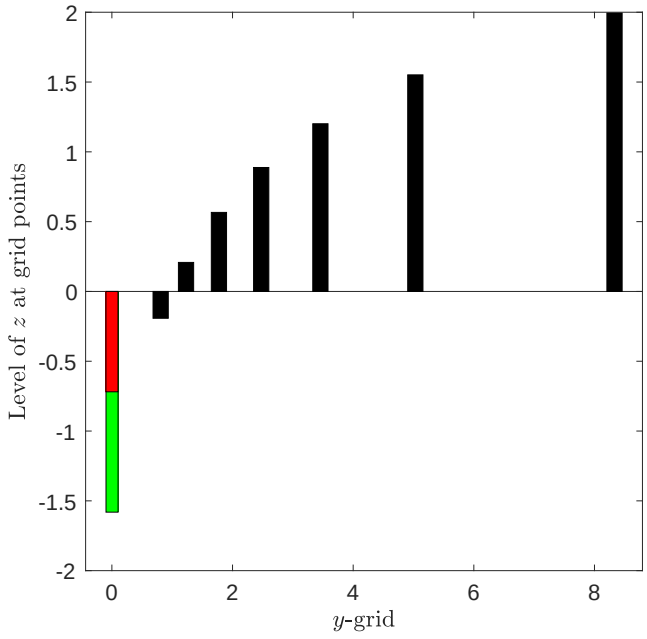
The optimal time-varying grids are visualized in Figure 4. The panel on the right displays how the grid points have a positive trend in age, and this is most notable in the grid point with the highest income level. This allows the right tail of the income distribution to get fatter over time if the stationary distribution over the grid points is reasonably constant across ages. As can be seen, the discretization method shows that it is optimal to have multiple zero-income states. In the case of $m = 9$, two unemployment states is optimal. As the number of grid points increases, the method adds grid points further in the tails, as well as more zero-income states (not visualized here). Having multiple states with an income level of zero results in heterogeneity in job-finding probabilities.

Figure E1 in the Appendix visualizes the age-dependent transition probability matrix. The first two rows represent the zero-income states, and by looking at the diagonal, we can see that these states differ in terms of their persistence, and that this persistence changes with age. The first row represents a less persistent non-employment state than the second row. The persistence of this state does increase 40 percentage points over the life-cycle, and at the end

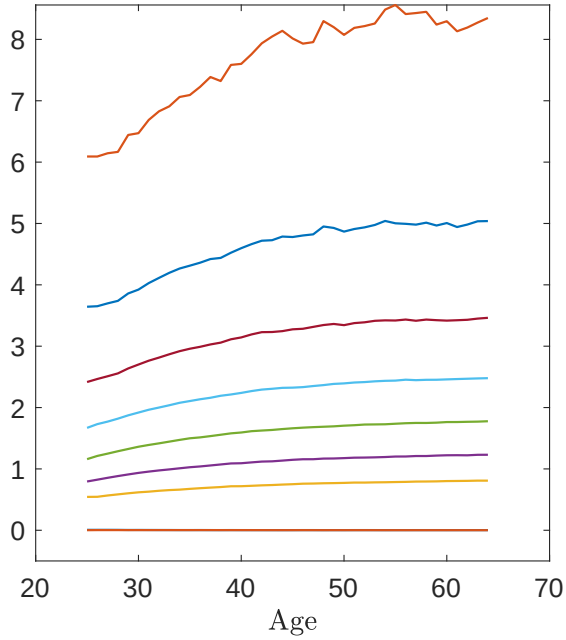
[7] Guvenen et al. (2021) remark that many grid points would be required to appropriately discretize their income process. Our ability to discretize the process well with only nine grid points comes from the optimal grid that follows from our method, and the choice of variables to which we apply the discretization method.

Figure 4: Visualisations of the optimal grid of the discretization of the stochastic process in Guvenen et al. (2021) with $m = 9$.

(a) Optimal grid at age $T = 65$.



(b) Optimal grid over the life-cycle



of the working life, ending up in this non-employment state implies an 80% probability of still being there next period. The second non-employment state is highly persistent already from the beginning of the life-cycle, and towards the end of life the only "escape" out of this state is by transitioning into the other non-employment state. This suggests that in the Guvenen et al. (2021) process, non-employment becomes an almost absorbing process towards the end of working life. Here it should be noted that Guvenen et al. (2021) do not differentiate between unemployment and non-employment, which explains why these transition probabilities out of the zero-income states are different from those we know from the unemployment duration literature.

4.2.1 Comparison between methods

To the best of our knowledge, this paper is the first to discretize the process in Guvenen et al. (2021). To demonstrate the performance of our method, we propose two different binning methods: (i) "simple" binning and (ii) "clever" binning. Both methods are explained in more detail in Appendix Section D. The simple binning method determines a grid with a non-employment state and then bins on quantiles for all $y > 0$, and computes the transition probabilities between these bins based on simulation. The clever binning method comes

down to having a equal-quantile-based tensor grid between the persistent component z, the employment status v and the transitory component ε. Just as in our discretization, this will generate heterogeneous transition probabilities out of non-employment.

Figure 5 displays how the unconditional moments of the earnings levels in the process in Equation (38) vary by age, and the extent to which our discretization method and the existing methods can replicate this. As these figures show, our discretization method captures both the unconditional mean, variance, skewness and kurtosis of the earnings levels well, and does so better than the two binning methods. The simple binning method performs better than the clever binning method in this dimension.

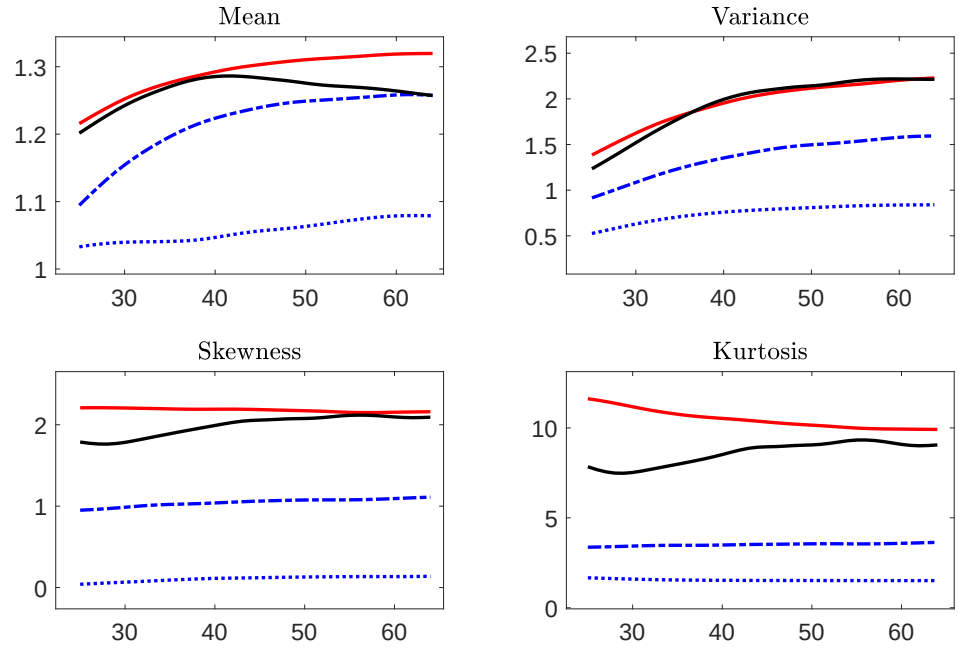
Figure 5b also shows the unconditional moments of the arc-changes in earnings y, $\frac{y_{i,t+1}-y_{i,t}}{(y_{i,t+1}+y_{i,t})/2}$, a statistic the paper by Guvenen et al. (2021) focuses on. Looking at the unconditional moments of $\frac{y_{i,t+1}-y_{i,t}}{(y_{i,t+1}+y_{i,t})/2}$ presented here, we see that the simple binning method performs best at these statistics, but we do perform better at capturing the variance, skewness and kurtosis over the life-cycle than the clever binning method. The simple binning method performs well at this statistic by construction, given that it matches period-to-period transitions.

In Figure 6, the life-cycle development of the non-employment dynamics are visualized for the three different discretizations employed. As can be seen, our discretization is able to both match the levels and the development of the two-period and three-period ahead conditional non-employment probabilities over the life-cycle better than both binning methods. The clever binning method performs better at these moments than the simple binning method because of its use of a tensor grid in the non-employment status. The simple binning method performs well at the one-period-ahead persistence of non-employment by construction, but fails to capture the long-run non-employment dynamics.

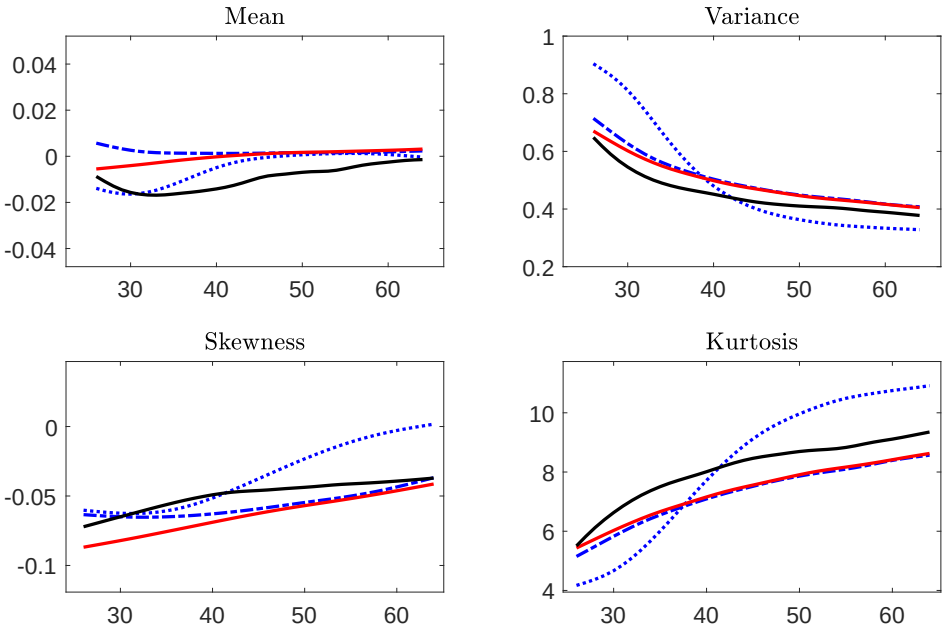
Overall, we conclude that the simple binning method captures some moments well by construction, but misses out on the longer-term non-employment dynamics as well as the cross-section of earnings. Our method's discretization performs consistently well across all moments, and, because of that, captures the overall riskiness of the process better than the two binning methods. Particularly capturing the longer-term non-employment risk well should matter for the estimates of the welfare cost of risk. The tensor-based clever binning method is outperformed in all dimensions by our method, most likely because nine grid points is not sufficient when using tensor grids.

Figure 5: Age-dependent moments, for three different discretizations with $m = 9$ grid points of the stochastic process by Guvenen et al. (2021).

(a) Unconditional moments of earnings y

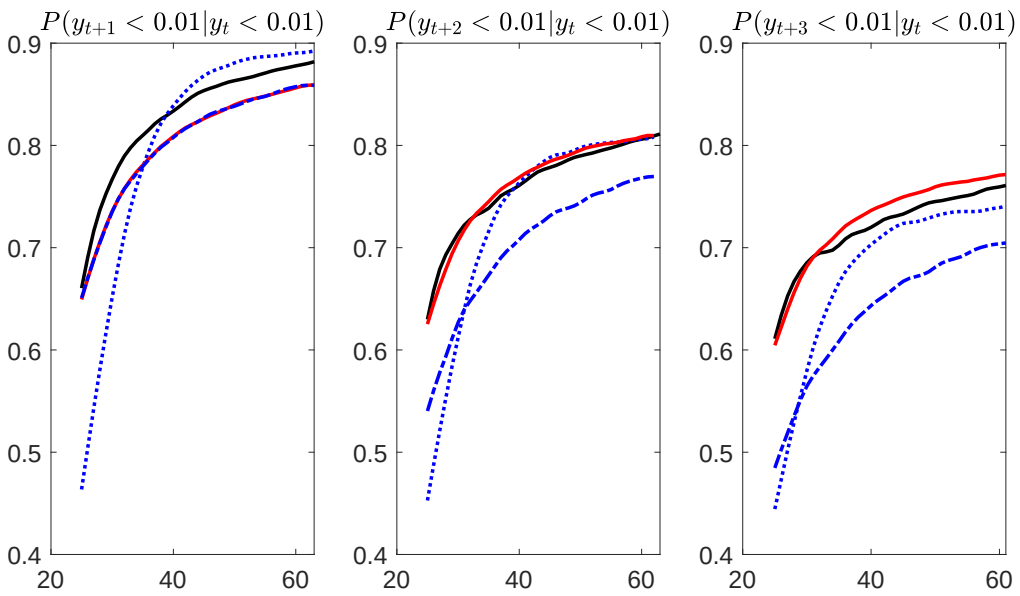


(b) Unconditional moments of arc changes in earnings $\frac{y_{i,t+1}-y_{i,t}}{(y_{i,t+1}+y_{i,t})/2}$



Solid red line represents the continuous-support process, solid black line is our discretization method, blue dotted line is the clever binning method, the blue dash-dot line is the simple binning method.

Figure 6: Non-employment dynamics for three different discretizations of the Guvenen et al. (2021) process with $m = 9$ grid points.



Solid red line represents the continuous-support process, solid black line is our discretization method, blue dotted line is the clever binning method, the blue dash-dot line is the simple binning method.

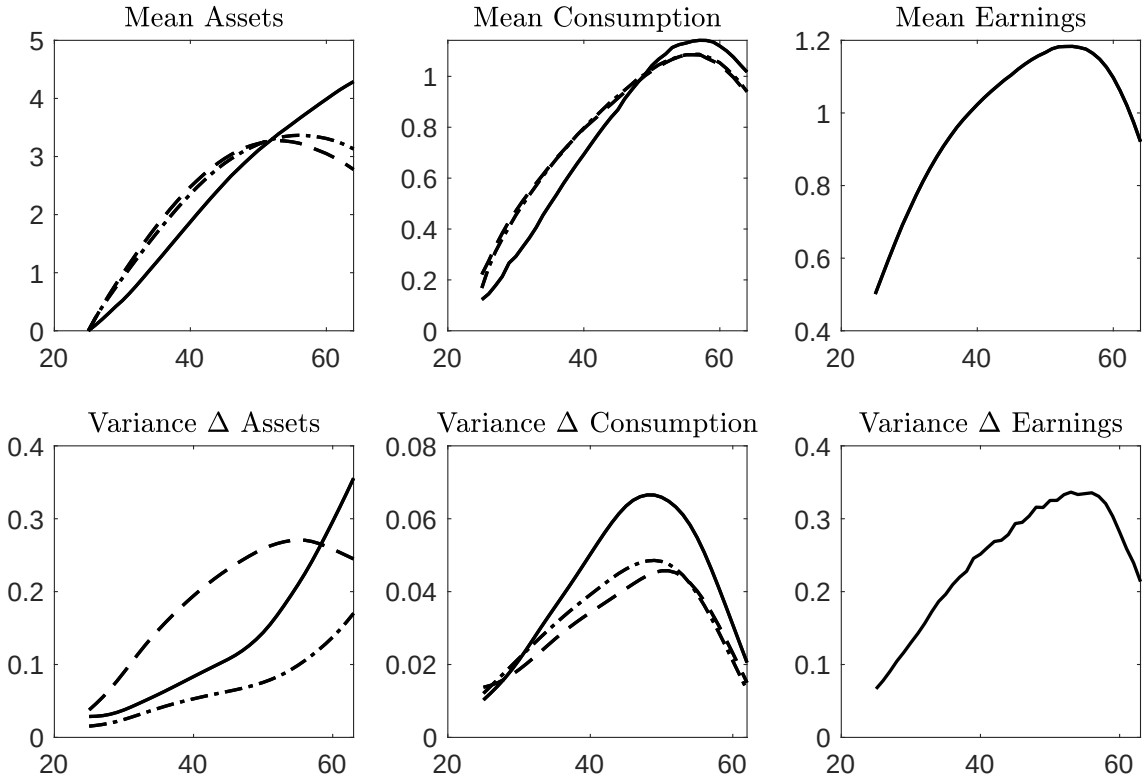
4.3 A life-cycle model with the earnings process of Guvenen et al. (2021)

Next, we illustrate the importance of the choice of the discretization method for the earnings process of Guvenen et al. (2021) through the lens of the life-cycle model described above. Figure 7 shows how the choice of the discretization method matters for the model solution, and leads to different implications of how assets and consumption develop over the life cycle of an individual.

The discretization method matters for how the variance in the change in assets from one period to another varies over the life cycle. At the end of the working life, the variance in asset changes is more than twice as large for our discretization method than for the simple binning method. For the variance in consumption changes we see a similar pattern, but this statistic peaks around age 45-50 and goes down again towards the end of the life-cycle. However, the variance of consumption changes implied by our method is again higher than that of the binning methods.

Table 5 summarizes some additional statistics computed from the life-cycle model solutions. Looking at the certainty equivalent (CEV) in this table, the lifetime consumption an individual

Figure 7: Simulations from the life-cycle model for three different discretizations of the earnings process of Guvenen et al. (2021). Assets, consumption and earnings over the life-cycle (age on the x-axis).



Solid line represents our discretization method, the dashed line is the clever binning method, the dashed-dot line is the simple binning method. The last panel depicts the continuous-support earnings process. Data is simulated on the grid, and the solution is computed using the four different discretizations.

Table 5: Summary statistics computed from simulations from the life-cycle model for three different discretizations of the earnings process of Guvenen et al. (2021). $m = 9$ is used for the discretization. Rescaled such that mean income for all processes is the same.

Method	Janssens-McCrary	Clever binning	Simple binning
Variance c_{it}	0.808	0.295	0.427
Variance Δc_{it}	0.043	0.031	0.033
Variance a_{it}	11.065	2.986	4.749
Variance Δa_{it}	0.139	0.205	0.078
Covariance c_{it}, y_{it}	0.873	0.297	0.446
Covariance Δc_{it}, Δy_{it}	0.072	0.072	0.056
CEV	0.921	0.683	0.752
ψ^P_{BPP}	0.378	0.394	0.431

would be willing to sacrifice in order to remove all risk, we see that this value varies across discretization methods. The CEV that follows from a model solved with our discretization method is higher than when using a binning-based discretization. This is not surprising given that we capture the long-term non-employment dynamics better – an important source of risk in this model – as well as the overall cross-sectional distribution of earnings. The difference of 17 percentage points is large considering typical policy experiments only cause the CEV to change by 1-2 percentage points.

Note that a CEV of 0.92 is a large number. This comes from the highly-persistent non-employment state in the stochastic process of Guvenen et al. (2021) that the individuals in our life-cycle model, as we will demonstrate in the next paragraphs. In reality, non-employment is – in part – a labor supply choice, and part involuntary unemployment. The Guvenen et al. (2021) process does not distinguish between the two. Therefore, not all aspects of this process are truly exogenous, and the CEV overestimates the actual earnings risk in this economy. Re-estimating the process in Guvenen et al. (2021) such that v_t only captures involuntary non-employment goes beyond the scope of this paper, and our goal is merely to demonstrate the ability of our method to discretize highly non-standard processes such as Guvenen et al. (2021), and, furthermore, that the discretization method one chooses matters for the welfare implications in a life-cycle model.

4.3.1 Main sources of risk in Guvenen et al. (2021)

In this subsection, we use our discretization method to study what the main sources of risk are in the earnings process of Guvenen et al. (2021). As discussed by Guvenen et al. (2021), there are two main features of the earnings process that are important in matching several moments of the earnings data they use, being (i) non-employment shocks and (ii) innovations to earnings being drawn from a normal mixture. We analyze the relative contribution of these features to the overall risk an individual faces by shutting down the non-employment shocks.

We compute the certainty equivalent for the discretized process of Guvenen et al. (2021) (i) as in Equation (38) and (ii) when $v_t^i = 0$ for all i, t. These results are summarized in Table 6. As can be seen, the non-employment risk is the most important source of risk individuals face, and removing this risk decreases the CEV from 0.92 to 0.51 when using our discretization method, and from 0.75 to 0.34 when using simple binning with the same number of gridpoints. In both cases, the discretization based on simple binning estimates the CEV to be 17 percentage points lower than our discretization. Adding more grid points to the binning method increases its CEV when considering the process without non-employment risk, but does not help as

Table 6: Certainty equivalent (CEV) for a variation on the earnings process of Guvenen et al. (2021). Rescaled such that mean income for all processes is the same.

	Guvenen et al.	Guvenen et al. No nonempl. risk
CEV JM $m = 9$	0.921	0.505
CEV simple binning $m = 9$	0.752	0.339
CEV simple binning $m = 60$	0.775	0.413

Note that we do not report CEV's for the clever binning method when there is no non-employment risk, because this method takes a tensor grid in the non-employment state v_t.

much in the presence of non-employment risk. This insight is comparable to what we find in Appendix Section F, where Table F2 shows that a binning-based method underestimates the CEV of a process with fat tails.

4.4 Discretizing Arellano et al. (2017)

Next, we discuss how we discretize the nonparametric earnings process in Arellano et al. (2017). As in Arellano et al. (2017), let Y_{it} denote $\log y_{it}$, where y_{it} is unexplained pre-tax labor earnings. Decompose Y_{it} as follows:

$$Y_{it} = \eta_{it} + \varepsilon_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T.$$

Here η_{it} denotes the persistent component and ε_{it} denotes the transitory component. The transitory component has mean zero and is independent over time and from all η_{is}. The persistent component η_{it} follows a general first-order Markov process, with its τth conditional quantile given $\eta_{i,t-1}$ by $Q_t(\eta_{i,t-1}, \tau)$ for each $\tau \in (0, 1)$, i.e., without loss of generality:

$$\eta_{it} = Q_t(\eta_{i,t-1}, u_{it}), \quad (u_{it}|\eta_{i,t-1}, \eta_{i.t-2}, ...) \sim \text{Uniform}(0, 1), \quad t = 2, ..., T$$

This model allows for nonlinear dynamics of earnings, and in particular, generates nonlinear persistence:

$$\rho_t(\eta_{it-1}, \tau) = \frac{\partial Q_t(\eta_{i,t-1}, \tau)}{\partial \eta}, \quad \rho_t(\tau) = \mathbb{E}\left[\frac{\partial Q_t(\eta_{i,t-1}, \tau)}{\partial \eta}\right]$$

ρ measures persistence of earnings histories and individuals with similar income levels but different values of ρ may have very different expectations over their next period's earnings. Arellano et al. (2017) estimate this model non-parametrically, approximating Q using low-

order products of Hermite polynomials and limiting time-dependence to age-dependence, i.e.,

$$Q_t(\eta_{i,t-1}, \tau) = Q(\eta_{i,t-1}, \text{age}_{it}, \tau) = \sum_{k=0}^{K} a_k^Q(\tau)\phi_k(\eta_{i,t-1}, \text{age}_{it}).$$

One advantage of our method is that it only requires a simulated sample from the true stochastic process, therefore it can also be applied non-parametric methods as those by Arellano et al. (2017). We focus on the discretization of η_{it} only, as the transitory component ε_{it} is i.i.d. over time. The simulated values from the stochastic process are noisy, so we truncate the data at four age-dependent standard deviations about the mean.[8] We use 14 grid points, because, as we show below, that is sufficient to be able to capture the process well. The grids and transition probability matrices vary at each age. We visualize the time-varying transition probabilities and grid in Figures E2 and E3 in the Appendix.[9]

4.4.1 Comparison across methods

We compare the performance of our discretization method with the method De Nardi et al. (2020) propose to discretize the Arellano et al. (2017) process. In particular, their method adapts Adda and Cooper (2003) and uses simulation-based binning for both the persistent component η_{it} and the transitory component ε_{it} and then uses a tensor grid to obtain a discretization for y_{it}. The innovation of their method is to add bins in the tails of the process. Note that their discretization originally was applied to a re-estimated version of Arellano et al. (2017) that uses after-tax earnings. Their discretization for η_{it} uses 18 grid points. For details we refer to their paper.

Figure 8 visualizes the moments of η_t and $\Delta\eta_t$ for the Arellano et al. (2017) process, our discretization and the binning method of De Nardi et al. (2020) that we below will refer to as 'tail binning'. As can be seen, our discretization method does a good job at capturing the mean, skewness and kurtosis of the levels of η_t. For the variance, our method is better at the younger ages, while the tail binning method does better at ages 50 and higher. Overall, the tail-binning method misses the gradual increase in skewness and kurtosis over the entire life cycle, and instead catches up by rapidly increasing around age 45-50.

For the differences $\Delta\eta_t$, we observe that our method and the tail binning method perform similarly well at the mean, but our performance is better at the variance, skewness and

[8] For the simulations from their earnings process, we use the publicly-available codes that accompany their publication.

[9] $m = 10$ is chosen here for readability.

kurtosis. We do still underestimate the excess kurtosis and skewness compared to the true process, but less than the tail binning method.

4.5 A life-cycle model with the ABB earnings process

Next, we use the earnings process of Arellano et al. (2017) in our life-cycle model. Figure 9 visualizes how the mean of assets and consumption and the variance of changes in assets and consumption evolve in the models solved using the two different discretizations. As can be seen, the development of the mean over the life-cycle is fairly similar, but there are considerable differences in the variance of asset and consumption changes. The differences can be well-explained by the differences we observe in Figure 8, where we visualize how our and the tail-binning method by De Nardi et al. (2020) perform at capturing the moments of the underlying process. The tail-binning method fails to capture the gradual increase in skewness and kurtosis of the process over the lifecycle, but sees a rapid increase around age 50, and an overestimation of those moments at the end of the life-cycle. This is reflected in the variance of asset changes that follow from their discretization, which is consequently too low at the earlier ages, and too high at the ages above 55. A similar pattern is visible for consumption changes.

Table 7 summarizes several other moments simulated from the life-cycle model, and we see that our solution generates more asset and consumption inequality than the tail-binning-based solution. We also compute the welfare cost of risk, the CEV, and find that ours is 0.233, while the tail-binning method suggests a CEV of 0.166, implying we get an estimate that is 6.7 percentage points, that is 40%, larger.

We believe that these differences can also explain the results found in De Nardi et al. (2020), concluding that the CEV of a canonical earnings process is higher than the highly-nonlinear earnings process of Arellano et al. (2017). Using a discretization method that better captures the development of the higher-order moments over the life-cycle seems to generate higher CEV estimates. We should add that the paper of De Nardi et al. (2020) uses a different process than we consider here, and that we have to be careful when extrapolating our conclusions here to the process considered in their paper.

Figure 8: Moments of η_t and $\Delta\eta_t$ for the process of Arellano et al. (2017). The red line is data simulated from the Arellano et al. (2017) process, the black line follows from our discretization method, and the blue dotted line is based on the tail binning methhod of De Nardi et al. (2020).

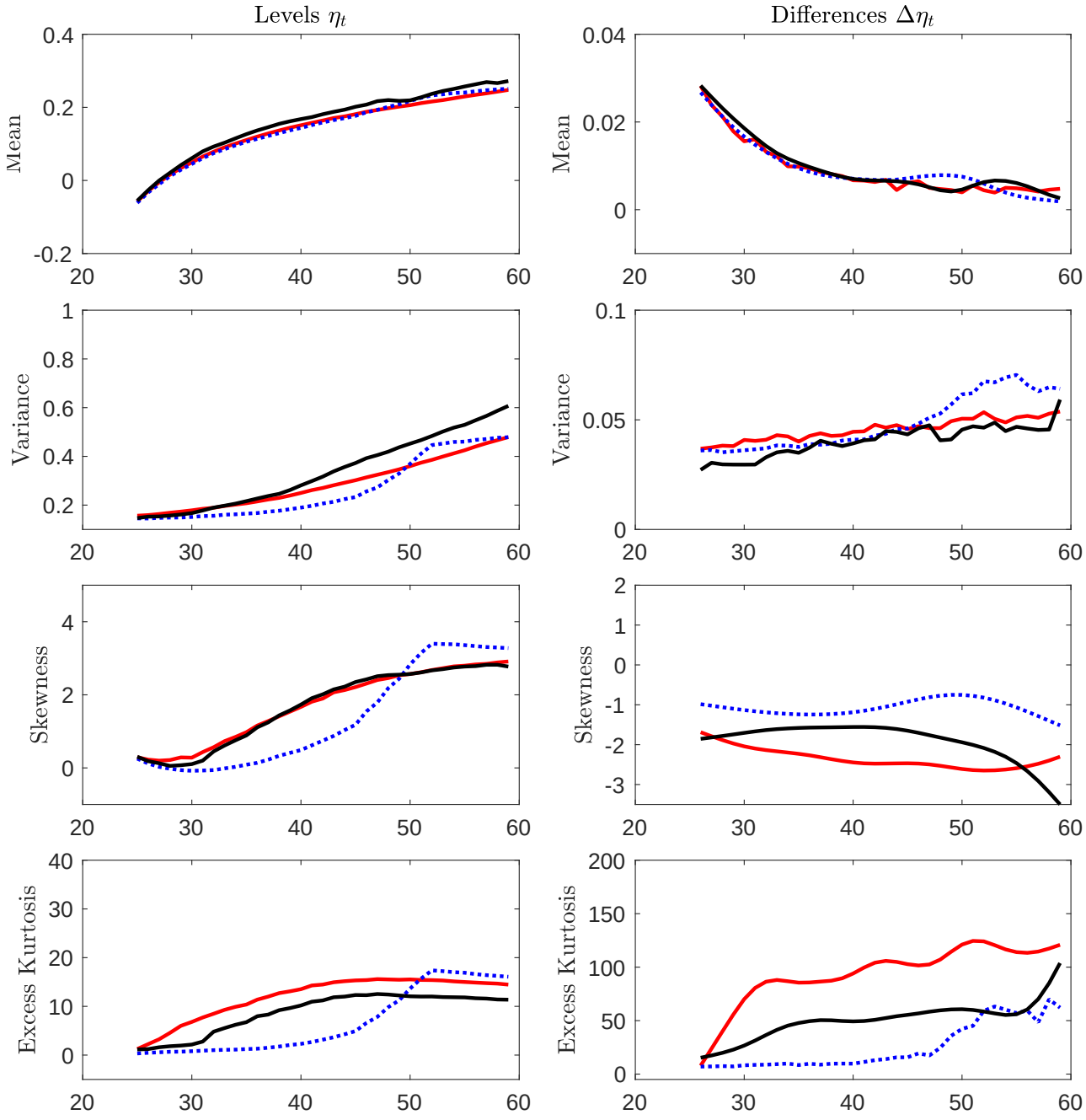


Figure 9: Two discretizations of Arellano et al. (2017) in a life-cycle model. Solid line is the model solution using our discretization, dashed line uses the tail-binning method of De Nardi et al. (2020).

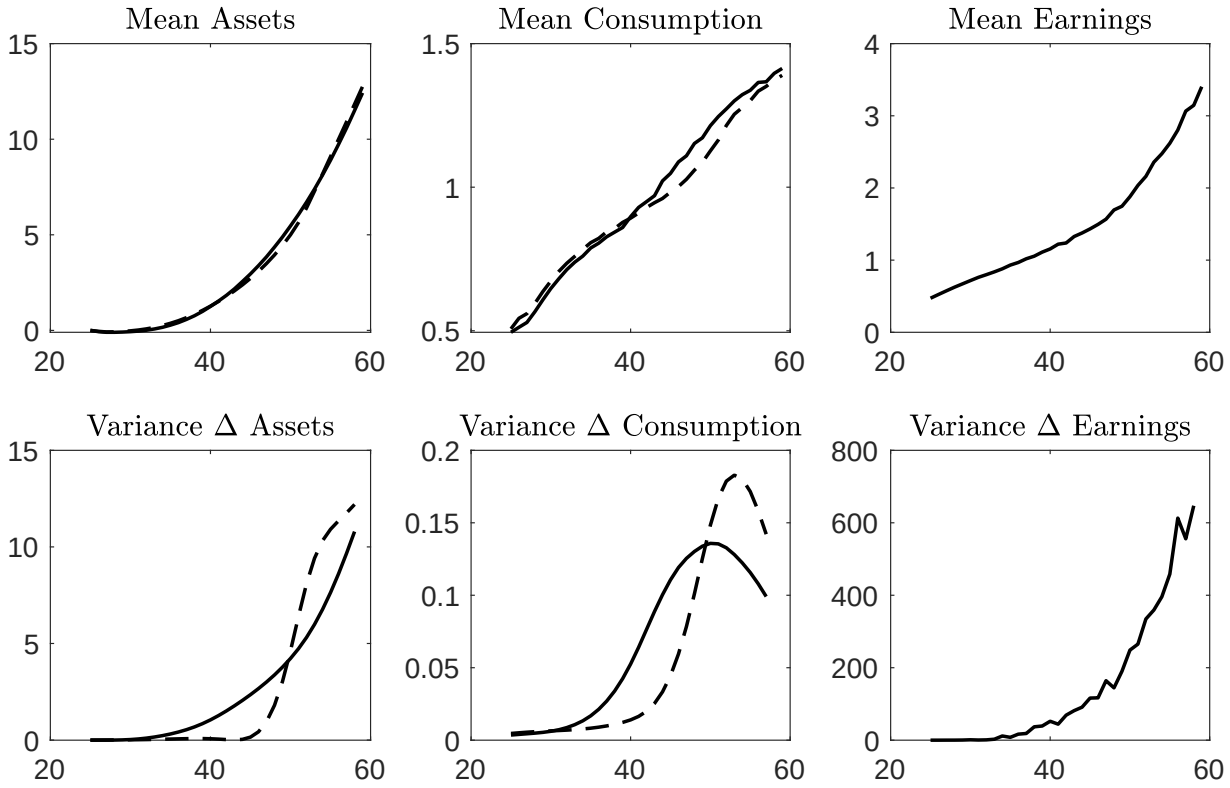


Table 7: Summary statistics computed from the life-cycle model for two different discretizations of the ABB earnings process. $m = 14$ is used for JM, binning uses 18.

Discretization method	Janssens-McCrary	Tail-binning
Variance c_{it}	1.939	0.988
Variance Δc_{it}	0.067	0.061
Variance a_{it}	240.3	139.7
Variance Δa_{it}	3.007	3.108
Covariance c_{it} and y_{it}	3.241	2.111
Covariance Δc_{it} and Δy_{it}	0.163	0.160
CEV	0.233	0.166
ψ^P_{BPP}	0.660	0.678

4.6 Comparing Guvenen et al. (2021) and Arellano et al. (2017) and their implications for wealth inequality

Our discretized processes allow for a direct comparison between the process of Guvenen et al. (2021) and Arellano et al. (2017). First of all, we note that the time-varying grids of Arellano et al. (2017), as visualized in Appendix Figure E3, has grid points much further in the tails than the process by Guvenen et al. (2021).[10] There are also considerable differences in the life-cycle dynamics of the transition probabilities between the two processes. All life-cycle dynamics of the transition probabilities in Guvenen et al. (2021) occurs in the non-employment and low-earnings states. In contrast, Arellano et al. (2017) is estimated on a sample of employed married males in the PSID, and the transition probabilities of the low-earnings states are fairly constant. The transition probabilities of the high-earnings states do vary over the life-cycle. As can be seen in Figure E2, the persistence of the highest earnings state increases over the life-cycle, while the persistence of the second and third-highest earnings state decrease.

Finally, we use the different discretizations of the processes of Guvenen et al. (2021) and Arellano et al. (2017) in the life-cycle model to analyze their implications for wealth inequality. Table 8 summarizes wealth shares of the United States as reported in Krueger, Mitman, and Perri (2016) and as follow from the life-cycle model for the different earnings processes and discretizations.

Table 8 compares the wealth distribution using our discretization of the ABB and GKOS processes in the life-cycle model versus existing methods. Note that a better discretization does not necessarily lead to a better model fit in terms of the wealth distribution, but our method generates wealth shares that are very similar to those observed in the Panel Study of Income Dynamics (PSID) and the Survey of Consumer Finances (SCF) data sets in 2006 and 2007. The tail binning discretization method proposed by De Nardi et al. (2020) overestimates the wealth shares of the lower quantiles, and underestimates the wealth shares of the higher quantiles. For example, it underestimates the wealth share of the top 20% by 9.7 percentage points, while our discretization is only 2.3 percentage points below the wealth share found in the PSID. For the top 1%, our discretization generates a wealth share that is only 1.4 percentage points above the share found in the SCF. The Gini coefficient that follows from our discretization is equal to the Gini coefficient Krueger et al. (2016) obtain from the SCF, while using the discretization method proposed by De Nardi et al. (2020) leads to a Gini coefficient that is 0.09 lower. The GKOS process does not generate as much income inequality as the Arellano et al. (2017) process, and therefore can not generate enough wealth inequality to match the wealth

[10] Note that the grid in Figure E3 still has to be exponentiated for a comparison with Figure 4b.

Table 8: Wealth inequality measures. Data on from Krueger et al. (2016).

	Data		Model + Guvenen		
% Share held by:	PSID, 06	SCF, 07	Janssens-McCrary	Clever binning	Simple binning
Q1	-0.9	-0.2	-1.1	-1.3	-1.2
Q2	0.8	1.2	-0.1	3.3	2.9
Q3	4.4	4.6	7.5	14.1	12.0
Q4	13.0	11.9	21.9	31.1	26.0
Q5	82.7	82.5	71.8	52.8	60.3
T1%	30.9	33.5	11.5	4.0	6.2
Gini	0.77	0.78	0.71	0.56	0.61
	Data		Model + ABB		
% Share held by:	PSID, 06	SCF, 07	Janssens-McCrary	Tail-binning	
Q1	-0.9	-0.2	-0.4	-0.4	
Q2	0.8	1.2	0.8	1.7	
Q3	4.4	4.6	5.5	8.2	
Q4	13.0	11.9	13.6	17.3	
Q5	82.7	82.5	80.4	73.0	
T1%	30.9	33.5	34.9	28.4	
Gini	0.77	0.78	0.78	0.69	

shares observed in household surveys. Our discretization of GKOS does generate more wealth inequality than the two binning methods. For example, we underestimate the share of the top 1% by about 19-22 percentage points, while the binning methods miss the wealth shares of the top 1% by 24-29 percentage points. Consequently, the Gini coefficient that follows from our method is 0.06-0.07 below the SCF estimate, and the other methods result in Gini coefficients that are even lower. It follows that if one were using the wealth distribution to assess the performance of the standard incomplete markets model, they could erroneously conclude the model provides a poor fit without an accurate discretization of the earnings process.

5 Conclusion

This paper proposes a novel discretization method for a large class of stochastic processes which provides both an optimally selected grid and transition probability matrix. The method is based on minimizing the information loss of an individual using our discretized process to make its decisions rather than the true continuous-support process. We compare its performance to existing methods such as those by Farmer and Toda (2017), Rouwenhorst (1995),

Tauchen (1986), and Adda and Cooper (2003). Our discretized process provides a closer fit to the continuous-support process, and, a lower mean squared forecast error than when using any of the other methods. We apply our method to a large set of continuous-support processes, and find that the gain of using our method is largest in case of multivariate processes such as the AR(1) process with stochastic volatility, and processes with a life-cycle component like those proposed by Arellano et al. (2017) and Guvenen et al. (2021).

We apply and compare our method in two main applications. The first application is an asset-pricing model with stochastic volatility, which, as shown by De Groot (2015) has a closed-form analytical solution. We use this analytical solution as our benchmark to the solutions based on different discretization methods. We find that our method results in numerical solutions closer to this benchmark, especially with regards to the welfare cost of risk.

We also evaluate the effect of the choice of discretization method on the solutions that follow from a life-cycle model with a variety of different earnings processes, and find that the method matters greatly for several statistics, such as the pass-through of income risk to consumption, and the welfare cost of risk. Given that both are commonly reported statistics, the fact that they depend so strongly on the choice of the discretization method is an important insight for applied modeling. Finally, we find the choice of the different method matters for the amount of wealth inequality that a model can generate, and we show that a discretization of Guvenen et al. (2021) and Arellano et al. (2017) using our method can generate wealth shares and other wealth inequality statistics closer to those observed in the data than when using other methods.

Discretized stochastic processes have many more applications than the asset-pricing and life-cycle models we use to benchmark out method. Our method's generality provides a tool to bridge the gap between reduced form statistical processes and applied quantitative modeling. We hope this opens the door to the use of richer statistical processes in structural economic models.

References

Adda, J., and Cooper, R. W. (2003). *Dynamic Economics: Quantitative Methods and Applications*. MIT press.

Altonji, J. G., Hynsjö, D. M., and Vidangos, I. (2022, May). *"individual earnings and family income: Dynamics and distribution"* (Working Paper No. 30095). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w30095 doi: 10.3386/w30095

Arellano, M., Blundell, R., and Bonhomme, S. (2017). "Earnings and Consumption Dynamics: a Nonlinear Panel Data Framework". *Econometrica*, *85*(3), 693–734.

Bansal, R., and Yaron, A. (2004). "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles". *Journal of Finance*, *59*(4), 1481–1509.

Benabou, R. (2002). "Tax and Education Policy in a Heterogenous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?". *Econometrica*, *70*(2), 481-517.

Blundell, R., Pistaferri, L., and Preston, I. (2008). "Consumption Inequality and Partial Insurance". *American Economic Review*, *98*(5), 1887–1921.

Civale, S., Díez-Catalán, L., and Fazilet, F. (2016). "Discretizing a Process with Non-Zero Skewness and High Kurtosis". *Available at SSRN 2636485*.

De Groot, O. (2015). "Solving Asset Pricing Models with Stochastic Volatility". *Journal of Economic Dynamics and Control*, *52*, 308–321.

De Nardi, M., Fella, G., and Paz-Pardo, G. (2020). "Nonlinear Household Earnings Dynamics, Self-Insurance, and Welfare". *Journal of the European Economic Association*, *18*(2), 890–926.

Douc, R., and Moulines, E. (2012). "Asymptotic Properties of the Maximum Likelihood Estimation in Misspecified Hidden Markov Models". *The Annals of Statistics*, *40*(5), 2697–2732.

Farmer, L. E., and Toda, A. A. (2017). "Discretizing Nonlinear, Non-Gaussian Markov Processes with Exact Conditional Moments". *Quantitative Economics*, *8*(2), 651–683.

Fella, G., Gallipoli, G., and Pan, J. (2019). "Markov-Chain Approximations for Life-Cycle Models". *Review of Economic Dynamics*, *34*, 183–201.

Galindev, R., and Lkhagvasuren, D. (2010). "Discretization of Highly Persistent Correlated AR (1) Shocks". *Journal of Economic Dynamics and Control*, *34*(7), 1260–1276.

Gordon, G. (2021). "Efficient VAR Discretization". *Economics Letters*, *204*, 109872.

Gospodinov, N., and Lkhagvasuren, D. (2014). "A Moment-Matching Method for Approximating Vector Autoregressive Processes by Finite-State Markov Chains". *Journal of Applied Econometrics*, *29*(5), 843–859.

Guvenen, F., Karahan, F., Ozkan, S., and Song, J. (2021). "What do Data on Millions of US Workers Reveal About Lifecycle Earnings Dynamics?". *Econometrica*, *89*(5), 2303–2339.

Kopecky, K. A., and Suen, R. M. (2010). "Finite State Markov-Chain Approximations to Highly Persistent Processes". *Review of Economic Dynamics*, *13*(3), 701–714.

Krueger, D., Mitman, K., and Perri, F. (2016). "Macroeconomics and Household Heterogeneity". In *Handbook of macroeconomics* (Vol. 2, pp. 843–921). Elsevier.

Krueger, D., and Wu, C. (2021). "Concumption Insurance against Wage Risk: Family Labor Supply and Optimal Progressive Income Taxation". *American Economics Journal: Macroeconomics*, *13*(1), 79-113.

McLachlan, G. J., Lee, S. X., and Rathnayake, S. I. (2019). "Finite Mixture Models". *Annual Review of Statistics and Its Applications*, *6*, 355-378.

Mitchell, O. S., and Phillips, J. W. (2006). "Social Security Replacement Rates for Alternative Earnings Benchmarks". *Benefits Quarterly*, *4*, 37-47.

Rouwenhorst, K. G. (1995). "Asset Pricing Implications of Equilibrium Business Cycle Models". In T. F. Cooley (Ed.), *Frontiers of business cycle research* (pp. 294–330). Princeton University Press.

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

Tauchen, G., and Hussey, R. (1991). "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models". *Econometrica: Journal of the Econometric Society*, 371–396.

Terry, S. J., and Knotek II, E. S. (2011). "Markov-chain Approximations of Vector Autoregressions: Application of General Multivariate-Normal Integration Techniques". *Economics Letters*, *110*(1), 4–6.

A A multi-step EM algorithm for HMM

In this section, we outline the multi-step EM algorithm we use for the estimation of the HMM in case of life-cycle dynamics, where the transition matrix Π_t and grid μ_t are allowed to vary by age. The large number of parameters to be estimated here requires N to be large, and the EM algorithm has to converge for many parameters. A multi-step algorithm provides more stability.

Assume a panel of $y_{it} \in \mathbb{R}^k$, $t = 1, ..., T$ and $i = 1, ..., N$. Assume a given grid size m. Initialization:

- Estimate a Gaussian Mixture Model on y_{i1}, $i = 1, ..., N$. This gives a grid for the first time period and iteration, μ_1^1, stationary probabilities δ_1^1 and the filtered probabilities α_1^1. Set iteration $j = 1$.

We have a forward and backward step. For the forward step, set $t = 1$ and:

- Estimate the HMM of Section 2.2 for (y_{it}, y_{it+1}), $i = 1, ..., N$, restricting the grid of time period t to μ_t^j, the stationary probabilities of time period t to δ_t^j, the forward probabilities to α_t^j (except for $t = 1$, in which case they follow from Equation (11)). For $j > 1$, also restrict the backward probabilities for $t + 1$ to those obtained from the backward step, β_{t+1}^{j-1}, else set to 1. Estimate and store the grid μ_{t+1}^j, the transition probability matrix Π_t^j, stationary probabilities δ_{t+1}^j, and forward probabilities α_{t+1}^j. Set $t = t + 1$ and repeat up until and including $t = T - 1$.

For the backward step, set $t = T$ and:

- Estimate the HMM of Section 2.2 for (y_{it-1}, y_{it}), $i = 1, ..., N$, restricting the grid of time period t to μ_t^j, the stationary probabilities of time period t to δ_t^j, the forward probabilities to α_t^j, the backward probabilities to β_t^j (for $t < T$). When $t = T$, all of these (except the backward probabilities) come from the last time period of the forward step. Estimate and store the grid μ_{t-1}^j, the transition probability matrix Π_{t-1}^j, stationary probabilities δ_{t-1}^j, and backward probabilities β_{t-1}^j. Set $t = t - 1$ and repeat up until and including $t = 2$.

Once can iterate multiple times between the forward and backward step until they stabilize. In that case, update $j = j + 1$.

B Discretization of linear Gaussian processes

B.1 Discretizing an AR(1) process with Gaussian innovations

We consider an autoregressive process of order one with Gaussian errors:

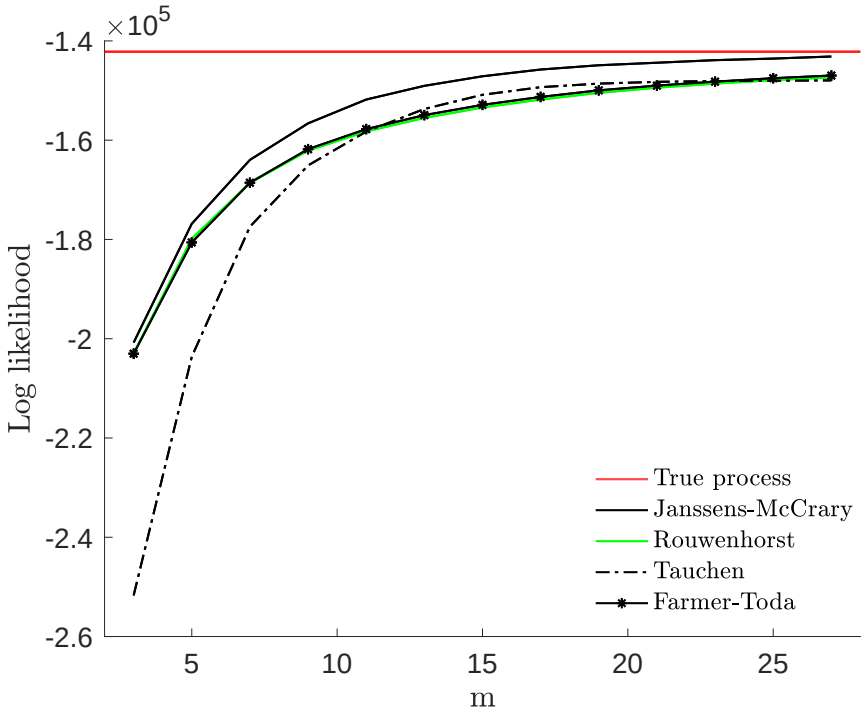
$$y_t = \rho y_{t-1} + \sigma_\varepsilon \varepsilon_t, \quad \varepsilon_t \sim N(0,1), \quad t = 1, ..., T. \tag{B.1}$$

We simulate data from the process in Equation (29). We impose both restriction (26) and (27) because of the symmetry of the process. Figure 1 in the main paper visualizes the elbow plot for this process. In Figure B1 we visualize the same log likelihood as in Figure 1 of the main paper, but we also add the log likelihoods that we obtain when we interpret the grids and transition probabilities of the existing methods as restrictions of our misspecified HMM. This visualizes the relative information loss of all methods relative to the true process. We can see that our method always produces the smallest information loss, and converges to the true process log likelihood faster. The information loss in the Rouwenhorst method and Farmer-Toda method is (almost) identical. The Tauchen method loses more information for low grid sizes m, but catches up on the Rouwenhorst and Farmer-Toda method from 11 grid points onwards.

We can use Figure B1 to analyze how much more parsimonious our discretization is in terms of information loss than the existing methods. For example, we can obtain a similar information loss with nine grid points where the existing methods need thirteen.

Comparing our method to the Tauchen (1986), Rouwenhorst (1995) and Farmer and Toda (2017) method, we obtain the results summarized in Table B1. For both grid sizes, the mean-squared forecast error indicates that our discretization has a much better fit to the data simulated from the continuous-support process than the discretized processes implied by the other methods. Focusing on conditional and unconditional moments, we see that for $m = 7$, our method does better at capturing the conditional skewness and conditional kurtosis than the other methods. Given that the Rouwenhorst (1995) targets the first two moments and the autocorrelation, it is not unexpected that we do not outperform this method in this dimension. Consistent with Figure B1, as m increases, the Tauchen (1986) method provides a better discretization, and actually captures the conditional skewness and kurtosis better than our method, and outperforms the discretizations that follow from applying Rouwenhorst (1995) and Farmer and Toda (2017) with regards to the MSFE.

Figure B1: Log likelihoods for the misspecified HMM of an AR(1) process, discretized by our method and existing methods for different values of grid size m.



B.2 Discretizing an AR(1) with a mixture of Gaussian innovations

We now consider an autoregressive process of order one with a mixture of Gaussian errors:

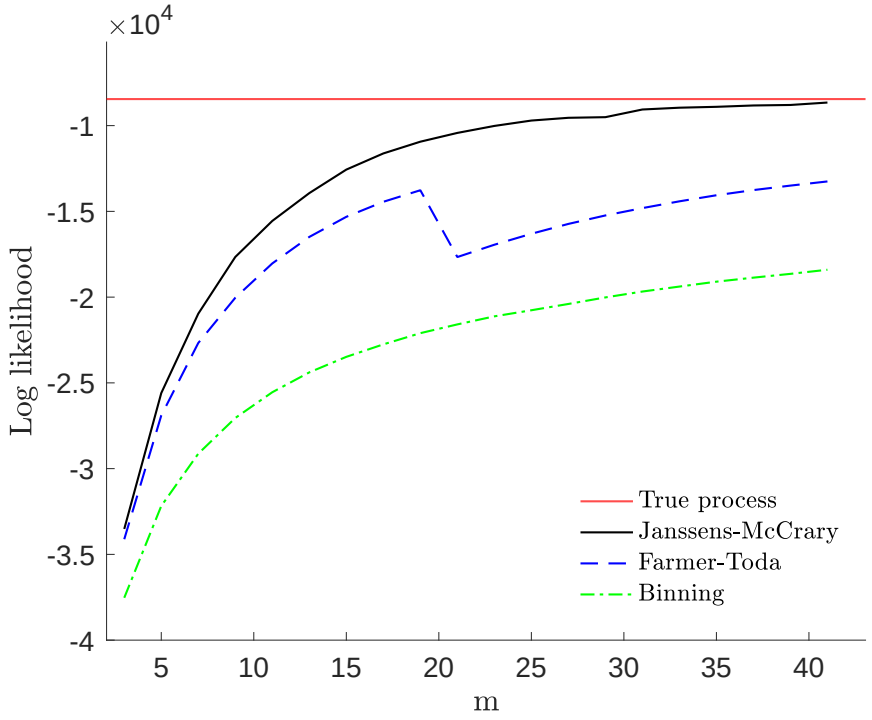
$$y_t = \rho y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \begin{cases} N(0, \sigma_1^2) & \text{with probability } p_1 \\ N(0, \sigma_2^2) & \text{with probability } p_2 \\ N(0, \sigma_3^2) & \text{with probability } 1 - p_1 - p_2. \end{cases} \tag{B.2}$$

Table B1: Comparison AR(1), lowest values in bold.

Method	Janssens-McCrary	Tauchen	Rouwenhorst	Farmer-Toda
m = 7				
Abs dev. uncond. mean y	$<$ **0.001**	$<$ 0.001	$<$ 0.001	$<$ 0.001
% dev. uncond. variance y	-11.1	52.8	$<$ **0.001**	$<$ 0.001
% dev. autocorrelation y	0.31	1.24	0.06	**0.04**
Abs. dev. uncond. skewness y	0.005	0.019	$<$ **0.001**	-0.043
% dev. uncond. kurtosis y	**-5.33**	-7.53	-11.6	-7.19
Abs. dev. cond. mean y	1.25	5.88	$<$ **0.001**	$<$ 0.001
% abs. dev. cond. variance y	12.80	18.2	$<$ **0.001**	$<$ 0.001
Abs. dev. cond. skewness y	**0.65**	1.11	1.42	2.11
% abs. dev. cond. kurtosis y	**95**	151	195	735
MSFE y	**1.41**	1.83	1.51	1.51
m = 11				
Abs dev. uncond. mean y	$<$ 0.001	$<$ **0.001**	$<$ 0.001	$<$ 0.001
% dev. uncond. variance y	-3.5	25.8	$<$ **0.001**	$<$ 0.001
% dev. autocorrelation y	1.11	-0.105	0.117	**0.096**
Abs. dev. uncond. skewness y	0.019	-0.030	0.014	**-0.004**
% dev. uncond. kurtosis y	**-0.707**	-6.05	-8.05	-1.04
Abs. dev. cond. mean y	1.45	3.21	$<$ **0.001**	$<$ 0.001
% abs. dev. cond. variance y	17.2	26.2	$<$ **0.001**	$<$ 0.001
Abs. dev. cond. skewness y	0.303	**0.217**	1.05	1.05
% abs. dev. cond. kurtosis y	29.37	**7.61**	117	571
MSFE y	**1.20**	1.29	1.32	1.32

Parametrization of AR(1): $\rho = 0.95$, $\sigma = 1$. $T = 5,000$.

Figure B2: Elbow plot for an AR(1) process with a mixture of Gaussian innovations



Parametrization: $\rho = 0.9$, $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.2$, $\sigma_3^2 = 0.8$, $p_1 = 0.2$, $p_2 = 0.6$. $T = 40,000$.

Table B2: Comparison for an AR(1) process with a mixture of Gaussian innovations

Method	Janssens-McCrary	Farmer-Toda	Binning
m = 7			
Dev. uncond. mean y	**< 0.001**	< 0.001	-0.036
% dev. uncond. variance y	-6.51	**-1.06**	-22.2
% dev. autocorrelation y	1.44	**-0.124**	-2.37
Abs. dev. uncond. skewness y	**-0.013**	0.023	-0.108
% dev. uncond. kurtosis y	**-9.79**	-16.4	-40.3
Abs. dev. cond. mean y	0.020	**<0.001**	0.016
% abs. dev. cond. variance y	24.45	**<0.001**	10.59
% abs. dev. cond. skewness y	**0.92**	1.14	1.00
% abs. dev. cond. kurtosis y	24.4	**20.2**	90.7
MSFE y	**0.192**	0.199	0.222
m = 15			
Dev. uncond. mean y	**< 0.001**	< 0.001	−0.030
% dev. uncond. variance y	-3.12	**-1.06**	-11.0
% dev. autocorrelation y	0.959	**-0.077**	-0.693
Abs. dev. uncond. skewness y	**-0.008**	0.047	-0.086
% dev. uncond. kurtosis y	**-2.13**	3.29	-26.4
Abs. dev. cond. mean y	0.028	**<0.001**	0.012
% abs. dev. cond. variance y	15.7	**<0.001**	9.71
% abs. dev. cond. skewness y	**0.651**	1.01	1.00
% abs. dev. cond. kurtosis y	**26.3**	135	90.7
MSFE y	**0.166**	0.174	0.181

Parametrization: $\rho = 0.9$, $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.2$, $\sigma_3^2 = 0.8$, $p_1 = 0.2$, $p_2 = 0.6$. The Farmer-Toda method here matches the first four conditional moments, but does not manage to match them at each grid point. In that case, it only matches the first two or three.

Figure B2 shows the elbow plot to determine the optimal number of grid points. For the specific parametrization of $\rho = 0.9$, $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.2$, $\sigma_3^2 = 0.8$, $p_1 = 0.2$, $p_2 = 0.6$, 9 grid points is suggested. We compare our discretization method to the method by Farmer and Toda (2017) and the binning method by Adda and Cooper (2003) and also visualizes the log likelihood that is obtained when using the obtained grid for their discretization methods as restrictions in our HMM model. As can be seen, the relative information loss between our method and the binning method is large. The method by Farmer and Toda (2017) does better, but at $m = 19$ falls down, because from $m = 19$ onwards, it is able to match more moments, but this seems to go at the cost of the overall fit to the true process.

Figure B2 also shows that we can be more parsimonious than the other two discretizations. When $m = 19$ for Farmer and Toda (2017), we can get to the same information loss to the true process using 14 grid points instead. For binning, the difference is even larger, and when binning uses $m = 13$, we can get to the same information loss with 6 grid points.

Next, Table B2 summarizes several statistics to compare our discretization method with the method by Farmer and Toda (2017). The method by Farmer and Toda (2017) aims to match both conditional and unconditional first to fourth order moments, but we see that their method only outperforms at the unconditional variance, autocorrelation, and the conditional mean and variance. The reason is that their method drops moment-restrictions it cannot match, and therefore sometimes does worse at those. We perform better with respect to all other moments (except the conditional kurtosis at $m = 7$).

B.3 Discretizing VAR models

In this subsection, we demonstrate the performance of our method for discretizing a bivariate VAR model of the form

$$y_t^1 = \beta_{11} y_{t-1}^1 + \beta_{12} y_{t-1}^2 + \varepsilon_t^1 \tag{B.3}$$

$$y_t^2 = \beta_{21} y_{t-1}^1 + \beta_{22} y_{t-1}^2 + \varepsilon_t^2, \tag{B.4}$$

where $\varepsilon_t \sim N(0, \Sigma)$.

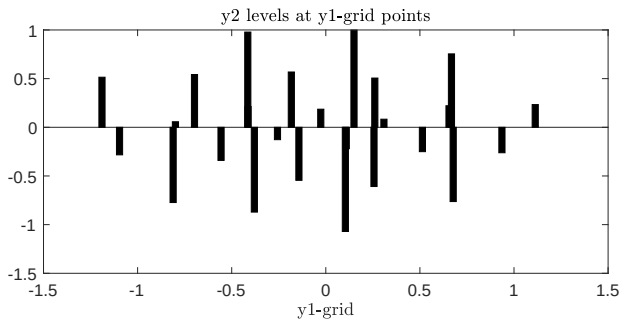
We consider three different parametrizations but keep the grid size fixed to $m = 17$ to show how our discretization method optimally selects the grid. The optimal grids are visualized in Figure B3. As can be seen, as opposed to a tensor grid, our optimal grid incorporates the structure of the process into the grid. For example, in a VAR model where both variables are positively correlated ($\beta_{12} = \beta_{21} > 0$), if y_1 is large, y_2 is also likely large. Figure B3b shows how this is reflected in our optimal grid, while a standard tensor grid as in Figure B3d does not reflect this co-dependence.

Table B3 summarizes the performance of our discretization compared to the discretization of Farmer and Toda (2017) for two different parametrizations of the VAR model in Equation (B.3). In their discretization, Farmer and Toda (2017) target the first two conditional moments.[11] As we can see, they outperform our discretization method on the first two conditional and

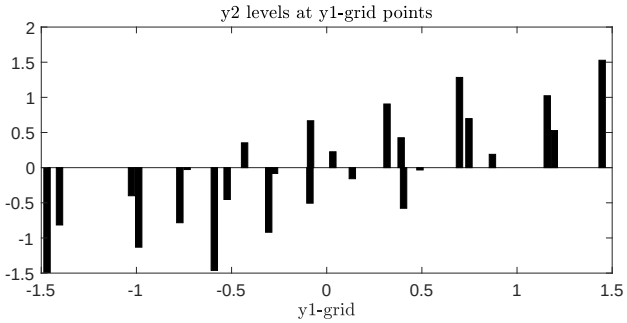
[11] Their optimization procedure cannot target higher-order moments for this specific model and parametrization.

Figure B3: Visualisation of optimal grid for three different parametrizations of the data generating process in Equation (B.3)*,* $m = 25$*.*

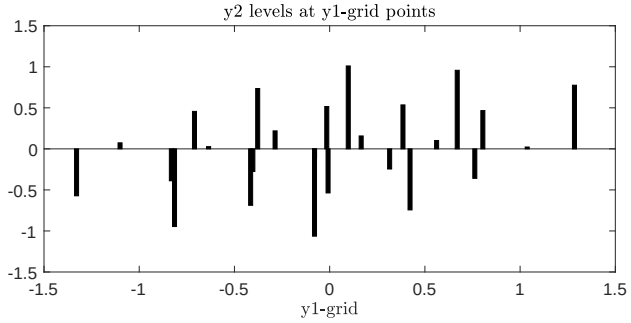
(a) Symmetric and uncorrelated



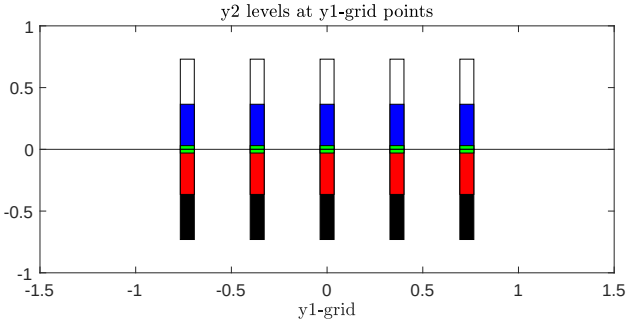
(b) Symmetric and positively correlated



(c) Asymmetric



(d) Symmetric and positively correlated: Farmer-Toda



For all three parametrizations, $\Sigma = \text{diag}(0.1)$.
Panel (a): $\beta_{11} = 0.7$, $\beta_{12} = 0$, $\beta_{21} = 0$, $\beta_{22} = 0.7$
Panel (b) and (d): $\beta_{11} = 0.7$ $\beta_{12} = 0.2$, $\beta_{21} = 0.2$, $\beta_{22} = 0.7$
Panel (c): $\beta_{11} = 0.7$ $\beta_{12} = 0.1$, $\beta_{21} = 0.0$, $\beta_{22} = 0.7$.

unconditional moments, but for all other moments, our method is closer to the true process. Our method also has a smaller mean squared forecast error.

Table B3: Comparison for VAR model in Equation (B.3) *for* $m = 25$ *(*$m_{y_1} = 5,\ m_{y_2} = 5$ *for Farmer-Toda).*

Method	Janssens-McCrary	Farmer-Toda
Parametrization 1		
Abs. dev. uncond. mean y	0.087	$<$ **0.001**
% dev. uncond. variance y	-0.134	**0.005**
% dev. autocorrelation y	**0.106**	-0.436
Abs. dev. uncond. skewness y	**<0.001**	0.072
% dev. uncond. kurtosis y	**-0.037**	-0.100
Abs. dev. correlation(y_1, y_2)	0.031	**-0.002**
Abs. dev. cond. mean y	0.035	$<$ **0.001**
% abs. dev. cond. variance y	35.8	$<$ **0.001**
% abs. dev. cond. skewness y	**0.203**	0.490
% abs. dev. cond. kurtosis y	**13.1**	17.3
MSFE y	**0.104**	0.104
Parametrization 2		
Abs. dev. uncond. mean y	0.093	$<$ **0.001**
% dev. uncond. variance y	-0.142	**0.009**
% dev. autocorrelation y	**0.102**	-0.326
Abs. dev. uncond. skewness y	**-0.005**	0.082
% dev. uncond. kurtosis y	**-0.040**	-0.223
Abs. dev. correlation(y_1, y_2)	0.058	**-0.036**
Abs. dev. cond. mean y	0.035	$<$ **0.001**
% abs. dev. cond. variance y	34.8	$<$ **0.001**
% abs. dev. cond. skewness y	**0.293**	0.615
% abs. dev. cond. kurtosis y	**14.36**	29.97
MSFE y	**0.106**	0.117

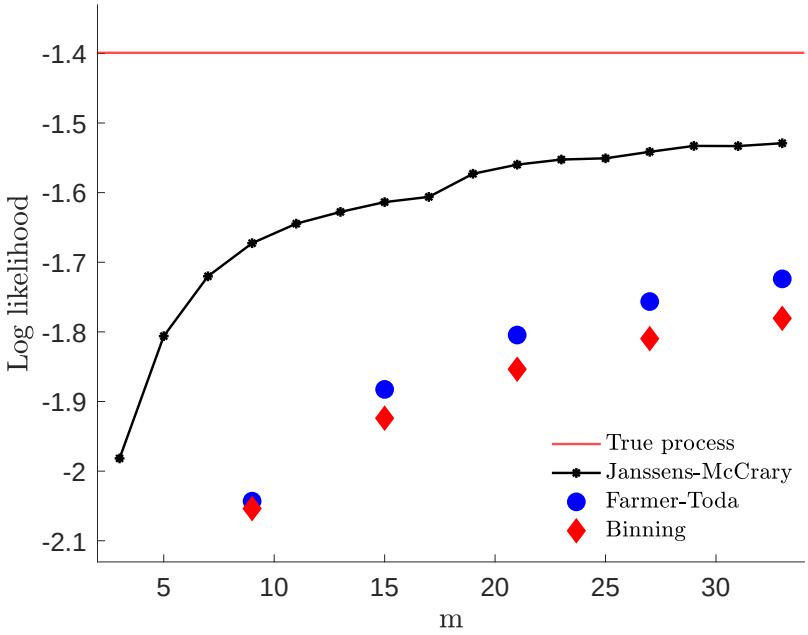
Parametrization 1: $\beta_{11} = 0.7\ \beta_{12} = 0.1,\ \beta_{21} = 0.0,\ \beta_{22} = 0.7$. Parametrization 2: $\beta_{11} = 0.7\ \beta_{12} = 0.1,\ \beta_{21} = 0.1,\ \beta_{22} = 0.7$. The statistics average over y_1 and y_2.

B.4 Discretizing an AR(1) with stochastic volatility

Consider the following stochastic process:

$$\begin{aligned} y_t &= \rho y_{t-1} + e^{h_t/2}\varepsilon_t, \quad \varepsilon_t \sim N(0,1) \\ h_t &= \mu + \phi(h_{t-1} - \mu) + \sigma_h \eta_t, \quad \eta_t \sim N(0,1). \end{aligned} \tag{B.5}$$

Figure B4: Average log likelihood for the misspecified HMM of the AR(1)-SV process

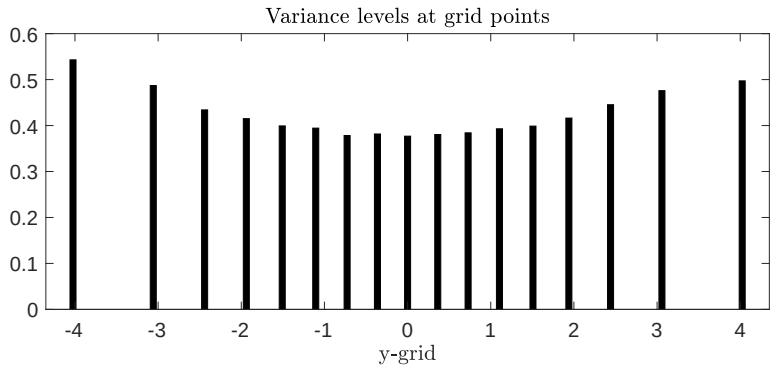


Parametrization: $\rho = 0.9$, $\mu = -1.9$, $\sigma_h^2 = 0.4228$, $\phi = 0.3$. T=150,000.

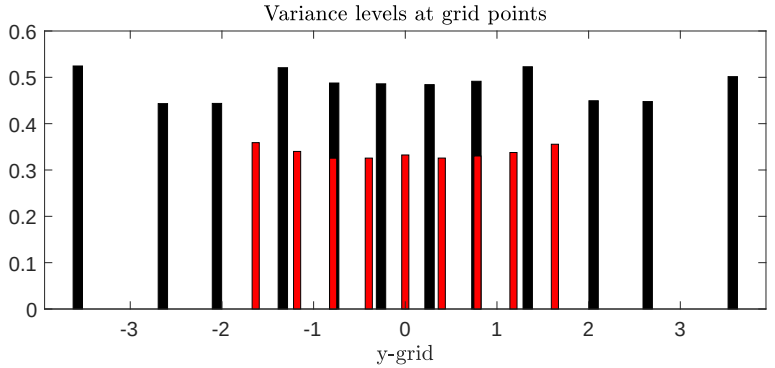
For the discretization, we treat the process in Equation (B.5) as a multivariate process in y and h. Figure B4 visualizes the log-likelihood of this model for different grid sizes. Interestingly, the log likelihood has a small jump at $m = 17$ rather than being monotonously increasing. This is not due to numerical imprecisions, but because of the way our discretization method selects its optimal grid, which we will elaborate upon below. Figure B4 also visualizes in red the log likelihood of the true process. As can be seen, our process gradually converges to the true process, but less fast than for a simple AR(1) process. We compute the implied log likelihoods of the competing methods too, interpreting their discretization as a restriction for our HMM. As follows, the information loss from using their discretization methods is substantial. We can obtain the same relative information loss as the Farmer-Toda process does in 27 grid points with only 7.

Figure B5: Visualisation of optimal grid for three different grid sizes ($m = 17, 21, 27$)*, where the data generating process is an AR(1) process with stochastic volatility as in Equation* (B.5).

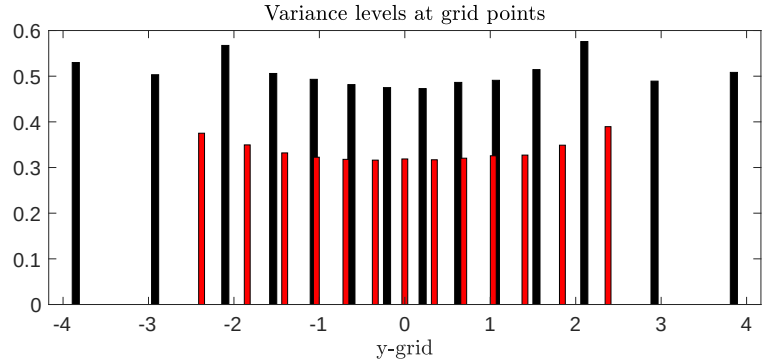
(a) $m = 17$



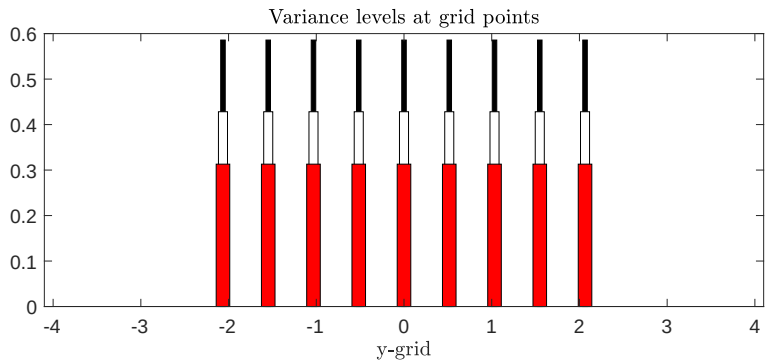
(b) $m = 21$



(c) $m = 27$



(d) Toda-Farmer method: $m = 27$ *(*$m_y = 9,\ m_h = 3$*)*



Parametrization: $\rho = 0.9$, $\mu = -1.9$, $\sigma_h^2 = 0.4228$, $\phi = 0.3$. Height of the bars depicts the variance level at the grid points, positioning on the x-axis of the bars depicts the level of y. The distinction between the red and black bars is to indicate that for m large, pairs of grid points arise where y has very similar values, but the value of the variance differs. Red bars are states that have a variance level below 0.4, black bars are above 0.4

Figure B5 visualizes the optimal grid for different grid sizes m. At $m = 17$ grid points, the method focuses on fitting y well, and, as can be seen, captures the fact that in the tails, the value of $\exp(h/2)$ is larger than in the middle. However, as more grid points are added, e.g. for $m = 21$ and $m = 27$, states will appear with very similar levels of y as other grid points, but other values for h. These grid points appear from $m = 19$ onwards, which explains the temporary decline in the explained variation after $m = 17$, which only measures the fit of y. On the other hand, the log likelihood, which measures the fit in both y and h, shows a steep increase from $m = 19$ to $m = 21$, capturing the fact that the process does a good job at modelling the joint dynamics of h and y.

Next, we compare our discretization method to other discretization methods. We use the method and codes of Farmer and Toda (2017), as well as a two-dimensional binning method adapted from Adda and Cooper (2003). For the two-dimensional binning method, we simulate data from the AR(1)-SV model, and create bins for y and h based on equal quantiles. The grid value is the median of the bin. The overall grid for (y, h) follows from the tensor grid. We compute the joint transitions between all bins to determine the transition probability matrix.

One important advantage of our discretization method is that ours does not rely on tensor grids, but, as visualized in Figure B5, optimally determines the joint grid of y and h. In contrast, panel (d) in Figure B5 shows that in case of a tensor product, each y-level will always have a fixed and pre-determined number of h-levels. Our process suggests that this is far from optimal, because, e.g., for very large (or small) levels of y, mostly large values of h are relevant.

We compare the performance of the different discretization methods in Table B4. Our discretization method does a better job at capturing the stationary distribution of y of the AR(1)-SV process, both for $m = 9$ and $m = 27$. Our discretization is closer to the unconditional mean, variance, skewness and kurtosis of the true process.

The discretization of Farmer and Toda (2017) is closer in terms of the autocorrelation of y. In terms of conditional moments, the method by Farmer and Toda (2017) performs well at the mean and variance, because these are moments their method explicitly targets. We do well at the conditional skewness. Binning does well at the conditional kurtosis. The difference between the MSFE's are large, especially for $m = 9$, but also at $m = 27$ the MSFE is 12% higher for Farmer and Toda (2017), and 23% for binning.

Table B4: Comparison for an AR(1) process with stochastic volatility

Method	Janssens-McCrary	Farmer-Toda	Binning
m = 9 ($m_y = 3$, $m_h = 3$ for Farmer-Toda and binning)			
Dev. uncond. mean y	$<$ **0.001**	$<$ 0.001	0.003
% dev. uncond. variance y	**-8.48**	-9.61	-40.9
% dev. autocorrelation y	2.70	**-0.021**	-10.5
Dev. uncond. skewness y	**-0.006**	-0.032	-0.016
% dev. uncond. kurtosis y	**-4.25**	-64.3	-54.7
% abs. dev. cond. mean y	0.026	$<$ **0.001**	0.061
% abs. dev. cond. variance y	25.6	**15.8**	34.3
Abs. dev. cond. skewness y	**0.493**	2.42	1.27
% abs. dev. cond. kurtosis y	139.8	292.6	**60.3**
MSFE y	**0.218**	0.365	0.402
m = 27 ($m_y = 9$, $m_h = 3$ for Farmer-Toda and binning)			
Dev. uncond. mean y	$<$ **0.001**	$<$ 0.001	0.003
% dev. uncond. variance y	**-3.78**	6.47	-15.7
% dev. autocorrelation y	1.66	**-0.166**	-2.09
Dev. uncond. skewness y	-0.032	-0.011	**-0.010**
% dev. uncond. kurtosis y	**-1.31**	-25.4	-32.0
% abs. dev. cond. mean y	0.024	$<$ **0.001**	0.020
% abs. dev. cond. variance y	27.7	**2.78**	11.1
Abs. dev. cond. skewness y	**0.733**	1.20	0.416
% abs. dev. cond. kurtosis y	226.8	288.5	**37.4**
MSFE y	**0.194**	0.218	0.240

Parametrization: $\rho = 0.9$, $\mu = -1.9$, $\sigma_h^2 = 0.4228$, $\phi = 0.3$. $T = 100,000$.

C An asset pricing model with stochastic volatility

C.1 A closed-form solution

From De Groot (2015), we obtain closed-form expressions for the asset pricing model with stochastic volatility presented in Equations (33)-(34). The solution for the price-dividend ratio is given by:

$$v_t = \sum_{i=1}^{\infty} \beta^i \exp(B_i y_t + C_i \bar{\eta} + D_i(\eta_t - \bar{\eta}) + H_i),$$

where

$$B_i = \left(\frac{1-\sigma}{1-\rho}\right)\rho(1-\rho^i)$$

$$C_i = \frac{1}{2}\left(\frac{1-\sigma}{1-\rho}\right)^2\left(i - 2\rho\frac{1-\rho^i}{1-\rho} + \rho^2\frac{1-\rho^{2i}}{1-\rho^2}\right)$$

$$D_i = \frac{\rho_\eta}{2}\left(\frac{1-\sigma}{1-\rho}\right)^2\left(\phi_1 + \phi_2\rho_\eta\rho_\eta^{i-1} + \phi_3\rho^{i-1} + \phi_4\rho^{2(i-1)}\right)$$

$$H_i = F_i\omega^2$$

where

$$F_i = \frac{1}{8}\left(\frac{1-\sigma}{1-\rho}\right)^4\left(i\phi_1^2 + \phi_2^2\frac{1-\rho_\eta^{2i}}{1-\rho_\eta^2} + \phi_3^2\frac{1-\rho^{2i}}{1-\rho^2} + \phi_4^2\frac{1-\rho^{4i}}{1-\rho^4}\ldots\right.$$
$$\ldots + 2\phi_1\phi_2\frac{1-\rho_\eta^i}{1-\rho_\eta} + 2\phi_1\phi_3\frac{1-\rho^i}{1-\rho} + 2\phi_1\phi_4\frac{1-\rho^{2i}}{1-\rho^2} + 2\phi_2\phi_3\frac{1-(\rho_\eta\rho)^i}{1-\rho_\eta\rho}\ldots$$
$$\left.\ldots + 2\phi_2\phi_4\frac{1-(\rho_\eta\rho^2)^i}{1-\rho_\eta\rho^2} + 2\phi_3\phi_4\frac{1-\rho^{3i}}{1-\rho^3}\right)$$

and

$$\phi_1 = \frac{1}{1-\rho_\eta}, \quad \phi_2 = \frac{-\rho_\eta(\rho_\eta+\rho)(1-\rho)^2}{(\rho^2-\rho_\eta)(\rho-\rho_\eta)(1-\rho_\eta)},$$
$$\phi_3 = \frac{-2\rho^2}{\rho-\rho_\eta}, \quad \phi_4 = \frac{\rho^4}{\rho^2-\rho_\eta}.$$

The conditional expected return on equity is defined as

$$\mathbb{E}_t R^e_{t+1} = \mathbb{E}_t \left(\frac{d_{t+1} + p_{t+1}}{p_t} \right) = \frac{\mathbb{E}_t \exp(y_{t+1}) + \mathbb{E}_t v_{t+1} \exp(y_{t+1})}{v_t}$$

The solution to this expression gives that

$$\mathbb{E}_t \exp(y_{t+1}) = \exp\left(\rho y_t + \frac{1}{2}\bar{\eta} + \frac{\rho_\eta}{2}(\eta_t - \bar{\eta}) + \frac{1}{8}\omega^2 \right)$$

and

$$\mathbb{E}_t v_{t+1} \exp(y_{t+1}) = \sum_{i=1}^{\infty} \beta^i \exp\Big((B_i + 1)\rho y_t + (C_i + \frac{1}{2}(B_i + 1)^2)\bar{\eta} + \frac{1}{2}(B_i + 1)^2 \rho_\eta (\eta_t - \bar{\eta}) + \ldots$$
$$(F_i + \frac{1}{2}(\frac{1}{2}(B_i + 1)^2 + D_i)^2)\omega^2 \Big).$$

As shown by De Groot (2015), there is a parameter restrictoin that guarantees a finite price-dividend ratio:

$$\beta \exp\left(\frac{1}{2} \left(\frac{1 - \sigma}{1 - \rho} \right)^2 \bar{\eta} + \frac{(1 - \sigma)^4}{8(1 - \rho)^4 (1 - \rho_\eta)^2} \omega^2 \right) < 1.$$

We chose our parametrization of β and σ such that this condition is satisfied, follow De Groot (2015) for the other parameters.

C.2 A discretized solution

Instead of solving the model using the continuous-support process in Equations (33)-(34), one can discretize the stochastic process and obtain approximate solutions for the price-dividend ratio, the conditional expected return on equity, and other objects of interest. If y_t follows a discrete-state-space first-order Markov process with states y_s, $s \in \{1, \ldots, m\}$ and transition probability matrix Π with elements $\Pi_{ss'} = P(y_{t+1} = y_{s'} | y_t = y_s)$, then we can rewrite Equation (35) as

$$v(y_s) = \beta \sum_{s'=1}^{m} \exp((1 - \sigma) y_{s'})(v(y_{s'}) + 1)\Pi_{ss'}$$

which solves to

$$v = \left(I_m - \beta \Pi \text{diag}(\exp(1 - \sigma) y) \right)^{-1} \beta \Pi \exp((1 - \sigma) y), \tag{C.1}$$

where m denotes the number of discrete states of y_t, y is an $s \times 1$ vector with all the levels y_t attains, and v is an $s \times 1$ vector with all discrete realizations of the price-dividend ratio in each discrete realization of y. Similarly, for the vector of conditional expected returns on equity at each value of the grid y_s, denoted $R^e(y_s)$, we have

$$R^e(y_s) = \Big(\sum_{s'} \Pi_{ss'} \exp(y_s)(1 + v(y_{s'})) \Big) / v(y_s). \tag{C.2}$$

The reason why we are interested in the performance of capturing $\mathbb{E}_t R^e_{t+1}$ is because of its non-linear dependence on v_t, which is also approximated. The approximation errors will compound in a non-trivial way, and we are interested in how accurate our discretization method is when these errors accumulate.

D Two binning methods for Guvenen et al. (2021)

D.1 Simple binning

We adapt Adda and Cooper (2003) such that it can deal with the large number of 0-states that the Guvenen et al. (2021) process generates, as well as the life-cycle dependence. Choose $m-1$ quantile levels, typically with equal distance, and denote these by $x_1, ..., x_{m-1}$. We define bins b_i, $i = 1, ..., m$:

$$\begin{aligned} b_1 = [0,\ 0],\ b_2 = [\text{quantile}_{x_1}(y|y>0),\ \text{quantile}_{x_2}(y|y>0)], \dots , \\ b_m = [\text{quantile}_{x_{m-1}}(y|y>0),\ +\infty] \end{aligned} \tag{D.1}$$

The grid is then given by $\mu_1 = 0$ for the first grid point, and $\mu_i = \text{quantile}_{(x_i+x_{i-1})/2}(y|y>0)$ for the others. To determine the transition probability matrix, simulate N life-cycles of length T (resulting in a panel of dimensions $(T \times N)$) from the process in Equation (38) and assign the simulated observations of y into bins. The transition probabilities at age t are computed by counting the transitions between bins:

$$P^t_{i,j} = \frac{\sum_{i=1}^N I\{y_{i,t+1} \in b_i | y_{i,t} \in b_j\}}{\sum_{i=1}^N I\{y_{i,t} \in b_j\}}.$$

D.2 Clever binning

In this binning method, we use more information about the process, and do not simply bin on y, but instead, first, bin on the persistent component of earnings, z. Therefore, first, select the discretization level for z, denoted m_z, and correspondingly have equal-distance quantiles, denoted by $x^z_1, ..., x^z_{m_z-1}$. The bins for z are defined as b^z_i, $i = 1, ..., m_z$: $b^z_1 = [-\infty, \text{quantile}_{x^z_1}]$, $b^z_i = [\text{quantile}_{x^z_{i-1}}(z_{i,t}),\ \text{quantile}_{x^z_i}(z_{i,t})]$ and $b^z_{m_z} = [\text{quantile}_{x^z_{m-1}}(z_{i,t}), +\infty]$. The corresponding grid for z is then given by the midpoints: $\mu^z_i = \text{quantile}_{(x^z_i+x^z_{i-1})/2}(z)$, for $i = 1, ..., m_z$. Simulating a series of z, the transition probabilities for z can then be computed by

$$P_{i,j} = \frac{\sum_{i=1}^N I\{z_{i,t+1} \in b_i | z_{i,t} \in b^z_j\}}{\sum_{i=1}^N I\{z_{i,t} \in b^z_j\}}. \tag{D.2}$$

Next, we double the grid, because for each value of z, one can either be unemployed with probability p^t_v, or employed with probability $1 - p^t_v$, see Equation (38). This gives a tensor grid of $(v_{it} \in \{0, 1\}) \otimes (z_{it} \in \{b^z_i\}^{m_z}_{i=1})$. To compute the transition probabilities of this tensor

grid, one needs to multiply $P_{i,j}$ of Equation (D.2) by p_v^{t+1} or $1 - p_v^{t+1}$ respectively, where this probability is defined in Equation (38).

At last, one should discretize the transitory component ε for the employed states. If the number of discretization levels for ε is given by m_ε with an equal-distant grid, the transition probabilities $P(z_{it+1} \in b_i^z, \varepsilon \in b_j^\varepsilon, v_{t+1} = 0 | (z_{it+1} \in b_k^z, \varepsilon \in b_l^\varepsilon, v_t = 0) = \frac{1}{m_\varepsilon} P(z_{it+1} \in b_i^z, v_{t+1} = 0 | z_{it+1} \in b_k^z, v_t = 0)$.

E Age-dependent transition probabilities and grids

Figure E1: Visualisation of the age-dependent transition probabilities for a $m = 9$ discretization of the stochastic process in Guvenen et al. (2021). The order of the matrix corresponds with a sorted (low-to-high) earnings grid, where the two lowest states are zero-earnings states.

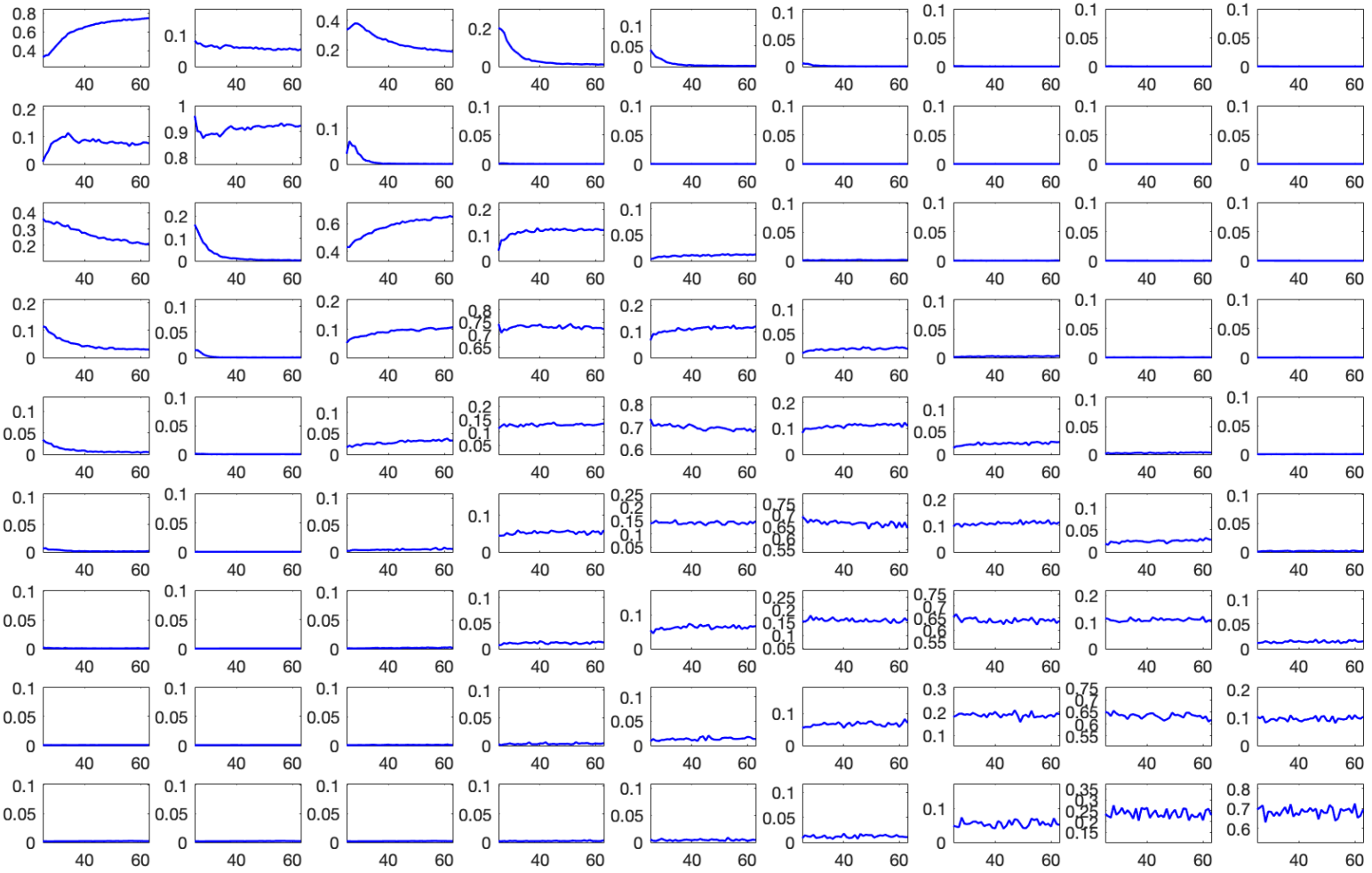


Figure E2: Visualisation of the age-dependent transition probabilities for a $m = 10$ discretization of the stochastic process in Arellano et al. (2017). The order of the matrix corresponds with a sorted (low-to-high) earnings grid.

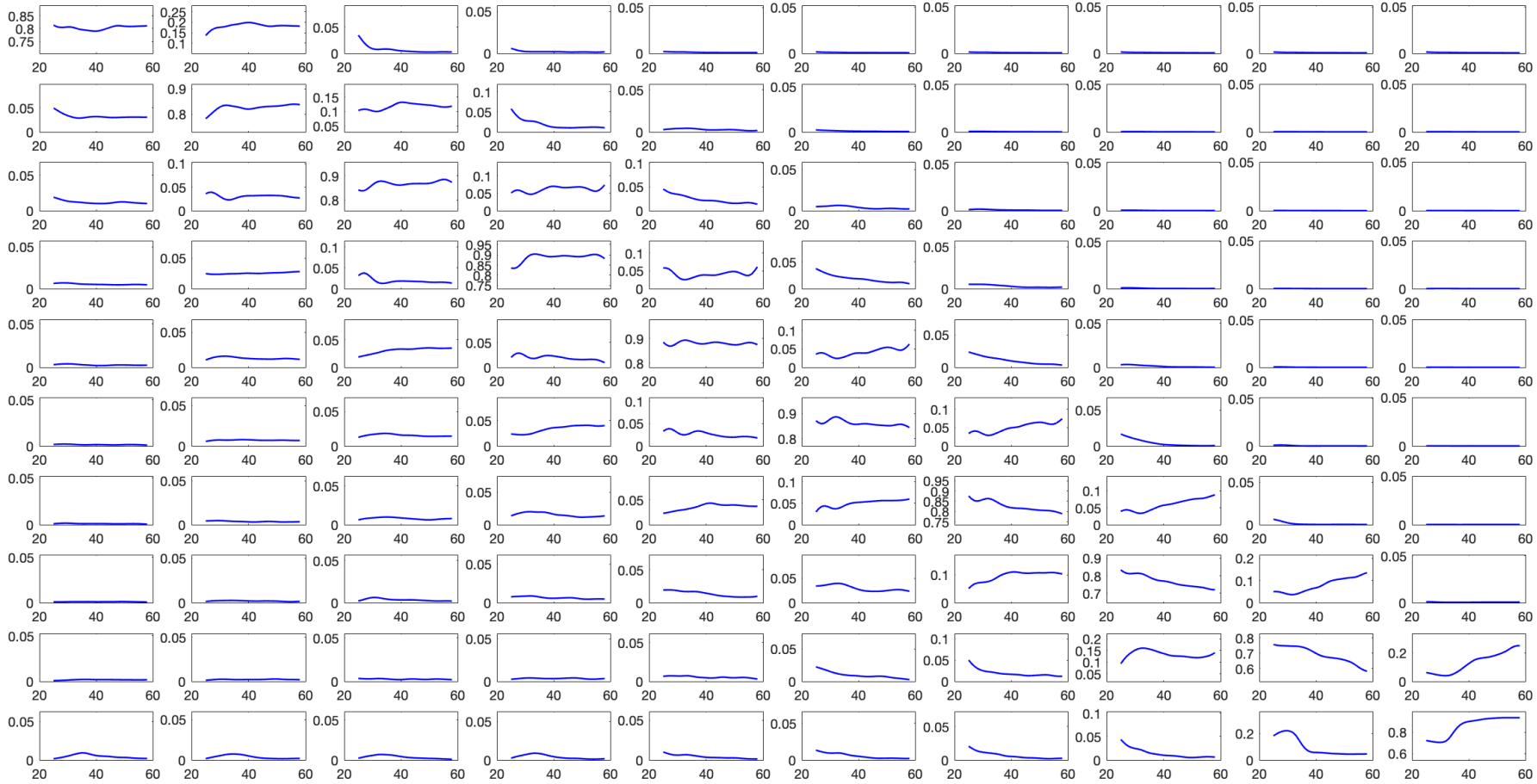
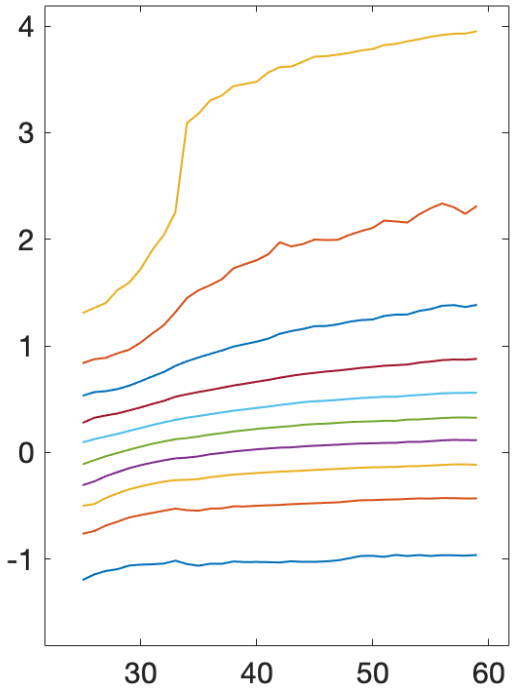


Figure E3: Visualisation of the age-dependent grid of a $m = 10$ discretization of the stochastic process in Arellano et al. (2017).



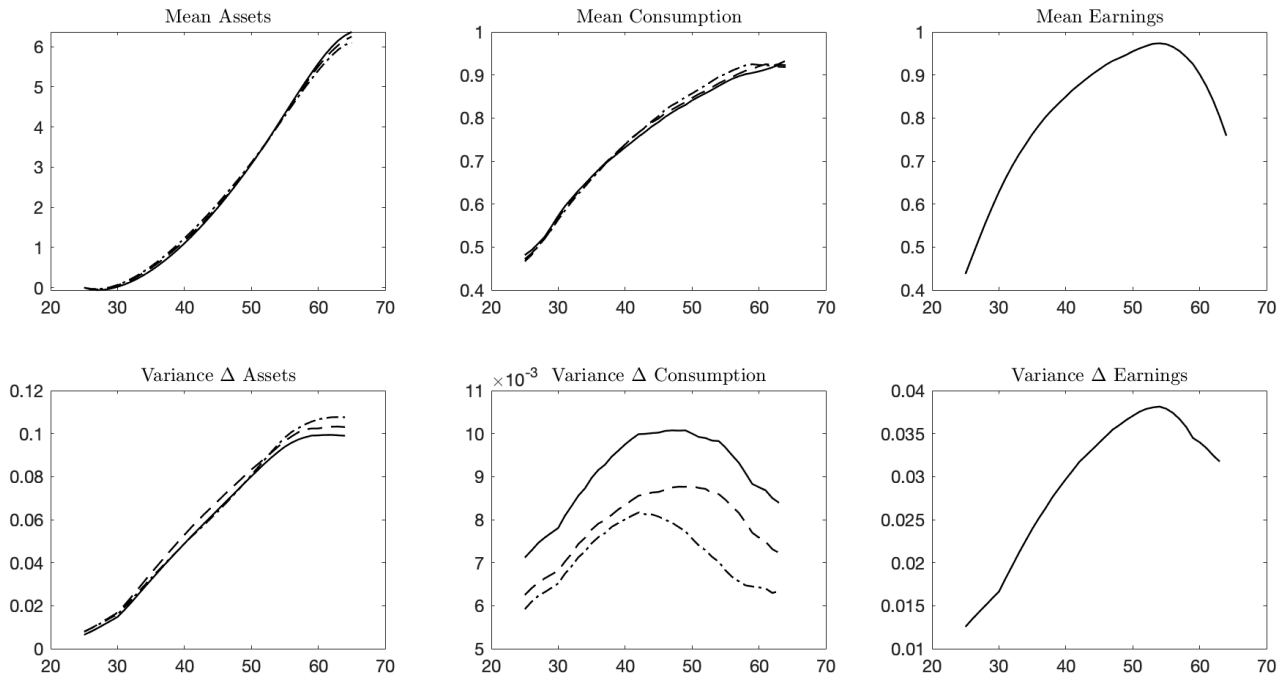
F Life-cycle model with other stochastic processes

F.1 A life-cycle model with an AR(1) process for earnings

In this subsection we show that even when the earnings process is given by a simple AR(1) process, the choice of the discretization method matters. We compare our method with the method of Farmer and Toda (2017) and Tauchen (1986). We do not consider the Rouwenhorst method in this section, because the discretization is almost identical to one by Farmer and Toda (2017) for the AR(1) process.

Some summary statistics are provided in Table F1, and visualized in Figure F1.

Figure F1: Simulations from the life-cycle model for three different discretizations of an AR(1) process. $m = 7$ for all methods. Assets, consumption and earnings over the life-cycle (age on the x-axis).



Solid black line represents our discretization method, the dashed line is the Farmer-Toda method, the dash-dot line is the Tauchen method. Note that in all cases, the data is simulated from the continuous-support earnings process, but the solution is computed using the four different discretizations, which is why in the last column, the three lines coincide, but in the other graphs, the lines are different. Process scaled such that mean of y_t is 1 for all discretizations.

Table F1: Summary statistics computed from simulations from the life-cycle model for three different discretizations of an AR(1) process. $m = 7$ is used for the discretization. Parametrization of $\log y_t$*: $\rho = 0.95$, $\sigma = 0.2$. Process scaled such that mean of y_t is 1 for all discretizations.*

Discretization method	Janssens-McCrary	Farmer-Toda	Tauchen
Variance c_{it}	0.16	0.16	0.16
Variance Δc_{it}	0.009	0.008	0.007
Variance a_{it}	16.4	16.8	15.9
Variance Δa_{it}	0.07	0.07	0.07
Covariance c_{it} and y_{it}	0.19	0.19	0.19
Covariance Δc_{it} and Δy_{it}	0.01	0.01	0.01
CEV	0.40	0.40	0.49
ϕ^P_{BPP}	0.45	0.49	0.50

Both Figure F1 and Table F1 show that the discretization method matters for the implications that follow from a life-cycle model, even when discretizing an AR(1) process. Our discretization implies welfare costs similar to those of the Farmer-Toda method, but a lower partial insurance to persistent income shocks coefficient, as measured by ϕ^P_{BPP}, defined in Blundell et al. (2008). This also leads to a lower variance in assets for our method. Our method captures the higher order moments of the process better, as seen in Table B1, which can explain this difference. The CEV for the Tauchen method is considerably higher, and the variance in assets lower. Recall from Appendix Section B1 that the Tauchen method does not perform well for lower values of m.

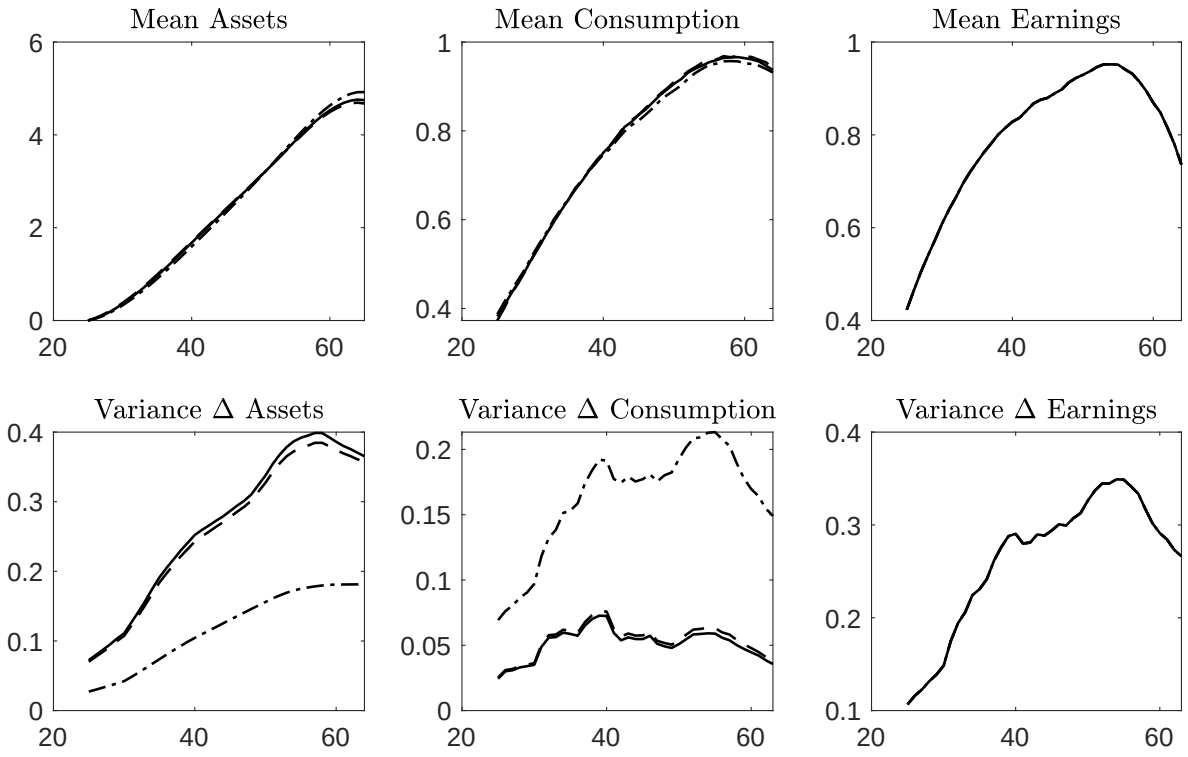
F.2 A life-cycle model with an AR(1) earnings process with fat tails

Next, we analyze the implications of using our discretization method versus other discretization methods when the earnings process is an AR(1) process where the innovations have fat tails. As can be seen, our discretization method is similar in terms of conclusions as the one by Farmer and Toda (2017). The Farmer and Toda (2017) method aims to match the first four moments of the continuous process. Using a binning method, however, does lead to different results. For example, the variance in consumption changes can be up to four times as large when using the binning method, see Figure F2. As can be seen in Table F2, the certainty equivalent value is lower when using binning. This suggests that the binning method does not capture the amount of risk in the process well. Recall from Table B2 in Appendix A that the binning performed poorly at most moments of the stochastic process, even for a large grid. The MSFE of the binning method is 10% larger than our method for $m = 15$, and they

method performs poorly at matching both the conditional and unconditional kurtosis of the continuous-support process.

Another important difference between the methods is the partial insurance to persistent income shocks, measured by ψ^P_{BPP}. The binning method suggests a partial insurance coefficient that is half the size of ours and the coefficient that follows from the discretization of Farmer and Toda (2017). The same holds for. the covariance between consumption changes and earnings changes, which for the binning method is double in size.

Figure F2: Simulations from the life-cycle model for three different discretizations of an AR(1) process with fat tails. $m = 17$ for all methods. Assets, consumption and earnings over the life-cycle (age on the x-axis).



Solid black line represents our discretization method, the dashed line is the Farmer-Toda method, the dash-dot line is the binning method. Note that in all cases, the data is simulated from the continuous-support earnings process, but the solution is computed using the four different discretizations, which is why in the last column, the four lines coincide, but in the other graphs, the lines are different. Parametrization: $\rho = 0.9$, $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.2$, $\sigma_3^2 = 0.8$, $p_1 = 0.2$, $p_2 = 0.6$.

Table F2: Summary statistics computed from simulations from the life-cycle model for three different discretizations of an AR(1) process with fat tails. $m = 17$ is used for the discretization. Parametrization of $\log y_t$*: $\rho = 0.9$, $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.2$, $\sigma_3^2 = 0.8$, $p_1 = 0.2$, $p_2 = 0.6$. Processes rescaled such that y_t has a mean of one.*

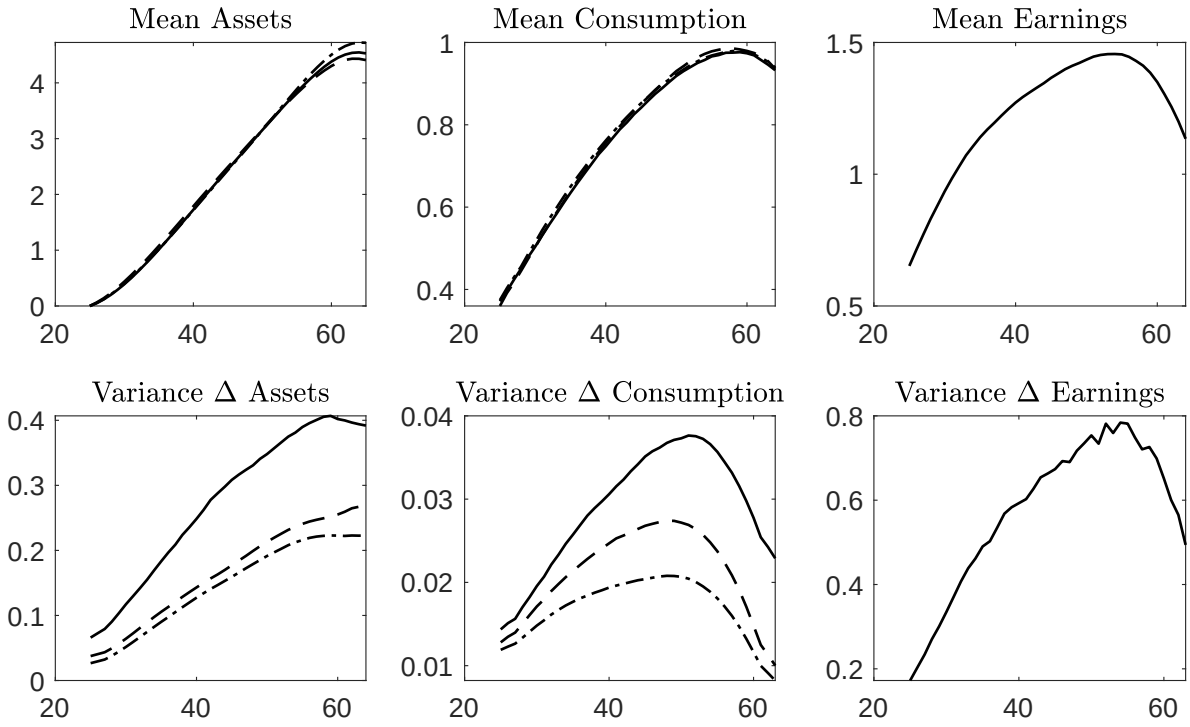
Discretization method	Janssens-McCrary	Farmer-Toda	Binning
Variance c_{it}	0.38	0.39	0.52
Variance Δc_{it}	0.05	0.05	0.16
Variance a_{it}	25.7	24.8	15.5
Variance Δa_{it}	0.45	0.46	0.60
Covariance c_{it} and y_{it}	0.45	0.46	0.60
Covariance Δc_{it} and Δy_{it}	0.09	0.10	0.19
CEV	0.56	0.56	0.51
ψ^P_{BPP}	0.62	0.61	0.25

F.3 A life-cycle model with an AR(1)-SV earnings process

In Figure F3 we visualize several summary statistics of simulations from a life-cycle with an AR(1)-SV earnings process, where we compare three different discretization methods: (i) our proposed method, (ii) the method by Farmer and Toda (2017) and (iii) a two-dimensional binning method adapted from Adda and Cooper (2003).

The choice of the discretization method matters for the asset and consumption choices of individuals. Although the mean consumption and asset holdings over the life-cycle are similar across discretization methods, this is not the case for the variances. For example, when using our discretization method, individuals face a variance of their consumption changes that is up to two times as large than when using the other discretization methods. The differences between the discretization methods also can be seen from Table F3. The choice of the discretization method also matters for the covariance between consumption and income (as well as between their first differences), and this covariance is considerably higher when using our discretization method. Our CEV is similar to the one found by the Farmer-Toda method, but we do find a higher partial insurance coefficient ψ^P_{BPP}.

Figure F3: Simulations from the life-cycle model for three different discretizations of an AR(1)-SV process. $m = 27$ for all methods. Assets, consumption and earnings over the life-cycle (age on the x-axis).



Solid black line represents our discretization method, the dashed line is the Farmer-Toda method, the dash-dot line is the binning method. The last panel depicts the continuous-support earnings process. Data is simulated on the grid, and the solution is computed using the four different discretizations. Parametrization: $\rho = 0.9$, $\mu = -1.9$, $\sigma_h^2 = 0.4228$, $\phi = 0.3$.

Table F3: Summary statistics computed from simulations from the life-cycle model for three different discretizations of an AR(1)-SV earnings process. $m = 27$ is used for the discretization.

Discretization method	Janssens-McCrary	Farmer-Toda	Binning
Variance c_{it}	0.43	0.35	0.24
Variance Δc_{it}	0.03	0.02	0.02
Variance a_{it}	26.4	16.3	13.2
Variance Δa_{it}	0.28	0.17	0.15
Covariance c_{it} and y_{it}	0.49	0.39	0.28
Covariance Δc_{it} and Δy_{it}	0.06	0.05	0.04
CEV	0.61	0.61	0.54
ψ^P_{BPP}	0.67	0.60	0.63