Federal Reserve Bank of Cleveland Working Paper Series

The Art of Temporal Approximation: An Investigation into Numerical Solutions to Discrete & Continuous-Time Problems in Economics

Keyvan Eslami and Thomas M. Phelan

Working Paper No. 23-10

May 2023

Suggested citation: Eslami, Keyvan and Thomas M. Phelan. 2023. "The Art of Temporal Approximation: An Investigation into Numerical Solutions to Discrete & Continuous-Time Problems in Economics." Working Paper No. 23-10. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202310.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

The Art of Temporal Approximation: An Investigation into Numerical Solutions to Discrete- & Continuous-Time Problems in Economics*

Keyvan Eslami[†] Thomas Phelan[‡]

May 9, 2023

Abstract

A recent literature within quantitative macroeconomics has advocated the use of continuous-time methods for dynamic programming problems. In this paper we explore the relative merits of continuous-time and discrete-time methods within the context of stationary and nonstationary income fluctuation problems. For stationary problems in two dimensions, the continuous-time approach is both more stable and typically faster than the discrete-time approach for any given level of accuracy. In contrast, for convex lifecycle problems (in which age or time enters explicitly), simply iterating backwards from the terminal date in discrete time is superior to any continuous-time algorithm. However, we also show that the continuous-time framework can easily incorporate nonconvexities and multiple controls—complications that often require either problem-specific ingenuity or nonlinear root-finding in the discrete-time context. In general, neither approach unequivocally dominates the other, making the choice of one over the other an art, rather than an exact science.

Keywords: Markov chain approximation, dynamic programming, numerical methods, income fluctuation problems.

JEL codes: C63, E21.

*The views stated herein are those of the authors and are not necessarily those of the Federal Reserve Bank of Cleveland or the Board of Governors of the Federal Reserve System. We thank Pierlauro Lopez, Kurt Lunsford, Ben Moll, Sergio Ocampo, Pontus Rendahl, and participants at the 2022 meetings of the Southern Economics Association for helpful comments. Code is written in Python 3.8.0 and can be found at https://github.com/tphelanECON/The_Art_of_Temporal_Approximation_WP.

[†]Toronto Metropolitan University, Department of Economics, Email: k.eslami@torontomu.ca.

[‡]Federal Reserve Bank of Cleveland. Email: tom.phelan@clev.frb.org.

1 Introduction

A vast literature within both applied mathematics and economics has developed tools for the numerical solution to optimal control problems in both continuous-time and discrete-time settings. In a recent contribution, Achdou et al. (2022) have illustrated the potential for substantial speed gains when applying continuous-time techniques (specifically, the method of finite-differences) to standard macroeconomic problems.[1] Given this work, when the most urgent concern is the computational intensity of the numerical solution to the problem at hand (and not a conceptual feature of the model that necessitates the choice of one environment), should the researcher ever use discrete-time methods, or do continuous-time methods always dominate? The purpose of this paper is to provide some guidance on the costs and benefits of each approach using examples from the quantitative macroeconomics literature.

We follow Achdou et al. (2022) and focus on *income fluctuation problems* (IFPs), in which an agent self-insures against idiosyncratic income risk by saving in a risk-free bond. Variations of such problems form the backbone of many incomplete-market models, and hence of much of modern macroeconomics. For the continuous-time problems, we discretize using the Markov chain approximation (MCA) method of Kushner and Dupuis (2001) in such a way that the policy function may be updated in closed form, and for the discrete-time problems, we update the policy function using the endogenous grid method (EGM) of Carroll (2006). In both cases, we then compare the performance of value function iteration (VFI), policy function iteration (PFI), and modified policy function iteration (MPFI) to update the value function.[2] Although these techniques are not specific to IFPs, it is within this class of problems that we aim to provide practical guidance. For stationary problems, the continuous-time approach is superior, as it is typically faster than the fastest implementation of EGM and is guaranteed to converge. In contrast, for nonstationary (or lifecycle) problems, the situation is more complicated. For a concave lifecycle model, simply iterating backwards from the terminal date in discrete time using the EGM is much faster than any continuous-time algorithm. However, we illustrate with an example that the continuous-time framework is more flexible and requires no problem-specific ingenuity in more complicated settings with nonconvexities and multiple controls. We therefore believe that one cannot assert the general superiority of one method over the other.

In some sense, continuous-time and discrete-time formulations describe fundamentally different environments, and so we wish to emphasize upfront precisely how we compare the

[1] See also Candler (2001) for an earlier application of finite-difference methods to economics.

[2] Kushner and Dupuis (2001) show that finite-difference approximations can be used to construct approximating chains, while Phelan and Eslami (2022) show that the implicit finite-difference method favored by Achdou et al. (2022) is equivalent to using a limiting case of PFI for a particular chain. We therefore believe that MCA methods encompass the algorithms available in the continuous-time framework.

accuracy of various algorithms. In both the continuous-time and discrete-time frameworks, we discretize the asset space and thereby reduce the agent's problem to a finite-space decision problem. For each framework, we then fix the processes governing income (and age, if relevant) and regard the policy and value functions computed on a fine grid for assets as the "true" values. All references to accuracy are made in comparison to these values.[3]

We deliberately do not explore the accuracy of various discretization methods for income, as such comparisons already exist in the literature and the pros and cons of different discretizations seem to be orthogonal to our purposes.[4] We instead fix the income process used in each framework as we vary the fineness of the asset grid and suppose that the income process in the discrete-time case is consistent with the continuous-time process in the following sense. In the continuous-time case, the agent's income follows a Markov chain constructed with the MCA approach with a small timestep, and the income process in discrete-time is chosen to be consistent with this same chain sampled at a lower frequency.[5]

We wish to provide intuition on the source of the differences across the frameworks, hoping that this may guide researchers in other settings. To this end, we emphasize that all of the algorithms employed in this paper consist of two distinct steps:

(i) updating the policy function given a guess of the value function; and,

(ii) updating the value function given a guess of the policy function, and (possibly) a guess of the value function.

We will refer to step (i) as *policy updating*, and step (ii) as *value updating*. The continuous-time framework can potentially aid in both steps: the local nature of the transitions implies that policy updating can often be performed in closed form, and the sparse nature of the transitions often renders policy function iteration rapid in the value updating step.

We first consider stationary problems, in which the agent is infinitely lived and the state variables are simply income and wealth. For such problems, Achdou et al. (2022) show that the implicit finite-difference method is faster (for a given level of accuracy) than the method of Carroll (2006), in which at each step one updates the policy function using the EGM and iterates once on the Bellman operator. The rapid convergence of the implicit method is unsurprising given its close connection with PFI, since Puterman and Brumelle (1979) show that PFI converges quadratically near the solution and typically requires a small number of iterations. Further, in two dimensions, the local nature of the transitions in continuous

[3]This is similar to the exercise conducted in Section 5.6 and Appendix F.1 of Achdou et al. (2022), except that we also allow for modified policy function iteration and policy function iteration in the comparison and also consider lifecycle models.

[4]For an analysis of such discretizations, see, e.g., Kopecky and Suen (2010), Gospodinov and Lkhagvasuren (2014) and Farmer and Toda (2017).

[5]We consider the more familiar Tauchen (1986) discretization in the appendix and show that our main findings do not change.

time ensures that sparse solvers can be used to efficiently solve the linear system defining the updated value function.

Rendahl (2022) recognizes that this sparsity can also be exploited in discrete time, and that for a two-state income process, combining ideas from Young (2010) with PFI can dramatically close the gap in run times between the continuous-time and discrete-time approaches. We extend Rendahl (2022) in two respects. First, we systematically compare the performance of VFI and PFI with MPFI. This is a simple but useful extension, because in discrete time the potentially nonlocal nature of income transitions implies that sparse solvers become less attractive as the income grid becomes finer.[6] Second, we document the speed versus accuracy tradeoff in each framework by combining these run times with two measures of accuracy: mean and maximum percentage error in the policy function. This is distinct from documenting run times by grid sizes because the discrete-time framework is more accurate than the continuous-time framework for every grid considered.[7] In this setting, we find that it is not the case that one framework always dominates the other for all levels (and measures) of accuracy. However, for most of the grids we consider, we find that the continuous-time analysis is typically faster for a given level of accuracy. Further, we emphasize that the continuous-time analysis is more stable than the (fastest implementation of) the discrete-time framework. The reason for this is that although Puterman and Brumelle (1979) and Santos and Rust (2004) show that PFI is globally convergent, it does not preserve concavity of the value function even when the true value function is concave, which leads to instability in any algorithm (such as the EGM) that relies on first-order conditions. Consequently, we believe that the continuous-time approach is superior in a two-dimensional setting, although for the algorithms considered here the gains in speed are smaller than those documented in Achdou et al. (2022).

This discussion of the role of sparsity hints at why the relevant tradeoffs can potentially be very different in nonstationary problems. When the agent's problem depends explicitly on age, the transition matrix appearing in the problem remains sparse. However, as we noted in a different context in Phelan and Eslami (2022), sparse solvers slow down dramatically as the dimension of the problem increases, which renders a (naive) application of PFI much less attractive for age-dependent problems. Further, employing PFI ignores two trivial (but important) aspects of lifecycle problems that ought to be exploited wherever possible: age increases monotonically, and the value at death is known *a priori*. These two observations are naturally exploited in the discrete-time setting when we begin at the terminal value and iterate backwards using the EGM, a procedure that we show is much faster than employing PFI or VFI within the MCA framework for the timesteps that ensure convergence.

[6]PFI and MPFI also have essentially identical coding difficulty, because the latter just replaces an inversion step with a finite sum, leaving all other implementation details unchanged.

[7]This is not a new observation and is consistent with Appendix F.1 of Achdou et al. (2022).

We then show that it is possible to jointly exploit the monotonicity of age and the sparseness of the transition matrices within the continuous-time framework using an algorithm that we term *sequential* policy function iteration (SPFI).[8] The basic idea here is to assume that age increases stochastically in a finite grid, and for each value of age, to solve a control problem in which the values for higher ages are viewed as exogenous. This amounts to applying PFI in sequence (hence our name) to each age group separately, and so makes use of the rapid convergence properties of PFI on a series of simpler (two-dimensional) problems instead of ever attempting to invert a large (three-dimensional) system.

Equipped with this algorithm, we then arrive at interesting tradeoffs between continuous time and discrete time in the nonstationary setting. Although literally applying PFI or VFI to a continuous-time MCA problem is never better than iterating backwards using EGM, the comparison with SPFI is more subtle. On the one hand, for a concave income fluctuation problem with a finite life-span, employing the method of Carroll (2006) and iterating backwards from the terminal date is typically faster than any continuous-time method. However, because it relies on the ability to eliminate a root-finding step, it is also less generally applicable. We illustrate this point by enriching the lifecycle problem in the spirit of Hall and Jones (2007) to allow the agent to affect her life-span through expenditures on healthcare and to make a binary labor-leisure choice. In this setting, the problem of the agent at each date is not necessarily jointly concave in consumption and healthcare expenditures, even if the continuation value is concave in wealth (which also is not guaranteed). Consequently, the first-order conditions in the discrete-time problem may fail to characterize the optimal control. In contrast, no such subtleties arise in the continuous-time approach, because the state only transitions to adjacent points and the concavity and regularity of the value function are irrelevant. It is here that the flexibility of the MCA approach becomes particularly useful. We show that among the (infinitely) many Markov chains that may be used to approximate the continuous-time value function, we can choose one that ensures additive separability of the agent's objective in consumption and healthcare. Further, this procedure does not require delicate choices of grids (which are unchanged relative to the simpler problem) or any interpolation step. For lifecycle problems, we therefore conclude that discrete-time methods are faster for (simple) concave problems but also less general for more complicated extensions.

Related literature The general theory of discrete-time dynamic programming and the solution methods that we use in this paper are outlined in the seminal contributions of Bellman (1954), Howard (1960), Puterman and Brumelle (1979), and Puterman and Shin (1978). For textbook treatments with a particular focus on economic applications, we refer

[8]This is a particular case of what Kushner and Dupuis (2001) refer to as an implicit method, and to the best of our knowledge has not been exploited within the economics literature.

the reader to Stokey et al. (1989) and Stachurski (2009). Within applied mathematics, a vast literature has studied numerical methods for the solution of partial differential equations, such as the Hamilton-Jacobi-Bellman (HJB) equations that characterize the solutions to optimal control problems. In a general and abstract setting, Barles and Souganidis (1991) provide sufficient conditions for the convergence of finite-difference schemes to solutions of Hamilton-Jacobi-Bellman equations. Candler (2001) is perhaps the earliest application of finite-difference methods to economic models. More recently, Achdou et al. (2022) apply finite-difference schemes to a canonical macroeconomic model with incomplete markets in the spirit of Huggett (1993) and Aiyagari (1994).

Turning to discrete-time solution methods, Carroll (2006) introduces the endogenous grid method (EGM) and applies it to a neoclassical growth model. The EGM eliminates the need to solve a nonlinear equation when updating the policy function and so is substantially faster than brute force VFI. For this reason, when comparing frameworks, we will use the EGM as a benchmark for discrete-time methods. Many papers have extended Carroll (2006) to more complicated settings. Barillas and Fernández-Villaverde (2007) incorporate labor-leisure choice, Hintermaier and Koeniger (2010) allow for occasionally binding constraints, Fella (2014) considers a problem with durable consumption, while White (2015), Druedahl and Jørgensen (2017), and Iskhakov et al. (2017) consider multi-dimensional problems.

Phelan and Eslami (2022) show that the implicit finite-difference scheme may be viewed as a limiting form of PFI applied to a particular Markov chain in the MCA approach. In Phelan and Eslami (2022) we fix a discretization and compare the performance of PFI, MPFI, and a novel generalization of the latter, and do not explore the costs and benefits of different Markov chains or compare with discrete-time methods. Despite the extensive literature on both approaches, relatively few papers attempt to explicitly compare continuous-time and discrete-time approaches or provide guidance to the quantitative researcher in economics. One recent exception is the aforementioned Rendahl (2022). We build on the insights of Rendahl (2022) and highlight the role of nonstationarity, because this (natural) extension can substantially change the speed versus accuracy tradeoff facing the modeler.

Specification note. All computations in this paper were performed using a Intel Core i7-8650U processor with no parallelization, using the standard libraries in Python.[9] Since absolute running times are potentially subject to substantial idiosyncratic variation, we will focus primarily on relative run times as we vary the algorithms used.

Outline of paper Section 2 considers a stationary income fluctuation problem in both discrete and continuous time; Section 3 considers a nonstationary (lifecycle) problem, com-

[9] Further details on the implementation may be found with the replication files located at https://github.com/tphelanECON/The_Art_of_Temporal_Approximation_WP.

paring the continuous- and discrete-time approaches, and provides an example with endogenous mortality and health expenditures. Section 4 concludes.

2 Stationary problems

In this section, we consider stationary income fluctuation problems, in which an infinitely lived agent discounts at a constant rate and makes consumption-savings decisions facing a stationary income process. Section 2.1 considers a discrete-time setting, and Section 2.2 considers the continuous-time counterpart. Section 2.3 computes examples and records the speed and accuracy of various numerical methods.

2.1 Discrete time

We first suppose that time is discrete and assumes values in the set $\{0, \Delta_t, 2\Delta_t, \ldots\}$ for some fixed $\Delta_t > 0$. We explicitly allow the timestep Δ_t to be a parameter in order to facilitate comparison with the continuous-time environment considered in Section 2.2.[10] An infinitely lived agent has preferences over sequences of consumption $c := (c_n)_{n=0}^{\infty}$ given by

$$U(c) = \mathbb{E}\left[\sum_{n=0}^{\infty} e^{-\rho n \Delta_t} \Delta_t u(c_n)\right]. \tag{2.1}$$

In the objective function in (2.1), the flow utility function u is assumed to take a constant relative risk aversion (CRRA) form, so that $u(c) = c^{1-\gamma}/(1-\gamma)$ for some $\gamma > 0$. Each period's subjective discount factor is $e^{-\rho}$, so that the agent discounts by $e^{-\rho\Delta_t}$ between successive periods.

The agent's income during period n is $y(z) = \overline{y}e^{z_n}$ for some $\overline{y} > 0$, where $(z_n)_{n=1}^{\infty}$ is a Markov chain assuming values in a compact set $\mathcal{Z}$. We denote the transition kernel of this chain by Γ, so that for any $z \in \mathcal{Z}$, the future value z' is distributed according to the measure $\Gamma(z, dz')$.

The agent can either consume her income or save it in a risk-free asset, which cannot fall below an *ad hoc* exogenous level $\underline{b}$, and evolve according to

$$b_{n+1} = (1 + \Delta_t r)[b_n + \Delta_t(\overline{y}e^{z_n} - c_n)],$$

provided that consumption is chosen such that $b_{n+1} \geq \underline{b}$ for all $n \geq 1$ almost surely.[11]

[10] This also allows us to use the same timestep in our lifecycle problems, ensuring that the choice of the temporal grid is not driving the results on relative performance.

[11] We therefore interpret b_n as the level of assets at the beginning of period n, prior to the realization of income. Consumption is chosen after z_n is realized and uniquely determines future assets b_{n+1}.

The problem of an agent with assets b and log income z is, then

$$
\begin{aligned}
V(b,z) = \max_{(c_n,b_n)_{n=0}^{\infty}} \quad & \mathbb{E}\left[\sum_{n=0}^{\infty} e^{-\rho n \Delta_t}\Delta_t u(c_n)\right] \\
\text{s.t.} \quad & b_{n+1} = (1+\Delta_t r)\left[b_n + \Delta_t(\overline{y}e^{z_n} - c_n)\right], \\
& b_{n+1} \geq \underline{b},\ z_{n+1} \sim \Gamma(\cdot \mid z_n), \\
& (b_0, z_0) = (b,z).
\end{aligned}
\tag{2.2}
$$

In what follows we will refer to V as the *value function* of the agent.

We will assume that $r < \rho$, which ensures that for some $\overline{b} > \underline{b}$ we have dissaving for all $b > \overline{b}$, and so there is no loss in assuming that the state space for the agent is the compact set $\mathcal{G} := \mathcal{B} \times \mathcal{Z}$ where $\mathcal{B} := [\underline{b}, \overline{b}]$.[12] If we denote by $\mathcal{C} := C(\mathcal{G})$ the set of real-valued continuous functions on $\mathcal{G}$, then an application of the Principle of Optimality to problem (2.2) implies that V is the unique function in $\mathcal{C}$ that solves the *Bellman functional equation* (BFE):

$$
\begin{aligned}
V(b,z) = \max_{c\geq 0, b' \in \mathcal{B}} \quad & \left\{\Delta_t u(c) + e^{-\rho\Delta_t}\mathbb{E}_z\left[V(b',z')\right]\right\} \\
\text{s.t.} \quad & b' = (1+\Delta_t r)\left[b + \Delta_t(\overline{y}e^{z} - c)\right],
\end{aligned}
\tag{2.3}
$$

where $\mathbb{E}_z[V(b',z')] = \int_{\mathcal{Z}} V(b',z')\Gamma(z,dz')$ denotes the conditional expectation operator.

We can write the functional equation in equation (2.3) as

$$
V(b,z) = B[V](b,z) \tag{2.4}
$$

where the Bellman operator B is defined according to

$$
\begin{aligned}
B[V](b,z) = \max_{c\geq 0, b' \in \mathcal{B}} \quad & \left\{\Delta_t u(c) + e^{-\rho\Delta_t}\mathbb{E}_z\left[V(b',z')\right]\right\} \\
\text{s.t.} \quad & b' = (1+\Delta_t r)\left[b + \Delta_t(\overline{y}e^{z} - c)\right].
\end{aligned}
\tag{2.5}
$$

The operator B is a contraction with modulus $e^{-\rho\Delta_t}$ on $\mathcal{C}$ by Blackwell's conditions (see Blackwell (1965)), and so by the contraction mapping theorem, there exists a unique solution to the functional equation (2.4), that coincides with the value function of the agent.

The fact that the state space is continuous implies that we must somehow discretize it in order to approximate a solution. To this end, we construct finite grids for the state space and, if necessary, replace the Markov process for log income with a finite-state Markov chain. In order to facilitate comparison with the continuous-time approach, we index all quantities in this approximation by a parameter $h > 0$ indicating the fineness of the approximation.

[12] Since the utility function is of the CRRA form, this dissaving for sufficiently high asset values follows from Proposition 4 in the working paper version of Aiyagari (1994).

For a pair of integers $N = (N_b^h, N_z^h)$, we define the increments in assets $\Delta_b^h := (\bar{b} - \underline{b})/N_b^h$ and log income $\Delta_z^h := (\bar{z} - \underline{z})/N_z^h$, and use these to define equispaced grids for assets and log income,

$$\begin{aligned} \mathcal{B}^h &:= \left\{ \underline{b} + i\Delta_b^h \,\middle|\, i = 0, 1, \dots, N_b^h \right\} \\ \mathcal{Z}^h &:= \left\{ \underline{z} + j\Delta_z^h \,\middle|\, j = 0, 1, \dots, N_z^h \right\} \end{aligned} \tag{2.6}$$

and replace the continuous-state process for log income with a finite-state process on $\mathcal{Z}^h$ with transition kernel Γ^h. Note that because we include the endpoints in (2.6) of the state space in assets and income, we have N_b^h+1 and N_z^h+1 points in each dimension, respectively. We write $\mathcal{G}^h := \mathcal{B}^h \times \mathcal{Z}^h$ and denote the approximate value function defined on $\mathcal{G}^h$ by V^h. We are then searching for a solution to the *finite-state* BFE

$$V^h(b, z) = B^h[V](b, z), \tag{2.7}$$

where B^h now represents the discrete Bellman operator

$$\begin{aligned} B^h[V](b, z) = \max_{c \geq 0, b' \in \mathcal{B}} \quad & \left\{ \Delta_t u(c) + e^{-\rho \Delta_t} \mathbb{E}^h[\tilde{V}^h(b', z')] \right\} \\ \text{s.t.} \quad & b' = (1 + \Delta_t r) \left[b + \Delta_t (\bar{y} e^z - c) \right], \end{aligned} \tag{2.8}$$

and $\tilde{V}^h$ is the linear interpolant of V^h and $\mathbb{E}^h[\tilde{V}^h(b', z')] = \sum_{z' \in \mathcal{Z}} \Gamma^h(z, z')\tilde{V}^h(b', z')$ is the relevant expectation operator. The Bellman equations for the discrete-time income fluctuation problems that we solve are all of the form given in equations (2.7) and (2.8). In our applications, we will either use VFI, MPFI or PFI to update the value function, and the endogenous grid method of Carroll (2006) to update the policy function.[13]

2.2 Continuous time

We now turn to a continuous-time formulation of the income fluctuation problem. The agent's preferences over a consumption process $(c_t)_{t \geq 0}$ are now represented by

$$U(c) = \mathbb{E}\left[\int_0^\infty e^{-\rho t} u(c_t) dt \right], \tag{2.9}$$

for the same discount rate ρ and flow utility function u as in equation (2.1).

As before, we assume that income is given by $y(z_t) = \bar{y} e^{z_t}$ for some constant $\bar{y} > 0$ and Markov process $(z_t)_{t \geq 0}$, but, now, we assume that the latter is a diffusion process restricted

[13] These algorithms are well-known and so formal statements are relegated to Appendix A.1.

to some compact interval $[\underline{z}, \overline{z}]$ evolving according to

$$dz_t = \mu(z_t)dt + \sigma d\omega_t, \tag{2.10}$$

where $(\omega_t)_{t\geq 0}$ is a standard Brownian motion, μ is a real-valued function representing the drift of the process, and $\sigma \geq 0$ is a constant capturing the diffusion.

Given assets b and log income z, the problem of the agent can be written as

$$\begin{aligned} V(b,z) = \max_{(c_t)_{t\geq 0}} \quad & \mathbb{E}\left[\int_0^\infty e^{-\rho t}u(c_t)dt\right] \\ \text{s.t.} \quad & db_t = [rb_t + y(z_t) - c_t]dt, \\ & dz_t = \mu(z_t)dt + \sigma d\omega_t, \\ & b_t \geq \underline{b}. \end{aligned} \tag{2.11}$$

As with the discrete-time framework, we will refer to V as the *value function* of the agent. Assuming, as before, that $r < \rho$, the agent dissaves for assets exceeding some $\overline{b} > \underline{b}$. Therefore, there is again no loss in restricting attention to the compact interval $[\underline{b}, \overline{b}]$.[14] This ensures that the above problem is finite valued and that there are no subtleties in the application of the Principle of Optimality.

It can be shown (see, e.g., Achdou et al. (2022)) that the individual's value function is the unique (viscosity) solution to the following partial differential equation (PDE) known as the *Hamilton-Jacobi-Bellman equation* (HJB):

$$\rho V(b,z) = \max_{c\geq 0}\left\{u(c) + [rb + y(z) - c]V_b(b,z) + \mu(z)V_z(b,z) + \frac{\sigma^2}{2}V_{zz}(b,z)\right\}, \tag{2.12}$$

where V_b, V_z and V_{zz} denote partial derivatives of function V with respect to its first and second arguments, respectively. A common approach in the numerical analysis literature is to discretize the state space and to replace the derivatives in (2.12) by their one-sided or two-sided finite-difference approximations over this discrete grid. Kushner and Dupuis (2001) show that this approach is a special case of approximating the stochastic process governing the states with a *locally consistent* discrete-time Markov chain and then solving a control problem in which the states are governed by this approximating Markov chain.

To construct this approximating chain, one has to discretize the *temporal* state-space—*i.e.*, time—and construct a Markov chain that *locally* resembles the original continuous process. We will only outline the approximation procedure here and will refer the reader to Kushner and Dupuis (2001) for details. In this section, we outline one possible construction

[14]The fact that the agent dissaves for sufficiently large assets is part 1. of Proposition A.2 in Appendix G.1 of Achdou et al. (2022).

for the above income fluctuation problem. We first describe the requirements of "local consistency" in a general framework before specializing to the current setting.

To this end, we write the law of motion of the state as

$$dx_t := \mu(x_t, c_t)dt + \Sigma d\omega_t, \tag{2.13}$$

where x_t is the individual's state vector, c_t is consumption, μ and Σ are the drift and diffusion of the state vector, respectively, and ω is standard two-dimensional Brownian motion. The state vector x takes values in the set $\mathcal{G} = \mathcal{B} \times \mathcal{Z}$ and the control can take values in some subset $\mathcal{U} \subseteq \mathbb{R}$. An approximation of a continuous-path stochastic process with a discrete process requires a discretization of the state-space. In order to facilitate comparison, we choose the same discretization as in the discrete-time setting of Section 2.1, and again index all quantities by $h > 0$. The increments in assets and log income are again denoted by Δ_b^h and Δ_z^h, respectively, and the associated grids are defined as in (2.6).

Our next goal is to approximate, for each process $(c_t)_{t\geq 0}$, the state variable $x_t = (b_t, z_t)'$ by a controlled, discrete-time Markov chain, $(\xi_n^h)_{n=1}^{\infty} = (\xi_{b,n}^h, \xi_{z,n}^h)_{n=1}^{\infty}$, assuming values in the set $\mathcal{G}^h$ under an admissible control process denoted by χ_n^h. In this representation, χ_n^h is a random variable that determines the control, c, at any discrete time n, and is once again assumed to lie in the set $\mathcal{U}$. We will assume that the approximating chain and control process change values at discrete times indexed by n. The length of each time interval can be a function of the state and control, and will be denoted $\Delta t^h : \mathcal{G}^h \times \mathcal{U} \mapsto \mathbb{R}_{++}$.

Intuitively, one can think of this exercise as an attempt to "mimic" the sample paths of the original process in (2.13) probabilistically. Two cross-sections are separated by a time interval of length $\Delta t^h(x, c)$—which we will refer to as the *interpolation interval.* For this to be the case, the approximating Markov chain must have the same "local behavior" as the diffusion process, in the following sense,

$$\mathbb{E}_{x,n}^{h,c}\left[\Delta\xi_n^h\right] = \mu(x, c)\Delta t^h(x, c) + o\Big(\Delta t^h(x, c)\Big), \tag{2.14a}$$

$$\mathbb{E}_{x,n}^{h,c}\left[\widehat{\Delta\xi_n^h}\widehat{\Delta\xi_n^h}'\right] = \Sigma\Sigma'\Delta t^h(x, c) + o\Big(\Delta t^h(x, c)\Big), \tag{2.14b}$$

where $\Delta\xi_n^h := \xi_{n+1}^h - \xi_n^h$ and $\widehat{\Delta\xi_n^h} = \Delta\xi_n^h - \mathbb{E}_{x,n}^{h,c}[\Delta\xi_n^h]$ are the original and demeaned increments, respectively, and $\mathbb{E}_{x,n}^{h,c}$ denotes expectations conditional on $(\xi_n^h, \chi_n^h) = (x, c)$. Kushner and Dupuis (2001) refer to equations (2.14a) and (2.14b) as *local consistency conditions*, and show that the value function for the discrete problem, V^h, converges to the value function for the original problem, V, as $h \to 0$, provided that the local consistency conditions are satisfied for the associated sequence of chains.

As emphasized above, there are many ways in which one can construct a finite-state

Markov chain satisfying these local consistency conditions. One possibility is to define, at time n, for any current value of $\xi_n^h = x \in \mathcal{G}^h$ and control c, the transition probabilities

$$
\begin{aligned}
p^h\left(b \pm \Delta_b^h, z \,\middle|\, c\right) &= \left[\frac{\Delta t^h(x,c)}{\Delta_b^h}\right][rb + y(z) - c]^{\pm}, \\
p^h\left(b, z \pm \Delta_z^h \,\middle|\, c\right) &= \left[\frac{\Delta t^h(x,c)}{(\Delta_z^h)^2}\right]\left[\sigma^2/2 + \Delta_z^h[\mu(z)]^{\pm}\right],
\end{aligned} \tag{2.15}
$$

where for any scalar x we write $[x]^{\pm} \coloneqq \max\{\pm x, 0\}$. The probability of no transition is then defined in such a way that the transition probabilities sum to unity at every point:

$$
p^h(b, z \mid c) = 1 - \sum_{(b',z') \in \mathcal{G}^h \setminus (b,z)} p^h\left(b', z' \mid c\right). \tag{2.16}
$$

The interpolation interval can be any positive function, as long as the transition probabilities remain nonnegative.[15] In view of the expressions in (2.15), the requirement that the transition probabilities be nonnegative evidently places an upper bound on the possible values of the interpolation interval. This restriction is reminiscent of the Courant-Friedrichs-Lewy (CFL) condition that plays an important role in the finite-difference literature.[16]

If the agent's state vector is governed by the Markov chain defined by (2.15) and her control action remains constant over discrete time intervals, her value function must solve

$$
V^h(x) = B^h[V^h](x), \tag{2.17}
$$

for all $x \in \mathcal{G}^h$, where the operator B^h is defined by

$$
B^h[V^h](x) = \max_{c \geq 0} \left\{ \Delta t^h(x,c) u(c) + e^{-\rho \Delta t^h(x,c)} \sum_{x' \in \mathcal{G}^h} p^h\left(x, x' \mid c\right) V^h\left(x'\right) \right\}, \tag{2.18}
$$

The algorithms we employ to solve Bellman equations of the form (2.18) are standard and so are relegated to Appendix A.2.

The above procedure replaces the original continuous-time control problem with a discrete-time control problem. Researchers unfamiliar with this method may then wonder why this is useful, given that we could have simply started with the discrete-time problem in Section 2.1 and the more familiar Bellman operator in equation (2.8). The key point here is that the global "shape" or "regularity" of the value function is entirely irrelevant to the optimal choice of consumption on the right-hand side of (2.18), because only the current and

[15] In our applications, we will set the interpolation interval to a constant.

[16] See, e.g., Candler (2001) for further discussion. We do not draw upon finite-difference (FD) arguments in this paper but we do wish to emphasize that both the FD and MCA approaches require similar restrictions on the relative sizes of the increments in time and the state.

adjacent values $V^h\big(b-\Delta_b^h, z\big), V^h(b,z)$ and $V^h\big(b+\Delta_b^h, z\big)$ enter into the calculation. As the agent considers various choices of consumption, the possible future asset values remain fixed and the transition probabilities vary. Consequently, it is the shape of the exogenously known *transition functions* (and not the unknown *value function*) that determines the ease of solving the first-order conditions. However, in contrast with the discrete-time setting, some restrictions are necessary on the size of the timestep in order to ensure that this is a well-defined control problem. This last point will become particularly relevant in our analysis of lifecycle problems.

2.3 Numerical results

In this section, we illustrate the performance of the discrete-time and continuous-time algorithms. We first plot an example, before considering accuracy and run times as we vary grid sizes and the timestep, holding all other parameters fixed.

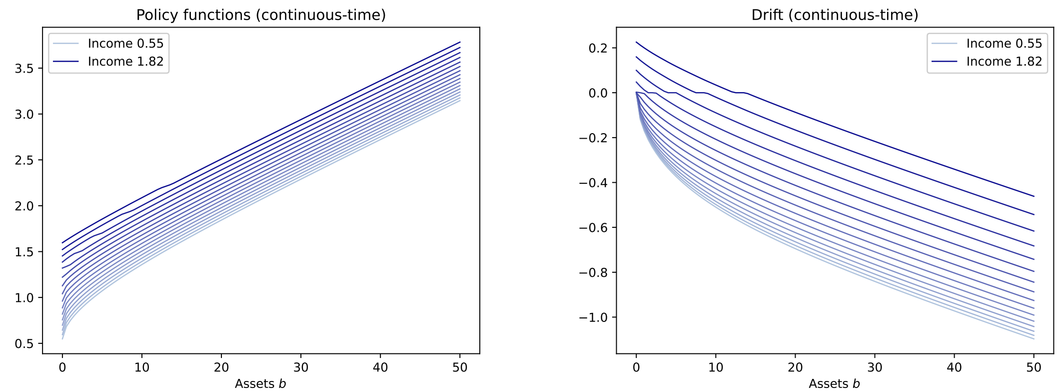
The parameters for this example are taken from Achdou et al. (2022) wherever possible. For the utility function, $u(c) = c^{1-\gamma}/(1-\gamma)$, we choose a coefficient of relative risk aversion of $\gamma = 2$. For income, $y(z) = \overline{y}e^z$, we set $\overline{y} = 1$, and follow Achdou et al. (2022) in assuming that log income evolves according to a continuous-time AR(1) (Ornstein-Uhlenbeck) process satisfying $dz_t = -\overline{\mu}z_t dt + \sigma d\omega_t$ for some Brownian motion ω_t. In this case the stationary distribution of z_t is Gaussian with mean zero and variance $\nu^2 = \sigma^2/(2\overline{\mu})$. The parameters we choose are given in Table 1.

Table 1: Parameters used in numerical example

Parameter	Description	Value
γ	CRRA parameter	2
$\overline{\mu}$	coeff. of income drift	$-\ln(0.95)$
ν	s.d. of log income	0.2
ρ	subjective discount rate	$1/0.95-1$
r	risk-free rate	0.03
$\underline{b}$	borrowing constraint	0.0
$\overline{b}$	upper bound on assets	50
ϵ	tolerance	10^{-8}
$\overline{z}, \underline{z}$	log income bounds	$\pm 3\nu$
Δ_t	timestep in MCA	10^{-6}

We use $\Delta_t = 10^{-6}$ and $\Delta_t = 1$ for the "true" solutions in the continuous-time and discrete-time cases, respectively, and experiment with other timesteps below. We choose the income process in the discrete-time setting to be that implied by the continuous-time choice sampled at unit intervals (representing an annual frequency in our examples).[17] Figure 1 depicts the policy functions and drifts computed using PFI in the continuous-time framework with grid size $(N_b, N_z) = (100, 15)$.

Figure 1: Example policy function and drift



Note that there is a slight "kink" in the policy functions at points at which the associated drift is zero. This is to be expected, because the transition probabilities are nonsmooth in consumption when the drift is zero (due to the maximum operators in (2.15)). The drift is negative for all income levels at the upper bound, and so there is no loss in restricting attention to the interval [0,50]. Further, the borrowing constraint is binding for the lowest income level and the derivative of the drift appears to diverge at this point, which is consistent with Proposition A.1 in Appendix G.1 in Achdou et al. (2022). We now pursue a more systematic analysis of the accuracy and run times of various algorithms.

We first compare the accuracy of the discrete-time and continuous-time algorithms as we vary the timestep and the size of the grid. In order to assess accuracy, we follow Achdou et al. (2022), solve each problem on a fine grid for assets (here $N_b = 5000$) and regard the computed policy and value functions as the "true" values. For the discrete-time problem, we use brute force with $N_c = 5000$ consumption points between 10^{-6} and $2 \times \max_{\mathcal{G}^h} \{c_{\text{NS}}\}$, where c_{NS} denotes the level of consumption consistent with no saving. In both cases we use PFI to update the value function, which, by the results of Santos and Rust (2004), is guaranteed to converge in this setting. We then compare the policy functions computed on coarser grids to these "true" quantities and document the percentage differences in both the l_1 and l_∞ norms (mean and maximum differences).

For the discrete-time framework this requires only varying the asset grid, because the timestep grid is interpreted as a parameter of the model and not a choice made by the

[17] That is, the transition matrix in the discrete-time setting is obtained by iterating the transpose of the continuous-time transition matrix $1/10^{-6}$ times and taking the transpose again.

modeler. Table 2 performs this exercise for the discrete-time approach. As expected, the policy functions appear to converge monotonically to their respective "true" values as the asset grid becomes finer.

Grid size	$\|\|\Delta c\|\|_1$	$\|\|\Delta c\|\|_\infty$	$\|\|\Delta c(\%)\|\|_1$	$\|\|\Delta c(\%)\|\|_\infty$
(25, 15)	0.0215	0.0931	1.0022	9.0893
(50, 15)	0.0099	0.0580	0.4720	6.2129
(100, 15)	0.0032	0.0380	0.1711	4.0583
(250, 15)	0.0007	0.0168	0.0394	2.6622
(500, 15)	0.0004	0.0132	0.0192	2.0689

Table 2: Accuracy of discrete-time EGM ($\Delta_t = 1$)

Grid size	$\|\|\Delta c\|\|_1$	$\|\|\Delta c\|\|_\infty$	$\|\|\Delta c(\%)\|\|_1$	$\|\|\Delta c(\%)\|\|_\infty$
(25, 15)	0.0633	0.0942	2.9569	12.8105
(50, 15)	0.0327	0.0619	1.5401	9.2201
(100, 15)	0.0165	0.0415	0.7845	6.5738
(250, 15)	0.0065	0.0249	0.3109	4.1525
(500, 15)	0.0031	0.0170	0.1487	2.9153

Table 3: Accuracy of continuous-time approach ($\Delta_t = 10^{-6}$)

In the continuous-time approach we also face the additional choice of the time increment. Table 3 repeats the exercise conducted in Table 2 for the continuous-time framework with $\Delta_t = 10^{-6}$, the same value as in the "true" solution, and reveals that the discrete-time approach with EGM is more accurate in both the l_1 and l_∞ norms for a given grid.

We now turn to an analysis of the speed of convergence. For both the discrete-time and continuous-time frameworks, we consider VFI, MPFI, and PFI with a direct sparse solver, and record the average of 10 runs.[18] In all cases, we assume the initial guess of the policy function or value function is that corresponding to zero net savings. It is well-known that the speed of convergence in PFI remains bounded as the time-step vanishes. However, the same is not true for VFI and MPFI, the convergence of which becomes arbitrarily slow as the discount rate vanishes. Because we wish to compare the relative merits of VFI, MPFI, and PFI, we will also consider larger timesteps than the above value of $\Delta_t = 10^{-6}$. In order

[18] We use the standard sparse solver from the Scipy library (scipy.sparse.linalg.spsolve) for all of our sparse calculations.

for the expressions in (2.16) and (2.15) to define a transition kernel, the timestep must be sufficiently small that all probabilities lie within the unit interval, and because the optimal consumption policy is *a priori* unknown, so too is the maximum level of the timestep. For the parameters given in Figure 1, probabilities fail to remain in the unit interval for the finest grid above when $\Delta_t \approx 0.09$. In order to explore the (possible) benefits of MPFI in the continuous-time setting we will consider the slightly smaller value $\Delta_t = 0.05$.[19]

Table 4 documents the speed of convergence in the continuous-time model for this larger timestep. In this case, the Bellman operator is a contraction with modulus $e^{-\rho\Delta_t} \approx 1$, and so VFI converges slowly.[20] In contrast, convergence is rapid with PFI, and the speed of convergence is not particularly sensitive to the timestep when the latter is small. For every grid in Table 4, the most rapid convergence occurs using PFI with a direct sparse solver, and so the above discussion of the bounds on the timestep that ensure convergence is therefore moot in this context, because it is best to choose a minuscule timestep and employ a direct sparse solver with PFI.

Grid size	VFI	MPFI(10)	MPFI(50)	MPFI(100)	MPFI(200)	PFI
(25, 15)	10.656	1.154	0.304	0.180	0.119	0.021
(50, 15)	11.329	1.251	0.339	0.210	0.139	0.032
(100, 15)	13.017	1.473	0.421	0.262	0.186	0.042
(250, 15)	18.208	2.140	0.640	0.419	0.317	0.085
(500, 15)	29.400	3.512	1.066	0.725	0.546	0.166

Table 4: Speed of convergence (continuous-time, $\Delta_t = 0.05$)

Table 5 documents the time until convergence for the discrete-time problem on various different grids. In each case, the policy function is updated using the EGM, while the columns indicate different algorithms for updating the value function. Note that for every grid considered, PFI is more than an order of magnitude faster than VFI, and that there are some implementations of MPFI that outperform PFI.

[19]Table 10 and Table 11 in Appendix B.1.1 show that the accuracy of the continuous-time approach is not particularly sensitive to variation in the timestep.

[20]This is related to the observation of Achdou et al. (2022) that the explicit method is typically slow to converge. See Rendahl (2022) for further discussion relating the explicit method with VFI.

Grid size	VFI	MPFI(10)	MPFI(50)	MPFI(100)	MPFI(200)	PFI
(25, 15)	2.641	0.304	0.121	0.130	0.148	0.129
(50, 15)	2.843	0.333	0.136	0.150	0.177	0.143
(100, 15)	3.360	0.403	0.162	0.186	0.229	0.177
(250, 15)	4.642	0.587	0.240	0.288	0.385	0.287
(500, 15)	9.644	1.216	0.495	0.523	0.678	0.574

Table 5: Speed of convergence (discrete-time, $\Delta_t = 1$, EGM)

Speed versus accuracy The above tables document the time until convergence and accuracy attained on various grids. However, the relevant tradeoff to the practitioner is not speed versus grid size, but speed versus accuracy. This is an important distinction, because, as the above shows, the accuracy attained on a given grid differs across the two frameworks. Figure 2 depicts the accuracy attained within a particular time in each framework, for both mean and maximum percentage differences (i.e. the l_1 and l_∞ norms), using the "true" timestep $\Delta_t = 10^{-6}$ for the continuous-time case.

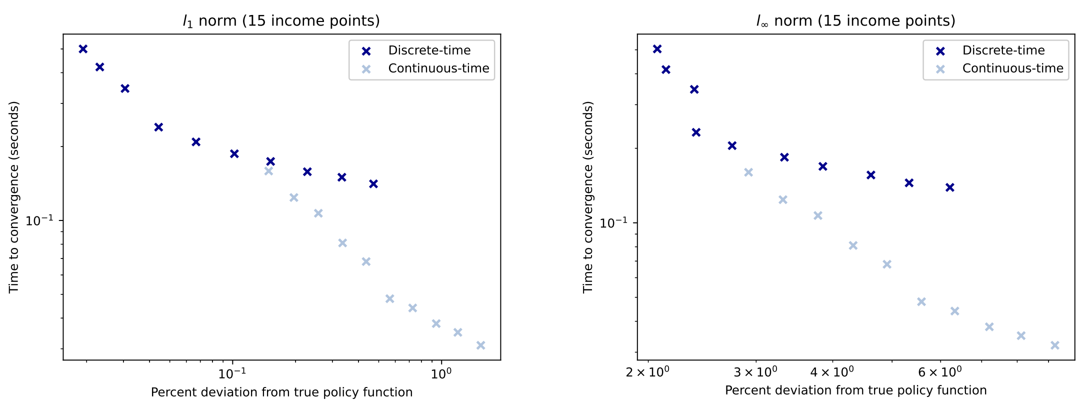


Figure 2: Time versus accuracy (15 income points)

It is important to emphasize some countervailing effects here. On the one hand, the continuous-time approach converges in a shorter time for any fixed grid. This seems intuitive to us, because the discrete-time approach requires an interpolation step that is unnecessary in the continuous-time case, and the transitions in assets and income are nonlocal, which slows down the application of the sparse solver. However, the discrete-time approach is more accurate for any fixed grid. Figure 2 shows that the continuous-time approach is faster for any given level of accuracy in this example (the light blue dots are typically below the dark blue dots), but this does not appear to be true for all parameters.[21] Since the continuous-time approach is slightly faster and guaranteed to converge while the discrete-time approach is not, we believe that the former is superior in this context.

Discussion As we noted in the introduction, all of the algorithms we consider in this paper consist of two distinct steps: updating the policy function and updating the value function. In the above continuous-time problems, the drift in wealth was linear in consumption, which meant that we could find transition probabilities that were also linear in consumption, and that led to simple first-order conditions to update the policy function. Further, for the above grids the direct sparse solver converged rapidly in the value updating step.

However, in the discrete-time setting the EGM allows us to update the policy function without solving a nonlinear equation, and Rendahl (2022) shows how one can use the observation of Young (2010) to represent the discrete-time BFE in terms of sparse matrices. Consequently, two of the benefits often cited for the continuous-time framework are sometimes shared by the discrete-time framework. While this is true, we emphasize that in the far right column of Table 5 we employed Algorithm 3, which combines PFI with the EGM. Although for the above parameters this algorithm converged, this is not guaranteed.[22] Indeed, we have observed a failure of convergence for lower γ, which appears intuitive to us because this reduces the curvature in the utility function. We therefore believe that for a stationary problem, the continuous-time framework is superior to the discrete-time framework, as it is (usually) faster and is guaranteed to converge.

However, a key point of this paper is that this intuition does not extend to all environments that often arise in quantitative macroeconomics. Indeed, we show in Section 3 that the discrete-time approach is often superior in settings in which age enters explicitly. We now provide a preview as to why this is the case. First, as we note in a different context in Phelan and Eslami (2022), standard sparse solvers' performance declines rapidly when the transition matrix grows larger or becomes less sparse. In Phelan and Eslami (2022), this occurred when we allowed for durable consumption or a richer (two-dimensional) process for log income, in which case the MPFI dominates PFI and VFI in terms of the speed of convergence. This lack of sparsity also occurs naturally when age enters separately as a state variable, as in Section 3. Consequently, treating age like any other state variable in the MCA construction and applying PFI is very time-consuming.

Second, there are two trivial (but important) features of lifecycle models that ought

[21] Appendix B.1.3 depicts the speed versus accuracy tradeoff for the cases of 5 and 25 income points, respectively, and finds that although the continuous-time approach is typically faster for a given level of accuracy, this is not true for all parameters and grids.

[22] Puterman and Brumelle (1979) show that PFI converges monotonically to the solution of the Bellman equation, but this does not mean that Algorithm 3 converges, because it updates the policy function by seeking the solution to first-order conditions, and the iterates generated by PFI need not be concave.

to be exploited in the numerical method, namely, that age always increases and that the value function at death is known. Both of these observations suggest that simply iterating backwards in discrete time from the terminal date may be superior, regardless of the sparsity structure of the system.

Third, we noted above that the iterates in PFI need not be concave, and that this causes a problem when combined with EGM. However, this not a problem in the lifecycle model, because we never use PFI in the discrete-time setting and instead iterate backwards. Two of the benefits of the continuous-time approach in the stationary setting therefore do not appear to carry over to the nonstationary setting.

3 Nonstationary problems

The foregoing examples show that the continuous-time framework is typically superior, from a computational intensity viewpoint, when dealing with stationary income fluctuation problems. The reason for this boils down to the fact that policy function iteration may be employed more easily in this setting. This may give one the impression that continuous-time formulations are always superior. In this section, we show that this intuition does not extend to *nonstationary* problems, in which time enters the problem explicitly.

There are several ways in which this could occur. For instance, the income process could depend on an individual's age.[23] To illustrate the role of nonstationarity in affecting the relative costs and benefits of continuous- and discrete-time problems in as simple a way as possible, we consider the problem of an agent with a deterministic finite life-span, running from $A = 0$ to $A = \overline{A}$, who has no descendants. We consider this class of problems for two reasons. First, lifecycle problems often arise in quantitative macroeconomics. Second, and more directly related to this paper, some of the superiority of the continuous-time approach does not obviously apply to this situation, and we wish to explain why and to provide some practical guidance for the choice of a solution method.

We emphasize that nonstationarity does not pose a problem for the general theory of Kushner and Dupuis (2001). The problem still admits a recursive structure, and we can always treat age as if it were just another state variable (like assets or income) and apply the methodology described in Section 2. However, the addition of age can greatly affect the relative speeds of various algorithms. As Phelan and Eslami (2022) note, the appeal of sparse solvers rapidly decreases in three dimensions, suggesting that we might wish to avoid solving linear systems when age enters explicitly. Further, treating age just like any other state and employing PFI does not exploit the (trivial, but important) fact that age

[23]Many such examples recur throughout the quantitative macroeconomics literature. See, e.g., Fella et al. (2019) for a recent example.

only increases. This is useful because the value function at death is known exogenously (and here is zero), and it seems natural to choose an algorithm that exploits this fact.

This section explores the above intuition and documents the speed and accuracy of various solution algorithms. Section 3.1 considers the discrete-time formulation; Section 3.2 considers the continuous-time counterpart and introduces an algorithm termed *sequential* PFI that exploits both the sparse nature of transitions and the monotonicty of age; and Section 3.3 records results for speed and accuracy.

Note that the explicit dependence on age in this environment now makes the comparison of discrete-time and continuous-time environments more delicate. The set of ages attained in the discrete-time environment is a primitive of the problem, but in the continuous-time setting, it is a choice made by the modeler. Thus, to make the two frameworks more comparable, we define the age increment by $\Delta_A = (\overline{A} - 0)/N_A$, for a given integer $N_A \geq 1$, and use the same age grid,

$$\mathcal{A}^h = \{0 + k\Delta_A \mid k = 0, \ldots, N_A\}, \tag{3.1}$$

in both settings, in what follows.

3.1 Discrete time

We first consider the analogue of Section 2.1 in which the agent lives for a finite time. Preferences of the agent over sequences of consumption $(c_n)_{n=0}^{N_A-1}$ are given by

$$U(c) = \mathbb{E}\left[\sum_{n=0}^{N_A-1} e^{-\rho n \Delta_A} \Delta_A u(c_n)\right]. \tag{3.2}$$

Note that we might not have $A = N_A$ because we need not necessarily have $\Delta_A = 1$. In this setting, the timestep and the age step necessarily coincide.[24]

Following the stationary settings, we assume that the agent earns income according to a stochastic process, governed by Γ, while alive, and can save her income in a risk-free asset, denoted by b. We normalize utility upon death to zero.

[24] Note that the preferences in (3.2) indicate that we assume death occurs at $\overline{A}$ not $\overline{A} + \Delta_A$.

The problem of an agent at age $A = N\Delta_A$ with assets b and log income z is then

$$
\begin{aligned}
V(b,z,A) = \max_{\{c_n, b_{n+1}\}_{n=N}^{N_A-1}} \quad & \mathbb{E}\left[\sum_{n=N}^{N_A-1} e^{-\rho n \Delta_A} \Delta_A u(c_n)\right] \\
\text{s.t.} \quad & b_{n+1} = (1+\Delta_A r)\left[b_n + \Delta_A(\overline{y}e^{z_n} - c_n)\right], \\
& b_{n+1} \geq \underline{b},\ z_n \sim \Gamma(\cdot \mid z_{n-1}), \\
& (b_N, z_N) = (b, z).
\end{aligned} \tag{3.3}
$$

The analysis of the discrete-time nonstationary problem requires little discussion. There is now no need to appeal to the Principle of Optimality: Instead we can use the terminal condition, $V(\cdot,\cdot,\overline{A}) \equiv 0$, and iterate backwards in age using the relationship

$$
V(b,z,A) = B[V](b,z,A), \tag{3.4}
$$

for $A \in \mathcal{A}$ with $A \neq \overline{A}$, where the Bellman operator B is defined as

$$
\begin{aligned}
B[V](b,z,A) = \max_{c \geq 0, b' \geq \underline{b}} \quad & \left\{\Delta_A u(c) + e^{-\rho\Delta_A}\int_{\mathcal{Z}} V\big(b', z', A+\Delta_A\big)\Gamma\big(z, dz'\big)\right\} \\
\text{s.t.} \quad & b' = (1+\Delta_A r)\left[b + \Delta_A(\overline{y}e^{z} - c)\right].
\end{aligned} \tag{3.5}
$$

When approximating the value function of the agent, we will iterate backwards from the terminal date, using the endogenous grid method to obtain the optimal policy function at each age.

To illustrate the qualitative features of the problem, before turning to the continuous-time environment and a comparison of the solution methods, we first calculate an example. The parameters are all fixed at the values given in Table 1 in Section 2.3, with the sole additional choice of $\overline{A} = 60$. One may think of this choice as assuming that the agent makes her first consumption-savings decisions at age 20 and dies at age 80.

Figure 3 depicts the difference in assets implied by the optimal consumption function for the discrete-time model, in both the initial period of the nonstationary problem and for the stationary problem. The drift in assets for young agents is qualitatively similar to the analogous figure for the stationary problem given in Section 2. However, the situation is different for older agents, who wish to reduce their wealth to zero immediately prior to death, and so the change in assets is sometimes an order of magnitude larger than that depicted in Figure 3. These large changes in the drift in assets throughout the lifetime of the agent are irrelevant for the discrete-time approach, in which attention need not be confined to local transitions. However, as we shall see in the following sections, this complicates the analysis of the continuous-time problem.

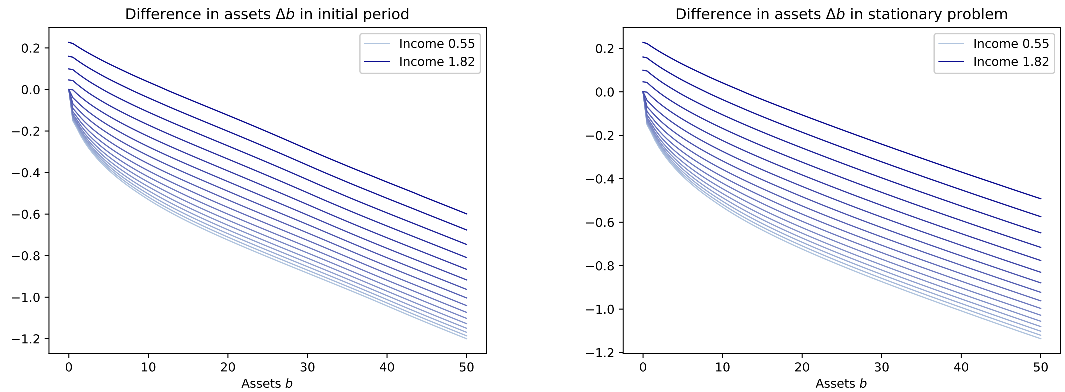


Figure 3: Evolution of assets when young and in a stationary environment

3.2 Continuous time

It is straightforward to extend the environment of the previous section (and that in Section 2.2) to a continuous-time setting. Given assets b and log income z at age A, the agent's problem, in this alternative environment, will be

$$\begin{aligned} V(b,z,A) = \max_{(c_t)_{t\in[A,\overline{A}]}} \quad & \mathbb{E}\left[\int_A^{\overline{A}} e^{-\rho s} u(c_s) ds\right] \\ \text{s.t.} \quad & db_s = [rb_s + y(z_s) - c_s]ds, \\ & dz_s = \mu(z_s)ds + \sigma d\omega_s, \\ & b_s \geq \underline{b}. \end{aligned} \tag{3.6}$$

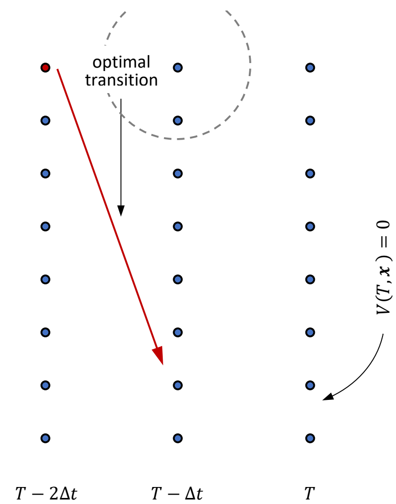
Relative to the stationary problem in Section 2.2, we now face an additional choice as to how to treat age when reducing the above to a finite-state problem. We consider two different approaches. In the first, we assume that age increases with certainty at each step, while in the second, we treat age just as we treat the other state variables, and assume that it only changes values probabilistically.[25]

3.2.1 Deterministic age

We first consider an approach in which age is treated like time and increases deterministically by $\Delta_A = \Delta_t$ at each step of the chain. The remainder of the construction (i.e., the transition probabilities in assets and income) is identical to that given in Section 2.2, so that the transition kernel over assets and income is defined by equations (2.15) and (2.16) with

[25] These two approaches are termed, respectively, "explicit" and "implicit" Markov chain approximation methods by Kushner and Dupuis (2001), following their close relationship with the explicit and implicit finite-difference methods in classical numerical analysis.

Figure 4: Optimal transition for large b, for t near T



$\Delta_t = \Delta_A$. It is straightforward to check that the resulting Markov chain satisfies the local consistency conditions of Kushner and Dupuis (2001).

The associated value function V_{da}^h—where the superscript indicates the dependence on the approximation parameter h and the subscript indicates the deterministic treatment of age—for this discrete-state problem then solves the BFE,

$$V_{\text{da}}^h(b, z, A) = \max_{c \geq 0} \left\{ \Delta_A u(c) + e^{-\rho \Delta_A} \mathbb{E}_{b', z'}^h \left[V_{\text{da}}^h(b', z', A + \Delta_A) \right] \right\}, \tag{3.7}$$

together with the terminal condition $V_{\text{da}}^h\left(\cdot, \cdot, \overline{A}\right) \equiv 0$. The subscripts on the right-hand side of equation (3.7) indicate that the expectation is taken over assets and income but not age.

Although this construction is a natural approach to incorporating the monotonicity of age, it suffers from a serious drawback. As we emphasized in Section 2.2, in order for the MCA approach to lead to a well-defined finite-state problem, the transition probabilities must always lie within the unit interval for any control, which places an upper bound on the size of the timestep. This restriction to adjacent points can cause a problem when the agent wishes to rapidly decumulate assets late in life. In the continuous-time environment, the transition probabilities for assets are proportional to the drift in wealth and the agent decumulates wealth rapidly when old. Figure 4 illustrates why this rapid decumulation causes a problem for the continuous-time approach: the optimal transition is often very far from the current point. This means that either the asset grid must be very coarse (leading to low accuracy) or the timestep must be very small (leading to slow convergence).

For these reasons, iterating backwards from the terminal condition in the above manner is typically a bad idea, as convergence is likely to be slow.

3.2.2 Treating age like any other state variable

An alternative to the approach in Section 3.2.1 is to treat age like any other state variable and assume it transitions stochastically. We choose our transition probabilities such that, if the chain is at point $(b, z, A) \in \mathcal{B}^h \times \mathcal{Z}^h \times \mathcal{A}^h$ at time t, then the possible values at time $t + \Delta_t$ lie in the set

$$\{(b, z, A), (b \pm \Delta_b, z, A), (b, z \pm \Delta_z, A), (b, z, A + \Delta_A)\}.$$

For any control c, we choose the following analogue of (2.15),

$$\begin{aligned} p^h(b \pm \Delta_b, z, A) &= \left(\frac{\Delta_t}{\Delta_b}\right)[rb - c + \overline{y}e^z]^{\pm}, \\ p^h(b, z \pm \Delta_z, A) &= \left(\frac{\Delta_t}{\Delta_z^2}\right)\left[\frac{\sigma^2}{2}\chi(z) + \Delta_z[-\mu z]^{\pm}\right], \\ p^h(b, z, A + \Delta_A) &= \frac{\Delta_t}{\Delta_A}, \end{aligned} \tag{3.8}$$

where $\chi(z) := \mathbb{1}_{z \notin \{\underline{z}, \overline{z}\}}$, provided that all of the transition probabilities lie in the unit interval. Note that this Markov chain satisfies the local consistency conditions by construction. The state constraints are imposed by restricting the possible consumption choices at the boundary points, just as in the stationary case.

The BFE for this discrete problem is then

$$V^h(b, z, A) = \max_{c \geq 0} \left\{\Delta_t u(c) + e^{-\rho\Delta_t}\mathbb{E}^h_{b', z', A'}\left[V^h(b', z', A')\right]\right\}, \tag{3.9}$$

for all $(b, z, A) \in \mathcal{B}^h \times \mathcal{Z}^h \times \mathcal{A}^h$, where the subscripts in the expectation operator on the right-hand side of equation (3.9) make explicit the fact that the expectation is now taken over b', z' and A', in contrast to the BFE given in equation (3.7).

As we noted above, the agent decumulates assets rapidly near the end of her life, and so we may need to choose Δ_t very small to ensure that transition probabilities lie within the unit interval. However, in contrast to Section 3.2.1, there is now no need for the timestep Δ_t to coincide with the age step Δ_A, and we can always make the timestep sufficiently small that the transition probabilities in (3.8) remain within the unit interval.

There remains the question as to how to solve the BFE in (3.9). There are several choices available to us. We can always use VFI and iterate backwards from the terminal payoff, just as in the discrete-time case. However, because of the rapid decumulation of assets in the final period, the timestep necessary to ensure that the transition probabilities remain in the unit interval might need to be minuscule, leading to slow convergence. In light of the analysis of stationary problems considered in Section 2, one alternative approach would be to apply PFI. However, as we alluded to earlier, for the current problem this is probably a bad idea (as we verify below), as it requires solving a three-dimensional linear system at each stage of the algorithm. In what follows, we will call this approach "naive PFI" for clarity, because it applies PFI without any regard for the structure of the problem.

We now describe an algorithm that exploits the monotonicity of age and successively solves a series of two-dimensional control problems by iterating backwards in age. We will refer to this as "sequential PFI" and state it in the main text because to the best of our knowledge, it does not appear to have been used in the economics literature.

Algorithm 1 (Sequential policy function iteration)**.** For a given tolerance level $\epsilon > 0$, sequential policy function iteration is defined as follows:

(i) Given the terminal condition $V^h(\cdot,\cdot,\overline{A}) \equiv 0$, find the function $V^h(\cdot,\cdot,\overline{A}-\Delta_A)$ by applying Algorithm 5 with tolerance ϵ to a control problem with state space $\mathcal{B}^h \times \mathcal{Z}^h$, flow utility function $\tilde{u}$, transition probabilities $\tilde{p}^h$ and discount rate $\tilde{\rho}$ defined by

$$\begin{aligned} \Delta_t \tilde{u}(b,z,c) &:= \Delta_t u(c) + e^{-\rho\Delta_t} p^h(b,z,\overline{A}) V^h(b,z,\overline{A}), \\ \tilde{p}^h(b',z') &:= p^h(b',z')/\overline{p}, \\ e^{-\tilde{\rho}\Delta_t} &:= e^{-\rho\Delta_t}\overline{p}, \end{aligned} \tag{3.10}$$

for $(b,z),(b',z') \in \mathcal{B}^h \times \mathcal{Z}^h$, where $\overline{p} := \sum_{(b',z')\in\mathcal{B}^h\times\mathcal{Z}^h} p^h(b',z',A)$ and the transition probabilities p^h are given by (3.8).

(ii) Return to Step (i) with $V^h(\cdot,\cdot,\overline{A}-\Delta_A)$ in place of $V^h(\cdot,\cdot,\overline{A})$ and repeat until $A = 0$ is reached.

We have chosen the name "sequential policy function iteration" for Algorithm 1 because the algorithm proceeds by solving for each slice $V^h(\cdot,\cdot,A)$ sequentially, using the values at the higher age together with Algorithm 5. Note that it follows immediately from the original analysis of Puterman and Brumelle (1979) that this algorithm is convergent at each age level. Also recall that Algorithm 5 (PFI) takes as given an initial guess for the policy function upon which it successively iterates. One obvious choice for the initial guess of the policy function when age equals A in Algorithm 1 is to choose the optimal policy associated with $A + \Delta_A$. We adopt this choice in our implementation.

3.3 Numerical results

We now turn to our numerical results for the nonstationary setting. The structure of this setting will parallel that given in Section 2.3. We first document the accuracy of continuous-time and discrete-time methods on various different grids, before turning to an analysis of the speed of convergence. The parameters coincide with those given in Section 2.3, together with the assumption that age runs from $A = 0$ to $A = \overline{A} = 60$. Note that throughout, the timestep (or, equivalently, the age step) in the discrete-time problem is a parameter of the problem, and not a quantity chosen by the modeler.

Recall that in Section 2.3, we assessed accuracy by computing the policy and value functions on a fine grid in assets and viewing this as the "true" solution for each environment. Quantities computed on coarser asset grids were then compared to these "true" quantities. The notion of accuracy is more delicate in this nonstationary setting, because the analogue of the "true" continuous-time solution in this nonstationary setting must also use many values for age. The true discrete-time and continuous-time quantities are therefore no longer defined on the same grids, and when holding fixed the income process (as we do throughout), there are now two ways in which we can increase the accuracy of the continuous-time quantities: increase the size of the asset grid (as in Section 2.3), or increase the size of the age grid. When documenting the accuracy of the discrete-time and continuous-time approaches on coarser grids, we simply suppose that $N_A^h = 60$ in both the discrete-time and continuous-time analysis in order to ease the comparison.

In this setting, the policy functions can potentially differ substantially between the discrete-time and the continuous-time settings, especially toward the end of the agent's life. This should not be surprising, because in contrast to the discrete-time problem, in the continuous-time problem the agent does not wish to completely exhaust her entire wealth at the penultimate age $\overline{A} - \Delta_A$. We do not view this discrepancy as a strength or weakness of either analysis, as the two approaches are simply describing different environments. In each case we continue to compare quantities computed on coarse grids with their corresponding "true" values computed on a fine asset grid.

With these considerations in mind, Table 6 reports the accuracy of the discrete-time approach when $\Delta_t = 1$, and Table 7 repeats the exercise for the continuous-time approach with $\Delta_t = 10^{-6}$ using the sequential PFI algorithm given in Algorithm 1.[26]

[26]Note that the (naive) PFI algorithm and Algorithm 1 literally solve the same finite system of equations. Verification that they return near-identical values is only a check on the code (and not on the accuracy of different discretizations) and so is relegated to Appendix B.3.

Grid size	$\|\|\Delta c\|\|_1$	$\|\|\Delta c\|\|_\infty$	$\|\|\Delta c(\%)\|\|_1$	$\|\|\Delta c(\%)\|\|_\infty$
(25, 15)	0.0529	0.4215	1.8220	17.2187
(50, 15)	0.0184	0.1676	0.7266	10.5580
(100, 15)	0.0059	0.0835	0.2596	5.7510
(250, 15)	0.0019	0.0441	0.0760	4.2350
(500, 15)	0.0014	0.0221	0.0484	3.0425

Table 6: Accuracy of discrete-time EGM ($\Delta_t = 1$)

Grid size	$\|\|\Delta c\|\|_1$	$\|\|\Delta c\|\|_\infty$	$\|\|\Delta c(\%)\|\|_1$	$\|\|\Delta c(\%)\|\|_\infty$
(25, 15)	0.2128	3.1176	4.8730	36.4760
(50, 15)	0.1162	1.6650	2.6608	27.8550
(100, 15)	0.0613	0.8654	1.4063	20.7426
(250, 15)	0.0251	0.3500	0.5773	13.6159
(500, 15)	0.0122	0.1692	0.2812	9.7168

Table 7: Accuracy of continuous-time approach ($\Delta_t = 10^{-6}$)

Table 6 and Table 7 show that both frameworks are substantially less accurate in the nonstationary setting than in the stationary setting. Most of this loss of accuracy appears to occur near the end of the agent's life, where the consumption function changes substantially from the stationary setting and the agent rapidly decumulates assets.

Table 8 gives the time until convergence for the sequential PFI, naive PFI, and the EGM. In contrast to the stationary setting in Section 2, the EGM converges more quickly on every grid. Further, both the EGM and the sequential PFI are much faster than (naive) PFI. We omit the run time for naive PFI on larger grids as the sparse solver becomes very slow and clearly should never be used.

Grid size	EGM	Seq. PFI	Naive PFI
(25, 5)	0.064	0.683	0.482
(50, 5)	0.078	0.740	1.721
(100, 5)	0.108	0.877	5.050
(250, 5)	0.174	1.170	-
(500, 5)	0.273	1.652	-

Grid size	EGM	Seq. PFI	Naive PFI
(25, 15)	0.139	0.835	6.901
(50, 15)	0.172	1.036	31.203
(100, 15)	0.223	1.356	118.109
(250, 15)	0.386	2.476	-
(500, 15)	0.687	5.099	-

Table 8: Speed of convergence for nonstationary problem

Speed versus accuracy As in Section 2.3, we emphasize that the relevant tradeoff is speed versus accuracy, and not speed versus grid-size. For this reason, we now combine in Figure 5 the above results in a scatterplot that depicts the degree of accuracy attained within a particular time in each framework (discrete time versus continuous time). Appendix B.2 documents the speed versus accuracy tradeoff with Tauchen (1986) transition probabilities and finds that the discrete-time framework continues to outperform the continuous-time framework. For both discretizations, we find that the discrete-time approach converges much more rapidly than the continuous-time approach for a given level of accuracy, and that sequential PFI is far superior to naive PFI. The incorporation of age as a state variable can therefore effectively reverse the relative speed of the discrete-time and continuous-time approaches.

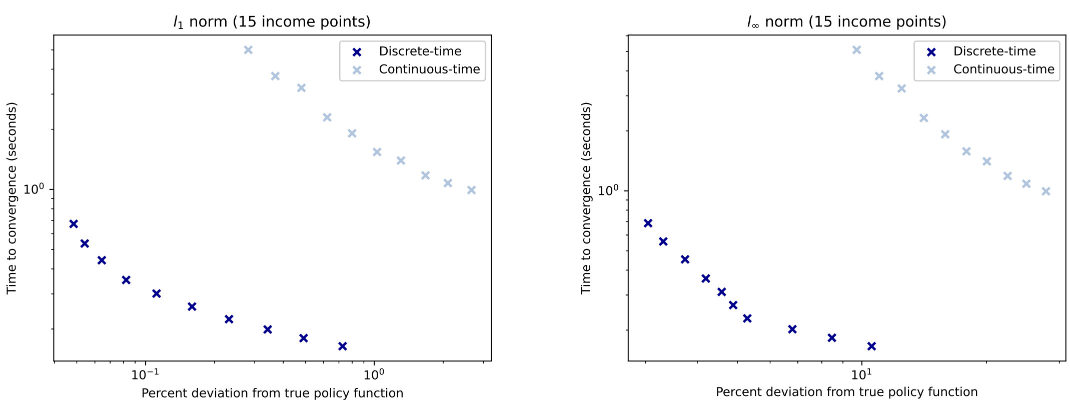


Figure 5: Time versus accuracy (15 income points)

3.4 Extension: Health expenditures and labor choice

The foregoing has shown that, for an income fluctuation problem with concave preferences, the continuous-time approach is superior for a stationary problem, while the discrete-time approach is superior for a lifecycle problem. We believe that this is a simple but useful point to emphasize because such problems recur throughout the macroeconomics literature. However, we do not argue that the continuous-time approach should never be used in *any* lifecycle model, because the above superiority of the discrete-time approach relied on the use of the EGM to eliminate nonlinear root-finding. Although the EGM proved stable for the problem in Section 3.1, the appropriate implementation of the EGM in richer environments often requires problem-specific ingenuity in order to produce reliable results.

We therefore conclude this paper with an example that illustrates the *flexibility* and *simplicity* of the continuous-time approach, which extends easily to more complicated settings with essentially no changes. We consider a problem similar in form to those studied in Hall and Jones (2007), Ales et al. (2012) and Eslami and Karimi (2019) in which an agent may affect her life-span through medical expenditures m and makes an extensive labor-leisure choice, in addition to the usual consumption-savings choice.

If the agent purchases m units of health services at age A, she dies at rate $\lambda(A, m)$ for some function λ that is convex in m for each age. For simplicity we restrict m to lie in $[0, \overline{m}]$ for some $\overline{m} \geq 0$. The utility upon death is denoted D and the (stochastic) age of death is denoted A_D. In addition, we assume that the agent makes a binary choice $l_t \in \{0, 1\}$ whether to work or not, where $l_t = 0$ denotes no work.

Flow utility at age A is now assumed to be $u(c, 1, A) = (c\eta(A))^{1-\gamma}/(1-\gamma)$ and $u(c, 0) = c^{1-\gamma}/(1-\gamma)$ for some $\eta(A) \leq 1$ (so that $\eta = 1$ corresponds to inelastic labor supply).

The agent's problem, in continuous time, can be written as

$$\begin{aligned} V(b, z, A) = \max_{(c_t, m_t, l_t)_{t \in [A, \overline{A}]}} \quad & \mathbb{E}\left[\int_A^{\overline{A}} e^{-\int_A^s [\rho + \lambda(\tau, m_\tau)] d\tau} u(c_s, l_s, A_s) ds + e^{-\rho A_D} D\right] \\ \text{s.t.} \quad & db_s = [rb_s + \overline{y} e^{z_s} l_s - c_s - m_s] ds, \\ & dz_s = \mu(z_s) ds + \sigma d\omega_s, \\ & b_s \geq \underline{b}. \end{aligned} \tag{3.11}$$

To construct a locally consistent chain, it is useful to imagine that at age A the agent faces the probability of death $p^h(\text{death}) = \Delta_t \lambda(A, m)$ and that conditional on survival, we construct a locally consistent chain for assets and income. For the latter, we could use the analogue of (3.8) with the drift now $rb - c - m + \overline{y} e^z l$. However, this leads to complicated first-order conditions due to the presence of the $1 - \lambda(A, m)\Delta_t$ term and to the appearance of medical expenditures in the maximum operators.

It is here that the flexibility of the continuous-time approach can prove useful: with a slight modification of the transition probabilities in (3.8), one can choose

$$\begin{aligned}
p_D^h(b+\Delta_b, z, A) &= \frac{\Delta_t[rb - c + \overline{y}e^z l]^+}{\Delta_b[1-\lambda(A,m)\Delta_t]},\\
p_D^h(b-\Delta_b, z, A) &= \frac{\Delta_t\big([rb - c + \overline{y}e^z l]^- + m\big)}{\Delta_b[1-\lambda(A,m)\Delta_t]},\\
p_D^h(b, z\pm\Delta_z, A) &= \frac{\Delta_t\big[\sigma^2\chi(z)/2 + \Delta_z[-\mu z]^\pm\big]}{\Delta_z^2[1-\lambda(A,m)\Delta_t]},\\
p_D^h(b, z, A+\Delta_A) &= \frac{\Delta_t/\Delta_A}{1-\lambda(A,m)\Delta_t},
\end{aligned} \tag{3.12}$$

and then define $p_D^h(b.z, A)$ so that probabilities sum to unity. It is easy to show that the system (3.12) defines a locally consistent Markov chain for the true law of motion, because the denominator $1-\lambda(A,m)\Delta_t$ tends to unity with Δ_t.

Thus, the optimization step in the resulting Bellman functional equation for the continuous-time formulation, now, involves solving

$$\max_{\substack{c,m\geq 0\\ l\in\{0,1\}}} \Big\{\Delta_t u(c,l,A) \ + \ e^{-\rho\Delta_t}\Big[[1-\Delta_t\lambda(A,m)]\mathbb{E}^h_{b',z',A'}\Big[V^h(b',z',A')\Big] + \Delta_t\lambda(A,m)D\Big]\Big\}, \tag{3.13}$$

where the expectation is taken with respect to the probabilities in (3.12) over assets, log income, and age, conditional on survival.

Now the choice in (3.12) becomes clearer: by omitting terms on the right-hand side of (3.13) that are independent of the controls and dividing by $\Delta_t e^{-\rho\Delta_t}$, the consumption and labor choices must maximize

$$e^{\rho\Delta_t}u(c,l) + \lambda(A,m)\Big[D - V^h\Big] + [rb - c + \overline{y}e^z l]^+ V^h_{bF} - \big[[rb - c + \overline{y}e^z l]^- + m\big]V^h_{bB}. \tag{3.14}$$

This reveals the simplicity of the continuous-time approach. Despite the two additional controls and the discrete choice in flow utility, the policy updating step can be obtained in closed form. For consumption we simply maximize (3.14) for each choice of labor and then perform an additional pointwise maximization. Because we assume that $\lambda(A,m)$ is convex in m for every age A, if $V^h_{bB} > 0$, the optimal m is zero if $D > V^h$, and otherwise is the maximum of zero and the solution to $V^h_{bB} = \lambda_m(A,m)\big[D - V^h\big]$.

Example We assume that $\lambda(A,m) = \Lambda_0(A)e^{-\Lambda_1(A)\cdot m}$ for some $\Lambda_0(A)$ and $\Lambda_1(A)$, so that $\lambda_m(A,m) = -\Lambda_1\Lambda_0 e^{-\Lambda_1 m}$. The first-order condition, thus, reduces to $m = \Lambda_1^{-1}\ln(\Lambda_1\Lambda_0[V^h - D]/V^h_{bB})$. Note that Λ_0 may be interpreted as the chance of dying in the absence of any medical expenditures. We assume that this varies from 0.1 percent to 6 percent between the youngest and oldest agents and is cubic in age, so that $\Lambda_0(A) = \underline{\Lambda}_0 + (\overline{\Lambda}_0 - \underline{\Lambda}_0)(A/\overline{A})^3$ for $(\underline{\Lambda}_0, \overline{\Lambda}_0) = (0.001, 0.06)$.[27]

The remaining parameters are simply chosen such that health expenditures are non-negligible but small relative to consumption. We assume that Λ_1 and η are linear in A, so that there exist constants $\underline{\Lambda}_1, \overline{\Lambda}_1, \underline{\eta}$ and $\overline{\eta}$ such that $\Lambda_1(A) \equiv \underline{\Lambda}_1 + (\overline{\Lambda}_0 - \underline{\Lambda}_1)A/\overline{A}$ and $\eta(A) \equiv \underline{\eta} + (\overline{\eta} - \underline{\eta})A/\overline{A}$. The complete set of parameters is specified in Table 1 and Table 9.

Table 9: Parameters used in medical example

Parameter	Value
$\underline{\Lambda}_0, \overline{\Lambda}_0$	$0.001, 0.06$
$\underline{\Lambda}_1, \overline{\Lambda}_1$	$3, 0$
$\underline{\eta}, \overline{\eta}$	$0.9, 0.6$
$\underline{D}, \overline{m}$	$-100, 1.0$
(N_b, N_z)	$(100, 15)$

Figure 6 plots the policy functions for consumption and medical expenditures for the youngest age group. For these agents, the probability of death is low and the upper bound on medical expenditures is never attained. Labor supply is identically $l = 1$ for the youngest age group and is therefore omitted.

Figure 6: Policy functions in initial period

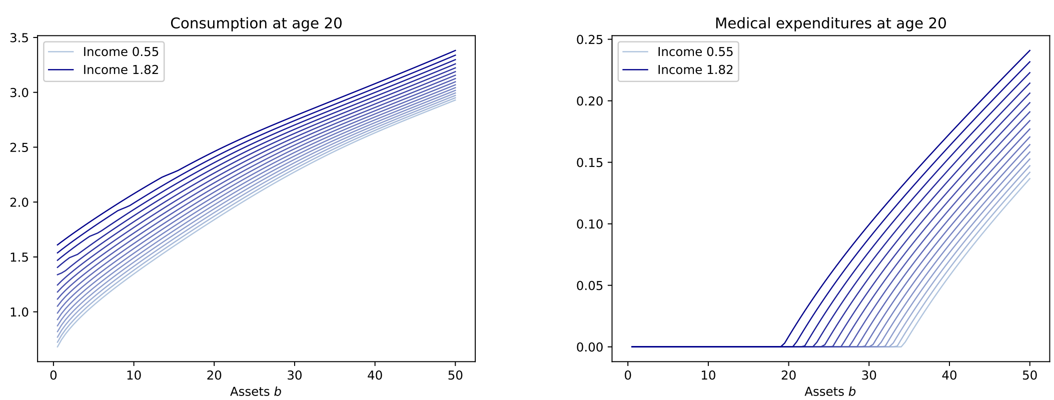


Figure 7 plots the policy functions for consumption and medical expenditures for older agents. In this case the risk of death is higher and the upper bound on medical care is binding

[27] This range is approximately consistent with death probabilities for individuals age 20 and 80 as reported by the Social Security Administration. See https://www.ssa.gov/oact/STATS/table4c6.html.

for high asset values. Further, the consumption function now has several discontinuities, which, as Figure 8 shows, occur at the points at which the agent chooses to stop working. This abruptly alters the marginal utility of consumption, leading to abrupt changes in the optimal consumption.

Figure 7: Policy functions in middle age

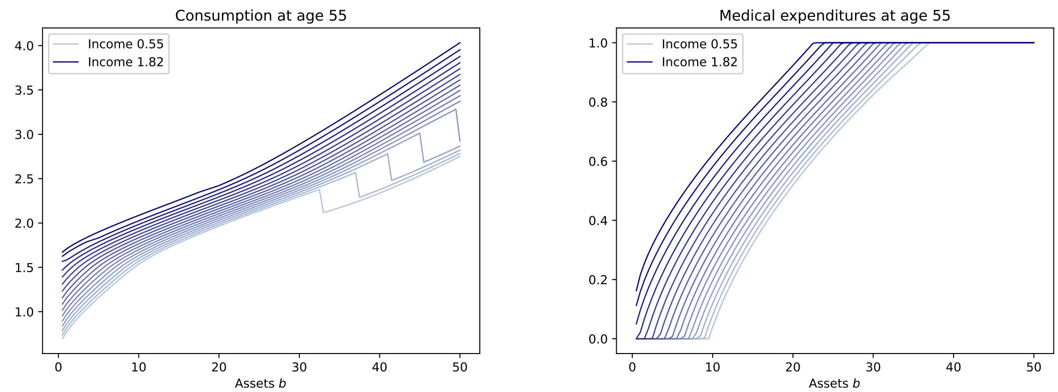
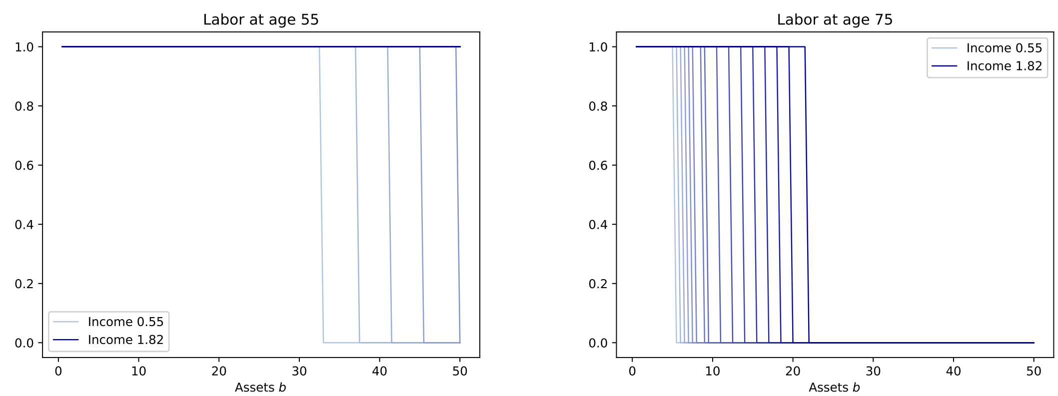


Figure 8: Labor supply at various ages



As we noted in the introduction, there are many extensions of the EGM in the literature designed to deal with nonconvexities, and it is possible that such techniques could be applied to our medical example. The point of this final example is that no additional ingenuity or coding is necessary in the continuous-time framework when we add both an additional continuous choice (medical expenditures) and a discrete choice (labor). Although it is hard to quantify programming difficulty, we believe that it is an important consideration to the researcher and that this section therefore provides some practical guidance. Specifically, we recommend that the researcher use discrete-time techniques in a lifecycle problem if there exists a simple way to implement the EGM, but that if such an extension is complicated or not covered by the existing literature, then they may wish to consider the continuous-time approach. However, in the latter case, we emphasize that the sequential PFI given in Algorithm 1 is much faster than a naive application of PFI, and produces the same solution (within numerical tolerance).

4 Conclusion

We do not believe it is possible or desirable to attempt to give an exhaustive set of recipes for when the continuous-time approach dominates the discrete-time approach. Our goal in this paper has been to focus on a series of problems that commonly arise in macroeconomics (stationary and nonstationary income fluctuation problems) and to provide guidance within this setting. We have shown that, for stationary problems, continuous-time methods appear superior to discrete-time methods, as they are both (typically) faster and have more stable convergence properties. In contrast, for nonstationary problems, such as an income fluctuation problem with a finite lifetime, discrete-time methods are superior for concave problems. However, the continuous-time approach is more flexible and easily adaptable to more complex environments in which nonlinear root-finding is more difficult to eliminate.

References

Yves Achdou, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach. *The Review of Economic Studies*, 89(1):45–86, January 2022. doi:10.1093/restud/rdab002.

S. Rao Aiyagari. Uninsured Idiosyncratic Risk and Aggregate Saving. *The Quarterly Journal of Economics*, 109(3):659–684, 1994. doi:10.2307/2118417.

Laurence Ales, Roozbeh Hosseini, and Larry E. Jones. Is There "Too Much" Inequality in Health Spending Across Income Groups? Working Paper 17937, National Bureau of Economic Research, March 2012.

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

G. Barles and P. E. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Analysis*, 4(3):271–283, January 1991. doi:10.3233/ASY-1991-4305.

Richard Bellman. The theory of dynamic programming. *Bulletin of the American Mathematical Society*, 60(6):503–515, 1954. doi:10.1090/S0002-9904-1954-09848-8.

David Blackwell. Discounted Dynamic Programming. *The Annals of Mathematical Statistics*, 36(1):226–235, 1965. ISSN 0003-4851.

Graham V. Candler. Finite-Difference Methods for Continuous-Time Dynamic Programming. In Ramon Marimon and Andrew Scott, editors, *Computational Methods for the Study of Dynamic Economies*, pages 172–194. Oxford University Press, October 2001. doi:10.1093/0199248273.003.0008.

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

Jeppe Druedahl and Thomas Høgholm Jørgensen. A general endogenous grid method for multi-dimensional models with non-convexities and constraints. *Journal of Economic Dynamics and Control*, 74:87–107, January 2017. doi:10.1016/j.jedc.2016.11.005.

Keyvan Eslami and Sayed M. Karimi. Health spending: Necessity or luxury? evaluating health care policies using an estimated model of health production function, 10 2019. Manuscript.

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

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

Giulio Fella, Giovanni Gallipoli, and Jutong Pan. Markov-chain approximations for life-cycle models. *Review of Economic Dynamics*, 34:183–201, October 2019. doi:10.1016/j.red.2019.03.013.

Nikolay Gospodinov and Damba Lkhagvasuren. A Moment-Matching Method for Approximating Vector Autoregressive Processes by Finite-State Markov Chains. *Journal of Applied Econometrics*, 29(5):843–859, 2014. doi:10.1002/jae.2354.

Robert E. Hall and Charles I. Jones. The Value of Life and the Rise in Health Spending. *The Quarterly Journal of Economics*, 122(1):39–72, February 2007. doi:10.1162/qjec.122.1.39.

Thomas Hintermaier and Winfried Koeniger. The method of endogenous gridpoints with occasionally binding constraints among endogenous variables. *Journal of Economic Dynamics and Control*, 34(10):2074–2088, October 2010. doi:10.1016/j.jedc.2010.05.002.

Ronald A. Howard. *Dynamic Programming and Markov Processes*. John Wiley, Oxford, England, 1960.

Mark Huggett. The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control*, 17(5):953–969, September 1993. doi:10.1016/0165-1889(93)90024-M.

Fedor Iskhakov, Thomas H. Jørgensen, John Rust, and Bertel Schjerning. The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks. *Quantitative Economics*, 8(2):317–365, 2017. doi:10.3982/QE643.

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

Harold J. Kushner and Paul Dupuis. *Numerical Methods for Stochastic Control Problems in Continuous Time*, volume 24. Springer Science & Business Media, 2001. doi:10.1007/978-1-4613-0007-6.

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

Martin L. Puterman and Shelby L. Brumelle. On the Convergence of Policy Iteration in Stationary Dynamic Programming. *Mathematics of Operations Research*, 4(1):60–69, 1979. doi:10.1287/moor.4.1.60.

Martin L. Puterman and Moon Chirl Shin. Modified Policy Iteration Algorithms for Discounted Markov Decision Problems. *Management Science*, 24(11):1127–1137, 1978. doi:10.1287/mnsc.24.11.1127.

Pontus Rendahl. Continuous vs. discrete time: Some computational insights. *Journal of Economic Dynamics and Control*, 144:104522, November 2022. doi:10.1016/j.jedc.2022.104522.

Manuel S. Santos and John Rust. Convergence Properties of Policy Iteration. *SIAM Journal on Control and Optimization*, 42(6):2094–2115, January 2004. doi:10.1137/S0363012902399824.

John Stachurski. *Economic Dynamics: Theory and Computation*. MIT Press, January 2009. ISBN 978-0-262-01277-5.

Nancy L. Stokey, Robert E. Lucas Jr, and Edward C. Prescott. *Recursive Methods in Economic Dynamics*. Harvard University Press, October 1989.

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

Matthew N. White. The method of endogenous gridpoints in theory and practice. *Journal of Economic Dynamics and Control*, 60:26–41, November 2015. doi:10.1016/j.jedc.2015.08.001.

Eric R. Young. Solving the incomplete markets model with aggregate uncertainty using the Krusell–Smith algorithm and non-stochastic simulations. *Journal of Economic Dynamics and Control*, 34(1):36–41, January 2010. doi:10.1016/j.jedc.2008.11.010.

List of Figures

List of Tables

A Review of solution methods

In this appendix, we review the solution methods adopted in the main text. All of the following are standard, but are included here in order to remove any ambiguity in the algorithms used. Appendix A.1 covers the discrete-time analysis, and Appendix A.2 covers the continuous-time analysis.

A.1 Discrete-time framework

In this appendix we review and formally state the algorithms that we use to solve the discrete-time Bellman equations considered in Section 2.1. Recall that in the stationary setting, whenever the Principle of Optimality is applicable, the value function of the agent is the unique solution to the functional equation $V \equiv B[V]$, where the Bellman operator B was defined by declaring

$$\begin{gathered} B[V](b,z) = \max_{c \geq 0, b' \geq \underline{b}} \Delta_t u(c) + e^{-\rho \Delta_t} \int_{\mathcal{Z}} \Gamma(z, dz') V(b', z') \\ \text{s.t. } b' = (1 + \Delta_t r)(b + \Delta_t(\overline{y} e^z - c)) \end{gathered} \tag{A.1}$$

for any $(b, z) \in [\underline{b}, \overline{b}] \times \mathcal{Z}$. Since we cannot evaluate the value function at every point in the continuum state space $\mathcal{G} := \mathcal{B} \times \mathcal{Z}$, we must discretize the state space. As in the main text, we suppose that $\mathcal{B}^h$ and $\mathcal{Z}^h$ are finite grids for assets and log income, respectively, define $\mathcal{G}^h := \mathcal{B}^h \times \mathcal{Z}^h$ for the state space for the discrete problem, and assume that log income now evolves according to a finite-state chain on $\mathcal{Z}^h$ with transition kernel Γ^h. We then consider the discrete-state Bellman equation

$$V^h = B^h[V^h] \tag{A.2}$$

where B^h is the discrete Bellman operator given by

$$B^h[V^h](b,z) = \max_{c\geq 0, b'\geq \underline{b}} \Delta_t u(c) + e^{-\rho\Delta_t} \sum_{z'\in\mathcal{Z}} \Gamma^h(z,z')\tilde{V}^h(b',z') \\ b' = (1+\Delta_t r)(b + \Delta_t(\overline{y}e^z - c)) \tag{A.3}$$

and $\tilde{V}^h$ is the linear interpolant of V^h. We can subtract the identity from both sides of the Bellman equation and obtain the functional equation

$$0 = \max_{c\geq 0} \Delta_t u(c) + T^h(c)\left[V^h\right](x). \tag{A.4}$$

Our purpose in this paper has been to compare the performance of discrete- and continuous-time approaches to solving optimal control problems. In order to make this comparison "fair," it is natural to consider the approaches most commonly used in each paradigm. For instance, it would not be informative to illustrate that finite-difference methods dominate a brute force approach to a discrete-time problem, because the latter is slow and can often be avoided. For this reason, when updating the guess for the policy function in discrete-time problems, we avoid nonlinear root-finding by employing the *endogenous grid method* (EGM) of Carroll (2006).[28] We now briefly review this method.

First suppose that one wished to evaluate the maximization on the right-hand side of (A.3) for an arbitrary guess of the value function V^h on the grid $\mathcal{G}^h := \mathcal{B}^h \times \mathcal{Z}^h$. One approach would be to choose the consumption level that solves the first-order condition, which rearranges to

$$u'(c) = (1+\Delta_t r)e^{-\rho\Delta_t} \sum_{z'\in\mathcal{Z}} \Gamma^h(z,z')\frac{\partial\tilde{V}^h}{\partial b}((1+\Delta_t r)(b+\Delta_t(\overline{y}e^z - c)), z') \tag{A.5}$$

For an arbitrary guess for V^h, the equation (A.5) is nonlinear and does not admit a closed form solution. Solving the nonlinear equation (A.5) is often a major bottleneck in discrete-time dynamic programming.

The insight of Carroll (2006) was that one could eliminate the need to solve a nonlinear equation when updating the value function by judiciously varying the grid at each step. Instead of fixing the current asset value and finding the optimal future level of assets, we fix future assets and find the current asset level at which (A.5) is satisfied. To describe this formally, denote the current guess for the value function by V^h, and for any future level of assets $b' \in \mathcal{B}^h$ and log income $z \in \mathcal{Z}^h$, we find the current asset value b for which

[28] See also Barillas and Fernández-Villaverde (2007) for a generalization.

condition (A.5) holds. If we denote this by $b^h(b', z)$, then the associated consumption is

$$c = (b^h\big(b', z\big) - b'/(1 + \Delta_t r))/\Delta_t + y(z).$$

The first-order condition (A.5) will be satisfied at this consumption point if

$$u'\Big((b^h\big(b', z\big) - b'/(1 + \Delta_t r))/\Delta_t + y(z)\Big) = (1 + \Delta_t r)e^{-\rho\Delta_t} \sum_{z' \in \mathcal{Z}^h} \Gamma^h\big(z, z'\big) \frac{\partial \tilde{V}^h}{\partial b}\big(b', z'\big), \tag{A.6}$$

and the above rearranges to give an explicit representation for current assets

$$b^h\big(b', z\big) = \max\Big\{\min\Big\{\tilde{b}^h(b', z), \bar{b}\Big\}, \underline{b}\Big\} \tag{A.7}$$

where $\tilde{b}^h(b', z)$ is a candidate choice of assets found by inverting equation (A.6),

$$\tilde{b}^h(b', z) = \Delta_t\big(u'\big)^{-1}\left((1 + \Delta_t r)e^{-\rho\Delta_t} \sum_{z' \in \mathcal{Z}^h} \Gamma^h\big(z, z'\big) \frac{\partial \tilde{V}^h}{\partial b}\big(b', z'\big)\right) + b'/(1 + \Delta_t r) - \Delta_t y(z). \tag{A.8}$$

Given the expression in equation (A.7), we have an associated consumption choice at the point $b^h(b', z)$ given by rearranging the budget constraint

$$c^h\Big(b^h(b', z), z\Big) = \frac{1}{\Delta_t}(b^h(b', z) - b'/(1 + \Delta_t r)) + y(z). \tag{A.9}$$

We then interpolate these consumption values for each z to obtain guesses for consumption. The above method of Carroll (2006) provides us with a way to update the policy function without ever solving a nonlinear equation. In principle, this can be paired with various different ways of updating the value function. For this reason, we outline two distinct algorithms.

Algorithm 2 (Value function iteration with EGM)**.** Given a tolerance level $\epsilon > 0$, value function iteration with the endogenous grid method is the following.

(i) Fix an arbitrary guess $V_0^h : \mathcal{G}^h \to \mathbb{R}$ for the value function on the grid $\mathcal{G}^h$, chosen such that the interpolant $\tilde{V}_0^h$ is concave in its first argument for every $z \in \mathcal{Z}$.

(ii) For each $(b', z) \in \mathcal{G}^h$ define consumption at the point $b_0^h(b', z)$ in equation (A.7) using the expression for consumption $c_0^h\big(b_0^h(b', z), z\big)$ in equation (A.9) together with a central-difference approximation of the derivative.

(iii) Extend c_0^h to every point $(b, z) \in \mathcal{G}^h$ using linear interpolation.

(iv) For each $(b,z) \in \mathcal{G}^h$ update the value function according to

$$V_1^h(b,z) = \Delta_t u(c_0^h(b,z)) + e^{-\rho\Delta_t} \sum_{z' \in \mathcal{Z}} \Gamma(z,z') \tilde{V}_0^h((1+\Delta_t r)(b + \Delta_t(\overline{y}e^z - c_0^h(b,z))), z'). \tag{A.10}$$

(v) Return to Step (i) with V_1^h in place of V_0^h and repeat until $||V_{n+1}^h - V_n^h|| < \epsilon$.

Algorithm 2 assumes that the first-order condition for optimality characterizes the optimum at every stage in the algorithm. The following is the same as Algorithm 2 except that we update the value function using policy function iteration in the analogue of step (iv).

Algorithm 3 (Policy function iteration with EGM)**.** Given a tolerance level $\epsilon > 0$, policy function iteration with the endogenous grid method is the following.

(i) Fix an arbitrary guess $V_0^h : \mathcal{G}^h \to \mathbb{R}$ for the value function on the grid $\mathcal{G}^h$, chosen such that the interpolant $\tilde{V}_0^h$ is concave in its first argument for every $z \in \mathcal{Z}$.

(ii) For each $(b',z) \in \mathcal{G}^h$ define consumption at the point $b_0^h(b',z)$ in equation (A.7) using the expression for consumpton $c_0^h\big(b_0^h(b',z),z\big)$ in equation (A.9).

(iii) Extend c_0^h to every point $(b,z) \in \mathcal{G}^h$ using linear interpolation.

(iv) Find the value function associated with adhering to the policy function c_0^h indefinitely, by solving the system of equations

$$V^h(b,z) = \Delta_t u(c_0^h(b,z)) + e^{-\rho\Delta_t} \sum_{z' \in \mathcal{Z}} \Gamma(z,z') \tilde{V}^h((1+\Delta_t r)(b + \Delta_t(\overline{y}e^z - c_0^h(b,z))), z') \tag{A.11}$$

for V^h.

(v) Return to Step (i) with V_1^h in place of V_0^h and repeat until $||V_{n+1}^h - V_n^h|| < \epsilon$.

A.2 Continuous-time framework

The Bellman equations in this paper that arise from a discretization of continuous-time control problems are all of the form

$$V^h(x) = B^h\Big[V^h\Big](x), \tag{A.12}$$

where the operator B^h is defined by

$$B^h\Big[V^h\Big](x) = \max_{c \geq 0} \Delta t^h(x,c) u(c) + e^{-\rho \Delta t^h(x,c)} \sum_{x' \in \mathcal{G}^h} p^h\big(x, x' \mid c\big) V^h\big(x'\big). \tag{A.13}$$

Note that if $N = |\mathcal{G}^h|$ is the cardinality of the state space, then the domain of the operator B^h in (A.13) is simply $\mathbb{R}^N$. Subtracting the identity matrix from both sides of (A.13) leads to the equivalent formulation of the fixed-point as a solution to

$$0 = \max_{c \geq 0} \Delta t^h(x, c)u(c) + T^h(c)\Big[V^h\Big](x). \tag{A.14}$$

The operators B^h and T^h defined in equations (A.13) and (A.14) are the building blocks for value function iteration and policy function iteration outlined below.

Algorithm 4 (Value function iteration)**.** For a given tolerance level $\epsilon > 0$, value function iteration is defined as follows:

(i) Begin with an initial guess $V_0 \in \mathbb{R}^N$ for the value function.

(ii) Update the value function by defining $V_1 = B^h[V_0]$, where B^h is defined in (A.13).

(iii) Return to Step (i) with V_1 in place of V_0 and repeat until $||V_{n+1} - V_n|| < \epsilon$.

Value function iteration is sometimes referred to as *the method of successive approximations*. If the transition probabilities p^h remain within the unit interval, Blackwell's conditions together with the contraction mapping theorem imply that $(V_n^h)_{n=1}^{\infty}$ converges to V^h, the solution to the BFE in equation (A.12), as $n \to \infty$. Note that value function iteration begins with an arbitrary guess for the *value* function. In contrast, the following begins with an arbitrary guess for the *policy* function.

Algorithm 5 (Policy function iteration)**.** For a given tolerance level $\epsilon > 0$, value function iteration is defined as follows:

(i) Begin with an initial guess $c_0 \in \mathbb{R}^N$ for the policy function.

(ii) Update the value function by defining V_1 to be the value of adhering to c_0 indefinitely, which is the solution to $0 \equiv \Delta t^h(x, c_0)u(c_0) + T^h(c_0)[V_1](x)$.

(iii) Update the policy function by defining c_1 to be the policy function that attains the maximum on the right-hand side of equation (A.14) for $V^h = V_1$.

(iv) Return to Step (i) with c_1 in place of c_0 and repeat until $||V_{n+1} - V_n|| < \epsilon$.

Policy function iteration is sometimes referred to as *approximation in the policy space* in the numerical analysis literature. Within the economics literature, it is sometimes referred to as Howard's improvement algorithm in honor of Howard (1960). Puterman and Brumelle (1979) show that policy function iteration is guaranteed to converge monotonically to the fixed-point of the equation (A.12). Further, the convergence is quadratic near the solution and so PFI typically requires a small number of iterations. However, each time we update the value function, we must solve a linear system of dimension the size of the state space.

An alternative procedure that in some sense lies between VFI and PFI, often referred to as the *modified policy function iteration method* (MPFI), combines ideas of these two methods. Instead of solving equation (A.14) exactly, MPFI performs a number of Jacobi relaxations on this equation. In other words, instead of solving the functional equation, in MPFI, one only finds an approximate solution to this equation by performing a number of iterations in the value space. Puterman and Shin (1978) provide formal convergence proofs for this class of algorithms.

B Miscellaneous figures and tables

This appendix contains miscellaneous figures and tables.

B.1 Stationary problems

B.1.1 Accuracy

Table 10 and Table 11 show that the accuracy of the continuous-time method is not particularly sensitive to the choice of timestep, at least within the range of values for which the transition probabilities remain in the unit interval.

Grid size	$\|\|\Delta c\|\|_1$	$\|\|\Delta c\|\|_\infty$	$\|\|\Delta c(\%)\|\|_1$	$\|\|\Delta c(\%)\|\|_\infty$
(25, 15)	0.0632	0.0942	2.9511	12.8098
(50, 15)	0.0325	0.0619	1.5341	9.2197
(100, 15)	0.0164	0.0415	0.7783	6.5735
(250, 15)	0.0064	0.0249	0.3047	4.1523
(500, 15)	0.0030	0.0170	0.1424	2.9152

Table 10: Accuracy of continuous-time approach ($\Delta_t = 0.005$)

Grid size	$\|\|\Delta c\|\|_1$	$\|\|\Delta c\|\|_\infty$	$\|\|\Delta c(\%)\|\|_1$	$\|\|\Delta c(\%)\|\|_\infty$
(25, 15)	0.0618	0.0942	2.8985	12.8036
(50, 15)	0.0311	0.0619	1.4796	9.2153
(100, 15)	0.0150	0.0415	0.7228	6.5705
(250, 15)	0.0049	0.0249	0.2486	4.1506
(500, 15)	0.0015	0.0170	0.0862	2.9138

Table 11: Accuracy of continuous-time approach ($\Delta_t = 0.05$)

B.1.2 Iterations

Table 12 and Table 13 record the number of iterations necessary for convergence for the stationary problem studied in Section 2.

Grid size	VFI	MPFI(10)	MPFI(50)	MPFI(100)	MPFI(200)	PFI
(25, 15)	3620.0	382.0	92.0	49.0	27.0	6.0
(50, 15)	3488.0	373.0	91.0	50.0	27.0	7.0
(100, 15)	3409.0	372.0	93.0	50.0	28.0	7.0
(250, 15)	3365.0	375.0	94.0	51.0	29.0	8.0
(500, 15)	3352.0	377.0	94.0	52.0	29.0	8.0

Table 12: Number of iterations (continuous-time, $\Delta_t = 0.05$)

Grid size	VFI	MPFI(10)	MPFI(50)	MPFI(100)	MPFI(200)	PFI
(25, 15)	212.0	24.0	9.0	9.0	9.0	9.0
(50, 15)	210.0	24.0	9.0	9.0	9.0	9.0
(100, 15)	214.0	25.0	9.0	9.0	9.0	9.0
(250, 15)	216.0	26.0	9.0	9.0	9.0	9.0
(500, 15)	217.0	26.0	9.0	8.0	8.0	8.0

Table 13: Number of iterations (discrete-time, $\Delta_t = 1$)

B.1.3 Speed versus accuracy

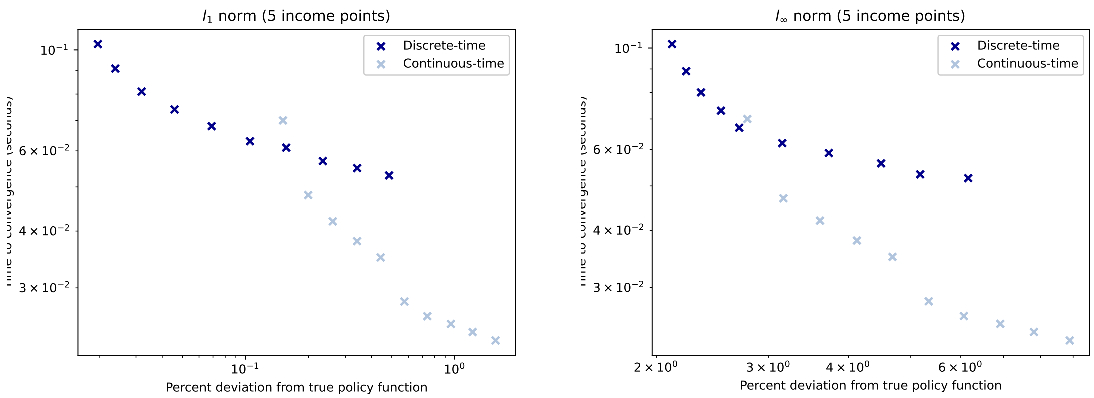


Figure 9: Time versus accuracy (5 income points)

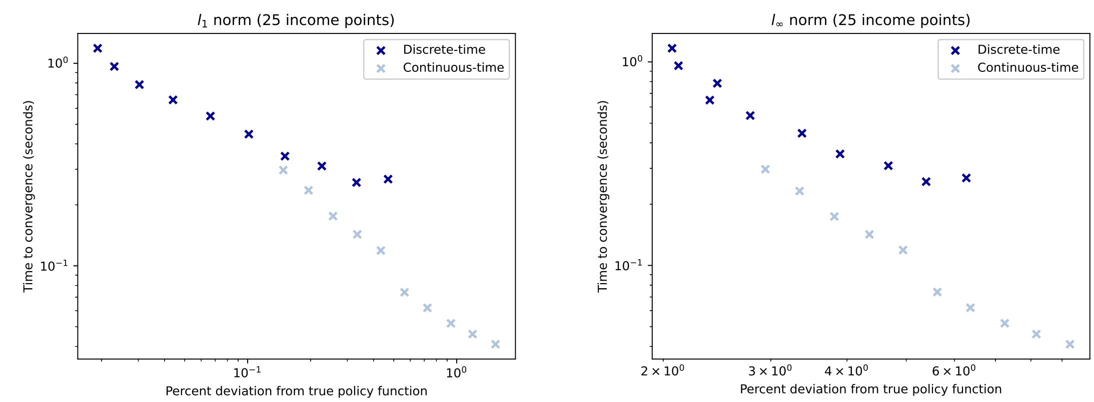


Figure 10: Time versus accuracy (25 income points)

B.2 Alternative discretization

In this appendix we review the discretization approach of Tauchen (1986). We enumerate the grid in log income by $z_0, \ldots, z_{N_z+1}$. Because assets will typically be indexed by i, we index current and future log income by j and k, respectively. In the main text, we considered an Ornstein-Uhlenbeck process for log income z_t with mean $\overline{z} = 0$ and diffusion term σ,

$$dz_t = -\overline{\mu} z_t + \sigma d\omega_t \tag{B.1}$$

where $\overline{\mu}$ gives the rate of mean reversion. In this case the discrete-time analogue is

$$z_{t+1} = (1 - \Delta_t \overline{\mu}) z_t + \sigma \sqrt{\Delta_t} \epsilon_t \tag{B.2}$$

where $(\epsilon_t)_{t=0}^{\infty}$ is normally distributed, i.i.d. over time with mean zero and unit variance. Denote by p_{jk} the probability of transitioning from point z_j to point z_k, and suppose that Δ_z is the size of the increment in z. We then define, for any $j \in \{0, \dots, N_z + 1\}$ and $k \neq 0, N_z + 1$, the probabilities

$$p_{jk} := P(z' \in (z_k - \Delta_z/2, z_k + \Delta_z/2))$$

while for the boundary points the probabilities are

$$\begin{aligned} p_{j0} &= P(z' \leq z_0 + \Delta_z/2) \\ p_{j,N_z+1} &= P(z' \geq z_{N_z+1} - \Delta_z/2). \end{aligned}$$

Writing Φ for the CDF of a standard normal distribution and substituting in the explicit expression for z' in (B.2), we have (dropping time subscripts from random variables for clarity),

$$\begin{aligned} p_{jk} &= P\Big(z_k - \Delta_z/2 - (1 - \Delta_t \overline{\mu}) z_j \leq \sigma \sqrt{\Delta_t} \epsilon \leq z_k + \Delta_z/2 - (1 - \Delta_t \overline{\mu}) z_j\Big) \\ &= \Phi\Big((z_k + \Delta_z/2 - (1 - \Delta_t \overline{\mu}) z_j)/[\sqrt{\Delta_t}\sigma]\Big) - \Phi\Big((z_k - \Delta_z/2 - (1 - \Delta_t \overline{\mu}) z_j)/[\sqrt{\Delta_t}\sigma]\Big) \end{aligned} \tag{B.3}$$

for $k \neq 0, N_z + 1$, and for the boundary points

$$\begin{aligned} p_{j0} &= \Phi\Big((z_0 + \Delta_z/2 - (1 - \Delta_t \overline{\mu}) z_j)/[\sqrt{\Delta_t}\sigma]\Big) \\ p_{j,N_z+1} &= 1 - \Phi\Big((z_{N_z+1} - \Delta_z/2 - (1 - \Delta_t \overline{\mu}) z_j)/[\sqrt{\Delta_t}\sigma]\Big). \end{aligned} \tag{B.4}$$

The transition matrix for log income is completely specified by (B.3) and (B.4).

Time versus accuracy. We now complement the main text by documenting the speed versus accuracy tradeoff using the discretization of Tauchen (1986). The results of this exercise are given in Figure 11. The relative ordering of the two approaches (continuous-time versus discrete-time) is unchanged relative to the main text.

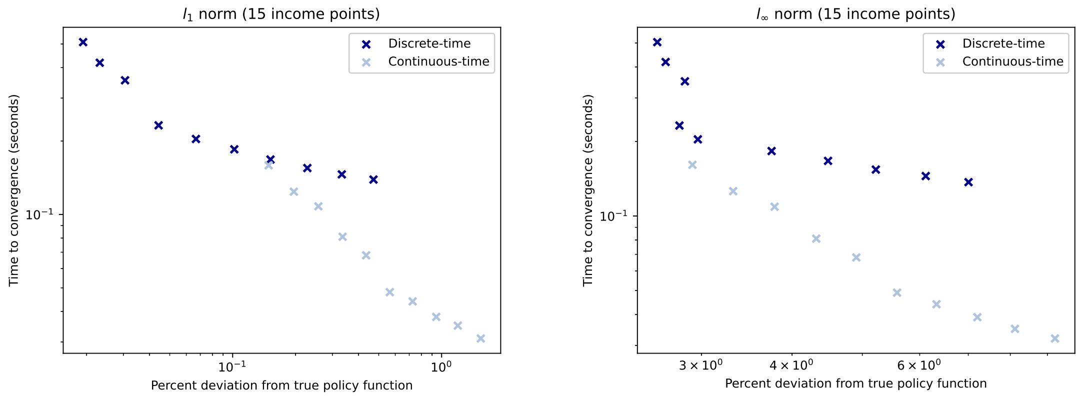


Figure 11: Time versus accuracy (15 income points, Tauchen)

B.3 Naive and sequential policy function iteration

In this appendix, we verify that the naive PFI algorithm and the sequential PFI algorithm given in Algorithm 1 return approximately the same values. These two algorithms solve the same system of equations, and so this is only a check on our code and not on the accuracy of a particular method. Because the naive PFI is prohibitively slow for large grids, we only verify on small grids. Reassuringly, all of the quantities in Table 14 are minuscule.

Grid size	$\|\|\Delta c\|\|_1$	$\|\|\Delta c\|\|_\infty$	$\|\|\Delta c(\%)\|\|_1$	$\|\|\Delta c(\%)\|\|_\infty$
(50, 10)	1.346799e-09	1.576741e-08	4.761515e-08	5.258059e-07
(100, 10)	1.783527e-09	2.436333e-08	6.479301e-08	9.284047e-07
(150, 10)	2.052531e-09	3.533098e-08	7.622061e-08	1.675828e-06
(200, 10)	2.221267e-09	3.596195e-08	8.347897e-08	1.605795e-06

Table 14: Comparison of naive and sequential PFI in a nonstationary environment