An Endogenous Gridpoint Method for Distributional Dynamics*

Christian Bayer†, Ralph Luetticke‡, Maximilian Weiss§ and Yannik Winkelmann¶

March 29, 2025

Abstract

Modeling continuous choices in heterogeneous agent models as "lotteries" over a discretized state space is standard practice (Young, 2010), but renders the distributional dynamics linear in optimal policies. We present a novel, simple method that captures nonlinearities and solves the distributional dynamics with interpolation instead of integration using the idea of an endogenous grid. Our approach solves for a stationary equilibrium as quickly as the lottery method for a given precision, outperforms it for linear dynamics, and accommodates nonlinear dynamics and aggregate risk. We demonstrate its efficacy by studying a model with aggregate investment risk with a third-order perturbation solution.

Keywords: Numerical Methods, Distributions, Heterogeneous Agent Models, Linearization

JEL-Codes: C46, C63, E32

*We thank Anastasiia Antonova, Thomas Hintermaier and an anonymous referee for helpful comments. Christian Bayer gratefully acknowledges support by the DFG through CRC-TR224 EPoS, project C05 and in the framework of the German Excellence Strategy-EXC2126/1-390838866. Ralph Luetticke gratefully acknowledges funding by the European Union (ERC, AIRMAC, 101114991). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

†Universität Bonn, CEPR, and IZA, *email:* christian.bayer@uni-bonn.de

‡Universität Tübingen, CEPR, and CFM, *email:* ralph.luetticke@uni-tuebingen.de

§Universität Tübingen, *email:* maximilian.weiss@uni-tuebingen.de

¶Universität Tübingen, *email:* yannik.winkelmann@uni-tuebingen.de

1 Introduction

A large class of heterogeneous agent models has the evolution of the distribution of agents at its core. In this paper, we propose a novel method for implementing this evolution numerically. Our method exploits the fact that policy functions in many such models are monotone. This allows us to express the evolution of the distribution without relying on either linearized mappings (Reiter, 2009; Young, 2010) or full integration (Krusell & Smith, 1998). Instead, we extend the idea of endogenous gridpoints (Carroll, 2006) to distributional dynamics.

We show that our distributional endogenous gridpoint method, hereafter *DEGM*, is as fast and tractable as the "lottery method" proposed by Young (2010)—the standard in the literature. We also show that our method converges faster as the number of gridpoints increases, even when solving for a stationary equilibrium and studying linear dynamics. Importantly, it preserves nonlinearities and is thus suitable for higher-order perturbation solutions of macroeconomic models with heterogeneous agents.

We illustrate our method with two applications. We start with the Aiyagari (1994) economy and document the numerical efficiency gains over the lottery method when solving for stationary distributions. Both methods converge to the same solution as the number of gridpoints increases, but *DEGM* reaches this limit an order of magnitude faster. Our method works directly on the cumulative distribution function, parsimoniously capturing its curvature through shape preserving interpolation. Importantly, updating the distribution function is not costly because the novel endogenous gridpoint approach works without integration.

We then propose a Krusell and Smith (1998) model with investment risk (depreciation shocks) as a new baseline model to study aggregate nonlinearities with household heterogeneity. This overcomes the approximate linearity in aggregate capital of the original Krusell and Smith (1998) model, while still being as parsimonious. We extend higher-order perturbation techniques to heterogeneous agent models, following Andreasen et al. (2018) and Levintal (2017). We solve our model up to third order and study asymmetric investment risk calibrated as in Barro (2006). The third-order solution becomes possi-

ble by combining the fast convergence of DEGM with state space reduction techniques developed in Bayer and Luetticke (2020) and Bayer et al. (2024). For the second-order solution, we also solve the unreduced model and show that the reduction techniques yield identical results.

For first-order perturbations, *DEGM* gives the same solution as the lottery method in the limit, but again converges to the true impulse responses faster in terms of the number of gridpoints. However, there is a significant difference for higher-order perturbations, where the lottery method does not capture all nonlinear effects, as already argued by Bhandari et al. (2023). The lottery method *overstates*, in particular, the distributional responses to shocks. At the same time, it *understates* the average long-run increase in wealth inequality as a consequence of the presence of investment risk. Using a third-order perturbation solution with *DEGM*, we find that aggregate investment risk lowers the capital stock by 5 to 11 basis points and increases wealth inequality by up to 11 basis points, depending on the calibration of idiosyncratic income risk. Aggregate investment risk increases inequality by reducing the incentive to save, especially for poor households. The lottery method, by ignoring nonlinear terms in the distributional dynamics, predicts lower wealth inequality than *DEGM* in the presence of investment risk.

Similar to Angeletos (2007), the introduction of risky returns to savings reduces aggregate savings through a negative substitution effect that dominates over the income effect for the majority of households. Angeletos discusses these channels, but in a stylized model where all agents save the same proportion of their lifetime wealth but have different ex-post returns, so that the wealth distribution becomes non-stationary.[1] Since investment risk is not idiosyncratic in our model, the model does not generate enough inequality within the top 1%. However, we find that adding aggregate investment risk to the original Aiyagari economy does increase wealth inequality and in particular wealth holdings at the very top.

This finding complements recent work showing how the expectation of lower as-

[1]Benhabib et al. (2024) also study the effects of investment risk in a model with heterogeneous agents. By making a distributional assumption about idiosyncratic investment risk, their model generates a wealth distribution with a realistic (Pareto-like) right tail.

set returns can increase wealth inequality through heterogeneous household portfolios (Fagereng et al., 2020; Fernández-Villaverde & Levintal, 2024; Gomez, 2024). While these studies emphasize the role of "passive" saving through rising asset prices, we show that aggregate uncertainty increases wealth inequality through "active" saving by the wealthy, as left-skewed investment risk has a stronger income effect for them.

Our method is directly implementable in the established sequence and state-space approaches for solving heterogeneous agent models with aggregate shocks (Auclert et al., 2021; Bayer et al., 2024). The idea of approximating the cumulative distribution function is not entirely new and can be found in Ríos-Rull (1997), Heer and Maussner (2009), and also in the special issue Den Haan et al. (2010). Our reformulation with an endogenous grid approach with nonlinear interpolation makes it differentiable, tractable and fast. The parsimonious representation of nonlinear distribution dynamics is key for higher-order perturbations.

The rest of the paper is organized as follows: Section 2 describes the distributional dynamics in terms of a difference equation of the distribution and policy functions, and presents our proposed method for solving this equation. Section 3 applies the method to the solution of an Aiyagari (1994) economy. Section 4 then uses an up to third order perturbation solution to the dynamic version of this economy with capital depreciation shocks as the source of aggregate risk. Section 5 concludes.

2 Problem and Method

Consider an economy in discrete time with a distribution of agents (of mass 1) over two variables a and y. We assume that $\ln(y)$ follows a stationary AR(1) process with normally distributed innovations ϵ and persistence $0 < \rho < 1$:

$$\ln(y_{t+1}) = \rho \ln(y_t) + \epsilon_{t+1}, \quad \epsilon_{t+1} \sim N(0, \sigma_\epsilon^2). \tag{1}$$

The continuous endogenous variable a is determined by the agent's policy function $a^*(a, y)$, which we assume to be strictly monotone in a (or composed of a constant part and

strictly monotone part) and, when discussing perturbation solutions, that is continuously differentiable both in a and y up to at least the order of perturbation.[2] The cumulative joint distribution in a and y at time t is given by $F_t(a, y) := P(x_t \leq a, z_t \leq y)$. We denote $f_t(a \mid y) = f_t(a, y)/dF_t(y)$ as the conditional density of non-mass point a for a given income y.[3] Because the process for y is exogenous and stationary, we focus on the case where the marginal density in y is the time-constant stationary one, $dF_t(y) = dF(y)$, which we know in closed form.

2.1 Distributional Dynamics in Discrete Time

The evolution of the distribution F is then given by the time-discrete Kolmogorov forward equation (making use of x and z' being independent):

$$F_{t+1}(a', y') = \int_{z' \leq y'} \int_{z} \int_{\{x | a' \geq a^*(x,z)\}} f_t(x \mid z)\, \phi(z, z') dx\, dz\, dz', \tag{2}$$

where $\phi(z, z')$ is the density of a transition from today's income, z, to tomorrow's income z', so that $\pi(z \mid z') := \frac{\phi(z,z')}{dF(z')}$ is the conditional density of having been at z in t conditional on being in $t+1$ at z'.

A brute-force approach to solving the equation for F_{t+1} would therefore require an integral approximation. This is computationally expensive. Originally, economists often used Monte Carlo methods to solve the Equation (2). To avoid this, Young (2010) suggests replacing the continuous distribution in a and y with a discrete counterpart. One defines a grid for a and a grid for y and represents the y-process by a discrete Markov chain and replaces the policy function $a^*(a, y)$ by lotteries over the gridpoints closest to $a^*(a, y)$. Therefore, this method is commonly known as the "lottery" method.[4] We denote the discrete probabilities from this "lottery" method (henceforth: LM) by the vector $\hat{f}(a, y)$. The simultaneous transitions along the a and y dimensions can then be summarized by a

[2] This is the standard case in many economic models, see Carroll (2006).

[3] We prove in Appendix C.3 that the distribution in a is indeed continuous above the borrowing constraint.

[4] Sometimes this is also referred to as "histogram method". We will use the latter term, however, for a related but not identical interpretation of the distribution as piece-wise linear interpolants over bins with transitions being uniformly distributed from one bin to the other, see Reiter (2009).

single transition matrix $\mathbf{A}^*$. This gives the discretized and stacked Kolmogorov forward equation:

$$\hat{f}_{t+1} = \hat{f}_t \mathbf{A}^*. \tag{3}$$

While this allows to calculate $\hat{f}_{t+1}$ very quickly, it not only creates an approximation error due to discretizing continuous densities, but also forces the Equation (3) to be linear in the optimal policies a^*.

2.2 The Endogenous Gridpoint Method for Distributions (DEGM)

Instead, we propose to use an endogenous gridpoint method analogous to the one proposed by Carroll (2006). To do so, we write the problem in terms of the distribution in a *conditional* on y, technically $F_t(a \mid y) = \left(\frac{\partial}{\partial y} F_t(a, y)\right) / dF_t(y)$, and divide each period into two sub-periods, a first one related to asset choices and a second one related to income changes. Because y is exogenous and strictly stationary, we can assume its marginal distribution to be constant over time and equal to its ergodic distribution $dF_t(y) = dF(y)$. This renders working with the conditional distribution or the joint distribution equivalent in the following, but also renders the former easier to handle. Concretely, we obtain the split in sub-periods by defining the distribution at the end of period t, after asset choices but before income transitions, as

$$\tilde{F}_t(a' \mid y) := \int_{\{x \mid a^*(x,y) \leq a'\}} f_t(x \mid y) dx, \tag{4}$$

such that we obtain the asset distribution at the beginning of period $t+1$, given new income level y', as

$$F_{t+1}(a' \mid y') = \int_z \tilde{F}_t(a' \mid z) \pi(z \mid y') dz. \tag{5}$$

Importantly, there are standard ways, such as Tauchen (1986), to find optimal quadrature weights to calculate the latter integral in form of a discretization of the y process by

creating a partitioning $\{\mathbb{Y}_1, \dots \mathbb{Y}_{n_y}\}$ and evaluating $\tilde{F}$ at the conditional expectations $\mathcal{Y}_k := E(y \mid y \in \mathbb{Y}_k)$. We can then replace the conditional transition densities by discrete probability counterparts:

$$F_{t+1}(a' \mid \mathcal{Y}_k) \approx \textstyle\sum_{j=1}^{n_y} \tilde{F}_t(a' \mid \mathcal{Y}_j)\Pi_{j,k}; \quad \Pi_{j,k} := P(y_t \in \mathbb{Y}_j \mid y_{t+1} \in \mathbb{Y}_k). \tag{6}$$

These discretizations converge to the theoretical integral in (5) as n_y increases.

In addition to our partitioning in y, we now specify grids $\{\mathcal{A}_i\}_{i=1\dots n_a}$ for a and calculate the associated policies $\mathcal{A}^*_{i,j} := a^*(\mathcal{A}_i, \mathcal{Y}_j)$. With these objects at hand, we first consider the case where $a^*(\cdot, y)$ is strictly monotone everywhere. This assumption allows us to simplify, for the *endogenous gridpoints*, $\mathcal{A}^*_{i,j}$, the set over which we integrate to $\{x|a^*(x, \mathcal{Y}_j) \leq \mathcal{A}^*_{i,j}\} = \{x|a^*(x, \mathcal{Y}_j) \leq a^*(\mathcal{A}_i, \mathcal{Y}_j)\} = \{x|x \leq \mathcal{A}_i\}$, where the last equation results from the monotonicity of a^*. This allows us to move from F_t to $\tilde{F}_t$ without explicit integration for these endogenous gridpoints as

$$\tilde{F}_t(\mathcal{A}^*_{i,j} \mid \mathcal{Y}_j) = \int_{\{x|a^*(x,\mathcal{Y}_j)\leq \mathcal{A}^*_{i,j}\}} f_t(x \mid \mathcal{Y}_j)dx = \int_{x\leq \mathcal{A}_i} f_t(x \mid \mathcal{Y}_j)dx = F_t(\mathcal{A}_i \mid \mathcal{Y}_j).$$

Therefore, the set of tuples $\left\{\left(\mathcal{A}^*_{i,j}, F_t(\mathcal{A}_i \mid \mathcal{Y}_j)\right)\right\}_{i=1\dots n_a}$ is on the graph of $\tilde{F}_t(\cdot \mid \mathcal{Y}_j)$. Consequently, we can construct an interpolant $\hat{\tilde{F}}_t$ for each $\mathcal{Y}_j$, since $\{\mathcal{A}^*_{i,j}\}_{i=1\dots n_y}$ is an ordered set, as a^* is strictly monotone in a. Replacing $\tilde{F}_t$ by $\hat{\tilde{F}}_t$ in (5), as well as replacing the integration by the quadrature over the finite income grid, then allows to evaluate F_{t+1} at any a' without integration:

$$F_{t+1}(a' \mid \mathcal{Y}_k) \approx \sum_j \quad \hat{\tilde{F}}_t(a' \mid \mathcal{Y}_j) \quad \Pi_{j,k}. \tag{7}$$

Finally, the definition of a cumulative distribution function implies two extrapolation rules. First, $\hat{\tilde{F}}_t(a' \mid \mathcal{Y}_j)$ is set to zero for any $a' < \min_i \left\{\mathcal{A}^*_{i,j}\right\}$. These are future endogenous states that are lower than the smallest optimal policy and hence are never reached. Second, $\hat{\tilde{F}}_t(a' \mid \mathcal{Y}_j)$ is set to $\lim_{a\to\infty} F_t(a \mid \mathcal{Y}_j) = 1$ for all $a' > \max_i \left\{\mathcal{A}^*_{i,j}\right\}$. The largest optimal policy is lower than these future endogenous states and hence the probability to

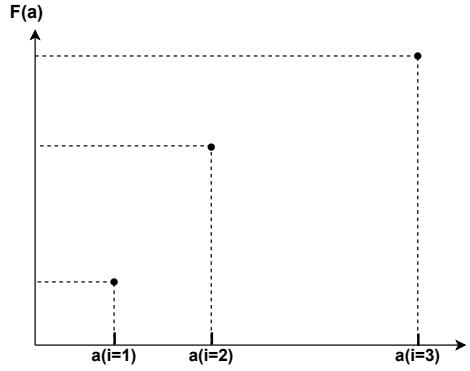
observe an agent with a lower than this endogenous state, given income $\mathcal{Y}_j$, is one.

Handling the Borrowing Constraint. When a^* is a savings function, however, it is typically only weakly monotone. It has a constant initial part at the borrowing constraint and is strictly monotone, for a given $\mathcal{Y}_j$, for a greater than some threshold of asset holdings $\underline{a}_j$. Our method can be easily adapted to account for this. Simply make the grid $\mathcal{A}_i$ start at $\underline{a}_j$, the EGM-solution (Carroll, 2006) corresponding to the borrowing constraint, to restore strict monotonicity. Because of weak monotonicity, and because we are working with cumulative distributions, evaluating $F_t(\underline{a}_j, \mathcal{Y}_j)$ gives the mass point at the borrowing constraint.

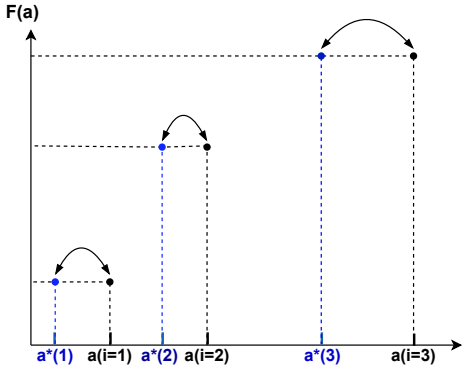
Another concern stemming from a mass point at the borrowing constraint is that such mass points can propagate through the distribution, especially when using any discrete approximation of the income process. Although there exist examples of policy functions that yield a distribution characterized by a finite number of mass points (see, e.g., Challe and Ragot (2016)), Appendix C demonstrates that this behavior is not generic. In fact, the discrete approximation of the income process produces a distribution function that is never truly "flat" — the set of points with nonzero mass is dense. The distribution function may still exhibit sharp increases — echoes of the borrowing constraint — which, however, diminish as the number of income states increases. In this light, a smooth interpolant can be expected to approximate the true distribution more accurately than a gridded step function; even when such echoes occur, a sufficiently flexible smooth interpolant will capture the "jump" effectively.

For a continuous income process, these echoes will be completely smoothed, and the distribution function will inherit differentiability from the differentiability of the policy function a^*. Importantly, this implies that the constructed interpolant from a finite set of income states approximates the wealth distribution well as we let the number of income states go to infinity, so that the quadrature approximations Π to the integrals converge. In practice, this convergence in income gridpoints is fast, and as few as ten income gridpoints smooth the echoes of mass at the borrowing constraint.

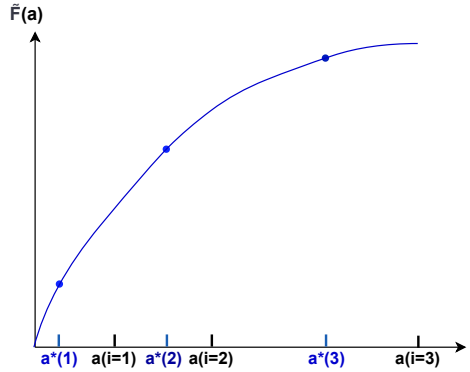
Figure 1: Illustration of DEGM with an interpolation over an endogenous grid



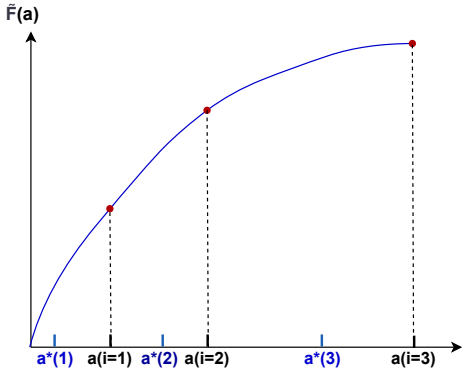
(a) Beginning of period CDF $\mathbf{F}_t(1:3 \mid \mathcal{Y}_j)$



(b) Graph $\left\{\left(\mathcal{A}^*_{1:3,j}, \mathbf{F}_t(1:3 \mid \mathcal{Y}_j)\right)\right\}$



(c) Construction of interpolant $\hat{\hat{F}}_t(\cdot \mid \mathcal{Y}_j)$



(d) Evaluation of interpolant $\hat{\hat{\mathbf{F}}}_t(1:3 \mid \mathcal{Y}_j)$

2.3 Numerical Implementation of DEGM

To provide a practical guide to implementation, we conclude with a summary of the proposed algorithm, assuming that the dynamic programming problem leading to the policy function is solved w.l.o.g. on the grid $\{\mathcal{A}_i\}$. This means that $\left\{\mathcal{A}^*_{i,j}\right\}$ is readily available as a discretized representation of the policy function.

Algorithm 1. *Start with the cumulative joint distribution (in a) at time t given by $F_t(a \mid y)$ that is discretized on the grids $\{\mathcal{A}_i\}, \{\mathcal{Y}_j\}$, see Figure 1 (a). The following assumes the values of this are stored in the matrix $\mathbf{F}_t = [F_t(\mathcal{A}_i \mid \mathcal{Y}_j)]_i^j$.*

1. *For each exogenous state with index j, $y = \mathcal{Y}_j$,* ***create the interpolant*** $\hat{\hat{F}}_t^j(a)$.

   (a) *Find the largest endogenous $\underline{a}_j$ s.t. $a^*(\underline{a}_j|y) = a_0$. This means, find the last asset state for which the policy a^* is a constant (based on the end-of-period marginal value of a_0 and the budget constraint as in Carroll (2006)'s EGM).*

(b) Define

$$\bar{\mathcal{A}}_{i,j} = \begin{cases} \{\underline{a}_j\} \cup \{a \in \mathcal{A}_i : a > \underline{a}_j\} & \text{if } \underline{a}_j > a_0 \\ \mathcal{A}_i & \text{else} \end{cases}$$

and the set of corresponding choices $\bar{\mathcal{A}}^*_{i,j}$

(c) Find the set of according points on the graph of $\tilde{F}^j_t$*:* $G := \{(\bar{\mathcal{A}}^*_{i,j}, \mathbf{F}_t(i,j))\}$*, see Figure 1 (b).*

(d) Create an interpolant $\hat{\tilde{F}}^j_t$ *based on the set* G*, see Figure 1 (c).*

2. *Loop through all* i, j *to* ***evaluate the interpolant*** *for each* $\mathcal{A}_i$ *from the fixed grid* $\{\mathcal{A}_i\}$ *and each* $\mathcal{Y}_j \in \{\mathcal{Y}_j\}$ *to calculate:*

$$\hat{\tilde{\mathbf{F}}}_t(i,j) = \begin{cases} 0 & \text{if } \mathcal{A}_i < \min\{\bar{\mathcal{A}}_{i,j}\} \\ \mathbf{F}_t(end, j) & \text{if } \mathcal{A}_i > \max\{\bar{\mathcal{A}}_{i,j}\} \\ \hat{\tilde{F}}^j_t(\mathcal{A}_i) & \text{else} \end{cases}$$

and collect this in a matrix $\hat{\tilde{\mathbf{F}}}_t$*. This yields the CDF in* a *on the fixed grid* $\{\mathcal{A}_i\}$ *prior to the exogenous Markov transitions, see Figure 1 (d).*

3. *Apply the* ***exogenous*** *Markov* ***transition*** *matrix* Π *to obtain* $\mathbf{F}_{t+1}$ *as:*

$$\mathbf{F}_{t+1} = \hat{\tilde{\mathbf{F}}}_t \Pi'$$

Practical implementation requires the choice of an interpolation routine. Since cumulative distribution functions are monotone, the interpolant should preserve this property. Both linear interpolation and piecewise cubic hermitian splines do. However, the linear interpolant does not preserve differentiability everywhere. Note that the linear interpolant is neither equivalent to the histogram nor to the lottery method, since we interpolate the CDF with optimal policy choices being the interpolation nodes.[5]

[5]See Appendix A.1 for details.

2.4 Nonlinear Distributional Dynamics

Bhandari et al. (2023) have highlighted the fact that *LM* fails to fully capture the nonlinear dynamics of the distribution. Returning to Equation (4), the nonlinear effects of the distribution f_t on its transition dynamics derive entirely from its effect on the policy function a^*. Holding the policy function constant, the transition is a linear operator.[6]

We can thus sufficiently characterize the missing nonlinearities of the *LM* by analyzing the effects of changes in the optimal policy function. Let them be caused by some generic perturbation ξ_t, for example, aggregate shocks or changes in the mean of the distribution. The second-order derivative of the transition matrix $\mathbf{A}^*$ of the Kolmogorov forward equation to such a shock is generally composed of two terms and is given by

$$\frac{\partial \mathbf{A}^*(k,l)}{\partial a_k^*}\frac{\partial^2 a_k^*}{\partial \xi_t^2} + \frac{\partial^2 \mathbf{A}^*(k,l)}{\partial a^{*2}_k}\left[\frac{\partial a_k^*}{\partial \xi_t}\right]^2, \tag{8}$$

where a_k^* denotes the optimal policy at wealth level $\mathcal{A}_k$. The first effect captures the direct nonlinearity of the policy function. The second effect reflects that the Kolmogorov forward equation is in principle nonlinear in policies. However, *LM* constructs $\mathbf{A}^*$ as (ignoring the exogenous state transitions for simplicity of notation)

$$\mathbf{A}^*(k,l) = \begin{cases} 1 - \frac{a_k^* - \mathcal{A}_l}{\mathcal{A}_{l+1} - \mathcal{A}_l} & \text{if } a_k^* \in [\mathcal{A}_l, \mathcal{A}_{l+1}) \\ \frac{a_k^* - \mathcal{A}_{l-1}}{\mathcal{A}_l - \mathcal{A}_{l-1}} & \text{if } a_k^* \in [\mathcal{A}_{l-1}, \mathcal{A}_l) \\ 0 & \text{else,} \end{cases} \tag{9}$$

which is linear in a^*. Therefore $\frac{\partial^2}{\partial a^{*2}_k}\mathbf{A}^*(k,l) = 0$. In Appendix A, we extend this analysis to the third-order derivative of Equation (4).

Our method, on the other hand, can capture all nonlinearities up to the order of the splines used to interpolate the CDF. Again, the second-order derivative, now of our

[6]The effect of the distribution on the policy function works through a market clearing condition, where higher aggregate demand for an asset, say, increases the market price.

interpolant $\hat{\tilde{F}}_t^j(\mathcal{A}_i)$, has the general form:

$$\frac{\partial^2 \hat{\tilde{\mathbf{F}}}_t(i,j)}{\partial \xi_t^2} = \frac{\partial \hat{\tilde{\mathbf{F}}}_t(i,j)}{\partial \mathcal{A}^*_{\cdot j}} \frac{\partial^2 \mathcal{A}^*_{\cdot j}}{\partial \xi_t^2} + \left[\frac{\partial \mathcal{A}^*_{\cdot j}}{\partial \xi_t}\right]' \frac{\partial^2 \hat{\tilde{\mathbf{F}}}_t(i,j)}{\partial \mathcal{A}^{*\,2}_{\cdot j}} \left[\frac{\partial \mathcal{A}^*_{\cdot j}}{\partial \xi_t}\right], \tag{10}$$

where, unlike (8), the second term is nonzero because $\mathcal{A}^*_{\cdot j}$ are the vectors of the interpolation nodes (and the derivatives are vector-valued). Therefore, the Hessian $\frac{\partial^2 \hat{\tilde{\mathbf{F}}}_t(i,j)}{\partial \mathcal{A}^{*\,2}_{\cdot j}}$ is generally nonzero. As also described in Bhandari et al. (2023), the second term in Equation (10) reflects second-order responses of the distributional dynamics to first-order changes in the optimal policy. What is more, if the continuous distribution has *curvature* at these pre-images, $\mathcal{A}^*_{\cdot j}$, approximation of $\frac{\partial^2 \hat{\tilde{\mathbf{F}}}_t(i,j)}{\partial \mathcal{A}^{*\,2}_{\cdot j}}$ requires a shape-preserving interpolation method, as illustrated in Figure 1 using cubic splines.[7]

3 Solving for Stationary Distributions

Our first application is the solution of an Aiyagari (1994) economy, where Equation (2) takes the special form of $F_t = F \; \forall t$ as an equilibrium condition. Specifically, we consider an economy with a continuum of households facing idiosyncratic risk in their human capital, h_t, which they rent out to firms at the wage rate, w_t. Households can self-insure by accumulating non-negative amounts of capital, k_t, which they rent out to firms at rate r_t. Capital depreciates at the rate δ_t. Human capital evolves according to a log-normal AR-1 process as in (1). We discretize this process in order to obtain quadrature weights for the transitions in form of the matrix Π using the Tauchen (1986) algorithm. Households enjoy utility from consumption, c_t, and solve the dynamic program:

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t) \tag{11}$$

$$\text{s.t. } c_t + k_{t+1} = (1 + r_t - \delta_t)\, k_t + h_t w_t \tag{12}$$

$$k_{t+1} \geq 0. \tag{13}$$

[7] Appendix A.3 shows that the continuous limit counterpart to $\frac{\partial^2 \hat{\tilde{\mathbf{F}}}_t(i,j)}{\partial \mathcal{A}^{*\,2}_{\cdot j}}$ is typically non-zero.

Table 1: Calibration

Parameters		Value	Parameters		Value
Calibration with persistent human capital process					
β	Discount factor	0.98	σ_ϵ	Std. of log-income shocks	0.14
γ	Rel. risk aversion	1.00	ρ	Persistence of log-income	0.98
α	Capital share	0.32	δ	Depreciation rate	0.02
Implied wealth distribution					
Mass at $k = 0$		0.04	Wealth Gini		0.66
Calibration with more transitory human capital process					
β	Discount factor	0.99	σ_ϵ	Std. of log-income shocks	0.18
γ	Rel. risk aversion	1.00	ρ	Persistence of log-income	0.88
α	Capital share	0.32	δ	Depreciation rate	0.02
Implied wealth distribution					
Mass at $k = 0$		0.01	Wealth Gini		0.42
Grid size used for calibration					
n_k	Gridpoints for k	160	n_h	Gridpoints for h	20

The wage and capital rates are given by the marginal products of labor and capital, respectively, where the production function is given by

$$Y_t = K_t^\alpha N^{1-\alpha}, \tag{14}$$

where N is the total amount of human capital supplied by households.

We seek an equilibrium in which prices are constant such that households form optimal policies given r, w, δ. These optimal policies are continuous choices of k_{t+1}. They depend on the continuous states k_t (endogenous) and h_t (exogenous). It is easy to show that for strictly concave felicity functions $u(c)$ the optimal policies are weakly monotone in k_t and strictly monotone outside the borrowing constraint. The problem thus fits the setup of Section 2.

We use this workhorse model as a laboratory to present our novel method and compare it to the widely used lottery method (LM). To do so, we follow in principle the calibration idea of Den Haan et al. (2010), see Table 1, but deviate by having a continuous human capital process with sufficient persistence to generate a significant fraction of credit-

constrained households. We also consider a version with larger income shocks and lower persistence, closer to the original Den Haan et al. (2010) setup. Conceptually, *DEGM* involves iterating the cumulative distribution function to find the stationary distribution.[8] We solve the model for $n_k = 160$ gridpoints in capital k and $n_h = 20$ gridpoints in human capital h.[9] As a baseline, we use our novel method to find the equilibrium. Going beyond 160 gridpoints for capital and 20 for income had no significant effect on the equilibrium (using the *DEGM* method), so we consider the distribution at $n_k = 160$ and up to $n_h = 20$ to be the "true" distribution.[10]

We perform two exercises. First, we isolate the quality of the approximation to the distribution by keeping prices and optimal policies fixed at the benchmark solution for the stationary equilibrium, that is, we use the *DEGM* solution with $n_k = 160$. We then select a subset of gridpoints for the sparser grid and use the associated policies to iterate the distribution to convergence for both the established *LM* and our new *DEGM*.[11] Second, we solve for the stationary equilibrium, including prices and policies, which more closely resembles the actual use case. *LM* finds the stationary distribution via eigendecomposition, while our method uses iteration. For the first exercise, we use the uniform distribution as a starting guess, which we update for the second exercise in each iteration on the equilibrium prices with the last converged distribution.

Table 2 shows the distance of two moments of the stationary distribution, average capital holdings and the Gini coefficient of capital holdings, for $n_k = 40, 80$, and 160 gridpoints in the asset dimension relative to the baseline solution using *DEGM* with 160 gridpoints for capital. All deviations are expressed by varying the asset grid but keeping the income grid fixed. For the income dimension, we consider three variants with $n_h = 5, 10$, and 20 gridpoints.

Panel A does this for the first exercise with constant prices and policies. Panel B

[8] We compute the aggregate capital stock as $E[X] = b \cdot F(b) - a \cdot F(a) - \int_a^b F(x)\,dx$.

[9] The grid is defined as $k_i = k_{min} + u_i^2$ with u_i uniformly distributed on $[0, \sqrt{200}]$

[10] Beyond 160 gridpoints, the *DEGM* solution no longer changes. For *LM*, convergence is achieved at about 320 gridpoints. Then there is no difference between the two methods in the solution of the stationary equilibrium. Also, going beyond 20 gridpoints for human capital does not significantly change the results.

[11] We use piecewise cubic Hermite splines to interpolate the cumulative distribution function.

Table 2: Convergence of stationary equilibria under *LM* and *DEGM* in n_h, n_k

		LM			DEGM	
	n_h \ n_k	**40**	**80**	**160**	**40**	**80**
Panel A: Stationary distribution [relative deviations in percent]						
Capital stock	5	1.72	0.56	0.12	0.09	0.03
	10	1.21	0.43	0.10	0.12	0.03
	20	0.85	0.32	0.09	0.17	0.03
Wealth gini	5	2.24	0.78	0.15	0.00	-0.01
	10	1.30	0.45	0.11	0.03	0.00
	20	0.84	0.30	0.08	-0.02	0.00
Panel B: Stationary equilibrium [relative deviations in percent]						
Capital stock	5	0.31	0.11	0.03	-0.03	0.00
	10	0.20	0.07	0.02	-0.01	0.00
	20	0.12	0.05	0.02	0.01	0.00
Wealth gini	5	2.66	0.83	0.17	0.20	0.02
	10	1.60	0.51	0.13	0.17	0.02
	20	1.09	0.36	0.09	0.11	0.02
Panel C: Computation times [in s]						
Time (s)	5	0.10	0.15	0.24	0.29	0.43
	10	0.18	0.26	0.45	0.41	0.65
	20	0.35	0.54	1.08	0.82	1.35

Notes: For each row, values represent percent deviations of the solutions with n_k gridpoints to the reference solution (*DEGM* with $n_k = 160$). For *LM* we use discrete aggregation methods, while we use continuous integration methods for *DEGM*. Values for the "persistent" income calibration from Table 1.
Panel A: Calculating stationary distribution using policies from the reference solution.
Panel B: Solving the stationary equilibrium including prices and policies.
Panel C: Time in seconds for solutions of Panel B. CPU with 16-cores, 3.3 GHz.

compares the two methods for the second stationary equilibrium exercise. Regardless of the size of the income grid n_h, we find that our method converges to the "true" distribution much faster, especially for cross-sectional moments.[12] For a given number of asset gridpoints, *LM* is faster in terms of computational time, mainly because it does not require iterations when updating the distribution. However, for a given accuracy ($n_k = 40$

[12]This mirrors the findings in Den Haan et al. (2010), which compares the approximation of the Kolmogorov forward equation by Monte Carlo simulation, the lottery method, or direct integration using a spline for the cumulative distribution function.

$DEGM \approx n_k = 160$ LM), our method is faster in solving for the stationary equilibrium (Panel C). The faster convergence reflects the fact that the distribution function is typically nowhere flat and the set of asset states that can be reached is dense, as discussed in Appendix C. LM approximates this necessarily by a step function for the distribution and for a coarse grid this approximation lacks precision.

4 Higher-Order Perturbations of Distributional Dynamics

Our second application is a setup with aggregate risk. As explained in Section 2, our method is able to capture nonlinearities in such setups. Specifically, we study an up to third-order perturbation solution of the Aiyagari model outlined above with capital depreciation shocks. Since we use third-order splines to interpolate the distribution, our method captures all nonlinear effects in both the distribution and the policy function.

To do this, we extend the state-space perturbation techniques for heterogeneous agent models to higher orders. Following Reiter (2009), we include the distribution and value functions in the state space. We then define a nonlinear differential equation on these objects, see Appendix B. The higher-order perturbation solution of this difference equation is then the same as for any differential equation reflecting a state-space system. Therefore, we can rely on established methods and solve the system with the algorithms developed by Andreasen et al. (2018) and Levintal (2017).[13] We also repeat the details of Levintal (2017)'s algorithm applied to heterogeneous agent models in Appendix B.

In addition, we show for the second-order perturbation that the Bayer and Luetticke (2020) reduction and its refinement in Bayer et al. (2024) yield the same results as solving the unreduced Reiter (2009) system. This reduces the state space of the model by writing the distribution in the form of a copula and marginals, and representing the value functions as sparse combinations of basis functions. For the third-order perturbation, we then exploit the sparseness of the reduced system. Currently, it is not feasible to solve

[13]The calculation of the first moments of the higher-order approximations using the Andreasen et al. (2018) method is equivalent to the method developed earlier by Rudebusch and Swanson (2012).

the full system beyond second order due to the increased memory requirements.

Solving the model up to third order allows us to implement asymmetric shocks. We follow Levintal in approximating the binomial distribution of a depreciation shock with a continuous distribution that has the same higher-order moments. Agents internalize the positive third moment of the shock, which acts as an investment risk. In particular, capital depreciation δ_t deviates from its steady-state value by following the process:

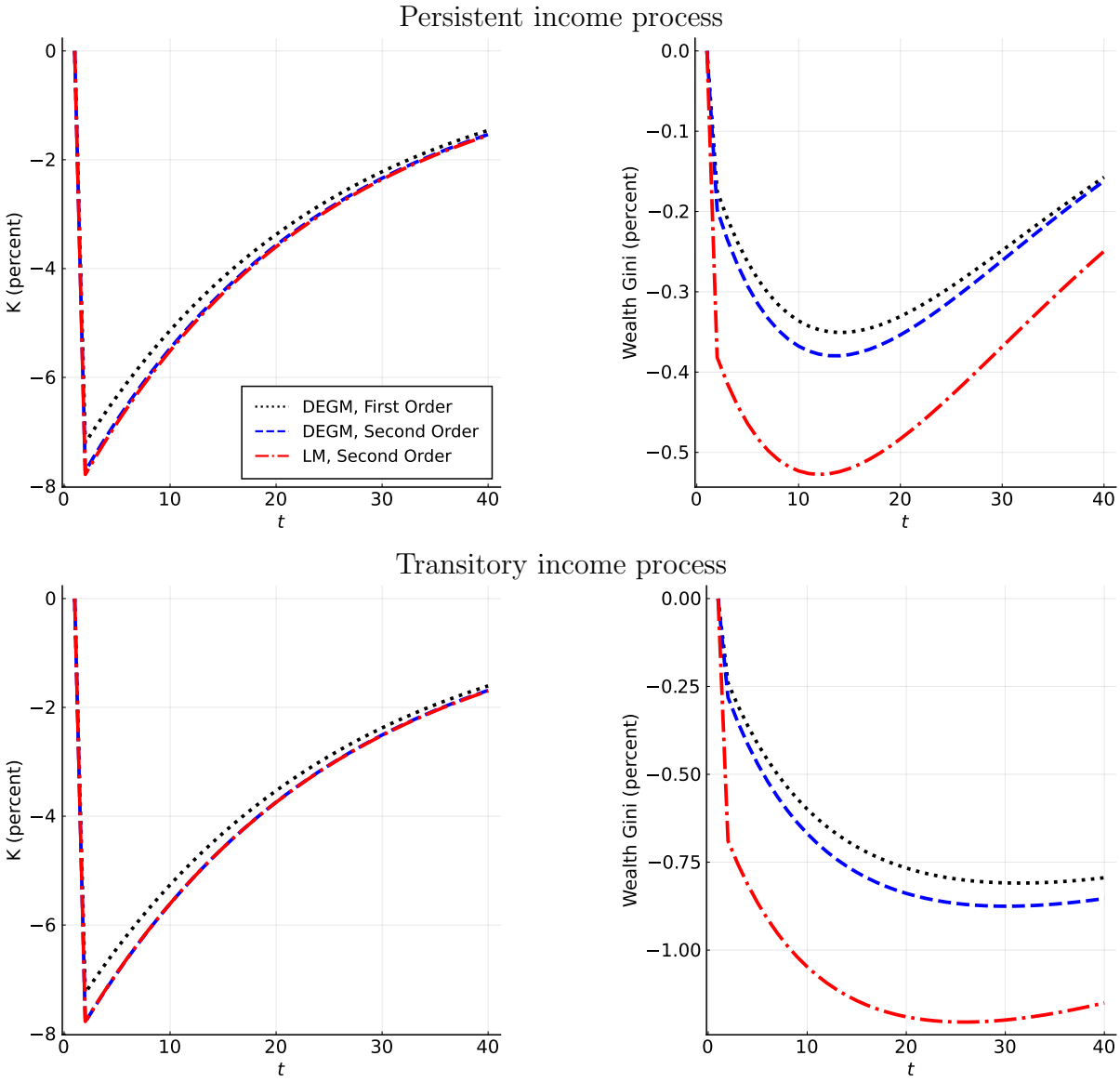
$$\delta_t = \delta + \nu_t,\ \nu_t \sim F^{\nu}(0, \sigma_\delta, \tau_\delta), \tag{15}$$

where σ_δ^2 and τ_δ^3 are the second and third moments of the depreciation shock distribution, respectively. In our calibration, $\sigma_\delta = 0.005$ and $\tau_\delta = 0.012$, which corresponds to a 0.4% chance that a disaster destroys 7.5% of the capital stock in a given period (quarter) and causes a 10% drop in annual GDP, consistent with the evidence in Barro (2006).

We propose this Krusell and Smith (1998)-style model with investment risk as a baseline model for studying aggregate nonlinearities with household heterogeneity. It overcomes the approximate linearity in aggregate capital of the original Krusell and Smith (1998) model while being equally parsimonious. Figure 2 compares the impulse responses to a one-time 7.5% destruction of the capital stock using *DEGM* to compute the dynamics with first- and second-order perturbation solutions. The first-order solution slightly understates the decline in aggregate capital and also the decline in the Gini coefficient of wealth in response to the capital depreciation shock. The difference in the second-order solution across methods is small for the response of the aggregate capital stock, but significant for the response of the Gini coefficient. This is true for both income processes considered. Taken together, this suggests that the distributional dynamics in this model are nonlinear with respect to aggregate shocks, but that the feedback from inequality to aggregates and equilibrium prices is modest.

Table 3 compares the solutions (without state-space reduction) for different grid sizes. *LM* and *DEGM* converge to the same first-order perturbation solution, just as they converge to the same stationary equilibrium (consistent with Bhandari et al., 2023, showing that *LM* has no bias up to first-order perturbations). In terms of IRFs of aggregates,

Figure 2: Impulse responses to a capital depreciation shock



Notes: Impulse responses of capital (left panels) and wealth Gini (right panels) to the shock $\nu_1 = 7.5$ *p.p.* from first-order (dotted black) and second-order (dashed blue) solutions using *DEGM*, and from second-order solution using *LM* (dash-dotted red), evaluated at the non-stochastic steady state ($n_y = 10, n_k = 160$). Top row: persistent calibration. Bottom row: transitory calibration. Y axis: Percent deviation from the non-stochastic steady state. X-axis: Quarter.

convergence is fast in n_k for both methods. In terms of the first-order dynamics of the wealth distribution, we find no significant differences in the $n_k = 40$ solution with DEGM to the 160-grid-point benchmark. For *LM*, the dynamics of the wealth distribution, in terms of IRFs of the wealth Gini, become the same as in our baseline only when we use $n_k = 160$ gridpoints. In other words, *DEGM* converges to the true solution much faster

Table 3: Convergence of 1st- and 2nd-order perturbation for *LM* and *DEGM*

		LM			*DEGM*	
	n_h \ n_k	**40**	**80**	**160**	**40**	**80**
Panel A: IRF statistic as in Bayer et al. (2024)						
Capital stock (FO)	5	1.00	1.00	1.00	1.00	1.00
	10	1.00	1.00	1.00	1.00	1.00
Capital stock (SO)	5	1.00	1.00	1.00	1.00	1.00
	10	1.00	1.00	1.00	1.00	1.00
Wealth Gini (FO)	5	0.89	0.99	1.00	1.00	1.00
	10	0.97	1.00	1.00	1.00	1.00
Wealth Gini (SO)	5	0.52	0.37	0.23	0.99	1.00
	10	0.91	0.85	0.78	1.00	1.00
Panel B: Second-order moments [relative deviations in basis points]						
Capital stock	5	-0.47	-0.46	-0.46	-0.09	-0.03
	10	-1.06	-0.77	1.36	-0.94	-0.05
Wealth Gini	5	-3.94	-4.18	-4.37	0.50	0.13
	10	-0.53	-1.32	-8.66	2.32	-0.24

Notes: In each row, the values represent the basis point deviations of the n_k gridpoints solutions from the reference (*DEGM* with $n_k = 160$). Parameters of the "persistent" calibration, Table 1.
Panel A shows the R^2 like statistics from Bayer et al. (2024) for an impulse response following a 7.5 p.p. shock to δ (over 100 periods) with a first order (FO) and second order (SO) perturbation solution. The R^2-like statistic is $1 - \sum_{h=1}^{H} \frac{(IRF_{DEGM,n_k=160,n_h}(h) - IRF_{method,n_k,n_h}(h))^2}{\sum_{h=1}^{H} IRF_{DEGM,n_k=160,n_h}(h)^2}$.
Panel B shows the ergodic moment solving the model with a second order perturbation. The deviation shown is the difference in percentage point deviation from each steady state over the percentage point deviation in the baseline solution (*DEGM*, $n_k = 160$). The aggregates under *LM* are derived using discrete aggregation methods, while continuous integration methods are used for *DEGM*.

for the first-order solution than *LM*. For the second-order solution, as expected, we find no convergence of the *LM* to the *DEGM* benchmark solution, neither in terms of ergodic means nor in terms of distributional IRFs.

The computing time (not reported) is now nearly identical for both methods for the same number of gridpoints. Both methods require one iteration of the Kolmogorov forward equation when solving for the perturbation solution, and these differ slightly, but compared to the total computational cost of solving the difference equation, this one iteration has a negligible computational cost so that variations therein are irrelevant. This gives *DEGM* a clear advantage in solving for aggregate dynamics over *LM*, because

Table 4: Comparison of second-order solution with different state-space reductions

Reduction	Persistent				Transitory			
	None	Copula	Copula +DCT	Copula +DCT +Factor	None	Copula	Copula +DCT	Copula +DCT +Factor
Panel A: Dimensions								
States	402	213	213	111	402	213	213	111
Controls	412	412	98	98	412	412	95	95
Total	814	625	311	209	814	625	308	206
Panel B: IRF statistic as in Bayer et al. (2024)								
Capital stock	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wealth Gini	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Panel C: Second-order moments [levels]								
Capital stock	25.54	25.54	25.54	25.54	25.29	25.29	25.29	25.29
Wealth Gini	0.61	0.61	0.61	0.61	0.43	0.43	0.43	0.43

Notes: "None" shows the results without reduction, "Copula" reduces the states for the distribution function, "Copula + DCT" additionally reduces the controls for the value function using the steady-state reduction approach of Bayer and Luetticke (2020), "Copula + DCT + Factor" derives a factor representation of the copula function from the first-order solution as in Bayer et al. (2024) and further reduces the dimensionality of the sparse set of basis functions to represent the copula, see Appendix B. The full state space refers to $n_k = 40$, $n_y = 10$ plus aggregate states and controls.
Panel A shows the number of states and controls that enter the difference equation representing the macroeconomic model and hence its dimensionality.
Panel B shows the R^2 like statistics from Bayer et al. (2024) for an impulse responses following a 7.5 p.p. shock to δ (over 100 periods) with second-order (SO) perturbation solution. The R^2 like statistics is $1 - \sum_{h=1}^{H} \frac{(IRF_{Full}(h) - IRF_{Reduced}(h))^2}{\sum_{h=1}^{H} IRF_{Full}(h)^2}$.
Panel C shows ergodic moments in levels for the second-order solution.

it requires less gridpoints for the same precision and computing time increases sharply with the number of gridpoints.

All of the above results do not use dimensionality reduction. Next, we check whether the Bayer and Luetticke (2020) method, which reduces the dimensionality of the difference equation, can be extended to higher-order perturbations. The method replaces the distribution function by a copula and marginals, and considers deviations from the steady state in copula and value functions only on a coarse grid / based on discrete cosine transformations with linear interpolations in between. The extension to higher-order perturbations and the use of *DEGM* require that the linear interpolants used by Bayer and Luetticke (2020) to represent copulas and value functions are also replaced by

spline interpolation. Table 4 evaluates the quality of the approximation introduced by the reduction step. We compare four variants. First, the no-reduction version ("None"), which defines both the distribution and the value functions on the full Cartesian product of the income and wealth grids.[14] Second, a version that reduces only the distribution function by writing it in terms of marginal distributions (defined on the full income and wealth grids) and a copula defined on a coarse wealth percentile grid. Third, we also use the DCT method of Bayer and Luetticke (2020) to reduce the number of controls representing the value function. Fourth, we apply the refinement developed in Bayer et al. (2024). This solves the model first at first order and finds a factor representation of the copula function based on its variance-covariance matrix obtained under the first-order representation.[15]

Panel A of Table 4 lists the number of states and controls, and hence the dimensionality, of the difference equation describing the macroeconomic model, see Appendix B. The maximum reduction removes 75% of all states and controls. This has no visible effect on the second-order IRFs, Panel B. Also, the ergodic moments hardly change, Panel C. The quality of the reduction is independent of whether the income process is persistent or transitory. This is an important result because the number of derivatives to compute, and hence the size of the matrices to store, scales with the number of states and controls to the power of the order of the perturbation solution plus one (quadratic for first order, cubic for second order, etc.).

For this reason, a third-order solution to the full model is not feasible on a machine with less than 20 terabytes of memory.[16] However, the strong reduction allows us to solve the reduced system at third order (still requiring almost 2 terabytes of memory). We use the fact that the reduction is practically lossless at the second order as a heuristic to

[14] In practice, since we use an EGM to solve for optimal policies, we write the difference equation in terms of policy functions instead of value functions.

[15] This requires in practice a sufficiently rich set of shocks such that all prices that vary in the higher-order solution are also varying in the first-order solution. Here we achieve this by having TFP and also time-preference shocks that drive a wedge between the interest rate and the ratio of expected marginal utilities.

[16] We estimate this memory requirement by extrapolating the relationship between system size and memory requirements for smaller models. We predict that 20 terabytes of memory will be required for the third-order solution of the full model at the lowest resolution we consider, $n_k = 40, n_y = 5$.

Table 5: Ergodic moments under second- and third-order solution with investment risk

Variable	Persistent				Transitory			
	LM		*DEGM*		*LM*		*DEGM*	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Panel A: Second-order solution [relative deviations in basis points]								
Output	-2.7	(60.5)	-2.7	(62.6)	-3.1	(62.8)	-2.9	(63.5)
Capital stock	-8.4	(189.0)	-8.4	(195.6)	-9.7	(196.2)	-9.2	(198.3)
Wealth Gini	-2.6	(19.0)	0.6	(15.3)	-18.2	(76.4)	1.1	(55.0)
Panel B: Third-order solution [relative deviations in basis points]								
Output	-3.6	(-)	-1.5	(-)	4.8	(-)	-3.7	(-)
Capital stock	-11.2	(-)	-4.7	(-)	15.0	(-)	-11.4	(-)
Wealth Gini	2.0	(-)	10.7	(-)	-149.7	(-)	2.1	(-)

Notes: $n_k = 40$, $n_y = 10$. Means and standard deviations (in brackets) across parameterization and methods are in basis point deviation from non-stochastic steady state. Moments are from closed-form solutions for pruned model dynamics (Andreasen et al., 2018). No standard deviations are reported for third-order perturbations because of memory requirements. Capital depreciation shock with $\sigma_\delta = 0.5\%$, $\tau_\delta = 1.2\%$.

expect a reasonable quality of approximation also at the third order.

Higher-order solutions not only provide a better approximation of the dynamics, but also, importantly, capture the response of households to aggregate risk. Table 5 documents how the ergodic distribution with aggregate risk differs from the stationary equilibrium without aggregate risk. Aggregate risk here refers to investment risk. The top panel shows the change for a second-order perturbation, the bottom panel for a third-order perturbation. Appendix A discusses that the third-order solution captures the effect of the skewed distribution of aggregate shocks, while the second-order solution captures only the effect of the variance. As in Angeletos (2007), we find that aggregate investment risk reduces the aggregate capital stock because for most households the substitution effect is stronger than the income effect. For the second-order perturbation, we find small effects on aggregates, and the distributional nonlinearities that *LM* misses seem to be of little importance for the first two central moments of the ergodic distribution, which is in line with the similarities of the IRFs documented before in Figure 2.

For the wealth distribution itself, in line with Bhandari et al. (2023), we find that

the two methods are already for the second-order perturbation qualitatively different (even though differences are small). *LM* always predicts a decrease in wealth inequality while *DEGM*, which avoids the Bhandari et al. (2023) criticism, predicts that inequality might increase in response to aggregate risk compared to the stationary equilibrium. The underaccumulation and increase in wealth inequality is driven by lower savings of less wealthy households. For these households, the substitution effect dominates as they have little capital income. Moreover, a lower capital stock implies a lower wage rate and a higher rate of return on capital. For wealthy households, however, the income effect is key and they have strong precautionary saving motives given the aggregate investment risk. Thus, while a capital depreciation shock upon realization compresses the distribution of wealth, as can be seen in Figure 2, the risk of such a shock can increase wealth inequality on average.

At higher orders, the interaction of income and substitution effect starts to play a role: how is the savings response to a change in income altered by a simultaneous change in investment opportunities? This is relevant when considering the risk of capital destruction, which induces a negative correlation between today's capital income, and future returns on investment. However, this correlation does not only matter for the individual household's optimization problem, but also for the interaction of optimal policies in the *cross-section*. As we show in Appendix A.2, *LM* misses several of the higher-order terms when computing IRFs and the ergodic distribution. The reason is that the higher-order effect of a change in the savings policy on the wealth distribution in $t+1$ depends on the curvature of the wealth distribution in t (in Appendix A.3 we derive this analytically). Since *LM* models a discrete distribution, it only accounts for the *level* of the "density" at a specific point, namely the size of the mass point. In contrast, *DEGM* models the continuous limit of the wealth distribution. Thereby, it accounts for *marginal* transition flows across wealth levels. The marginal flows are characterized by the slope and the curvature of the density.[17]

[17]For example, at a wealth level a where the density $f(a)$ is falling and convex, the marginal outflow when transitioning to other points on the distribution is dominated by households with wealth slightly below a. This means that the substitution effect is *marginally* more important for the change in distribution induced by the outflows from point a, if the substitution effect is more important for the poorer

5 Conclusion

We propose a novel endogenous gridpoint method for distributional dynamics (*DEGM*). Our method retains the tractability and speed of the lottery/histogram methods commonly used in the literature, while requiring significantly fewer gridpoints and capturing all nonlinear effects of distributional dynamics. By preserving the nonlinearities critical to heterogeneous agent models, *DEGM* provides an improved framework for studying models with household heterogeneity and aggregate risk. It allows for a straightforward implementation in the established sequence and state-space approaches for solving heterogeneous agent models with aggregate shocks (Auclert et al., 2021; Bayer et al., 2024). We provide an example of a state-space solution with a third-order perturbation. In particular, we propose a Krusell and Smith (1998) model with aggregate investment risk. In this model, we show that aggregate investment risk affects inequality.

References

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

Andreasen, M. M., Fernández-Villaverde, J., & Rubio-Ramírez, J. F. (2018). The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications. *The Review of Economic Studies*, *85*(1), 1–49.

Angeletos, G.-M. (2007). Uninsured idiosyncratic investment risk and aggregate saving. *Review of Economic Dynamics*, *10*(1), 1–30.

Auclert, A., Bardóczy, B., Rognlie, M., & Straub, L. (2021). Using the sequence-space jacobian to solve and estimate heterogeneous-agent models. *Econometrica*, *89*(5), 2375–2408.

Barro, R. J. (2006). Rare Disasters and Asset Markets in the Twentieth Century*. *The Quarterly Journal of Economics*, *121*(3), 823–866.

households, while conversely the income effect dominates for richer households.

Bayer, C., Born, B., & Luetticke, R. (2024). Shocks, frictions, and inequality in us business cycles. *American Economic Review*, *forthcoming*.

Bayer, C., & Luetticke, R. (2020). Solving discrete time heterogeneous agent models with aggregate risk and many idiosyncratic states by perturbation. *Quantitative Economics*, *11*(4), 1253–1288.

Benhabib, J., Cui, W., & Miao, J. (2024). Capital income jumps and wealth distribution. *mimeo, UCL*.

Bhandari, A., Bourany, T., Evans, D., & Golosov, M. (2023). A perturbational approach for approximating heterogeneous agent models. *NBER Working Paper No 31744*.

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

Challe, E., & Ragot, X. (2016). Precautionary saving over the business cycle. *Economic Journal*, *126*(590), 135–164.

Den Haan, W. J., Judd, K. L., & Juillard, M. (2010). Computational suite of models with heterogeneous agents: Incomplete markets and aggregate uncertainty. *Journal of Economic Dynamics and Control*, *34*(1), 1–3.

Fackler, P. L. (2019). Algorithm 993: Efficient computation with kronecker products. *ACM Trans. Math. Softw.*, *45*(2).

Fagereng, A., Holm, M., Moll, B., & Natvik, G. (2020). *Saving Behavior Across the Wealth Distribution: The Importance of Capital Gains* (CEPR Discussion Papers No. 14355). CEPR Discussion Paper.

Fernández-Villaverde, J., & Levintal, O. (2024). The distributional effects of asset returns. *mimeo, University of Pennsylvania and Reichman University*.

Folland, G. (1999). *Real analysis: Modern techniques and their applications*. Wiley.

Fritsch, F. N., & Butland, J. (1984). A method for constructing local monotone piecewise cubic interpolants. *SIAM Journal on Scientific and Statistical Computing*, *5*(2), 300–304.

Gomez, M. (2024). Wealth inequality and asset prices. *mimeo, Columbia University*.

Heer, B., & Maussner, A. (2009). *Dynamic general equilibrium modeling*. Springer.

Kim, J., Kim, S., Schaumburg, E., & Sims, C. A. (2008). Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models. *Journal of Economic Dynamics and Control*, *32*(11), 3397–3414.

Krusell, P., & Smith, A. A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, *106*(5), 867–896.

Levintal, O. (2017). Fifth-order perturbation solution to DSGE models. *Journal of Economic Dynamics and Control*, *80*, 1–16.

Reiter, M. (2002). Recursive computation of heterogeneous agent models. *mimeo, Universitat Pompeu Fabra*.

Reiter, M. (2009). Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control*, *33*(3), 649–665.

Ríos-Rull, J.-V. (1997). *Computation of equilibria in heterogeneous agent models* (tech. rep.). Federal Reserve Bank of Minneapolis.

Rudebusch, G. D., & Swanson, E. T. (2012). The bond premium in a dsge model with long-run real and nominal risks. *American Economic Journal: Macroeconomics*, *4*(1), 105–143.

Schaback, R., & Wendland, H. (2005). *Numerische mathematik*. Springer.

Tauchen, G. (1986). Finite state Markov-chain approximations to univariate and vector autoregressions. *Economics letters*, *20*(2), 177–181.

Young, E. R. (2010). Solving the incomplete markets model with aggregate uncertainty using the Krusell Smith algorithm and non-stochastic simulations [Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty]. *Journal of Economic Dynamics and Control*, *34*(1), 36–41.

A Perturbation of Distributional Dynamics

A.1 Comparison of Discretized Methods

We compare discretized methods[18] of distributional dynamics by abstracting from the stochastic transition across income levels, i.e., transitions across wealth levels are governed only by the optimal policy function. $\mathcal{A}$ is the wealth grid. $F_t(j)$ denotes the cumulative probability at wealth level $\mathcal{A}_j$ at time t, and $\mathbf{F}_t := [F_t(j)]_j$. $a_t^*(i) \in \mathcal{A}_t^*$ is the optimal policy at wealth level $\mathcal{A}_i$ at time t.

The "lottery" method (LM) represents the dynamics of F as

$$F_{t+1}(m) = \sum_{j=1}^{m} \sum_{i} \mathbf{A}_t^*(i,j) \left(F_t(i) - F_t(i-1)\right), \tag{16}$$

where $\mathbf{A}_t^*(i,j) = \mathbb{I}_{a_t^*(i)\in[\mathcal{A}_j,\mathcal{A}_{j+1})} \frac{\mathcal{A}_{j+1}-a_t^*(i)}{\mathcal{A}_{j+1}-\mathcal{A}_j} + \mathbb{I}_{a_t^*(i)\in[\mathcal{A}_{j-1},\mathcal{A}_j)} \frac{a_t^*(i)-\mathcal{A}_{j-1}}{\mathcal{A}_j-\mathcal{A}_{j-1}}$. Clearly, $\mathbf{A}_t^*(i,j)$ is linear in optimal policies a^*.

DEGM, instead, works through an interpolation

$$F_{t+1}(j) = \hat{\tilde{F}}(\mathcal{A}_j \mid \mathcal{A}_t^*, \mathbf{F}_t), \tag{17}$$

with values at interpolation nodes $\hat{\tilde{F}}(a_t^*(i) \mid \mathcal{A}_t^*, \mathbf{F}_t) = F_t(i)$. If the interpolator is piecewise linear (linear spline), *DEGM* has a structure similar to *LM*:

$$F_{t+1}(j) = \sum_{i:\, a_t^*(i-1)<\mathcal{A}_j\le a_t^*(i)} \mathbf{A}_t^{*,L}(i,j) \left(F_t(i) - F_t(i-1)\right) + F_t(i-1), \tag{18}$$

where $\mathbf{A}_t^{*,L}(i,j) = \frac{\mathcal{A}_j - a_t^*(i-1)}{a_t^*(i)-a_t^*(i-1)}$. Still, it is nonlinear in a^* as the optimal policies are the interpolation nodes in *DEGM*.

Instead, we use a cubic spline for interpolation as it captures nonlinearities of the continuous limit (see Appendix A.3). This adds smoothness conditions to the interpolation nodes. In practice, we use a piecewise cubic Hermite interpolating polynomial algorithm that preserves monotonicity (Fritsch & Butland, 1984), with a one-sided approximation

[18] We call a method "discretized" when a continuous distribution is represented by a finite vector.

of the slopes at the endpoints. A change at an interpolation node-value pair $(a_t^*(i), F_t(i))$ affects the entire interpolated function, so *DEGM* has the general form of Equation (17).

A.2 Generalization and Higher-Order Terms

To discuss the perturbation attributes of distributional dynamics, we propose the form

$$F_{t+1} = \mathbf{A}(a_t^*)F_t, \tag{19}$$

where we treat the distribution F and the optimal policies a^* as scalars for ease of exposition. This structure captures the "lottery" method and piecewise linear interpolation exactly, and *DEGM* with our implementation of cubic spline interpolation approximately.[19] We analyze the Taylor expansion of this dynamic with respect to the deviation $\hat{F}_t = F_t - \bar{F}$ and disturbances in aggregate variables that affect optimal policies. While in the main text, we consider a disturbance of aggregate state variable ξ_t, here we are more specific and account for the fact that state ξ_t affects optimal policies only through *control* variables P_t and ν_t.[20] Thus, we approximate around a steady state characterized by $\bar{F}$, $\bar{P}$, $\bar{\nu}$, and $\bar{a}^*$, where the disturbances $\hat{F}_t$, $\hat{P}_t$, and $\hat{\nu}_t$ are zero. The terms in red are zero for methods that are linear in optimal policies.

First-order approximation:

$$F_{t+1} \approx \bar{F} + \mathbf{A}(\bar{a}^*)\hat{F}_t + \frac{\partial \mathbf{A}(\bar{a}^*)}{\partial a^*}\bar{F}\left(\frac{\partial a_t^*}{\partial P_t}\hat{P}_t + \frac{\partial a_t^*}{\partial \nu_t}\hat{\nu}_t\right) \tag{20}$$

Second-order additional terms:

[19] Locally, $\hat{\bar{F}}(\mathcal{A}_j \mid \mathcal{A}_t^*, \mathbf{F}_t) \approx \hat{\bar{F}}_F(\mathcal{A}_j \mid \mathcal{A}_t^*, \bar{\mathbf{F}})\hat{\mathbf{F}}_t$. The simple cubic spline interpolation is exactly linear in the vector $\mathbf{F}_t$, but by taking the harmonic mean of the neighboring slopes at the interpolation nodes, which preserves monotonicity, our method loses this property. We abstract from this implementation detail.

[20] That is, $\frac{\partial a_t^*}{\partial \xi_t} = \frac{\partial a_t^*}{\partial P_t}\frac{\partial P_t}{\partial \xi_t} + \frac{\partial a_t^*}{\partial \nu_t}\frac{\partial \nu_t}{\partial \xi_t}$. P_t has the interpretation of market prices in t, while ν_t has the interpretation of marginal values at idiosyncratic states in period $t+1$, expected in t.

$$\underbrace{\frac{\partial \mathbf{A}(\bar{a}^*)}{\partial a^*}\hat{F}_t\left(\frac{\partial a_t^*}{\partial P_t}\hat{P}_t+\frac{\partial a_t^*}{\partial \nu_t}\hat{\nu}_t\right)}_{\textbf{(I)}}+\underbrace{\left(\frac{\partial \mathbf{A}(\bar{a}^*)}{\partial a^*}\frac{\partial^2 a_t^*}{\partial P_t \partial \nu_t}+\frac{\partial^2 \mathbf{A}(\bar{a}^*)}{\partial a^{*2}}\frac{\partial a_t^*}{\partial P_t}\frac{\partial a_t^*}{\partial \nu_t}\right)}_{\textbf{(II)}}\bar{F}\hat{P}_t\hat{\nu}_t+$$
$$+\underbrace{\frac{\bar{F}}{2}\left[\frac{\partial \mathbf{A}(\bar{a}^*)}{\partial a^*}\left(\frac{\partial^2 a_t^*}{\partial P_t^2}\hat{P}_t^2+\frac{\partial^2 a_t^*}{\partial \nu_t^2}\hat{\nu}_t^2\right)+\frac{\partial^2 \mathbf{A}(\bar{a}^*)}{\partial a^{*2}}\left(\left(\frac{\partial a_t^*}{\partial P_t}\right)^2\hat{P}_t^2+\left(\frac{\partial a_t^*}{\partial \nu_t}\right)^2\hat{\nu}_t^2\right)\right]}_{\textbf{(III)}} \tag{21}$$

Third-order additional terms:

$$\underbrace{\hat{F}_t\left(\textbf{(II)}\hat{P}_t\hat{\nu}_t+\frac{1}{2}\textbf{(III)}\right)}_{\textbf{(I*)}}+\underbrace{\frac{\bar{F}}{6}\hat{P}_t^3\left(\frac{\partial \mathbf{A}(\bar{a}^*)}{\partial a^*}\frac{\partial^3 a_t^*}{\partial P_t^3}+3\frac{\partial^2 \mathbf{A}(\bar{a}^*)}{\partial a^{*2}}\frac{\partial^2 a_t^*}{\partial P_t^2}\frac{\partial a_t^*}{\partial P_t}+\frac{\partial^3 \mathbf{A}(\bar{a}^*)}{\partial a^{*3}}\left(\frac{\partial a_t^*}{\partial P_t}\right)^3\right)}_{\textbf{(III*)}_P}+$$
$$+\underbrace{\frac{\bar{F}}{2}\hat{P}_t^2\hat{\nu}_t\left(\frac{\partial \mathbf{A}(\bar{a}^*)}{\partial a^*}\frac{\partial^3 a_t^*}{\partial P_t^2 \partial \nu_t}+\frac{\partial^2 \mathbf{A}(\bar{a}^*)}{\partial a^{*2}}\left(\frac{\partial^2 a_t^*}{\partial P_t^2}\frac{\partial a_t^*}{\partial \nu_t}+2\frac{\partial^2 a_t^*}{\partial P_t \partial \nu_t}\frac{\partial a_t^*}{\partial P_t}\right)+\frac{\partial^3 \mathbf{A}(\bar{a}^*)}{\partial a^{*3}}\frac{\partial a_t^*}{\partial \nu_t}\left(\frac{\partial a_t^*}{\partial P_t}\right)^2\right)}_{\textbf{(II*)}_{P^2\nu}}+$$
$$+\underbrace{\frac{\bar{F}}{6}\hat{\nu}_t^3\left(\cdots\right)}_{\textbf{(III*)}_\nu}+\underbrace{\frac{\bar{F}}{2}\hat{\nu}_t^2\hat{P}_t\left(\cdots\right)}_{\textbf{(II*)}_{\nu^2 P}} \tag{22}$$

Term **(I)** captures an interaction effect with the distribution: the distributional dynamics are different when the optimal policy is different from steady state. Term **(II)** captures an interaction effect between control variables: the simultaneous deviation of prices and value functions from steady-state affects the pass-through of the steady-state distribution via nonlinear optimal polices, and via an interaction of their respective first-order effects on optimal policies. Term **(III)** captures the effects of second-order fluctuations in control variables, which operate through the same channels as the interaction effect. Terms **(I*)**, **(II*)**:=**(II*)**$_{P^2\nu}$+**(II*)**$_{\nu^2 P}$ and **(III*)**:=**(III*)**$_P$+**(III*)**$_\nu$ are the analogons at third order, where **(III*)** captures skewness in control variables.

A.3 Nonlinearity of the Kolmogorov Forward Equation in a*

In Appendix A.2, we show that (discretized) distributional dynamics should be nonlinear in optimal policies to account for higher-order effects. We now analyze the distributional dynamics in the continuous limit, which provides an intuition for the cause of the nonlinearities and shows why it is crucial for the interpolation method of *DEGM* to account for the curvature of the distribution.

To see that even in the limit *LM* misses a potentially important nonlinearity, again use the monotonicity of the policy function to write the end-of-period distribution as:

$$\tilde{F}_t(a' \mid y) = \int_{x|a^*(x,y|\xi)\leq a'} f_t(x \mid y)dx = \int^{a^{*-1}(a',y|\xi)} f_t(x \mid y)dx. \tag{23}$$

Here we explicitly denote that optimal policies depend on some aggregate variable(s) ξ.

This implies that the first-order (Frechet) derivative of $\tilde{F}_t$ w.r.t. some variable ξ is

$$\frac{\partial \tilde{F}_t(a' \mid y)}{\partial \xi} = f_t(a^{*-1}(a', y \mid \xi) \mid y)\frac{\partial a^{*-1}(a', y \mid \xi)}{\partial \xi}, \tag{24}$$

so that the second-order derivative of $\tilde{F}_t$ w.r.t. some variable ξ_t is

$$\begin{aligned} \frac{\partial^2 \tilde{F}_t(a' \mid y)}{\partial \xi^2} &= f_t(a^{*-1}(a', y \mid \xi) \mid y)\frac{\partial^2 a^{*-1}(a', y \mid \xi)}{\partial \xi^2} \\ &+ \frac{\partial f_t(a^{*-1}(a', y \mid \xi) \mid y)}{\partial a}\left(\frac{\partial a^{*-1}(a', y \mid \xi)}{\partial \xi}\right)^2. \end{aligned} \tag{25}$$

This shows that the nonlinear effects of a change in ξ are composed of a nonlinear effect on the policies (here their inverse) and the derivative of the density w.r.t. a at the pre-image of a' times the squared linear effect of ξ on the (inverse) policy. The effect on the distribution F_{t+1} will be the average over income shocks. Thus, on average, the importance of the second term will be smaller for more symmetric distributions.

B Higher-Order Perturbation Solution of Heterogeneous Agent Models

B.1 Setting up the Model as a Difference Equation

We write the model as a difference equation $\Xi(\cdot) = 0$, where Ξ combines all equilibrium conditions from heterogeneous households, Ξ^i, and representative firms as well as market clearing, Ξ^A:

$$\Xi\left(F_t, S_t, \nu_t, P_t, F_{t+1}, S_{t+1}, \nu_{t+1}, P_{t+1}, \varepsilon_{t+1}\right) = \begin{bmatrix} \Xi^i(\cdot) \\ \Xi^A(\cdot) \end{bmatrix} \tag{26}$$

$$\Xi^i(\cdot) := \begin{bmatrix} F_{t+1} - L(F_t, h_t^k) \\ \nu_t - \left(u_{h_t^k} + \beta \mathbb{E}_{y'} \nu_{t+1}\right) \end{bmatrix} \tag{27}$$

$$\Xi^A(\cdot) := \begin{bmatrix} S_{t+1} - H(S_t) + \eta \varepsilon_{t+1} \\ \Phi_t(h_t^k, F_t) \\ \varepsilon_{t+1} \end{bmatrix} \tag{28}$$

s.t.

$$h_t^k(k, y) = \arg \max_{k' \in \Gamma(y,k;P_t)} u(y, k, k') + \beta \mathbb{E}_{y'|y} \nu_{t+1}(y', k'), \tag{29}$$

where F_t is the cumulative distribution function over idiosyncratic states (k, y), ν_t is the value function of households, $S_t \in \mathbb{R}^n$ denote the aggregate states in this economy other than the distribution of agents over their idiosyncratic states, and P_t denote aggregate prices (or other aggregate controls). We solve the system for the state dynamics $h : (F_t, S_t) \mapsto (F_{t+1}, S_{t+1})$ and the state-to-control mapping $g : (F_t, S_t) \mapsto (\nu_t, P_t)$. These functions are implicitly defined by

$$\mathbb{E}_t\left[\Xi\left((F_t, S_t), g(F_t, S_t), h(F_t, S_t), g(h(F_t, S_t)), \sigma \varepsilon_{t+1}\right)\right] = 0, \tag{30}$$

where σ is the perturbation parameter. Following Reiter (2002), we use a Taylor expansion around the non-stochastic steady state characterized by $\sigma = 0$, to approximate the

response to aggregate shocks ε_t. This requires differentiating Ξ up to the desired order of the Taylor approximation of h and g.

B.2 State-Space Reduction

In the main text, we show that four alternative ways to approximate the distribution and value functions lead to virtually the same results.

First, we consider a version where F and ν are represented by a set of values on the full idiosyncratic state space (Cartesian product).

Second, we consider a version where F is represented by marginal distributions and their copula, which again is represented by a set of values on a coarser grid of percentiles compared to the full idiosyncratic state space, but the value function ν remains unchanged. Here, we use 20 gridpoints along the asset dimension and the full grid along the human capital dimension to have a fine representation of the copula.

Third, we consider a version that reduces the dimensionality of the (marginal) value function by approximating $\nu_t \approx \mathcal{S}\hat{\nu}_t$, where the projection matrix $\mathcal{S}$ is obtained from a DCT of the steady-state value function as described in Bayer and Luetticke (2020). Here, we keep 99.99% of the variation of the steady-state value function in an R^2-sense.

Fourth, we implement the refinement from Bayer et al. (2024) that uses the first-order approximation of the model with a sufficiently rich set of shocks to obtain the resulting factor structure of the copula function, where we keep all eigenvalues larger than 1e-16 from the Jordan eigenvalue decomposition. Similarly, one could derive the factor representation of the value function. We do not need this for the third-order solution of our model, because the number of controls is already small enough, and numerically more expensive is the number of state variables. The first-order solution uses the Bayer and Luetticke (2020) reduction. The sufficiently rich set of shocks is necessary to approximate the dynamic effects of all current and future relative price changes that appear under the higher-order approximation. For example, in the first-order approximation with a fixed discount factor, there is a one-to-one relationship between the ratio of expected marginal utilities and the interest rate. In the higher-order solution, this relationship is not as tight

because of varying risk premia. In a model with just one asset, a shock to the discount factor can serve as a stand-in in the first-order solution.

In order to extend Bayer and Luetticke (2020) and Bayer et al. (2024) to higher-order perturbations and the use of *DEGM*, we first adjust the method by using splines instead of linear interpolants to represent copulas and value functions. Second, we write up the marginal distributions in terms of continuous CDFs instead of point-masses (and hence PDFs). When perturbing the CDFs, we make sure that the perturbation preserves essential CDF characteristics—namely, monotonicity and confinement within the unit interval.

B.3 Solution of Higher-Order Perturbations

Once the derivatives of Ξ are calculated, we solve for the derivatives of h and g by writing up a system of linear equations in the concise manner of Levintal (2017). Since our model is much larger than what Levintal solves, we innovate in terms of how to sparsely set up components like the permutation matrix and in terms of how to efficiently compute matrix Kronecker products. Finally, we use pruned dynamics when computing endogenous moments and generalized impulse responses, following Andreasen et al. (2018).

More in detail, to compute the second-order component of the Taylor approximations of h and g, denoted by h_{xx} and g_{xx}, respectively, we set up the condition

$$A_2 + \Xi_4 g_{xx} B_2 + \Xi_2 g_{xx} + (\Xi_3 + \Xi_4 g_x) h_{xx} = 0 \tag{31}$$

which follows from taking the total derivatives of Equation (30). We use the notation of Levintal (Equation (17) therein): A_2 is the Hessian of the system of model equations Ξ multiplied with (the product of) first-order derivatives of h and g, and B_2 contains the sum of the Kronecker-square of the first-order derivative of h and of shock variances. Note that the first-order derivatives of h and g are already taken as given here.

As outlined in Levintal (2017), Equation (31) can be split up into two blocks, where the first solves for higher-order system dynamics, and the second afterwards for the risk

correction terms $h_{\sigma\sigma}$ and $g_{\sigma\sigma}$. The first block can be rewritten in terms of a generalized Sylvester equation:

$$\mathcal{X} + \mathcal{A}^{-1}\mathcal{B}\mathcal{X}h_x^{\otimes 2} + \mathcal{A}^{-1}A_2 = 0 \tag{32}$$

with $\mathcal{X} = [h_{xx}, g_{xx}]$, $\mathcal{A} = [\Xi_3 + \Xi_4 gx \ \ \Xi_2]$ and $\mathcal{B} = [0 \ \ \Xi_4]$. We use that the lower rows contain conditions on g_{xx} only, due to the zero-columns in $\mathcal{B}$, while h_{xx} can be computed closed-form once g_{xx} is known. We solve the equation for g_{xx} iteratively, using the doubling algorithm proposed by Kim et al. (2008).

Levintal (2017) applies the method to models with about 20 variables, while we use our implementation to solve models with up to 3200 variables second order and with up to 220 variables third order. To achieve this, we innovate mainly on two accounts: We generate and store large matrices such as A_2 sparsely, and we avoid the direct computation of large Kronecker products by computing the product of a Kronecker product with another matrix in a two-step procedure:[21]

$$A(B \otimes B) = \text{reshape}\left(\underbrace{[\text{vec}(A(1)B) \cdots \text{vec}(A(n)B)]}_{=:M} B\right) \tag{33}$$

where $A(k)$ denotes the k-th column block of A with column-size n. M is computed first and the matrix product MB second. These numerical optimizations are necessary to avoid running out of memory. To see this, consider the Hessian of a model (without dimensionality reduction) with $n_y = 10$ and $n_k = 160$ gridpoints. This Hessian has roughly $3200 \times (2 \times 3200)^2 \approx 130$ billion entries,[22] at least half of which are typically zeros, half of which are known upfront, as outlined in Bayer and Luetticke (2020).

[21] This uses the property of Kronecker products that $A_1 \otimes A_2 = (A_1 \otimes I_{m_2})(I_{n_1} \otimes A_2)$, where m_2 is the number or rows of matrix A_2, and n_1 is the number of columns of matrix A_1. See, e.g., Fackler (2019) for a recent discussion on memory savings when multiplying a chain of Kronecker products with a matrix.

[22] The system has roughly $10 \times 160 = 1600$ states (the distribution) and $10 \times 160 = 1600$ controls (optimal value function over the distribution), and thus 3200 equilibrium conditions in total. Both present and future states and controls enter as variables in the system, which sums up to 2×3200 variables. For each equilibrium condition, the Hessian stores second-order derivatives and cross-derivatives of all variables, which amounts to $(2 \times 3200)^2$ entries by equilibrium condition. We keep symmetric cross-derivatives, as it simplifies setting up the Sylvester equation, following Levintal (2017).

B.4 Calculating Central Moments of the Ergodic Distribution of the Higher-Order Solution

We compute all first and second moments of the higher-order solutions using closed-form solutions. We describe the closed-form solution for the first moment of the states of the second-order solution below. We proceed analogously for the higher-order moments and solutions: Taking the matrices of the respective pruned state-space system from Andreasen et al. (2018), we iteratively go through the rows, starting at the bottom, and solve the respective matrix equations for the unconditional moments.

The state space system in Andreasen et al. (2018) for the second-order moment with pruning is:

$$E_t \begin{pmatrix} x^f_{t+1} \\ x^s_{t+1} \\ x^f_{t+1} \otimes x^f_{t+1} \end{pmatrix} = \begin{pmatrix} h_x & 0 & 0 \\ 0 & h_x & \frac{1}{2} H_{xx} \\ 0 & 0 & h_x \otimes h_x \end{pmatrix} \begin{pmatrix} x^f_t \\ x^s_t \\ x^f_t \otimes x^f_t \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{2} h_{\sigma\sigma} \\ (\eta \otimes \eta) vec(\Sigma_\epsilon) \end{pmatrix} \tag{34}$$

One can solve this directly in the unconditional expectation of $E[x^f; x^s; x^f \otimes x^f]^T$, and then $E[x^f] + E[x^s]$ is the unconditional first moment up to second order, *pruned* of all higher-than-second-order terms.

Due to its recursive block structure, one can also solve the system iteratively. This yields the same procedure as proposed by Rudebusch and Swanson (2012). First, rewrite the last row from the Kronecker-product notation into the (more familiar) second-moment notation. For this, note that $E[x^f \otimes x^f] = vec(\Sigma_{x^f})$, where Σ_x denotes the (uncentered) variance-covariance matrix of x. Then, the last row is

$$vec(\Sigma_{x^f}) = (h_x \otimes h_x) vec(\Sigma_{x^f}) + (\eta \otimes \eta) vec(\Sigma_\epsilon) \tag{35}$$

$$\Leftrightarrow \Sigma_{x^f} = h_x \Sigma_{x^f} h_x^T + \eta \Sigma_\epsilon \eta^T \tag{36}$$

This is exactly the equation Rudebusch and Swanson (2012) solve for the variance-

covariance matrix, which they call Σ_x. This is where the pruning happens, as this equation only includes first-order dynamics.

The second row in system (34) is then

$$E[x^s] = h_x E[x^s] + \frac{1}{2}\left(H_{xx} vec(\Sigma_{x^f}) + h_{\sigma\sigma}\right) \tag{37}$$

This is exactly the equation Rudebusch and Swanson (2012) solve for the unconditional second moments, after plugging in Σ_x as computed from first-order dynamics. As our system is defined in deviations from steady state, $E[x^f] = 0$ trivially.

Next, we describe the closed-form solution for the first moments of the states of the third-order solution with pruning. The third-order system extends the second-order system by three row-blocks and three column-blocks:

$$\begin{aligned} E_t \begin{pmatrix} x^f_{t+1} \\ x^s_{t+1} \\ {x^f_{t+1}}^{\otimes 2} \\ x^t_{t+1} \\ x^f_{t+1} \otimes x^s_{t+1} \\ {x^f_{t+1}}^{\otimes 3} \end{pmatrix} &= \begin{pmatrix} h_x & 0 & 0 & 0 & 0 & 0 \\ 0 & h_x & \frac{1}{2}H_{xx} & 0 & 0 & 0 \\ 0 & 0 & h_x \otimes h_x & 0 & 0 & 0 \\ \frac{3}{6}h_{\sigma\sigma x} & 0 & 0 & h_x & H_{xx} & \frac{1}{6}H_{xxx} \\ h_x \otimes \frac{1}{2}h_{\sigma\sigma} & 0 & 0 & 0 & h_x^{\otimes 2} & h_x \otimes \frac{1}{2}H_{xx} \\ 0 & 0 & 0 & 0 & 0 & h_x^{\otimes 3} \end{pmatrix} \begin{pmatrix} x^f_t \\ x^s_t \\ {x^f_t}^{\otimes 2} \\ x^t_t \\ x^f_t \otimes x^s_t \\ {x^f_t}^{\otimes 3} \end{pmatrix} \\ &+ \begin{pmatrix} 0 \\ \frac{1}{2}h_{\sigma\sigma} \\ (\eta \otimes \eta) vec(\Sigma_\epsilon) \\ \frac{1}{6}h_{\sigma\sigma\sigma} \\ 0 \\ (\eta^{\otimes 3}) E[\epsilon^{\otimes 3}] \end{pmatrix} \end{aligned} \tag{38}$$

We solve for the unconditional expectations of the last three row-blocks. First, we solve for the unconditional expectation of the triple Kronecker product of the first-order

dynamics:

$$E[x^{f\otimes 3}] = h_x^{\otimes 3} E[x^{f\otimes 3}] + (\eta^{\otimes 3}) E[\epsilon^{\otimes 3}] \tag{39}$$

where $E[\epsilon^{\otimes 3}]$ contains three types of non-zero elements: expectations of all triple-combinations of shocks, expectations of all tuple-combinations of shocks where one shock is squared, and third moments of all shocks. We use the doubling algorithm (Kim et al., 2008) to solve (39) for $E[x^{f\otimes 3}]$. In order to achieve this computationally, we use that $(h_x^{\otimes 3})^2 = (h_x^2)^{\otimes 3}$, and compute the Kronecker-matrix product $((h_x^2)^{\otimes 2} \otimes h_x^2) E[x^{f\otimes 3}]$ only indirectly, as described in Appendix B.3.

Next, we solve for the unconditional expectation of the Kronecker product of the first- and the additional second-order dynamics, using the same techniques:

$$E[x^f \otimes x^s] = h_x^{\otimes 2} E[x^f \otimes x^s] + (h_x \otimes \frac{1}{2} H_{xx}) E[x^{f\otimes 3}] \tag{40}$$

Note that this equation follows from $E[x^f] = 0$ dropping out of the second-to-last row block. Lastly, we set up the equation to solve for the unconditional expectation of the additional third-order dynamics, $E[x^t]$:

$$E[x^t] = h_x E[x^t] + H_{xx} E[x^f \otimes x^s] + \frac{1}{6}\left(H_{xxx} E[x^{f\otimes 3}] + h_{\sigma\sigma\sigma}\right) \tag{41}$$

The sum $E[x^s] + E[x^t]$ is the unconditional first moment up to third order, pruned of all higher-than-third-order terms, of the deviations of states from their values in non-stochastic steady state.

For computing the second moments of the second-order system, we follow "Method 3" of the Online Appendix of Andreasen et al. (2018). We need to compute the elements of the matrix

$$E[z_t z_t'] = E\left[\begin{pmatrix} x_t^f \\ x_t^s \\ x_t^{f\otimes 2} \end{pmatrix} \begin{pmatrix} (x_t^f)' & (x_t^s)' & (x_t^{f\otimes 2})' \end{pmatrix}\right] \tag{42}$$

First, we compute $E[(x_t^{f\otimes 2})(x_t^{f\otimes 2})']$. We use that ${x_{t+1}^f}^{\otimes 2} = h_x^{\otimes 2} x_t^{f\otimes 2} + v_{t+1}$, where we define v_t as the variable that contains all interactions with shocks in the system:

$$v_{t+1} = (h_x \otimes \eta)(x_t^f \otimes \epsilon_{t+1}) + (\eta \otimes h_x)(\epsilon_{t+1} \otimes x_t^f) + (\eta^{\otimes 2})(\epsilon_{t+1}^{\otimes 2}) \tag{43}$$

This leads to

$$\begin{aligned} E[(x_t^{f\otimes 2})(x_t^{f\otimes 2})'] = h_x^{\otimes 2} E[(x_t^{f\otimes 2})(x_t^{f\otimes 2})'](h_x^{\otimes 2})' + h_x^{\otimes 2} vec(\Sigma_x) E[v_{t+1}'] \\ + E[v_{t+1}] vec(\Sigma_x)'(h_x^{\otimes 2})' + E[v_{t+1} v_{t+1}'] \end{aligned} \tag{44}$$

Since $E[x_t^f \otimes \epsilon_{t+1}] = E[x_t^f \otimes E_t \epsilon_{t+1}] = 0$ by the Law of Iterated Expectations, we find that $E[v_{t+1}] = vec(\eta \Sigma_\epsilon \eta')$. To compute $E[v_{t+1} v_{t+1}']$, we first note that $E[(x_t^f \otimes \epsilon_{t+1})(\epsilon_{t+1}^{\otimes 2})'] = E[(x_t^f \otimes I_m)(I_n \otimes \epsilon_{t+1})(\epsilon_{t+1}^{\otimes 2})'] = 0$, since $E[x_t^f] = 0$. We then have

$$E[v_{t+1} v_{t+1}'] = (h_x \otimes \eta)(vec(\Sigma_x) \otimes vec(\Sigma_\epsilon))(h_x \otimes \eta)' \tag{45}$$

$$+ (h_x \otimes \eta) E[(x_t^f \otimes \epsilon_{t+1})(\epsilon_{t+1} \otimes x_t^f)'](\eta \otimes h_x)' \tag{46}$$

$$+ (\eta \otimes h_x) E[(\epsilon_{t+1} \otimes x_t^f)(x_t^f \otimes \epsilon_{t+1})'](h_x \otimes \eta)' \tag{47}$$

$$+ (\eta \otimes h_x)(vec(\Sigma_\epsilon) \otimes vec(\Sigma_x))(\eta \otimes h_x)' \tag{48}$$

$$+ (\eta \otimes \eta) E[(\epsilon_{t+1} \epsilon_{t+1}')^{\otimes 2}](\eta \otimes \eta)' \tag{49}$$

where $E[(\epsilon_{t+1} \epsilon_{t+1}')^{\otimes 2}]$ contains expectations of all quadruple-combinations of shocks, expectations of triple-combinations of shocks where one shock is squared, expectations of tuple-combinations of shocks where both shocks are squared, expectations of tuple-combinations of shocks where one shock is cubed, and the kurtosis of all shocks.[23] Then, we solve Lyapunov equation (44) for $E[(x_t^{f\otimes 2})(x_t^{f\otimes 2})']$ using the doubling algorithm, again by avoiding the direct computation of too large Kronecker products.

Setting up the Lyapunov equations for $E[x_t^s (x_t^{f\otimes 2})']$, $E[x_t^s (x_t^s)']$, and $E[x_t^f (x_t^s)']$ is more

[23] This shows that, even in a solution that is pruned up to the second order, higher-order shock moments enter into the variance of the system variables. The computed variance thus corresponds to the variance of a simulated second-order solution where shocks with non-zero skewness or kurtosis enter. See Andreasen et al. (2018) for a discussion.

straightforward, and we refer to the Online Appendix of Andreasen et al. (2018). When solving for $E[x_t^f({x_t^f}^{\otimes 2})']$, another set of unconditional expectations of shocks has to be computed, $E[\epsilon_{t+1}(\epsilon_{t+1}^{\otimes 2})']$, which includes the same non-zero elements as $E[\epsilon^{\otimes 3}]$ from above, including the shocks' skewness.

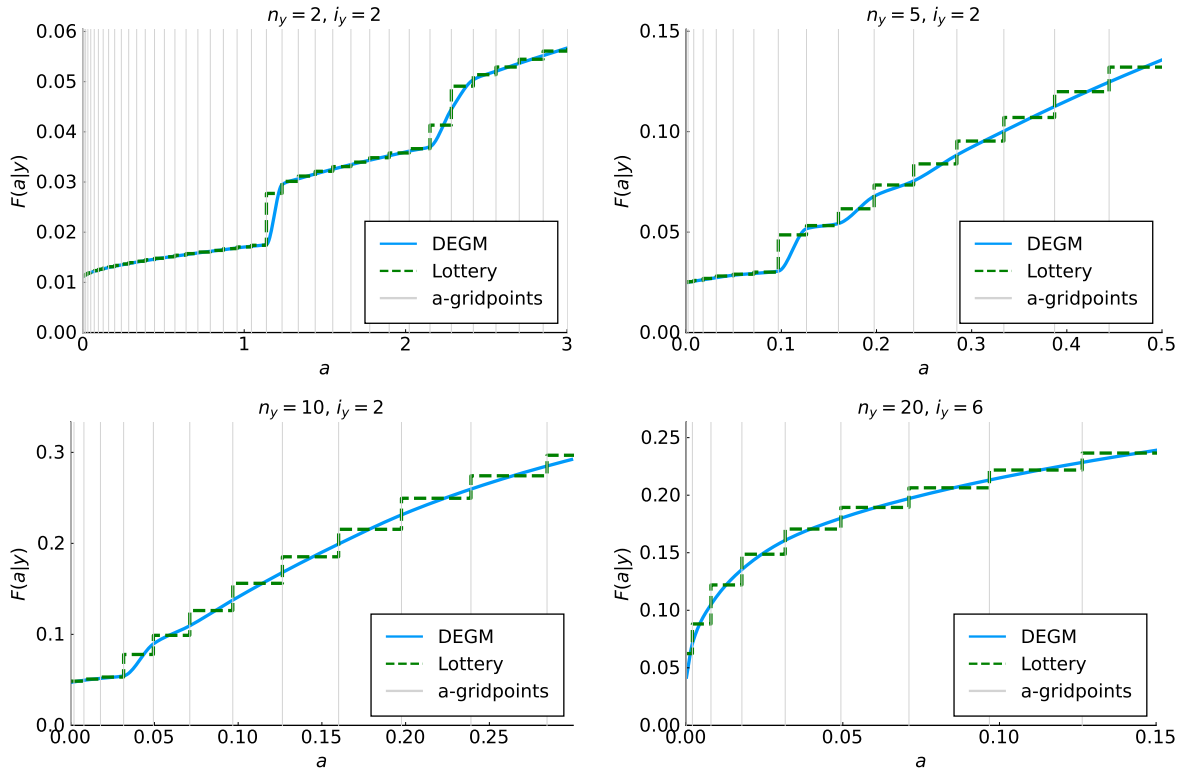
Finally, using $E[z] = \begin{pmatrix} 0 \\ E[x^s] \\ vec(\Sigma_x) \end{pmatrix}$, we compute the variance of z as

$$Var(z_t) = E[z_t z_t'] - E[z_t]E[z_t'] \tag{50}$$

From that, we obtain the variances of the control variables of the second-order system as $Var(y_t^s) = C_2 Var(z_t) C_2'$, with $C_2 = [g_x \; g_x \; \frac{1}{2} G_{xx}]$. The variances of the state variables are computed as $Var(x_t) = Var(x_t^f) + Var(x_t^s) + E[x_t^f(x_t^s)'] + E[x_t^s(x_t^f)']$.

Andreasen et al. (2018) do not provide the equations to solve recursively for the second moments of the third-order solution. As is clear, solving for the unconditional expectation $E[({x_t^f}^{\otimes 3})({x_t^f}^{\otimes 3})']$ is computationally challenging. Using our iterative solution technique, it requires the computation of terms $(h_x^{n\otimes 3})X(h_x^{n\otimes 3})'$ for a matrix X of size $n_s^3 \times n_s^3$, which we do not find feasible as of this writing.

Figure 3: Stationary distributions at selected income state



Notes: Conditional CDFs ($F(a|y) = P(x \leq a, z = y)/P(z = y)$) for discrete approximations of the income process with $n_y = 2, 5, 10, 20$ states. We plot the income state i_y, where the policy maps away from the constraint for the first time. The CDF at this income state inherits the biggest "echoes". "Lottery" refers to the method which assumes point-masses (Young, 2010).

C Properties of the wealth distributions for discrete and continuous income processes

In this appendix we prove properties of the distribution function in a generic consumption-savings problem with discrete or continuous income processes. The first property we show is that the stationary distribution is nowhere "flat" in the sense that the set of all points that are reached with strictly positive probability is dense. Second, we show that for the continuous income process limit, the distribution function inherits continuity and differentiability from the savings policy. Third, *DEGM* converges faster than *LM* to the true continuous distribution (in the number of gridpoints).

Figure 3 summarizes these results graphically. It shows the stationary distributions for different approximations of the income process ($n_y = 2, 5, 10, 20$) and solution meth-

ods. It conditions on the first income state i_y where households leave the borrowing constraint. We zoom in on the wealth level that the household chooses when leaving the borrowing constraint—i.e., we zoom in on the "echoes" from the mass point at the borrowing constraint.

We observe all of the above properties in Figure 3: First, even for two income states and both solution methods, we see that the distribution is nowhere "flat". Second, we observe the smoothing effect of increasing the number of income states. With ten income states, the echoes are almost gone, and with 20 income states they are practically invisible. Third, given that *DEGM* and *LM* converge to the same limit, we observe that *LM* with its steps overstates the distribution in between gridpoints. *DEGM* captures the continuous increase in mass also between gridpoints more accurately.

C.1 The set of attainable asset states is generically dense

We will show that for the simplest case of two income states, $\mathcal{A}^*$ is dense in a continuous interval, which is the support of the distribution. Here is the intuition up front: since we do *not* truncate histories, an infinite sequence of positive income shocks induces a *converging* wealth level, to a fixed point of the optimal policy. That implies density of $\mathcal{A}^*$ in infinitesimal neighborhoods left to that point. Then, we use the *inverse functions* of the optimal policies, to argue that any point in the support of the distribution could have been reached arbitrarily closely by some income sequence, starting from a wealth level in these neighborhoods. Of course, this does not hold for arbitrary policy functions but only for policy functions that adhere to some regularity conditions given below.

We have two income levels, y_l and y_h, $y_l < y_h$. This implies two optimal policy functions: $a_l(a) := a^*(a, y_l)$ and $a_h(a) := a^*(a, y_h)$, with $a \in \mathbb{R}_{\geq 0}$.

Assumption 1. ***Properties of the optimal policy functions***

The policy function a_l is strictly increasing for $a > \underline{a}_l$, and maps to 0 for $a \leq \underline{a}_l$. a_h is strictly increasing everywhere and $a_h(0) > 0$. Both a_l and a_h are continuous everywhere and $a_l \leq a_h$. There exists $\overline{a}_h > 0$ such that $\forall a > \overline{a}_h : a_h(a) < a$.

These assumptions typically hold in consumption-savings models as in Carroll (2006). The properties imply that there exists a fixed point a^f, i.e. $a_h(a^f) = a^f$. Let a^f denote the smallest such fixed point. Further, $a_l(a) < a$ for all $a \in [0, a^f]$, as $a_l(\underline{a}_l) = 0 < \underline{a}_l$ and $a_l(a^f) < a^f$, since $a_l < a_h$ everywhere.

We also assume that

Assumption 2. ***Dissaving at the top***

There exists some $\tilde{a}$ such that $a_h(a_l(a)) < a$ for all $a > \tilde{a}$, with $0 \leq \tilde{a} < a_l(a^f)$.

In words, we assume that for high enough wealth the household dissaves on net after a sequence of a low income realization followed by a high one. Conversely, the assumption implies $a_l^{-1}\left(a_h^{-1}(a)\right) > a$.

Assumption 3. ***Savings propensities strictly below one***

There exists a strictly positive number ϵ such that both inverse functions a_l^{-1} and a_h^{-1} have a slope above $1+\epsilon$ everywhere on $(a_h(0), a^f)$.

Note that this implies that the concatenation $x \mapsto a_l^{-1}(a_h^{-1}(x))$ also has a slope above $1+\epsilon$. We find that these properties hold for the policy functions that solve the consumption savings problem with a two income process for a typical calibration; see Figure 4 for the illustration of these properties for such an example.

$a_h^{\circ n}(a)$ denotes the n-times concatenation of the optimal policy map under a positive income shock, starting at a. The strict monotonicity and the existence of fixed point a^f implies that the sequence $\{a_h^{\circ n}(0)\}_n$ converges to a^f. a^f may not be in $\mathcal{A}^*$.

It is to show that $\mathcal{A}^*$ is dense in $[0, a^f)$.

Lemma 1. *$\mathcal{A}^*$ is dense in $(a_l(a^f), a^f)$.*

Proof by contradiction. Assume there exists an open interval $M_0 \subset (a_l(a^f), a^f)$ such that $\mathcal{A}^* \cap M_0 = \emptyset$. Then there must be no combination of income shocks that lead to an element of M_0. However, we can inversely construct the necessary sequence of income states iterating over sets M_i starting at M_0:

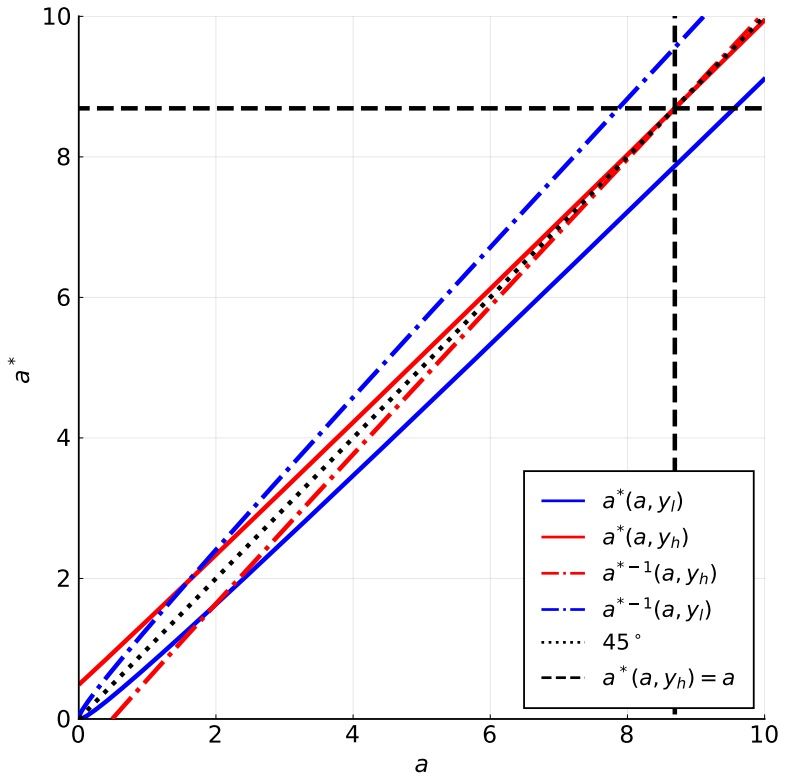


Figure 4: Policy functions for example income process

1. Map $M_i =: (x_i, z_i)$ to $\hat{M}_i = (\hat{x}_i, \hat{z}_i)$ by m times concatenating a_h^{-1}, such that $\hat{x}_i < a_l(a^f)$. This is always possible, as $a_h^{-1}(a) < a$ for all $a < a^f$. Choose smallest m that achieves this. Then define the next iteration $M_{i+1} := a_l^{-1}\left(\hat{M}_i\right) = \left(a_l^{-1}(\hat{x}_i), a_l^{-1}(\hat{z}_i)\right)$.

2. If $a^f \in M_{i+1}$, the proof is complete, as it cannot be that an open set around a^f has no overlap with $\mathcal{A}^*$ because a_h has a slope below one and a^f is reached only as the limit of a sequence of infinitely many high income realizations and thus $(a^f - \epsilon, a^f)$ contains points in $\mathcal{A}^*$ for any $\epsilon > 0$.

3. If $a^f \notin M_{i+1}$, then $z_{i+1} < a^f$ because $x_{i+1} < a^f$ by construction. At the same time, again by construction, the $m-1$-times concatenation of a_h^{-1} evaluated at x_i is above $a_l(a^f)$, and thus above $\tilde{a}$. Therefore, $x_{i+1} = a_l^{-1}(a_h^{-1}(a_h^{-1 \circ m-1}(x_i))) > a_l(a^f)$. Hence, $M_{i+1} := (x_{i+1}, z_{i+1}) \subset (a_l(a^f), a^f)$, and one returns to step 1.

The open intervals $\{M_i\}_i$ have no overlap with $\mathcal{A}^*$. Since a_h^{-1} has a slope above $1+\epsilon$, $\hat{z}_i - \hat{x}_i > (1+\epsilon)(z_i - x_i)$. Since $a_l^{-1} \circ a_h^{-1}$ has a slope above $1+\epsilon$, $z_{i+1} - x_{i+1} > (1+\epsilon)(\hat{z}_i - \hat{x}_i)$.

Hence, the intervals $\{M_i\}_i$ strictly increase in their size at least at rate $(1+\epsilon)^2$ in i. One can therefore choose a subsequence $\{i(k)\}_k$ of intervals where it holds that the upper bounds, $z_{i(k)}$, are strictly increasing, and where the intervals in the subsequence nest each other. Hence, if the iteration continues indefinitely, the sequence $\left\{z_{i(k)}\right\}_k$ converges to a^f. This is a contradiction to the existence of a convergent series in $\mathcal{A}^*$ with limit a^f: Choose an arbitrary index k of the subsequence of intervals $\left\{M_{i(k)}\right\}_k$ with upper limit $z_{i(k)}$. There must exist $a^* \in \mathcal{A}^* \cap (z_{i(k)}, a^f)$. But then, there is a $k' > k$ with $z_{i(k')} \in (a^*, a^f)$, and since intervals in the subsequence are nested, $x_{i(k')} < a^*$, which is contradictory. □

Lemma 2. *$\mathcal{A}^*$ is dense in $[0, a_l(a^f))$.*

Proof by contradiction. Assume there exists an open interval $M_0 := (x, z) \subset (0, a_l(a^f))$ with no overlap with $\mathcal{A}^*$. Construct a sequence of intervals $\{M_i\}_i$ with

$$M_i := \underbrace{a_l^{-1} \circ a_l^{-1} \circ \cdots \circ a_l^{-1}}_{\text{i times}}(M_0) \tag{51}$$

It must be that $(\cup_i M_i) \cap \mathcal{A}^* = \emptyset$. Denote the upper bound of M_i by z_i. Since by assumption, $a_l^{-1}(a) > a$ for $a \in (0, a^f]$, $\{z_i\}_i$ is a strictly increasing sequence. Suppose it was bounded above by $\kappa \leq a_l(a^f)$. This would imply that for all $\epsilon > 0$ small, $\kappa - \epsilon < a_l^{-1}(\kappa - \epsilon) < \kappa$. Since a_l^{-1} is continuous, this would imply $a_l^{-1}(\kappa) = \kappa$, which is a contradiction. Thus, $\cup_i M_i$ overlaps with $(a_l(a^f), a^f)$. This is a contradiction by Lemma 1. □

Combining Lemma 1 and Lemma 2, we have:

Theorem 1. *$\mathcal{A}^*$ is dense in $[0, a^f)$. If the borrowing constraint is binding, the set is countable.*

Proof. Denseness follows from Lemma 1 and 2. Countability from the borrowing constraint being binding. In that case, we can enumerate the sequence of income states since the borrowing constraint was last binding (coding a sequence like HLLH as 1001, HHLLH as 11001, etc.). □

C.2 Implications for approximation quality

Next, we discuss the ability of *DEGM*, as described in Section 2, to approximate the distribution over $\mathcal{A}^*$, which we define as F_{n_y}. That is, even if the income process is discrete, with a finite number n_y of income states, we show that *DEGM* with continuous interpolation approximates the true discrete distribution well. A crucial part of the proof is that $\mathcal{A}^*$ is dense in a continuous interval, which we have shown for the case $n_y = 2$, and which also holds for larger n_y.[24]

Theorem 2. *Let F'_{n_y} denote the distribution generated by one iteration of DEGM, starting at the true distribution F_{n_y} of $\mathcal{A}^*$, for a given wealth grid $\mathcal{A}$ and given optimal policy functions $a^*(a, y)$. For any $\delta > 0$, there exists $N > 0$ s.t. there is a finite set of points $\mathbb{A} \subset \mathcal{A}^*$, $|\mathbb{A}| \leq N$, s.t. $\forall\, a \in \mathcal{A}^* \setminus \mathbb{A}$, there exists $M > 0$ s.t. if $\mathcal{A}$ has at least size M, $|\mathcal{A}| \geq M$, it holds that*

$$|F'_{n_y}(a) - F_{n_y}(a)| < \delta \tag{52}$$

Proof. For a given $\delta > 0$, F_{n_y} can only have a finite number N of jump points where the jump is larger than $\delta/2$, i.e. points $\mathbb{A} := \{a_1, .., a_N\}$ s.t. for any $a_j \in \mathbb{A}$, $F_{n_y}(a_j - \epsilon) + \delta/2 < F_{n_y}(a_j)$ for all $\epsilon > 0$. This just follows from cumulative distribution function F_{n_y} being bounded above by 1. It implies that for any point $a \in \mathcal{A}^* \setminus \mathbb{A}$, there exists $\underline{\epsilon} > 0$ such that $F_{n_y}(a) - F_{n_y}(a - \underline{\epsilon}) \leq \delta/2$. Also, since $\mathbb{A}$ is not dense, there exists $\bar{\epsilon} > 0$ such that $F_{n_y}(a + \bar{\epsilon}) - F_{n_y}(a) \leq \delta/2$. Let $\epsilon := \min\{\underline{\epsilon}, \bar{\epsilon}\}$. Crucially, as $\mathcal{A}^*$ is dense in $\mathbb{R}$, the points $a \pm \epsilon$ can be assumed to be in the support of F_{n_y} w.l.o.g.

The interpolation nodes of *DEGM* are tuples consisting of optimal policy choices, $\mathcal{A}^*_{i,j}$, where $j = 1, .., n_y$ indexes the income level, and the cumulative probability of the corresponding on-grid wealth level $\mathcal{A}_i$. The wealth grid $\mathcal{A}$ with length M has the property that the maximum distance between neighboring gridpoints, $h := \max_i \{\mathcal{A}_i - \mathcal{A}_{i-1}\}_i$, falls in M. Since the optimal policy function is strictly monotone in wealth above the borrowing constraint, the maximum distance between interpolation nodes, $h^* :=$

[24] More income levels, stemming from a finer discretization of a continuous income process that remains the same, can only increase the variety of optimal policy choices, and thus the size of $\mathcal{A}^*$. Importantly, the properties of optimal policy functions used in the proof for the case $n_y = 2$ stay intact, as the underlying optimization problem of the household does not change.

$\max_i \left\{\mathcal{A}^*_{i,j} - \mathcal{A}^*_{i-1,j}\right\}_i$, also falls in M. For a given $a \in \mathcal{A}^* \setminus \mathbb{A}$, we can then increase M such that for each j there exists a smallest interval $(\mathcal{A}^*_{i-1,j}, \mathcal{A}^*_{i,j})$ that contains a, and where $(\mathcal{A}^*_{i-1,j}, \mathcal{A}^*_{i,j}) \subset (a-\epsilon, a+\epsilon)$.

We provide the proof for *DEGM* with linear spline interpolation. Cubic spline interpolation will only improve the approximation, as the degree of freedom of the interpolating polynomial increases. With linear spline interpolation, one *DEGM* step is given by

$$F'_{n_y}(a \mid \mathcal{Y}_k) = \sum_j \left\{ F_{n_y}(\mathcal{A}^*_{i,j} \mid \mathcal{Y}_j) \frac{a - \mathcal{A}^*_{i-1,j}}{\mathcal{A}^*_{i,j} - \mathcal{A}^*_{i-1,j}} + F_{n_y}(\mathcal{A}^*_{i-1,j} \mid \mathcal{Y}_j) \frac{\mathcal{A}^*_{i,j} - a}{\mathcal{A}^*_{i,j} - \mathcal{A}^*_{i-1,j}} \right\} \Pi_{j,k} \tag{53}$$

Repeatedly applying the triangle inequality, and using that the interpolation weights are bounded above by 1, we get

$$|F'_{n_y}(a \mid \mathcal{Y}_k) - F_{n_y}(a \mid \mathcal{Y}_k)| \leq \tag{54}$$

$$\sum_j \left\{ |F_{n_y}(\mathcal{A}^*_{i,j} \mid \mathcal{Y}_j) - F_{n_y}(a \mid \mathcal{Y}_j)| \frac{a - \mathcal{A}^*_{i-1,j}}{\mathcal{A}^*_{i,j} - \mathcal{A}^*_{i-1,j}} \right\} \Pi_{j,k} + \tag{55}$$

$$\sum_j \left\{ |F_{n_y}(\mathcal{A}^*_{i-1,j} \mid \mathcal{Y}_j) - F_{n_y}(a \mid \mathcal{Y}_j)| \frac{\mathcal{A}^*_{i,j} - a}{\mathcal{A}^*_{i,j} - \mathcal{A}^*_{i-1,j}} \right\} \Pi_{j,k} \tag{56}$$

$$< \frac{\delta}{2} 2 \sum_j \Pi_{j,k} = \delta \tag{57}$$

Integrating over income states gives the result for the marginal distribution:

$$|F'_{n_y}(a) - F_{n_y}(a)| \leq \sum_k |F'_{n_y}(a \mid \mathcal{Y}_k) - F_{n_y}(a \mid \mathcal{Y}_k)| P(\mathcal{Y}_k) < \delta \tag{58}$$

□

Theorem 2 shows that *DEGM* approximates the discrete distribution well with a continuous distribution. In fact, its approximation is likely better than the piecewise constant approximation of *LM* in most cases. The reason is that in between gridpoints, the distribution function is increasing (on a dense set). A piecewise constant approximation does not take this into account, so that for a fixed grid, there exists a small enough

δ where the set of points on which the deviation exceeds δ is dense itself. Confirming this intuition, in the next section we show that *DEGM* converges faster than *LM* to the limit distribution.

C.3 Continuous income processes: Continuity and differentiability

We denote the limit distribution of $\mathcal{A}^*$ for $n_y \to \infty$ as $\bar{F}$.

Theorem 3. *$\bar{F}$ is continuously differentiable on $(0, a^f]$ at the order of a^{*-1}.*

Proof. As in section 2, we assume log-income levels to follow an AR(1) process with normally distributed innovations, where $f(z)$ is the unconditional density of income level z, and ρ is the autocorrelation coefficient.

We first show the claim for the easiest case of $\rho = 0$. The optimal policy function a^* maps asset level a to asset level a', conditional on income state z. Since income shocks are iid, a^* only depends on z *through cash-on-hand* $c(a, z) := c_1 a + c_2 z$, where c_1, c_2 are positive, constant equilibrium prices. We therefore now define $a_c^* : c \mapsto a'$ as a map from cash-on-hand to next period's asset holding. By assumption, its inverse exists and is continuously differentiable at $a' \in (0, a^f]$ at the order of k. Since income shocks are iid, the distribution over wealth conditional on the *current realization* of income shocks equals the unconditional wealth distribution. We then can characterize the distribution through a one-step transition:

$$\bar{F}(a') = \int_0^\infty \bar{F}\left(\frac{a_c^{*-1}(a') - c_2 z}{c_1}\right) f(z) dz \tag{59}$$

$$= -\int_{a_c^{*-1}(a')/c_2}^{-\infty} \bar{F}\left(\tilde{z} c_2/c_1\right) f\left(a_c^{*-1}(a')/c_2 - \tilde{z}\right) d\tilde{z} \tag{60}$$

$$= \int_{\mathbb{R}} \bar{F}(\tilde{z} c_2/c_1) \mathbb{I}_{(-\infty, a_c^{*-1}(a^f)/c_2]}(\tilde{z}) h\left(a_c^{*-1}(a')/c_2 - \tilde{z}\right) d\tilde{z} \tag{61}$$

$$= (g * h)\left(a_c^{*-1}(a')/c_2\right) \tag{62}$$

where $g : \alpha \mapsto \bar{F}(\alpha c_2/c_1)\mathbb{I}_{(-\infty, a_c^{*-1}(a^f)/c_2]}(\alpha)$ is integrable, as the cdf $\bar{F}(\alpha)$ is bounded and vanishes for $\alpha < 0$, and $h(x) := f(x)$ if $x > 0$, $h(x) := 0$ o.w., inherits smoothness and

boundedness of its derivates from f.[25] Then, the convolution $g * h$ is well defined and, by Proposition 8.10 in Folland (1999), it is smooth. Note that in step (60), we change the integration variable. The result implies, by the chain rule, that $\bar{F}$ is continuously differentiable at the order of k on $a' \in (0, a^f]$, as was to show.

Next, we generalize the result for $\rho > 0$. Define $a_c^*(c, \nu(y))$ as the optimal policy function given cash-on-hand c, and the *additional* knowledge of income state y. Since incomes are persistent, y helps in forecasting future income, and thereby future marginal utility. Hence, y enters a^* independently of cash-on-hand only through ν, which signifies the future value function. We define conditional distributions such that the wealth distribution is conditioned on the *past* income level (what we denote as $\tilde{F}$ in the main text): $\bar{F}(a' \mid y)$, where $a' = a_c^*(c, \nu(y))$ for some c. Then, for some level of cash-on-hand c, we can write

$$
\begin{aligned}
&\int_0^\infty \bar{F}(a_c^*(c, \nu(y)) \mid y) f(y) dy \\
&\quad = \int_0^\infty \int_0^\infty \bar{F}\left(\frac{c - c_2 y}{c_1} \mid z\right) \pi(z \mid y) dz f(y) dy && (63) \\
&\quad = \int_0^\infty \int_{-\infty}^{c/c_2} \bar{F}\left(\tilde{y} c_2 / c_1 \mid z\right) \pi(z \mid c/c_2 - \tilde{y}) f(c/c_2 - \tilde{y}) d\tilde{y} dz && (64) \\
&\quad = \int_0^\infty (g_z * h_z)(c/c_2) dz =: \mu(c) && (65)
\end{aligned}
$$

for $g_z : \alpha \mapsto \bar{F}(\alpha c_2/c_1 \mid z) \mathbb{I}_{(-\infty, c(a^f, \bar{y})]}(\alpha)$, which is integrable, and $h_z(x) := \pi(z \mid x) f(x), x > 0$, $h_z(x) := 0$, o.w., which inherits smoothness and boundedness of its derivatives from the lognormal density. Therefore, the same proposition as used above applies and yields that the convolution $g_z * h_z$ is smooth. As $z \mapsto g_z * h_z$ is integrable, $\mu(c)$ inherits the smoothness.

We prove the final step by contradiction. Suppose the claim is for $k = 0$, i.e. continuity. If $\bar{F}$ is not continuous, there must exist y and $a' \in (0, a^f]$ s.t. a' is a jump point of $\bar{F}(\cdot \mid y)$. This implies that there exists a converging sequence $(a_n)_n \to a'$ s.t. $(\bar{F}(a_n \mid y))_n \not\to \bar{F}(a' \mid y)$. There is a corresponding sequence of cash-on-hand levels

[25] This is the case for the density of the lognormal distribution, as $\exp(-\ln(x)^2) x^{-n} \underset{x \to 0}{\longrightarrow} 0$ for all n.

$c_n = a_c^{*-1}(a_n, \nu(y))$. For all n, taking the derivative by y and evaluating at y on both sides of equation (65) gives

$$\bar{F}(a_n \mid y)f(y) = \frac{\partial}{\partial y}\mu(a_c^{*-1}(a_n, \nu(y))) \tag{66}$$

Taking the limit inside the differentiation, the right-hand side converges by continuity of a_c^{*-1} to $\frac{\partial}{\partial y}\mu(a_c^{*-1}(a', \nu(y))) = \bar{F}(a' \mid y)f(y)$. This is a contradiction. Contradictions for the case $k > 0$ can be constructed analogously, by additionally taking the respective number of derivatives by a on both sides of the equation. □

Corollary 1. *Let $\bar{F}^{DEGM}$ denote the distribution generated by one step* DEGM*, when the limit distribution $\bar{F}$ is the input. Let h denote the maximum distance between grid points $\mathcal{A}_i$, falling in M, the size of $\mathcal{A}$. Assume that a^{*-1} is at least twice continuously differentiable. Then, $\bar{F}^{DEGM}$ converges uniformly to $\bar{F}$ in M, at least at the rate at which $h^2 \to 0$. In contrast, $\bar{F}^{LM}$ converges to $\bar{F}$ only at the rate at which $h \to 0$.*

Proof. Let $h^* := max_{i,j}\left\{\mathcal{A}^*_{i,j} - \mathcal{A}^*_{i-1,j}\right\}$. Since a^* is differentiable, it is also Lipschitz continuous, hence there exists $L > 0$ s.t. $h^*(M) \leq Lh(M)$ for all grid sizes M. Assume that *DEGM* is done with linear Spline interpolation (note that higher-order interpolation has even better convergence properties). Since $\bar{F}$ is at least twice continuously differentiable by Theorem 3, we can apply classical results from interpolation theory, see e.g. Schaback and Wendland (2005), Folgerung 8.20 and Satz 11.3 therein:

$$|\bar{F}^{DEGM}(a) - \bar{F}(a)| \leq \frac{Lh^2||\bar{f}'||_\infty}{8} \tag{67}$$

for all $a \in (0, a^f]$, where $||\bar{f}'||_\infty$ denotes the supremum norm of the first derivative of the density of $\bar{F}$ on $(0, a^f]$. On the other hand, as F^{LM} is piecewise constant, i.e. interpolating polynomials are of order 0, theoretical convergence is only linear in the maximum distance

of grid points:[26]

$$|\bar{F}^{LM}(a) - \bar{F}(a)| \leq \frac{Lh||\bar{f}||_\infty}{8} \tag{68}$$

for all $a \in (0, a^f]$, where $||\bar{f}||_\infty$ denotes the supremum norm of the density of $\bar{F}$ on $(0, a^f]$. □

[26] As we point out in Appendix A.1, *LM* does not interpolate the CDF. With sufficient convergence, the *LM* solution crosses the true, continuous CDF in between gridpoints as well as at gridpoints, when it jumps (Figure 3 illustrates this, where *DEGM* solves the system assuming a continuous CDF, while *LM* assumes a discrete distribution). We can take the wealth levels where the crossings happen as the implicit "interpolation nodes" of *LM*. Interpreted in such a way, *LM* has double the amount of "interpolation nodes" as there are gridpoints. Hence, in order to map *LM* into the space of interpolation methods, we divide the maximum distance h by 2 when calculating the upper bound in Equation (68).

D Alternative Calibration

In the following we report convergence results for the transitory income calibration.

Table 6: Convergence under the Lottery method and DEGM with transitory calibration (calibration as in Table 1)

		Lottery			DEGM	
	n_h \ n_k	**40**	**80**	**160**	**40**	**80**
Panel A: Stationary distribution [relative deviations in percent]						
Capital stock	5	11.52	4.16	1.00	0.24	0.12
	10	8.37	2.99	0.78	0.28	0.10
	20	7.35	2.65	0.70	0.28	0.09
Wealth gini	5	10.66	4.61	1.15	-0.02	0.02
	10	7.86	3.22	0.89	-0.05	0.01
	20	6.78	2.75	0.79	-0.04	0.01
Panel B: Stationary equilibrium [relative deviations in percent]						
Capital stock	5	0.52	0.19	0.05	0.00	0.00
	10	0.47	0.17	0.05	0.01	0.00
	20	0.44	0.16	0.04	0.01	0.00
Wealth gini	5	11.27	4.65	1.21	0.22	0.02
	10	8.63	3.36	0.96	0.16	0.03
	20	7.60	2.91	0.87	0.17	0.03
Panel C: Computation times						
Time (s)	5	0.10	0.15	0.24	0.29	0.43
	10	0.18	0.26	0.45	0.41	0.65
	20	0.35	0.54	1.08	0.82	1.35

Notes: Values represent percent deviations of the solutions with n_k and n_h gridpoints to the reference solution, which is DEGM with $n_k = 160$ gridpoints for assets. The aggregates under the lottery methods are derived under discrete aggregation methods while continuous integration methods are used for DEGM. The table refers to the "transitory" income calibration from Table 1.
Panel A: calculating only the stationary distribution, using policies from the reference solution.
Panel B: solving the stationary equilibrium including prices and policies.
Panel C: time in seconds for solving the stationary equilibrium as in Panel B on a laptop with 16-core, 3.3 GHz CPU.

Table 7: Convergence at second order under the Lottery method and DEGM (transitory calibration as in Table 1)

		Lottery			DEGM	
	n_h \ n_k	**40**	**80**	**160**	**40**	**80**
Panel A: IRF statistic as in Bayer et al. (2024)						
Capital stock (FO)	5	1.00	1.00	1.00	1.00	1.00
	10	1.00	1.00	1.00	1.00	1.00
Capital stock (SO)	5	1.00	1.00	1.00	1.00	1.00
	10	1.00	1.00	1.00	1.00	1.00
Wealth Gini (FO)	5	1.00	1.00	1.00	1.00	1.00
	10	0.99	1.00	1.00	1.00	1.00
Wealth Gini (SO)	5	0.94	0.88	0.84	1.00	1.00
	10	0.94	0.88	0.84	1.00	1.00
Panel B: SO Moments [relative deviations in basis points]						
Capital stock	5	-0.87	-0.83	-0.37	-0.04	-0.03
Capital stock	10	-0.93	-0.90	-0.65	-0.28	-0.19
Wealth Gini	5	-20.21	-24.47	-31.11	2.75	0.75
Wealth Gini	10	-15.34	-18.11	-21.62	1.73	0.95

Notes: For each row, values represent basis point deviations of the solutions with n_k gridpoints to the reference solution (DEGM with $n_k = 160$). Values for the "transitory" income calibration from Table 1.

Panel A shows the R^2 like statistics from Bayer et al. (2024) for an impulse responses following a 7.5 p.p. shock to δ (over 100 periods) with first-order (FO) and second-order (SO) perturbation solution. The R^2 like statistics is $1 - \sum_{h=1}^{H} \frac{(IRF_{DEGM,n_k=160,n_h}(h) - IRF_{method,n_k,n_h}(h))^2}{\sum_{h=1}^{H} IRF_{DEGM,n_k=160,n_h}(h)^2}$.

Panel B shows ergodic moment solving the model with a second order perturbation. The deviation shown is the difference in the percentage point deviation from the respective steady state in excess of the percentage point deviation in the baseline solution (DEGM, $n_k = 160$). The aggregates under the Lottery methods are derived under discrete aggregation methods while continuous integration methods are used for DEGM.