An Endogenous Gridpoint Method for Distributional Dynamics*

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

April 25, 2024

Abstract

The "histogram method" (Young, 2010), while the standard approach for analyzing distributional dynamics in heterogeneous agent models, is linear in optimal policies. We introduce a novel method that captures nonlinearities of distributional dynamics. This method solves the distributional dynamics by interpolation instead of integration, which is made possible by making the grid endogenous. It retains the tractability and speed of the histogram method, while increasing numerical efficiency even in the steady state and producing significant economic differences in scenarios with aggregate risk. We document this by studying aggregate investment risk with a third-order solution using perturbation techniques.

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

JEL-Codes: C46, C63, E32

*We thank Anastasiia Antonova and Thomas Hintermaier for helpful comments. Christian Bayer gratefully acknowledges support by the DFG through CRC-TR224 EPoS, subarea C05 and in the framework of the German Excellence Strategy-EXC2126/1-390838866. Ralph Luetticke gratefully acknowledges support through the ERC-2023-STG project AIRMAC funded by the European Union's Horizon Program under grant agreement No. 101114991.

†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. 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 (Young, 2010; Reiter, 2009) or full integration (Krusell and Smith, 1998). Instead, we extend the idea of endogenous gridpoints (Carroll, 2006) to distributional dynamics.

We show that our distributional endogenous gridpoint method, henceforth *DEGM*, is as fast and tractable as the "histogram method" proposed by (Young, 2010), which has become the standard in the literature. We also show that our method converges faster in the number of gridpoints even when solving for the steady state 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 histogram 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 for studying 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).

For first order perturbations, *DEGM* gives the same solution as the histogram 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 histogram method does not capture all nonlinear effects, consistent with Bhandari et al. (2023). The histogram method overstates the aggregate and distributional responses to shocks. However, it understates the long-run aggregate and distributional consequences of the presence of investment risk. Using a third-order perturbation solution with *DEGM*, we find that the presence of aggregate investment risk hollows out the middle class and increases wealth inequality by 12%.

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 can also be found in the special issue Den Haan et al. (2010). Our reformulation with an endogenous grid approach makes it 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 y follows an exogenous discrete Markov process with transition probability matrix Π and set of states $\{\mathcal{Y}_j\}$. 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).[1] The cumulative joint distribution (in a) at time t is given by $F_t(a, y) := P(x \leq a, z = y)$, where $f_t(a, y)$ is the density (continuous in the a dimension, discrete in the y dimension).

2.1 Distributional Dynamics in Discrete Time

The evolution of the distribution F is then given by the time discrete Kolmogorov forward equation:

$$F_{t+1}(a', y') = \sum_j \int_{\{x|a' \geq a^*(x, \mathcal{Y}_j)\}} f_t(x, \mathcal{Y}_j)dx \quad \Pi(\mathcal{Y}_j, y'). \tag{1}$$

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 (1). To avoid this, Young (2010)

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

suggests replacing the continuous distribution in a with a discrete counterpart, what is commonly referred to as histogram method. This has a density given by the vector $\hat{f}$. One then replaces the policy function $a^*(a,y)$ by lotteries, so that the transition along both the a and y dimensions can be summarized by a single matrix $\mathbf{A}^*$. This gives the discretized stacked Kolmogorov forward equation:

$$\hat{f}_{t+1} = \hat{f}_t \mathbf{A}^*. \tag{2}$$

While this allows to solve $\hat{f}_t$ very fast, it not only creates an approximation error due to discretizing a continuous density, but also forces the Equation (2) to be linear in the optimal policies a^*.

2.2 An Endogenous Gridpoint Method for Distributional Dynamics

Instead, we propose to use an endogenous gridpoint method analogous to the one proposed by Carroll (2006). To do so, we divide each period into two subperiods, a first one related to asset choices and a second one related to income changes. Consequently, we define the distribution at the end of period t, after asset choices but before income transitions, as

$$\tilde{F}_t(a', \mathcal{Y}_j) := \int_{\{x|a' \geq a^*(x,\mathcal{Y}_j)\}} f_t(x, \mathcal{Y}_j) dx, \tag{3}$$

such that we obtain the asset-income distribution at the beginning of period $t+1$, i.e. after income transitions, as

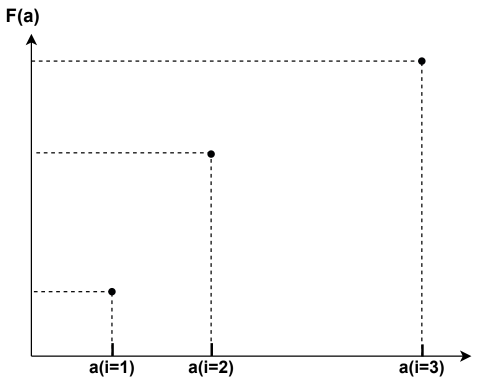
$$F_{t+1}(a', y') = \sum_j \quad \tilde{F}_t(a', \mathcal{Y}_j) \quad \Pi(\mathcal{Y}_j, y'). \tag{4}$$

With these objects at hand, we first consider the case where $a^*(\cdot, \mathcal{Y}_j)$ is strictly monotone everywhere and thus invertible. Now we consider the endogenous gridpoints $a' = a^*(a, \mathcal{Y}_j)$. For these points, the set over which we integrate simplifies to $\{x|a' \geq a^*(x, \mathcal{Y}_j)\} = \{x|a^*(a, \mathcal{Y}_j) \geq a^*(x, \mathcal{Y}_j)\} = \{x|a \geq x\}$, where the last equation results from the invertibility of a^*, given $\mathcal{Y}_j$. This, together with the definition of F_t, again implies that $\tilde{F}_t(a^*(a, \mathcal{Y}_j), \mathcal{Y}_j) = F_t(a, \mathcal{Y}_j)$.

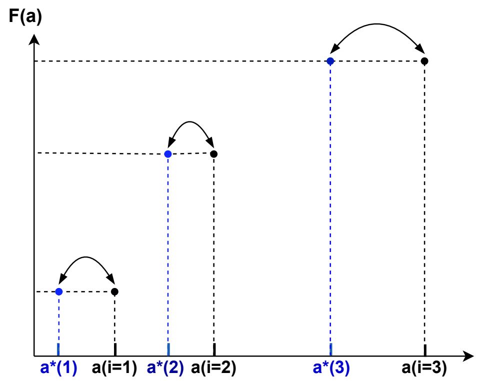
We now specify a grid $\mathcal{A}_i$ for a and compute the associated policies $\mathcal{A}^*_{i,j}$ when $y = \mathcal{Y}_j$. With these objects we know that the set $\left\{ \left(\mathcal{A}^*_{i,j}, F_t(\mathcal{A}_i, \mathcal{Y}_j) \right) \right\}$ is on the graph of $\tilde{F}_t$. This then allows us to construct an interpolant $\hat{\tilde{F}}_t$ for each $\mathcal{Y}_j$, since $\mathcal{A}^*_{i,j}$ is an ordered set, as a^* is strictly monotone. Replacing $\tilde{F}_t$ by $\hat{\tilde{F}}_t$ in (4) then allows to evaluate F_{t+1} at any a'

without integration. Finally, the definition of a cumulative distribution function implies two extrapolation rules. First, $\hat{\tilde{F}}_t$ is set to zero for any $a' < \min\left\{\mathcal{A}^*_{i,j}\right\}$. These are future endogenous states that are lower than the smallest optimal policy and hence are never reached. Second, $\hat{\tilde{F}}_t$ is set to $F_t(\infty, \mathcal{Y}_j) = P(y = \mathcal{Y}_j)$ for all $a' > \max\left\{\mathcal{A}^*_{i,j}\right\}$. 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 and income $\mathcal{Y}_j$ is equal to just observing said income. However, when a^* is a savings function, for example, 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 $\underline{a}_j$ 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.

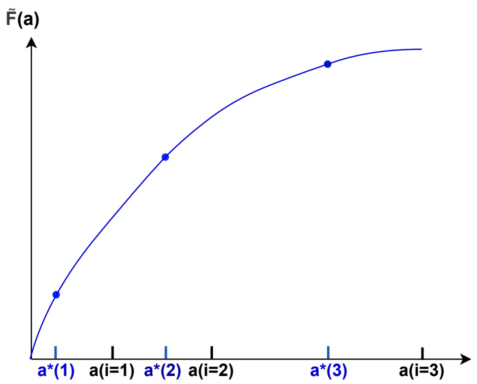
Figure 1: Illustration of Interpolation over Endogenous Grid



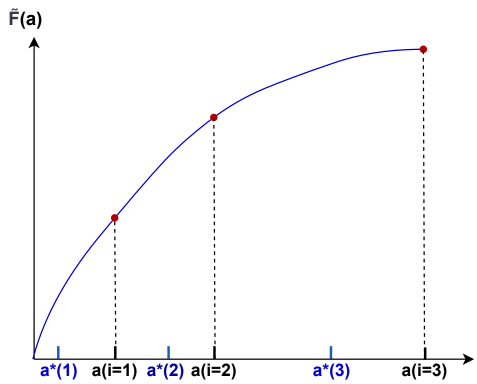
(a) Beginning of period CDF $\mathbf{F}_t(1:3, j)$



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



(c) Construction of interpolant $\hat{\tilde{F}}^j_t$



(d) Evaluation of interpolant $\hat{\tilde{\mathbf{F}}}_t(1:3, j)$

2.3 The Algorithm in a Nutshell

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* $F_t(a, y) := P(x \leq a, z = y)$ *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* $\mathbf{F}_t = [F_t(\mathcal{A}_i, \mathcal{Y}_j)]_i^j$.

1. *For each exogenous state with index* j*,* $y = \mathcal{Y}_j$*,* ***create the interpolant*** $\hat{\tilde{F}}_t^j$.
   (a) *Find* $\underline{a}_j$ *in the set* $\{\mathcal{A}_i\}$*, i.e. find the last endogenous state for which the policy* a^* *is a constant.*
   (b) *Discard all gridpoints below* $\underline{a}_j$ *from the set* $\{\mathcal{A}_i\}$ *and all corresponding choices from the set* $\left\{\mathcal{A}^*_{i,j}\right\}$ *and find the set of according points on the graph of* $\tilde{F}_t^j$*:* $G := \left\{\left(\mathcal{A}^*_{i,j}, \mathbf{F}_t(i,j)\right)\right\}$*, see Figure 1 (b).*
   (c) *Create an interpolant* $\hat{\tilde{F}}_t^j$ *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:*

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

   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 interpolant. Since cumulative distribution functions are monotone, it is advisable to use an interpolation routine that preserves monotonicity. Both linear interpolation and piecewise cubic hermitian splines have this property. However, the linear interpolant does not preserve differentiability everywhere. Note that the linear interpolant is not equivalent to the histogram method, since we interpolate the CDF with optimal policy choices being the interpolation nodes.[2]

2.4 Nonlinear Distributional Dynamics

Bhandari et al. (2023) have highlighted the fact that the histogram method fails to fully capture the nonlinear dynamics of the distribution. Returning to Equation (3), the nonlinear effects of the distribution f_t on its transition dynamics derive entirely from its effect on the optimal policy function a^*.[3] This is because, holding the policy function constant, the transition is a linear operator.

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

$$\frac{\partial \mathbf{A}^*(k,l)}{\partial a_k^*} \frac{\partial^2 a_k^*}{\partial D_t^2} + \frac{\partial^2 \mathbf{A}^*(k,l)}{\partial {a^*}_k^2} \left[\frac{\partial a_k^*}{\partial D_t} \right]^2, \tag{5}$$

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, the histogram method constructs $\mathbf{A}^*$ as (ignoring the exogenous state transitions for simplicity of notation)

$$\mathbf{A}^*(k,l) = \begin{cases} 1 - \frac{a_k^* - \mathcal{A}_l}{\mathcal{A}_{l+1} - \mathcal{A}_l} & \text{if } a_k^* \in [\mathcal{A}_l, \mathcal{A}_{l+1}) \\ \frac{a_k^* - \mathcal{A}_{l-1}}{\mathcal{A}_l - \mathcal{A}_{l-1}} & \text{if } a_k^* \in [\mathcal{A}_{l-1}, \mathcal{A}_l) \\ 0 & \text{else,} \end{cases} \tag{6}$$

which is linear in a^*. Therefore $\frac{\partial^2}{\partial {a^*}_k^2} \mathbf{A}^*(k,l) = 0$. In Appendix A, we extend this analysis to the third order derivative of Equation (3).

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

[2]See Appendix A.1 for details.

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

$$\frac{\partial^2 \hat{\hat{\mathbf{F}}}_t(i,j)}{\partial D_t^2} = \frac{\partial \hat{\hat{\mathbf{F}}}_t(i,j)}{\partial \mathcal{A}_{.j}^*} \frac{\partial^2 \mathcal{A}_{.j}^*}{\partial D_t^2} + \left[\frac{\partial \mathcal{A}_{.j}^*}{\partial D_t}\right]' \frac{\partial^2 \hat{\hat{\mathbf{F}}}_t(i,j)}{\partial {\mathcal{A}_{.j}^*}^2} \left[\frac{\partial \mathcal{A}_{.j}^*}{\partial D_t}\right], \tag{7}$$

where, unlike (5), the second term is nonzero because $\mathcal{A}_{.j}^*$ are the vectors of the interpolation nodes (and the derivatives are vector-valued). Therefore, the Hessian $\frac{\partial^2 \hat{\hat{\mathbf{F}}}_t(i,j)}{\partial {\mathcal{A}_{.j}^*}^2}$ is generally nonzero. As also described in Bhandari et al. (2023), the second term in Equation (7) 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, $\mathcal{A}_{.j}^*$, approximation of $\frac{\partial^2 \hat{\hat{\mathbf{F}}}_t(i,j)}{\partial {\mathcal{A}_{.j}^*}^2}$ requires a shape-preserving interpolation method, as illustrated in Figure 1 using cubic splines.[4]

3 Solving Stationary Distributions

Our first application is the solution of an Aiyagari (1994) economy, where Equation (1) 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 can take two values $\underline{h}, \overline{h}$ and transitions follow the matrix Π^h. Households enjoy utility from consumption, c_t, and solve the dynamic program:

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t) \tag{8}$$

$$\text{s.t. } c_t + k_{t+1} = (1 + r_t - \delta_t)\, k_t + h_t w_t \tag{9}$$

$$k_{t+1} \geq 0. \tag{10}$$

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

$$Y_t = K_t^{\alpha} N^{1-\alpha}, \tag{11}$$

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

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

Table 1: Calibration

Parameters		Value	Parameters		Value
β	Discount factor	0.99	n_k	Gridpoints for k	500
γ	Rel. risk aversion	2	n_h	Gridpoints for h	2
α	Capital share	0.36	Π^h	Transition prob.	$\begin{bmatrix} 0.6 & 0.4 \\ 0.045 & 0.956 \end{bmatrix}$
δ	Depreciation rate	0.02			

We search for 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 endogenous state k_t and the discrete exogenous state h_t. 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 therefore fits Section 2.

We use this workhorse model as a laboratory to present our novel method and compare it to the widely used histogram method. To do so, we follow the calibration idea of Den Haan et al. (2010), see Table 1. Conceptually, *DEGM* involves iterating on the cumulative distribution function to find the stationary distribution.[5] We solve the model for $n_k = 500$ gridpoints in capital k. As a baseline, we use our novel method to find the equilibrium. Going beyond 500 gridpoints had no effect on the equilibrium, so we consider the distribution at 500 gridpoints to be the "true" distribution.[6]

We perform two exercises. First, we isolate the quality of the approximation of the distribution by keeping prices and optimal policies fixed at the benchmark solution. We select a subset of gridpoints from this solution and iterate on the distribution until convergence for both the histogram method and our *DEGM*.[7] Second, we solve for the stationary equilibrium, including prices and policies, which more closely resembles the actual use case. The histogram method 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 50, 100, and 250 gridpoints relative to the baseline solution using *DEGM* with 500 gridpoints for capital.

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

[6] Beyond 500 gridpoints, the difference between the two methods in their solution of the stationary equilibrium becomes negligible.

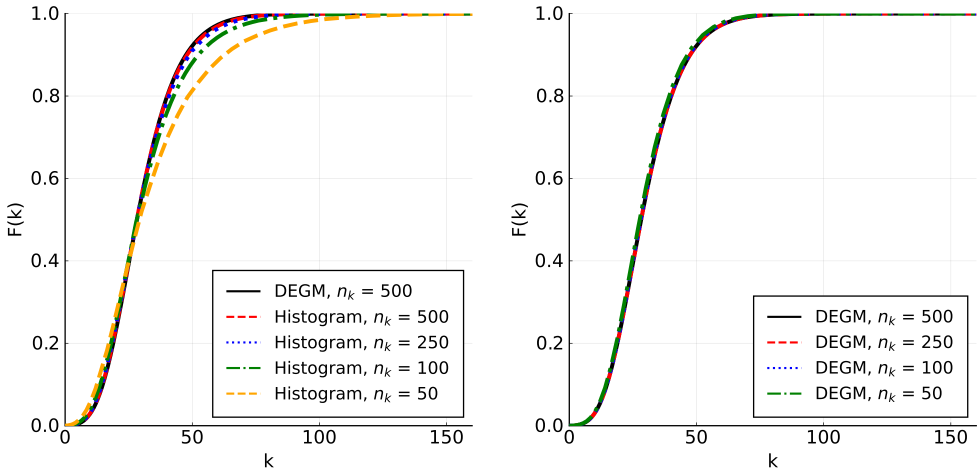
[7] We use piecewise cubic hermitian splines to interpolate the cumulative distribution function.

Table 2: Convergence under the histogram method and DEGM

	Histogram			DEGM		
n_k=	**50**	**100**	**250**	**50**	**100**	**250**
Panel A: Stationary distribution						
Capital stock	14.33	4.10	1.00	-2.33	-0.59	-0.04
Wealth gini	33.74	13.74	4.79	-2.61	-1.29	-0.25
Time	0.00	0.01	0.02	0.06	0.10	0.18
Panel B: Stationary equilibrium						
Capital stock	0.33	0.11	0.03	-0.09	-0.02	-0.00
Wealth gini	27.40	11.71	4.29	-0.86	-0.95	-0.26
Time	0.33	0.71	2.40	1.00	1.75	4.16
Panel C: Accumulated differences of impulse responses						
Capital stock (FO)	5.00	2.40	0.91	1.13	0.23	0.01
Wealth gini (FO)	17.93	11.40	2.76	0.60	0.34	0.21
Capital stock (SO)	11.13	6.98	4.70	10.94	5.44	1.23
Wealth gini (SO)	162.03	206.74	229.95	43.46	16.64	0.91
Time (FO)	0.01	0.03	0.30	0.01	0.03	0.32
Time (SO)	1.82	13.24	201.42	1.85	13.59	232.04

Notes: Values represent percent deviations of the solutions with n_k gridpoints to the "true" solution (DEGM and $n_k = 500$). **Panel A** shows the results with constant prices and policies from the "true" solution (only resolving the stationary distribution). **Panel B** shows the results for solving the stationary equilibrium including prices and policies. **Panel C** shows accumulated differences of impulse responses following a 7.5 p.p. shock to δ (over 300 periods) with first-order (FO) and second-order (SO) perturbation solution. Time refers to the computation time (in seconds) that it takes to solve for the solution on a laptop with 16-core, 3.3GHz CPU. Specifically, to solve for (A) the stationary distribution, (B) the stationary equilibrium and (C) the perturbation solution.

Figure 2: Comparison of stationary cumulative distribution in capital for different grid sizes with constant prices and policies from the "true" solution (DEGM and $n_k = 500$)



Panel A does this for the first exercise with constant prices and policies. Figure 2 graphically compares the implied stationary distribution in capital for different grid sizes in this case. Panel B compares the two methods for the second exercise of solving the stationary equilibrium. We find that our method converges to the "true" distribution much faster, especially for cross-sectional moments.[8] For a given number of gridpoints, the histogram method is faster in terms of computational time, mainly because it does not require iterations when updating the distribution. However, for a given accuracy, our method is faster when solving for the stationary equilibrium.

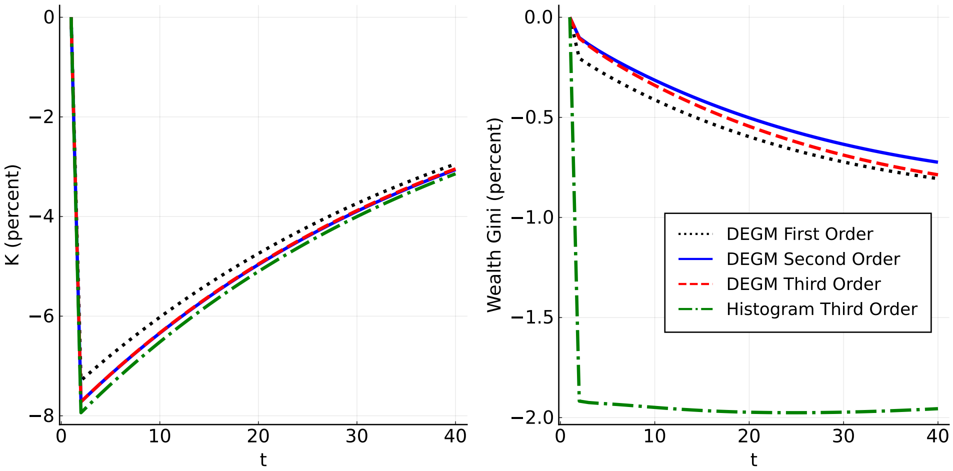
4 Solving Higher Order Distributional Dynamics

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 a 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 so, we extend state-space perturbation techniques for heterogeneous agent models from Bayer and Luetticke (2020) to higher orders, following Andreasen et al. (2018) and Levintal (2017). Solving the model up to third order allows us to implement

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

Figure 3: Impulse responses to capital depreciation shock



Notes: Impulse responses of capital (left panel) and wealth Gini (right panel) to shock $\nu_1 = 7.5$ *p.p.* from first-order (dotted black), second-order (solid blue), and third-order (dashed red) solutions using DEGM, and from third-order solution using the histogram method (dash-dotted green), evaluated at non-stochastic steady state ($n_k = 100$). Y-axis: Percent deviation from non-stochastic steady state. X-axis: Quarter.

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 investment risk.[9] In particular, capital depreciation δ_t deviates from its steady-state value by following the process:

$$\delta_t = \delta + \nu_t,\ \nu_t \sim F^{\nu}(0, \sigma_\delta, \tau_\delta), \tag{12}$$

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 and causes a 10% drop in annual GDP, in line with evidence collected by Barro (2006).

We propose this Krusell and Smith (1998)-style model with investment risk as a new 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 3 compares the impulse responses to a one-time 7.5% destruction of the capital stock, using *DEGM* to compute the dynamics with first-, second-, and third-order perturbation solutions.[10] The first-

[9]When simulating the model, we draw shock innovations from a normal-inverse Gaussian distribution that matches these moments.

[10]Appendix C summarizes the state-space system to which the perturbation is applied.

Table 3: Ergodic moments under second and third order solutions

	Steady state		**Second Order**		**Third Order**	
Variable	Hstgrm	DEGM	Hstgrm	DEGM	Hstgrm	DEGM
Output	2.95	2.95	-0.2 (0.7)	-0.2 (0.7)	-0.3 (0.7)	-0.5 (0.7)
Capital stock	30.06	30.02	-0.5 (2.2)	-0.6 (2.1)	-1.0 (2.2)	-1.6 (2.1)
Wealth Gini	26.95	23.90	-1.4 (1.3)	2.1 (0.7)	1.8 (1.4)	12.4 (0.6)

Notes: Non-stochastic steady state levels across methods ($n_k = 100$, columns 1-2). Means and standard deviations (in brackets) across perturbation order and methods, in percent deviation from non-stochastic steady state (columns 3-6). Moments are averages of simulated data generated from pruned model dynamics (Andreasen et al., 2018) for $T = 10.000$ periods. Depreciation shocks are drawn from normal-inverse Gaussian distribution $F^{\nu}(0, \sigma_\delta, \tau_\delta)$.

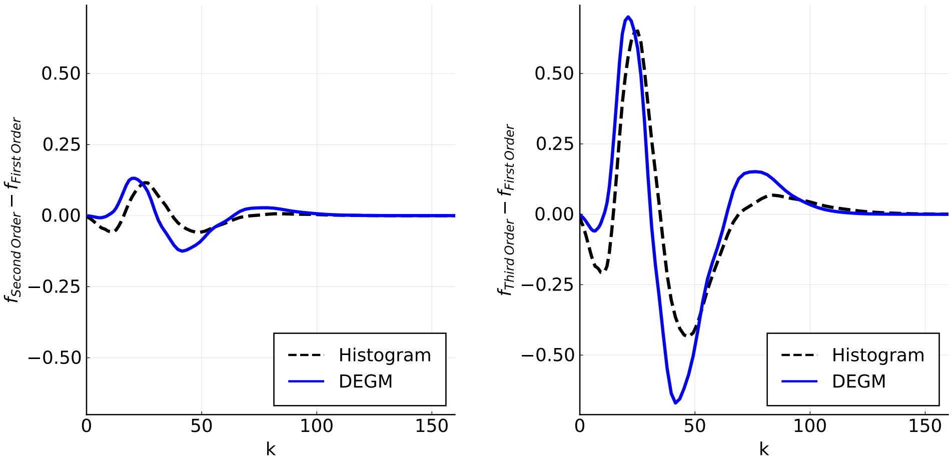
order solution slightly understates the decline in aggregate capital and overstates the decline in the Gini coefficient of wealth in response to the capital depreciation shock. Taken together, both reflect that the distributional dynamics in this model are nonlinear with respect to aggregate shocks, but that the feedback from inequality to equilibrium prices is modest.

As with the steady state, the histogram method and *DEGM* converge to the same first-order perturbation solution. Again, *DEGM* converges to the true solution faster in terms of number of gridpoints. Computation time is now close to identical across both methods for the same number of gridpoints.[11] Table 2 Panel C shows the absolute difference in the impulse response of the capital stock and the Gini coefficient to a 7.5 percentage point depreciation shock relative to the 500-gridpoint *DEGM* solution. Again, the performance of the *DEGM* with fewer gridpoints is better, especially for the cross-sectional moments. Table 2 Panel C also shows the lack of convergence for second order solutions. Both the results for first and second order solutions are consistent with Bhandari et al. (2023). As explained in Section 2, the histogram method misses the nonlinear response of the distribution to policy changes. Figure 3 also plots the third-order dynamics of capital and the Gini coefficient for the histogram method. It overestimates the decline in the capital stock and the Gini coefficient in response to a capital depreciation shock.

Higher-order solutions not only provide a better approximation of the dynamics, but also, importantly, capture the response of households to aggregate risk. Table 3 documents how the ergodic distribution with aggregate risk differs from the stationary distribution. Appendix A shows that the third-order solution captures the effect of the

[11] Both methods are updated via iteration when doing the perturbation solution.

Figure 4: Relative shift of the density of capital holdings with aggregate risk



Second Order Shift | Third Order Shift

Notes: Differences in density of capital holdings in the second-order (left panel) and third-order (right panel) ergodic distribution from first-order distribution for histogram method (dashed black) and DEGM (solid blue), solved with $n_k = 100$. Density computed as derivative of cubic spline-interpolated cumulative distribution function that is estimated from simulated data. Y-axis: Difference of density in percentage points. X-axis: Capital holdings.

skewed distribution of aggregate shocks, while the second-order solution only captures the effect of the variance. Aggregate risk here refers to investment risk. As in Angeletos (2007), this reduces the aggregate capital stock because for most households the substitution effect is stronger than the income effect. Using *DEGM*, we find a 1.6% reduction in capital. In comparison, the histogram method predicts only a 1% reduction in aggregate capital. Thus, the histogram method understates the impact of investment risk on economic activity.

The effect of under-accumulation of capital is driven by the lower savings of less wealthy households, so that wealth inequality increases. For these households the substitution effect dominates as they have little capital income. What is more, a lower capital stock implies a lower wage rate and a higher rate of return on capital. This also benefits wealthy households that rely mainly on dividend income. Thus, while a capital depreciation shock itself compresses the distribution of wealth, the risk of such a shock increases wealth inequality on average. Our method finds a 12% increase in the wealth Gini coefficient with aggregate risk, while the histogram method finds only a 2% increase, see Table 3. The difference between the methods is thus much more pronounced for cross-sectional moments than for aggregates.

Figure 4 zooms in and shows the changes that aggregate investment risk induces in the distribution of wealth. As expected, the changes are stronger for the third order solution, but the direction of change is the same between the second and third order solutions. The former captures the variance and the latter the skewness of investment risk. Economically important, we find that the introduction of aggregate investment risk hollows out the middle class.

5 Conclusion

We propose a novel endogenous gridpoint method for distributional dynamics (*DEGM*). Our method retains the tractability and speed of the histogram method 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. Specifically, we propose a Krusell and Smith (1998) model with aggregate investment risk. In this model, we show that aggregate investment risk has a strong impact on 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., and 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., and 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.

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

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

Bhandari, A., Bourany, T., Evans, D., and 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.

Den Haan, W. J., Judd, K. L., and 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.

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

Krusell, P. and 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.

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

A Perturbation of Distributional Dynamics

A.1 Comparison of Discretized Methods

We compare discretized methods[12] 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 histogram method is

$$F_{t+1}(m) = \sum_{j=1}^{m} \sum_{i} \mathbf{A}_t^*(i,j) \left(F_t(i) - F_t(i-1)\right), \tag{13}$$

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

$$F_{t+1}(j) = \hat{\hat{F}}(\mathcal{A}_j \mid \mathcal{A}_t^*,\, \mathbf{F}_t), \tag{14}$$

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

$$F_{t+1}(j) = \sum_{i:\, a_t^*(i-1) < \mathcal{A}_j \leq a_t^*(i)} \mathbf{A}_t^{*,L}(i,j) \left(F_t(i) - F_t(i-1)\right) + F_t(i-1), \tag{15}$$

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

Instead, we use a cubic spline for the interpolation as it captures nonlinearities of the continuous limit (see Appendix B). This adds smoothness conditions at the interpolation nodes. In practice, we use a piecewise cubic hermite interpolating polynomial algorithm that preserves monotonicity (Fritsch and Butland, 1984), with one-sided approximation of the slopes at the endpoints. A change at one interpolation node-value pair $(a_t^*(i), F_t(i))$ affects the entire interpolated function, so that *DEGM* has the general form of Eq. (14).

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

A.2 Generalization and Higher Order Terms

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

$$F_{t+1} = \mathbf{A}(a_t^*)F_t, \tag{16}$$

where we treat the distribution F and the optimal policies a^* as scalars for ease of exposition. This structure captures the histogram method and piecewise linear interpolation exactly, and *DEGM* with our implementation of cubic spline interpolation approximately.[13] We analyze the Taylor expansion of this dynamic with respect to the deviation $\hat{F}_t = F_t - \bar{F}$ and some aggregate disturbance D_t that affects optimal policies, around a steady state, characterized by $\bar{F}$ and $\bar{a}^*$, where these disturbances are zero. Terms in red are zero for methods that are linear in optimal policies.

First order approximation:

$$F_{t+1} \approx \bar{F} + \mathbf{A}(\bar{a}^*)\hat{F}_t + \frac{\partial \mathbf{A}(\bar{a}^*)}{\partial a^*}\frac{\partial a_t^*}{\partial D_t}\bar{F}D_t \tag{17}$$

Second order additional terms:

$$\underbrace{\frac{\partial \mathbf{A}(\bar{a}^*)}{\partial a^*}\frac{\partial a_t^*}{\partial D_t}}_{\textbf{(I)}}\hat{F}_t D_t + \underbrace{\frac{1}{2}\left(\frac{\partial \mathbf{A}(\bar{a}^*)}{\partial a^*}\frac{\partial^2 a_t^*}{\partial D_t^2} + \frac{\partial^2 \mathbf{A}(\bar{a}^*)}{\partial a^{*2}}\left[\frac{\partial a_t^*}{\partial D_t}\right]^2\right)}_{\textbf{(II)}}\bar{F}D_t^2 \tag{18}$$

Third order additional terms:

$$\underbrace{\frac{1}{2}\left(\frac{\partial \mathbf{A}(\bar{a}^*)}{\partial a^*}\frac{\partial^2 a_t^*}{\partial D_t^2} + \frac{\partial^2 \mathbf{A}(\bar{a}^*)}{\partial a^{*2}}\left[\frac{\partial a_t^*}{\partial D_t}\right]^2\right)}_{\textbf{(III)}}\hat{F}_t D_t^2 + \tag{19}$$

$$+\underbrace{\frac{1}{6}\left(\frac{\partial \mathbf{A}(\bar{a}^*)}{\partial a^*}\frac{\partial^3 a_t^*}{\partial D_t^3} + 3\frac{\partial^2 \mathbf{A}(\bar{a}^*)}{\partial a^{*2}}\frac{\partial a_t^*}{\partial D_t}\frac{\partial^2 a_t^*}{\partial D_t^2} + \frac{\partial^3 \mathbf{A}(\bar{a}^*)}{\partial a^{*3}}\left[\frac{\partial a_t^*}{\partial D_t}\right]^3\right)}_{\textbf{(IV)}}\bar{F}D_t^3 \tag{20}$$

[13] Locally, $\hat{\hat{F}}(\mathcal{A}_j \mid \mathcal{A}_t^*, \mathbf{F}_t) \approx \hat{\hat{F}}_F(\mathcal{A}_j \mid \mathcal{A}_t^*, \bar{\mathbf{F}})\hat{\mathbf{F}}_t$. 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.

A.3 Application: Asymmetric risk in our Krusell and Smith model

The aggregate risk in the economy arises from a capital depreciation shock that affects optimal saving choices through income and substitution effects. This induces four potential nonlinearities in the distributional dynamics. Term **(I)** captures a state-dependency: the distributional dynamics are different when the optimal policy is far from the steady state. Term **(II)** captures a *risk-adjustment*: nonlinearities and variance in optimal policies affect how the steady state distribution passes through. Term **(III)** captures a higher-order state dependence: nonlinearities and variance in optimal policies alter the dynamics of the distribution. Term **(IV)** is an *asymmetric risk adjustment*: higher-order nonlinearities and skewness in optimal policies affect how the steady-state distribution passes through.

Methods that are linear in optimal policies miss the effect of expected variation in households' saving decisions caused by asymmetric investment risk on the distributional dynamics and the ergodic distribution. Nevertheless, we find that the third-order risk correction to the distribution is large even for the histogram method, which implies that the dependence of optimal policies on third-order risk, $\frac{\partial^3 a_t^*}{\partial D_t^3}$, is significant. Note that D also captures endogenous aggregate states that differ from the steady state at higher orders, such as the capital stock. This creates feedback effects.

B Nonlinearity of the Kolmogorov Forward Equation in a^*

In Appendix A.2, we show that (discretized) distributional dynamics in optimal policies should be nonlinear 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 the "histogram" method misses a potentially important nonlinearity, again use the monotonicity of the policy function to write the end-of-period distribution as:

$$\tilde{F}_t(a', y) = \int_{x|a^*(x,y)\leq a'} f_t(x, y)dx = \int^{a^{*-1}(a',y)} f_t(x, y)dx. \tag{21}$$

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

$$\frac{\partial \tilde{F}_t(a', y)}{\partial D} = f_t(a^{*-1}(a', y), y)\frac{\partial a^{*-1}(a', y)}{\partial D}, \tag{22}$$

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

$$\frac{\partial^2 \tilde{F}_t(a', y)}{\partial D^2} = f_t(a^{*-1}(a', y), y)\frac{\partial^2 a^{*-1}(a', y)}{\partial D^2} + \frac{\partial f_t(a^{*-1}(a', y), y)}{\partial a}\left(\frac{\partial a^{*-1}(a', y)}{\partial D}\right)^2. \tag{23}$$

This shows that the nonlinear effects of a change in D are composed of a nonlinear effect on the policies (here their inverse) and the derivative of the density times the squared linear effect of D on the (inverse) policy.

Around the mode of the distribution (conditional on y) these effects will be small, because $\frac{\partial f_t(a^{*-1}(a',y),y)}{\partial a}$ is small. What is more, 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.

C Higher Order Perturbation Solution of Heterogenous Agent Models

Ξ combines all equilibrium conditions from households, Ξ^i, and firms as well as market clearing, Ξ^A:

$$\Xi\left(F_t, S_t, \nu_t, P_t, F_{t+1}, S_{t+1}, \nu_{t+1}, P_{t+1}, \varepsilon_{t+1}\right) = \begin{bmatrix} \Xi^i\left(\cdot\right) \\ \Xi^A\left(\cdot\right) \end{bmatrix} \tag{24}$$

$$\Xi^i\left(\cdot\right) := \begin{bmatrix} F_{t+1} - L(F_t, h_t^k) \\ \nu_t - \left(u_{h_t^k} + \beta \mathbb{E}_{y'} \nu_{t+1}\right) \end{bmatrix} \tag{25}$$

$$\Xi^A\left(\cdot\right) := \begin{bmatrix} S_{t+1} - H(S_t) + \eta\varepsilon_{t+1} \\ \Phi_t(h_t^k, F_t) \\ \varepsilon_{t+1} \end{bmatrix} \tag{26}$$

s.t.

$$h_t^k(k, y) = \arg \max_{k' \in \Gamma(y,k;P_t)} u(y, k, k') + \beta \mathbb{E}_{y'|y} \nu_{t+1}(y', k'), \tag{27}$$

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. We solve the system for the state dynamic $h : (F_t, S_t) \mapsto (F_{t+1}, S_{t+1})$ and the state-to-control mapping $g : (F_t, S_t) \mapsto (\nu_t, P_t)$. These functions are implicitly defined by

$$\mathbb{E}_t\left[\Xi\left((F_t, S_t), g(F_t, S_t), h(F_t, S_t), g(h(F_t, S_t)), \sigma\varepsilon_{t+1}\right)\right] = 0, \tag{28}$$

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. In practice, we follow Bayer and Luetticke (2020) and write the distribution function as a Copula function and its marginals. However, *DEGM* does not require this and works directly on the distribution function as well.

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 on how to sparsely set up components like the permutation matrix and on 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).