# DISCRETIZING THE INFINITE-DIMENSIONAL SPACE OF DISTRIBUTIONS TO APPROXIMATE MARKOV EQUILIBRIA WITH EX-POST HETEROGENEITY AND AGGREGATE RISK

Elisabeth Pröhl

*University of Geneva and Swiss Finance Institute*

June 26, 2017

Dynamic stochastic general equilibrium models with ex-post heterogeneity due to idiosyncratic risk have to be solved numerically. This is a nontrivial task as the cross-sectional distribution of endogenous variables becomes an element of the state space due to aggregate risk. Existing global solution methods assume bounded rationality in terms of a parametric law of motion of aggregate variables to reduce dimensionality. In this paper, we do not take that assumption and compute a fully rational equilibrium depending on the whole cross-sectional distribution. Dimensionality is tackled by polynomial chaos expansions, a projection technique for square-integrable random variables, resulting in a nonparametric law of motion. In contrast to existing methods, we establish theoretical convergence results and approximation error bounds. Economically, we find that idiosyncratic risk does not aggregate in our fully rational approximate equilibrium, which contrasts the well-known aggregation result for the bounded rational approximate equilibrium by Krusell and Smith (1998).

Keywords: Dynamic stochastic general equilibrium, Incomplete markets, Heterogeneous agents, Aggregate uncertainty, Convergence, Numerical solutions, Polynomial chaos.

JEL Classification: C63, D31, D52, E21.

## 1. INTRODUCTION

Economies consist of heterogeneous agents who are exposed to idiosyncratic risks, the most prominent example of which is labor income risk for households.

University of Geneva, Bd du Pont d'Arve 40, 1211 Geneva 4, Switzerland, elisabeth.proehl@unige.ch

I thank Johannes Brumm, René Carmona, Chris Carroll, David Childers, Jérôme Detemple, Rajna Gibson Brandon, Daniel Harenberg, Kenneth Judd, Greg Kaplan, Felix Kübler, Semyon Malamud, Benjamin Moll, Juan Sagredo, Yuliy Sannikov, Olivier Scaillet, Karl Schmedders, Harald Uhlig, Thomas Winberry, and participants at WEHIA 2017, the Workshop on Recent Developments in Numerical Methods with Applications in Statistics and Finance 2017, SNDE 2017, the RGS Doctoral Conference 2017, ESEM 2016, EcoMod 2016, CEF 2016, ESNASM 2016, the MFM Summer Camp 2016, the 12$^{\text{th}}$ Cowles Conference on General Equilibrium and its Applications, ISCEF 2016, the MFM Winter Meeting 2016, CEF 2015, the SFI Research Days 2015, PASC 2015, and seminar participants in Chicago, Geneva, Lausanne, Princeton and Zurich for fruitful comments on this work. Earlier versions of this work were disseminated under the name "Computing the Cross-Sectional Distribution to Approximate Stationary Markov Equilibria with Heterogeneous Agents and Incomplete Markets".

This was first modeled in a dynamic stochastic general equilibrium (DSGE) model by Bewley (1977) where agents face idiosyncratic income shocks affecting their wealth, and extended by Aiyagari (1994) to include a production technology. They show that individual precautionary savings contribute to aggregate savings because idiosyncratic risk cannot be fully insured. Hence, idiosyncratic risks generally affect aggregate variables in the economy. Other examples of idiosyncratic risks are firm-specific productivity shocks in models of firm exit and entry as in Hopenhayn (1992), or county-specific productivity shocks in real business cycle models as in den Haan et al. (2011).

These models, however, do not feature aggregate risk as this makes the equilibrium problem difficult to solve. The challenge in constructing a solution algorithm lies in handling the cross-sectional distribution of the agents' idiosyncratic variables, which becomes an infinite-dimensional element of the state space. This distribution changes over time depending on how the aggregate shocks realize. The aggregate variables evolve, in turn, depending on how the cross-sectional distribution changes. In their seminal paper, Krusell and Smith (1998) were the first to propose a global solution algorithm for the Aiyagari growth model with aggregate risk. They handle the dimensionality problem in assuming bounded rationality, which means that agents are not required to observe the whole cross-sectional distribution to predict the movement of aggregate variables. They rather use a parametric law of motion of the the aggregate variables depending on a finite number of moments. Given that assumption, they then solve the model by iterating on the following two steps: Firstly, solve for the optimal policies given a guess of parameters of the aggregate variables' law of motion, and secondly, estimate new parameters for the law of motion given a set of simulated data from the new optimal policy. The main economic result from this seminal work is that, given the bounded rationality assumption, adding higher moments than the mean to the parametric law of motion does not change the equilibrium solution. Hence, the idiosyncratic risk does not matter for aggregation. Various more recent papers improve the original algorithm mainly by eliminating the agent dimension in the simulation step, and by varying the parametric form of the law of motion. However, these works still rely on the bounded rationality assumption and a two-step iterative procedure with a simulation.

The existing methodology of global solution methods for heterogeneous agent models with aggregate risk has several drawbacks. Firstly, it is not clear whether the assumed parametric law of motion of the aggregates in the bounded rational expectations equilibrium is indeed close to its equivalent in the fully rational

expectations equilibrium. Generally, it is unknown whether the bounded rational solution is at all close to the fully rational solution since there is no theory on measuring their distance. Secondly, it is not clear a priori how many moments are necessary for the bounded rational equilibrium to exist. In fact, Kubler and Schmedders (2002) show that there are models, for which recursive equilibria depending only on aggregate wealth, i.e., the first moment of the cross-sectional distribution, do not exist. Thirdly, it is unclear whether the existing algorithms converge to the bounded rational equilibrium for every model setup as theoretical convergence results are lacking.

The contribution of this paper is to construct a global solution algorithm for DSGE models with heterogeneous agents and aggregate risk, which does not assume bounded rationality, for which convergence is proven, and error bounds, i.e., the distance to the fully rational expectations equilibrium, are established. Rather than assuming a parametric law of motion for the aggregate variables, we discretize the space of cross-sectional distributions with a projection technique called generalized polynomial chaos. This technique essentially extends the theory of projecting smooth functions on orthogonal polynomials to projecting square-integrable random variables on orthogonal polynomials. This distinction is important because the endogenous cross-sectional c.d.f. for models with hard borrowing constraints typically features jumps whose locations are unknown a priori. Furthermore, with the right choice of polynomials, the generalized polynomial chaos expansion converges quite fast. We find that a projection on polynomials up to first order is enough to obtain satisfactory precision of the solution. Overall, an advantage of discretizing the cross-sectional distribution is that the aggregate variables emerge automatically in a nonparametric fashion. Therefore, we do not require a separate step in the solution algorithm to estimate their law of motion. No simulation is necessary at all.

To obtain the theoretical convergence result, we leverage the convexity properties of the individual optimization problem. There is well established mathematical theory that an iteration on the inverse of a modification of the Lagrangian converges for convex optimization problems on Hilbert spaces. This results in the so called proximal point algorithm. We adapt this algorithm to accommodate the cross-sectional distributions. Hence, our algorithm converges to the fully rational equilibrium when updating the theoretical, i.e., non-discretized, optimal policies. As we have to discretize the policies for the implementation, we also derive approximation error bounds.

When comparing the results of our algorithm to existing methods for the bench-

mark Aiyagari growth model with aggregate risk, we find a significant improvement of precision for individual policies and aggregate variables in terms of Euler equation errors. Furthermore, there is a significant improvement in precision when truncating the polynomial chaos expansion in our algorithm at order one rather than at order zero. Note that the latter leads to optimal policies, which solely depend on the mean of the distribution, whereas the former leads to policies, which depend on an approximation of the whole cross-sectional distribution. This implies that idiosyncratic risk matters for aggregation in this fully rational equilibrium, and therefore, contrasts the aggregation result by Krusell and Smith (1998) for the bounded rational equilibrium. Truncating at higher orders does not lead to further significant improvement for the growth model at hand.

This paper is related to several strands of literature, first and foremost of course, the literature on numerical algorithms. In general, there are two types of algorithms: Local solution methods are based on perturbation techniques whereas global solution methods are based on projection techniques or a mixture of projection and simulation techniques. Our algorithm and the already mentioned seminal algorithm by Krusell and Smith (1998) belong to the latter group. The algorithm by Krusell and Smith (1998) has also been the subject of a special issue of the Journal of Economic Dynamics and Control in January 2010. This special issue presents various alternative algorithms, and compares them in den Haan (2010). They have in common that they assume bounded rationality, and hence, use a small finite number of moments instead of the full cross-sectional distribution to approximate the policy function and the law of motion of aggregate variables. One problem, which is addressed by Algan et al. (2008); Young (2010); Ríos-Rull (1997) and summarized in Algan et al. (2010), is the cross-sectional variation due to the simulation of a finite number of agents in Krusell and Smith (1998) when estimating the law of motion parameters. They use parametric and nonparametric procedures to get around this issue. However, the variation due to simulating over aggregate exogenous shocks remains. In contrast to the simulation approach, den Haan and Rendahl (2010) use direct aggregation to obtain the law of motion. Interestingly, Algan et al. (2008) and Reiter (2010a) parameterizes the cross-sectional distribution itself to obtain a better prediction of the law of motion but their parametric functional forms are somewhat ad hoc and not closed under the optimal savings policy. They do not span the space of square-integrable random variables. I use the algorithm by Reiter (2010a) in my numerical comparison and find that it performs significantly worse than the algorithm proposed herein.

Local solution methods based on perturbations do not assume bounded rational-

ity. To reduce dimensionality, they first solve for the optimal policy and stationary distribution of the model without aggregate shocks using projection methods, and then, perturb this solution to accommodate aggregate shocks. The most prominent perturbation algorithm goes back to Reiter (2009, 2010b). Childers (2015) investigates the theoretical underpinning of these perturbations. Mertens and Judd (2013) use perturbations for the law of motion. Winberry (2016) combines the low of motion approach in Algan et al. (2008) with the perturbation in Reiter (2009). He also presents a model where the aggregation result by Krusell and Smith (1998) does not hold. There are two major drawbacks for perturbation methods: Firstly, the perturbation in aggregate shocks often is only linear or at most quadratic. Therefore, any higher-order nonlinear effects of aggregate shocks are not accounted for. Secondly, as for all perturbation methods, the solutions are only accurate for small aggregate shocks. Crises scenarios in terms of a large aggregate shock or a long series of aggregate shocks in one direction cannot be analyzed with confidence.

Apart from the numerical literature, this paper builds on the existence literature. Generic existence of solutions to DSGE models has been shown by Duffie and Shafer (1985, 1986) and Duffie et al. (1994). However, these results only apply to models with ex-ante heterogeneity, i.e., where agents differ on finitely many model ingredients. Existence of a solution to the Aiyagari-Bewley growth model with aggregate risk, which features ex-post heterogeneity due to idiosyncratic risk, has long been an open research question. It has been first examined by Miao (2006). However, a flaw in the theoretical argument has been discovered in Cheridito and Sagredo (2016b) and corrected in Cheridito and Sagredo (2016a). They prove the existence of a fixed point in the law of motion in order to prove the existence of the equilibrium. By construction, our law of motion of aggregate variables coincides with the self-confirming aggregate capital predictions in Cheridito and Sagredo (2016a).

It is also worth pointing out the relation to the literature on mean field games and their numerical solutions because they are essentially continuous-time versions of DSGE models with ex-post heterogeneity. Achdou et al. (2014) show how to use partial differential equations to solve heterogeneous agent models. Kaplan et al. (2016) put forward a very interesting application of this methodology to monetary policy questions. However, their models incorporate only idiosyncratic shocks but no aggregate risk. Applying generalized polynomial chaos as in the algorithm presented herein to extend their framework to aggregate risk could yield interesting results.

The paper proceeds as follows. In the next section, we present the Aiyagari-Bewley growth model with aggregate risk, which serves as the benchmark model

for our algorithm throughout the paper. In Section 3, we introduce the methodology behind the algorithm. This section explains the concept of proximal point algorithms, which underlies our convergence result, and the polynomial chaos expansion, which is used to project the cross-sectional distribution. Section 4 then proves both convergence for the iteration of the theoretical solution in the algorithm, and approximation error bounds for the discretized solution. In Section 5, our numerical results are compared to three existing global solution methods and economic implications are analyzed. The last section concludes. The appendix contains all proofs.

## 2. THE MODEL

For illustration, we use the same growth model with aggregate shocks as in den Haan et al. (2010), which is used for a comparison of Krusell-Smith-style algorithms in the special issue of the Journal of Economic Dynamics and Control in January 2010. We consider a discrete-time infinite-horizon model with a continuum of agents of measure one. There are two kinds of exogenous shocks, an aggregate shock and an idiosyncratic shock. The aggregate shock characterizes the state of the economy with outcomes in $\mathcal{Z}^{ag} = \{0, 1\}$ standing for a bad and good state, respectively. The idiosyncratic shock with outcomes in $\mathcal{Z}^{id} = \{0, 1\}$ indicates that an agent is unemployed or employed, respectively. It is i.i.d. across agents conditional on the aggregate shock. We denote the compound exogenous process $\left(z_t^{ag}, z_t^{id}\right)_{t\geq 0}$ by $(z_t)_{t\geq 0} \in \mathcal{Z}$ with $\mathcal{Z} = \mathcal{Z}^{ag} \times \mathcal{Z}^{id}$. The transition probabilities are exogenously given by a four-by-four matrix.

The security market consists of a claim to aggregate capital $(K_t)_{t\geq 0}$. An agent's share of physical capital is denoted by $(k_t)_{t\geq 0}$. The aggregate endogenous variable $K$ is hence defined by

$$K_t = \sum_{z^{id}=0}^{1} \int_{-\infty}^{\infty} k d\mu_t \left(z^{id}, k\right) \; \forall \, t \geq 0, \tag{1}$$

where $\mu_t$ is the cross-sectional distribution of idiosyncratic endogenous variables at time $t$. It is simply the probability distribution of individual capital across the unemployed and the employed agents given the trajectory of aggregate shocks

$$\mu_t \left(z^{id}, k\right) = \mathbb{P}\left(\left\{z_t^{id} = z^{id}\right\} \cap \left\{k_t \leq k\right\} \middle| \, z_t^{ag}, \ldots, z_0^{ag}\right)$$

for all $t \geq 0$, $z^{id} \in \mathcal{Z}^{id}$ and $k \in \mathbb{R}$. The aggregate shocks cause the cross-sectional distribution to vary over time, which is indicated by the time subscript of $\mu_t$.

Each agent chooses her share of physical capital and consumption such that they satisfy certain constraints. Firstly, individual consumption must be positive at all times $c_t > 0$, $t \geq 0$, and capital holdings are subject to a hard borrowing constraint $k_t \geq 0$, $t \geq 0$. Secondly, given an initial capital endowment $k_{-1} \geq 0$ and an initial cross-sectional distribution $\mu_{-1}$[1] with non-negative support, each agent adheres to a budget constraint, which equates individual consumption and current capital stock to productive income and saved capital stock

$$k_t + c_t = I\left(z_t, k_{t-1}, K_{t-1}\right) + \left[1-\rho\right] k_{t-1} \,\forall\, t \geq 0. \tag{2}$$

The time line underlying this equation is clarified in Figure 1.[2] The parameters in

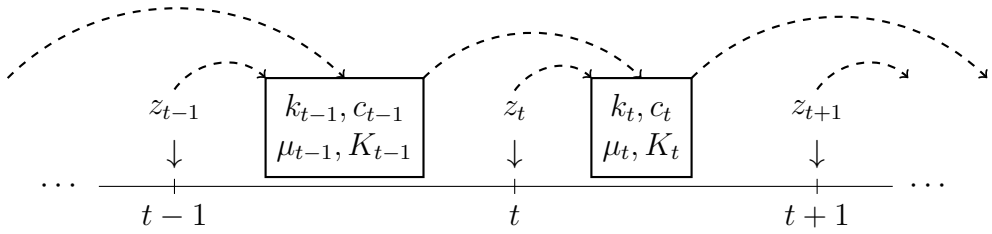


FIGURE 1. **Time line of events.** Before period $t$, the agent observes how much capital she saved in the previous period $k_{t-1}$ and what the cross-sectional distribution of individual capital savings $\mu_{t-1}$ and hence, aggregate capital $K_{t-1}$ is. At period $t$, the agent first observes the exogenous shocks $z_t$ and then decides how much to consume $c_t$ and how much capital $k_t$ to save in that period.

this budget constraint are defined as follows. The capital stock brought forward from period $t-1$ depreciates by a rate $\rho \in (0,1)$. The productive income is given by

$$\begin{aligned} I\left(z_t, k_{t-1}, K_{t-1}\right) &= R\left(z_t^{ag}, K_{t-1}\right) k_{t-1} \\ &\quad + z_t^{id} \pi \left[1-\tau_t\right] W\left(z_t^{ag}, K_{t-1}\right) + \left[1 - z_t^{id}\right] \nu W\left(z_t^{ag}, K_{t-1}\right). \end{aligned} \tag{3}$$

It is composed of, firstly, the return on capital stock, and secondly, labor income, which equals the individual's wage $W$ when the agent is employed and a propor-

[1] The initial cross-sectional distribution $\mu_{-1}$ does not only imply the initial aggregate capital $K_{-1}$, but also the initial aggregate economic state due to $p^e_{-1} = (1/K_{-1}) \int_0^\infty k d\mu_{-1}(1,k)$, which is an exogenously given quantity.

[2] Note that I specify the time line slightly differently than den Haan et al. (2010) and Krusell and Smith (1998). These authors substitute $k_t$ with $k_{t+1}$ in the budget constraint (2) because this is the capital, which is put forward as start capital to period $t+1$. In contrast to that notation, however, I want to emphasize the time period, at which the agent optimally chooses the magnitude of her capital savings. Taking this view, the optimal consumption and capital savings choice have the same time subscript. My time line therefore indicates, which filtration the endogenous variables are adapted to.

tional unemployment benefit $\nu W$ otherwise. The agent's wage is subject to a tax rate $\tau_t = \nu(1-p_t^e)/(\pi p_t^e)$ whose sole purpose it is to redistribute money from the employed to the unemployed. The parameter $\nu \in (0,1)$ denotes the unemployment benefit rate whereas $p_t^e = \mathbb{P}(z_t^{id} = 1 | z_t^{ag})$ is the employment rate at time $t$ and $\pi > 0$ is a time endowment factor. It is reasonable to assume $\nu/\pi < 1 - \tau_t \Leftrightarrow \nu < \pi p_t^e$ for all $t \geq 0$. The wage $W$ and the rental rate $R$ are derived from a Cobb-Douglas production function for the consumption good

$$\begin{aligned} W\left(z_t^{ag}, K_{t-1}\right) &= (1-\alpha)\left(1 + z_t^{ag} a - (1 - z_t^{ag})a\right)\left[\frac{K_{t-1}}{\pi p_t^e}\right]^{\alpha} \\ R\left(z_t^{ag}, K_{t-1}\right) &= \alpha\left(1 + z_t^{ag} a - (1 - z_t^{ag})a\right)\left[\frac{K_{t-1}}{\pi p_t^e}\right]^{\alpha-1}, \end{aligned}$$

where $a \in (0,1)$ is the absolute aggregate productivity rate and $\alpha \in (0,1)$ is the output elasticity parameter. Labor supply is defined by the employment rate $p_t^e$ scaled by the time endowment factor $\pi$.

We assume that all agents have time-separable CRRA utility with a risk aversion coefficient $\gamma > 1$ and time preference parameter $\beta \in (0,1)$. Then, given an agent's initial capital endowment $k_{-1} \geq 0$ and the initial cross-sectional distribution $\mu_{-1}$ with non-negative support, the individual optimization problem reads

$$\begin{aligned} (4) \qquad \max_{\{c_t, k_t\} \in \mathbb{R}^2} \; & \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\gamma} - 1}{1-\gamma}\right] \\ \text{s.t.} \; & k_t + c_t = I\left(z_t, k_{t-1}, K_{t-1}\right) + \left[1-\rho\right] k_{t-1} \, \forall \, t \geq 0 \\ & c_t > 0, \; k_t \geq 0 \, \forall \, t \geq 0 \end{aligned}$$

where the productive income $I$ is defined as in (3).

In a competitive equilibrium, the individual problems are solved subject to the market condition (1) that aggregate capital equals the expected optimal individual capital holdings. The question of existence of a competitive equilibrium, in particular one, which has a recursive form, has first been examined by Miao (2006) and has later been solved by Cheridito and Sagredo (2016a,b). To define a recursive equilibrium, let us switch to prime-notation for convenience, where a prime denotes variables in the current period and variables with no prime refer to the previous period.

Definition 1 (Recursive equilibrium) *A solution to the agents' individual optimization problems (4) subject to the market condition (1) given an initial cross-sectional distribution of individual capital $\mu_{-1}$ with non-negative support is called recursive if there exist functions $h_i : \mathcal{Z} \times \mathbb{R} \times \mathcal{P}(\mathcal{Z}^{id} \times \mathbb{R}) \to \mathbb{R}$, $i \in \{1,2\}$, such that, for any point in time, the current optimal consumption and capital savings*

*choices equal* $c' = h_1(z', k, \mu)$ *and* $k' = h_2(z', k, \mu)$ *for any agent with previous-period capital stock* $k$ *who observes the previous-period cross-sectional distribution* $\mu$ *and the current-period exogenous shock* $z' = (z^{ag'}, z^{id'})$.

Recursive equilibria of models with ex-post heterogeneity can rarely be computed in closed form such that they have to be numerically approximated. When designing a numerical solution algorithm, it is important to show theoretically that the algorithm converges to a true equilibrium. This is the goal of this work. To obtain the theoretical convergence result, we have to change perspective. In the existing literature, the optimal policy functions are approximated point-wise w.r.t. the idiosyncratic arguments whereas here, we view these functions in terms of distributions. Hence, I make the following assumption.

ASSUMPTION 2 (Square-integrability) (*i*) The initial idiosyncratic random variables distributed according to the initial cross-sectional distribution $(\zeta, \kappa) \sim \mu_{-1}$ are square-integrable.

(*ii*) Given any cross-sectional distribution $\mu$, the corresponding functions for the optimal consumption and savings choice $h_i(z^{ag'}, \zeta, \kappa, \mu)$, $i \in \{1, 2\}$, are square-integrable w.r.t. the idiosyncratic random variables $(\zeta, \kappa) \sim \mu$, i.e., in short-hand notation $h_1, h_2 \in L^2(\mathcal{Z}^{id} \times \mathbb{R}, \mathcal{B}(\mathcal{Z}^{id} \times \mathbb{R}), \mu)$.

In order to obtain a full description of equilibrium, we need to define the consistent law of motion of $\mu$ to $\mu'$. Given a fixed distribution $\mu$ over the cross-section of individual capital at the end of the previous period and a recursive equilibrium, the distribution in the current period changes in two steps $\mu \to \tilde{\mu}' \to \mu'$. In the first step, the new shocks $z'$ for all agents realize and shift the quantities of employed and unemployed agents depending on the outcome of the aggregate shock. Formerly employed agents either stay employed or become unemployed, the same holds for the formerly unemployed. Therefore, the distribution at the beginning of the current period $\tilde{\mu}'$ is given by

$$
\begin{aligned}
(5) \qquad \tilde{\mu}'\left(z^{id'}, k\right) &= \sum_{z^{id} \in \mathcal{Z}^{id}} \frac{p^{(z^{ag'}, z^{id'})|(z^{ag}, z^{id})}}{p^{z^{ag'}|z^{ag}}} \mu\left(z^{id}, k\right) \\
&= \sum_{z^{id} \in \mathcal{Z}^{id}} \frac{p^{(z^{ag'}, z^{id'})|(z^{ag}, z^{id})}}{p^{z^{ag'}|z^{ag}}} \mathbb{P}\left(\left\{\zeta = z^{id}\right\} \cap \left\{\kappa \leq k\right\} \middle| z^{ag}\right)
\end{aligned}
$$

for all $z^{id'} \in \mathcal{Z}^{id}$ and $k \in \mathbb{R}$. The multipliers in front of the previous distribution are the probabilities that the employment status changes from $z^{id}$ to $z^{id'}$ given the observed trajectory of $z^{ag}$ to $z^{ag'}$. In the second step, the agents implement their

optimal capital savings, which leads to the new current-period distribution

$$\mu'\left(z^{id'},k\right)=\sum_{z^{id}\in\mathcal{Z}^{id}}\frac{p^{(z^{ag'},z^{id'})|(z^{ag},z^{id})}}{p^{z^{ag'}|z^{ag}}}\mathbb{P}\left(\left\{\zeta=z^{id}\right\}\cap\left\{h_2\left(z^{ag'},z^{id'},\kappa,\mu\right)\leq k\right\}\middle|\, z^{ag'},z^{ag}\right). \tag{6}$$

From this definition of the new distribution, the new aggregate capital $K'$ follows immediately due to (1). Now that all model ingredients are defined, the next section lays out the methodology to compute the recursive equilibrium.

## 3. THE METHODOLOGY

The methodology proposed herein builds on the observation that the optimal policy functions $h_1$ for consumption and $h_2$ for capital savings solve the Euler equation, which is equivalent to the first-order condition of the following constrained optimization problem

$$\begin{aligned}\min_{\{h_1,h_2\}}\quad & -u(h_1)-\sum_{z''\in\mathcal{Z}}p^{z''|z'}\beta u\left(I\left(z'',h_2,K'\right)+\left[1-\rho\right]h_2-h_2'\right)\\ \text{s.t.}\quad & 0=I\left(z',k,K\right)+\left[1-\rho\right]k-h_1-h_2\\ & 0\geq -h_1,\ 0\geq -h_2.\end{aligned} \tag{7}$$

The utility function $u:\mathbb{R}_{>0}\to\mathbb{R}$, $c\mapsto\frac{1}{1-\gamma}(c^{1-\gamma}-1)$ is defined as in (4), $p^{z''|z'}$ is the exogenously given transition probability that $z'$ is followed by $z''$ and $I$ is as in (3).

The standard way of solving a constrained optimization problem is to set up the corresponding Lagrangian and to find a saddle point by minimizing over the policies and maximizing over the Lagrange multipliers. This can be done using the proximal point algorithm as explained subsequently. However, complications arise due to the ex-post heterogeneity, which introduces the cross-sectional distribution to the state space. We have to discretize the space of distributions, which is an infinite-dimensional object. A solution to this challenge is discussed after introducing the basic proximal point algorithm.

### 3.1. *The Proximal Point Algorithm*

Applying the proximal point algorithm to minimax problems on saddle functions like Lagrangians goes back to a series of papers by Rockafellar (1970, 1976a,b) who sets the theoretical base for the convergence of said algorithm. Let me first define what a saddle function is in this context.

DEFINITION 3 (Saddle function (see Rockafellar, 1970)) *(i) Let $\mathcal{C}$ and $\mathcal{D}$ be Hilbert spaces over $\mathbb{R}$. A saddle-function is an everywhere-defined function $L : \mathcal{C} \times \mathcal{D} \to [-\infty, \infty]$ such that $L(c, d)$ is a convex function of $c \in \mathcal{C}$ for any $d \in \mathcal{D}$ and a concave function of $d \in \mathcal{D}$ for any $c \in \mathcal{C}$.*

*(ii) A saddle function is called proper if there exists a point $(c, d) \in \mathcal{C} \times \mathcal{D}$ with $L(c, \tilde{d}) < +\infty$ for any $\tilde{d} \in \mathcal{D}$ and $L(\tilde{c}, d) > -\infty$ for any $\tilde{c} \in \mathcal{C}$.*

*(iii) The operator associated with the saddle function $L$ is defined as the set-valued mapping*

$$\begin{aligned}\mathbf{T}_L(c, d) = \{(v, w) | & L(\tilde{c}, d) - \langle \tilde{c}, v \rangle + \langle d, w \rangle \\ & \geq L(c, d) - \langle c, v \rangle + \langle d, w \rangle \\ & \geq L(c, \tilde{d}) - \langle c, v \rangle + \langle \tilde{d}, w \rangle \, \forall (\tilde{c}, \tilde{d}) \in \mathcal{C} \times \mathcal{D} \Big\},\end{aligned}$$

*where $\langle ., . \rangle$ denotes the Hilbert space inner product. A saddle point is a point $(c^*, d^*) \in \mathcal{C} \times \mathcal{D}$ such that $0 \in \mathbf{T}_L(c^*, d^*)$.*[3]

According to this definition, finding a recursive equilibrium translates into the following task: Given that a zero of the operator $\mathbf{T}_L$ associated with the saddle function of (7) exists (see Cheridito and Sagredo, 2016a), we want to construct a saddle point corresponding to the root of $\mathbf{T}_L$. This saddle point construction relies on an important mathematical property called maximal monotonicity.[4] The operator $\mathbf{T}_L$ associated with a saddle function $L$ possesses this property if the following corollary is satisfied.

COROLLARY 4 (Rockafellar (1970)) *Let $\mathcal{C}$ and $\mathcal{D}$ be Hilbert spaces over $\mathbb{R}$. If $L(c, d)$ is a proper saddle function on $\mathcal{C} \times \mathcal{D}$, which is lower semicontinuous in its convex element $c \in \mathcal{C}$ and upper semicontinuous in its concave element $d \in \mathcal{D}$, then the operator $\mathbf{T}_L$ associated with $L$ is maximal monotone.*

The reason for the importance of this property is that the resolvent[5] of a maximal monotone operator $\mathbf{T}$ is firmly nonexpansive.[6] This fact is due to Minty (1962).

[3] The operator $\mathbf{T}_L$ is closely related to the subdifferential of the saddle function $L$ as $v$ equals the subgradient of $L(., d)$ at $c \in \mathcal{C}$ and $w$ is the subgradient of $-L(c, .)$ at $d \in \mathcal{D}$.

[4] Maximal monotonicity (see e.g., Phelps, 1997; Bauschke and Combettes, 2011): Let $\mathcal{E}$ be a Hilbert space. An operator $\mathbf{T} : \mathcal{E} \to \mathcal{E}$ is called a monotone operator if for any two elements of its graph $(e, f), (\tilde{e}, \tilde{f}) \in G(\mathbf{T}) = \{(e, f) \in \mathcal{E}^2 | f \in \mathbf{T}(e)\}$ it holds that $\langle e - \tilde{e}, f - \tilde{f} \rangle \geq 0$. It is, additionally, called maximal monotone if any $(\tilde{e}, \tilde{f}) \in \mathcal{E}^2$ with $\langle e - \tilde{e}, f - \tilde{f} \rangle \geq 0 \, \forall (e, f) \in G(\mathbf{T})$ is necessarily also an element of the graph $(\tilde{e}, \tilde{f}) \in G(\mathbf{T})$.

[5] Resolvent (see e.g., Bauschke and Combettes, 2011): Let $\mathcal{E}$ be a Hilbert space. The resolvent of an operator $\mathbf{T} : \mathcal{E} \to \mathcal{E}$ is the operator $(\mathbf{Id} + \mathbf{T})^{-1}$ where $\mathbf{Id}$ is the identity operator.

[6] Nonexpansiveness (see e.g., Bauschke and Combettes, 2011): Let $\mathcal{E}$ be a Hilbert space. An

It is well known that any firmly nonexpansive operator is equivalent to a mixture $(1/2)\mathbf{Id} + (1/2)\mathbf{R}$ of the identity operator $\mathbf{Id}$ and a nonexpansive operator $\mathbf{R}$ (see e.g., Bauschke and Combettes, 2011, Remark 4.24 (iii)). Weak convergence of the iteration of such a mixture to its fixed point is well established (see e.g., Zeidler, 1986a, Proposition 10.16). This procedure is also known as damped fixed-point iteration.

Iterating on the resolvent of a maximal monotone operator yields the proximal point algorithm. This algorithm hence results in the fixed point of the resolvent, which is equivalent to a root of the operator $\mathbf{T}$ itself. Therefore, it leads to a recursive equilibrium when we consider the resolvent of the operator $\mathbf{T}_L$ associated with the saddle function of (7). To understand how such a resolvent is constructed, let us look at a simplified example first.

EXAMPLE (Resolvent of a subdifferential) Let $\mathcal{E}$ be a Hilbert space. Consider a lower semicontinuous proper convex function $F : \mathcal{E} \to [-\infty, \infty]$. It is well known that its subdifferential $\partial F$ is maximal monotone (see e.g., Bauschke and Combettes, 2011, Theorem 20.40). We are looking for a fixed point $e^* \in \mathcal{E}$ of the resolvent of $F$, which can be computed by simple iteration with iteration count $n$,

$$e_n \overset{n\to\infty}{\longrightarrow} e^* \text{ with } e_{n+1} = (\mathbf{Id} + \partial F)^{-1}(e_n).$$

The resolvent $(\mathbf{Id} + \partial F)^{-1}$ can be represented by

$$\begin{aligned} e_{n+1} = (\mathbf{Id} + \partial F)^{-1}(e_n) &\Leftrightarrow e_n = (\partial F + \mathbf{Id})(e_{n+1}) \\ &\Leftrightarrow 0 = (\partial F + \mathbf{Id})(e_{n+1}) - \mathbf{Id}(e_n) \\ &\Leftrightarrow e_{n+1} = \arg\min_{e\in\mathcal{E}} F(e) + \frac{1}{2}\|e - e_n\|^2. \end{aligned}$$

The latter is the update of the proximal point algorithm.[7]

This example shows that the proximal point algorithm in our case translates into an algorithm on augmented Lagrangians. To ensure convergence, a regularization term containing the previous iterate has to be added to the Lagrangian. We define the update of the proximal point algorithm for the Lagrangian of our agents in the growth model in the following.

operator $\mathbf{T} : \mathcal{E} \to \mathcal{E}$ is called nonexpansive if it is Lipschitz continuous with constant 1. It is called firmly nonexpansive if for all $e, \tilde{e} \in \mathcal{E}$ it holds that $\|\mathbf{T}(e) - \mathbf{T}(\tilde{e})\|^2 \leq \langle e - \tilde{e}, \mathbf{T}(e) - \mathbf{T}(\tilde{e})\rangle$.

[7] The proximal point update presented here is a simplified version. Rockafellar (1976a) proves convergence for a resolvent $(\mathbf{Id} + \lambda^n \mathbf{T})^{-1}$ where $\{\lambda^n\}_{n=1}^{\infty}$ is either constant and bounded away from zero or a series $0 < \lambda^n \nearrow \lambda^\infty \leq \infty$.

3.1.1. *The Proximal Point Algorithm for the Growth Model*

We follow Rockafellar (1976b) for defining the proximal point algorithm's update. This algorithm iterates on the resolvent of the operator associated with the Lagrangian of (7). Hence, each iteration on the resolvent updates the agents' optimal choices for consumption $h_1$ and individual capital $h_2$ as well as the three Lagrange multipliers $y_1$ for the equality constraint and $y_2$ and $y_3$ for the inequality constraints of (7). Similarly to the simplified example in the previous section, the $[n+1]$-th iterate of the agent's optimal choices, i.e., $h_1^{n+1}$ and $h_2^{n+1}$, is the minimizer of the Lagrangian, which is augmented by terms featuring the $n$-th iterate. The augmented Lagrangian is a function $L^A : \prod_{i=1}^{5} L^2(\mathcal{Z}^{id} \times \mathbb{R}, \mathcal{B}(\mathcal{Z}^{id} \times \mathbb{R}), \mu) \to [-\infty, \infty]$ given by

$$
\begin{aligned}
(8) \qquad L^A(h_1, h_2, y_1, y_2, y_3; h^n) = & -u(h_1) \\
& - \sum_{z'' \in \mathcal{Z}} p^{z''|z'} \beta u\left(I\left(z'', h_2, K'\right) + [1-\rho] h_2 - h_2'\right) \\
& + \frac{1}{2\lambda}\left(h_1 - h_1^n\right)^2 + \frac{1}{2\lambda}\left(h_2 - h_2^n\right)^2 \\
& + y_1\left(I\left(z', k, K\right) + [1-\rho] k - h_1 - h_2\right) \\
& + \frac{\lambda}{2}\left(I\left(z', k, K\right) + [1-\rho] k - h_1 - h_2\right)^2 \\
& + \begin{cases} -y_2 h_1 + \frac{\lambda}{2}(h_1)^2 & , h_1 \leq \frac{y_2}{\lambda} \\ -\frac{1}{2\lambda}(y_2)^2 & , h_1 > \frac{y_2}{\lambda} \end{cases} \\
& + \begin{cases} -y_3 h_2 + \frac{\lambda}{2}(h_2)^2 & , h_2 \leq \frac{y_3}{\lambda} \\ -\frac{1}{2\lambda}(y_3)^2 & , h_2 > \frac{y_3}{\lambda} \end{cases},
\end{aligned}
$$

where $h^n = (h_1^n, h_2^n)$ and $\lambda > 0$ is the step size parameter of the proximal point algorithm. Note that the next-period optimal capital savings are naturally given by the composition $h_2' = h_2^n \circ h_2$. The first two lines of the Lagrangian features the objective of (7), the fourth line contains its equality constraint with its Lagrange multiplier. The third and fifth line consist of the objective's and the equality constraint's proximal point augmentations, which transform the saddle-point operator into its resolvent. The last two lines correspond to the inequality constraints. They also consist of the Lagrange term and the augmentation but they are defined piecewise to account for the case of a binding constraint.

With the augmented Lagrangian as above, we now state the algorithm to approximate a recursive equilibrium of the growth model in Algorithm 1. Note that Rockafellar (1976a) shows that the proximal point algorithm converges to a sad-

**Algorithm 1** Proximal point algorithm for the growth model

▷ *A Initialization*

1: Set $n = 0$. Initialize the agents' choices of consumption and individual capital and the Lagrange multipliers $H^n = (h_1^n, h_2^n, y_1^n, y_2^n, y_3^n)$.
2: Set the parameter $\lambda > 0$.
3: Set the termination criterion small $\tau > 0$ and the initial distance larger $d > \tau$.

▷ *B Iterative procedure*

4: **while** $d > \tau$ **do**
5: Update $H^{n+1}$ by

$$
\begin{aligned}
h^{n+1} &\approx \arg\min_{h_1,h_2} L^A\left(h_1, h_2, y_1^n, y_2^n, y_3^n; h^n\right) \\
y_1^{n+1} &= y_1^n + \lambda\left(I\left(z', k, K\right) + \left[1-\rho\right]k - h_1^{n+1} - h_2^{n+1}\right) \\
y_2^{n+1} &= \max\left(0, y_2^n - \lambda h_1^{n+1}\right) \\
y_3^{n+1} &= \max\left(0, y_3^n - \lambda h_2^{n+1}\right)
\end{aligned}
$$

where $L^A$ is defined as in (8).
6: Compute the distance $d = \|H^{n+1} - H^n\|$.
7: Set $n = n + 1$.
8: **end while**

dle point of the Lagrangian even if the update of the optimal consumption and individual capital is only approximate. This is important as the minimizer of the Lagrangian is often not known in closed form, but it can be approximated with standard nonlinear solvers. Salzo and Villa (2012) extend this result to different concepts of approximation. Let me define which kind of approximation applies in this work.

Definition 5 (Resolvent approximation[8]) *Let $\mathcal{C}$ be a Hilbert space over $\mathbb{R}$. Consider the resolvent $(\mathbf{Id} + \lambda\mathbf{T}_L)^{-1}(c)$ of an operator $\lambda\mathbf{T}_L$ associated with a saddle function $L$ at $c \in \mathcal{C}$ with $\lambda > 0$. The approximation with $\epsilon$-precision of this resolvent at $c \in \mathcal{C}$ is defined as $\tilde{c} \in \left(\mathbf{Id} + \lambda\mathbf{T}_L^{\epsilon^2/(2\lambda)}\right)^{-1}(c)$ where*

$$
\mathbf{T}_L^{\epsilon^2/(2\lambda)}(c) = \left\{ v \,\middle|\, L(c) - L(\tilde{c}) + \langle \tilde{c} - c, v\rangle \leq \frac{\epsilon^2}{2\lambda}\ \forall \tilde{c} \in \mathcal{C} \right\}.
$$

*It is denoted by $\tilde{c} \approx (\mathbf{Id} + \lambda\mathbf{T}_L)^{-1}(c)$.*

The convergence rate of Algorithm 1 is $O(n^{-1})$ as is shown by Güler (1991).

[8] This definition corresponds to the type 2 approximation with $\epsilon$-precision in Salzo and Villa (2012). Note that the approximation operator is not an approximate saddle function operator but an approximate subdifferential operator. This is the case because we minimize the controls for fixed Lagrange multipliers rather than computing a minimax problem immediately in Algorithm 1.

The proximal point algorithm can, however, be accelerated, which goes back to Güler (1992). The convergence rate of the accelerated algorithm is $O(n^{-2})$, which was proven in Salzo and Villa (2012). In the following, I explain the acceleration.

#### 3.1.2. *The Accelerated Proximal Point Algorithm for the Growth Model*

The idea behind the acceleration is to approximate the highly nonlinear augmented Lagrangian with a sequence of simple convex quadratic functions $\{\phi^n\}_{n=1}^{\infty}$ such that the difference to the Lagrangian is reduced by a fraction $(1-\alpha^n) \in (0,1]$ in every iteration step

$$\phi^{n+1} - L^A \leq (1-\alpha^n)(\phi^n - L^A).$$

The update for the agents' optimal choices $h = (h_1, h_2)$ is then determined such that the following condition is satisfied

$$L^A\left(h^{n+1}, y_1^n, y_2^n, y_3^n; h^n\right) \leq \hat{\phi}^{n+1} = \min_h \phi^{n+1}(h),$$

where $\phi^{n+1}$ is of the form $\phi^{n+1}(h) = \hat{\phi}^{n+1} + (A^{n+1}/2)\|h - \nu^{n+1}\|^2$.

Salzo and Villa (2012) show that this is achieved by Algorithm 2. Furthermore, they show that this algorithm has a convergence rate of $O(n^{-2})$ if the resolvent approximation precision increases by $\epsilon^n = O(1/n^q)$ with $q > 3/2$.

### 3.2. *Discretizing the Space of Distributions*

So far, I just introduced the standard methodology of the proximal point algorithm. However, our model demands an extension. The recursive equilibrium, we want to solve for, depends on the cross-sectional distribution, which is an element of the state space. Therefore, we need to discretize the space of distributions. If we simply use a spline interpolation on the distribution, the discretized state space becomes very large very quickly. Another option would be a projection on orthogonal polynomials, which is widely used in computational economics, but a prerequisite is a smooth distribution. Due to the borrowing constraint, which is occasionally binding, however, the cross-sectional distribution exhibits mass points at the constraint and elsewhere as is proven in the following.

PROPOSITION 6 (A condition for mass points[9]) *Consider a recursive equilibrium as in Definition 1 with an explicit debt constraint $k \geq \delta$ with $\delta \in \mathbb{R}$. Suppose that there exists a $\hat{z} \in \mathcal{Z}$ with $p^{\hat{z}|\hat{z}} > 0$ and a $\hat{k} > \delta$ such that $h_2(\hat{z}, k, \mu) \leq k$ for all*

[9]The proof can be found in Appendix A.

**Algorithm 2** Accelerated proximal point algorithm for the growth model

▷ *A Initialization*

Set $n = 0$. Initialize the agents' choices of consumption and individual capital and the Lagrange multipliers $H^n = (h_1^n, h_2^n, y_1^n, y_2^n, y_3^n)$. Set $\nu^n = h^n = (h_1^n, h_2^n)$.

Set the parameters $\lambda > 0$, $A^n > 0$ and $b \in [0, 2)$.

Set the resolvent approximation precision $\{\epsilon^n\}_{n=0}^{\infty}$.

Set the termination criterion small $\tau > 0$ and the initial distance larger $d > \tau$.

▷ *B Iterative procedure*

**while** $d > \tau$ **do**

Update $\alpha^n = \frac{1}{2}\left(\sqrt{(b\lambda A^n)^2 + 4b\lambda A^n} - b\lambda A^n\right)$.

Update $x^n = (1 - \alpha^n)h^n + \alpha^n \nu^n$.

Update $H^{n+1}$ by

$$
\begin{aligned}
h^{n+1} &\approx \arg\min_{h_1, h_2} L^A\left(h_1, h_2, y_1^n, y_2^n, y_3^n; x^n\right) \\
y_1^{n+1} &= y_1^n + \lambda\left(I\left(z', k, K\right) + \left[1 - \rho\right]k - h_1^{n+1} - h_2^{n+1}\right) \\
y_2^{n+1} &= \max\left(0, y_2^n - \lambda h_1^{n+1}\right) \\
y_3^{n+1} &= \max\left(0, y_3^n - \lambda h_2^{n+1}\right)
\end{aligned}
$$

where $L^A$ is defined as in (8).

Update $A^{n+1} = (1 - \alpha^n)A^n$.

Update $\nu^{n+1} = \nu^n - \frac{\alpha^n}{(1-\alpha^n)\lambda A^n}(x^n - h^{n+1})$.

Compute the distance $d = \|H^{n+1} - H^n\|$.

Set $n = n + 1$.

**end while**

$k \in [\delta, \hat{k}]$. *Furthermore, assume that the optimal capital savings function has a kink at* $k^* := \max\{k \geq \delta \,|\, h_2(\hat{z}, k, \mu) = \delta\} > \delta$, *i.e., the debt constraint is binding, and that* $h_2$ *is strictly increasing in* $k \geq k^*$. *Then, the cross-sectional distribution has a mass point at the constraint* $\delta$. *If, additionally, there exists a* $\bar{z} \in \mathcal{Z}$ *with* $p^{\bar{z}|\hat{z}} > 0$ *and* $h_2(\bar{z}, \delta, \mu) > \delta$, *then the cross-sectional distribution has multiple mass points.*

This result implies jumps in the cross-sectional distribution $\mu$. Hence, standard orthogonal polynomial projection methods do not work here. There is, however, an efficient way of approximating distributions called polynomial chaos. This is a technique, which projects the distribution on orthogonal polynomials of random variables rather than the real line. As such, it can also handle discontinuous distributions. In the following, I will summarize polynomial chaos in general and subsequently, I explain how this technique is applied to our growth model.

3.2.1. *Polynomial Chaos*

The standard polynomial chaos expansion is an approach to represent random variables by a series of polynomials mapping basic random variables into the space of square-integrable random variables $L^2$. Originally, this approach yields the so-called Wiener-Hermite expansion, i.e., a projection onto Hermite polynomials, which take Gaussians as basic random variables. The well known Cameron-Martin theorem (see e.g., Ernst et al., 2012, Theorem 2.1) shows that this construction spans all square-integrable random variables, which are measurable w.r.t. the basic random variables. Xiu and Karniadakis (2002) extend this concept to sets of orthogonal polynomials mapping more general basic random variables, e.g., uniform, gamma or binomial variables, into $L^2$. The $L^2$-convergence result for these generalized polynomial chaos expansions is proven in Ernst et al. (2012). The main purpose of this generalization is the gain in convergence speed when the basic random variables are chosen such that they are similar to the approximated random variable. To summarize, given a basic random variable $\xi \in L^2$ with distribution $\xi \sim F$, which has finite moments of all orders, and a set of orthogonal polynomials $\{\Phi_i\}_{i=0}^{\infty}$, where $i$ denotes the order of each polynomial, we can represent any random variable $\kappa \in L^2$ with distribution $\kappa \sim \mu$ by

$$\kappa = \sum_{i=0}^{\infty} \varphi_i \Phi_i(\xi), \tag{9}$$

where $\varphi_i$ are constant projection coefficients.

It is important to note that there is a specific connection between the basic random variable and the set of orthogonal polynomial used. The orthogonality condition of the polynomials reveals this relation. For polynomials of order $i, j \in \{0, 1, \ldots\}$, it reads

$$\langle \Phi_i, \Phi_j \rangle = \int_{-\infty}^{\infty} \Phi_i(\xi)\, \Phi_j(\xi)\, dF(\xi) = \frac{\delta_{ij}}{a_i^2}, \tag{10}$$

where $\delta_{ij}$ denotes the Kronecker symbol and $a_i \neq 0$ are constants. One can see that the weighting function, which defines the orthogonal polynomials, has to equal the distribution of the basic random variable. Once a basic random variable is fixed, we can generate the corresponding orthogonal polynomials by the three-term recurrence relation (see e.g., Gautschi, 1982; Zheng et al., 2015)

$$\Phi_{i+1}(\xi) = (\xi - \theta_i)\, \Phi_i(\xi) - \omega_i \Phi_{i-1}(\xi), \; i \in \{0, 1, \ldots\}, \tag{11}$$

where the starting polynomials are defined as $\Phi_{-1}(\xi) = 0$ and $\Phi_0(\xi) = 1$ and $\theta_i, \omega_i \in \mathbb{R}$ are constant parameters with $\omega_i > 0$.

The projection coefficients in the polynomial chaos expansion of a random variable $\kappa \in L^2$ with distribution $\kappa \sim \mu$ are defined as usual by $\varphi_i = \langle \kappa, \Phi_i \rangle / \langle \Phi_i, \Phi_i \rangle$ for all $i \in \{0, 1, \ldots\}$. If $\kappa$ is not a direct function of the basic random variable $\xi$, one uses the fact that both c.d.f.s $\mu, F \sim \mathcal{U}[0,1]$ are uniform to compute the coefficients

$$\varphi_i = \frac{1}{\langle \Phi_i, \Phi_i \rangle} \int_{-\infty}^{\infty} \mu^{-1}\left(F\left(\xi\right)\right) \Phi_i\left(\xi\right) dF\left(\xi\right) \; \forall\, i \in \{0, 1, \ldots\}, \tag{12}$$

where $\mu^{-1}$ is the generalized inverse distribution function of $\kappa$. Hence, with the polynomial chaos expansion, we can translate any square integrable random variable $\kappa \sim \mu$ into a countable series of constant projection coefficients $\{\varphi_i\}_{i=0}^{\infty}$. For computational reasons, we later truncate the series of projection coefficients.

### 3.2.2. *Applying Polynomial Chaos to our Growth Model*

We apply the polynomial chaos expansion to discretize the cross-sectional distribution of our growth model. Let us first condition the distribution on the employment status, i.e., $\mu^0(k) = \mu(0,k)/p^{z^{id}=0|z^{ag}}$ denotes the cross-sectional distribution of the unemployed. Accordingly, we denote the basic random variable, the projection coefficients and the polynomials of the unemployed with superscript 0 and of the employed with superscript 1, respectively. The optimal consumption and capital savings choices can then be rewritten as $c' = h_1(z', k, \{\varphi_i^0\}_{i=0}^{\infty}, \{\varphi_i^1\}_{i=0}^{\infty})$ and $k' = h_2(z', k, \{\varphi_i^0\}_{i=0}^{\infty}, \{\varphi_i^1\}_{i=0}^{\infty})$. Similarly ,we can derive the law of motion of the projection coefficients from (12), (9) and (6). W.l.o.g. we will write down formulas only w.r.t. the distribution of the unemployed in the following. The law of motion for the projection coefficients of the unemployed reads

$$\varphi_i^{0\prime} = \frac{1}{\langle \Phi_i^0, \Phi_i^0 \rangle} \int_{-\infty}^{\infty} \left[\mu^{0\prime}\right]^{-1} \left(F^0\left(\xi^0\right)\right) \Phi_i^0\left(\xi^0\right) dF^0(\xi^0), \; \forall\, i \in \{0, 1, \ldots\}$$

with

$$\begin{aligned} \mu^{0\prime}(k) = & \frac{1}{p^{z^{id\prime}=0|z^{ag\prime}}} \sum_{j=0}^{1} \frac{p^{(z^{ag\prime},0)|(z^{ag},j)}}{p^{z^{ag\prime}|z^{ag}}} \mathbb{P}\left(\{\zeta = j\} \cap \right. \\ & \left. \left\{ h_2\left(z^{ag\prime}, 0, \sum_{i=0}^{\infty} \varphi_i^j \Phi_i^j\left(\xi^j\right), \{\varphi_i^0\}_{i=0}^{\infty}, \{\varphi_i^1\}_{i=0}^{\infty}\right) \leq k \right\} \middle| z^{ag\prime}, z^{ag} \right). \end{aligned} \tag{13}$$

3.2.3. *A Specific Choice of the Basic Random Variables*

It was illustrated in Xiu and Karniadakis (2002) that the speed of convergence significantly improves if the distribution of the basic random variable is not too far from the distribution we want to approximate. Since the cross-sectional distribution in our growth model is an endogenous object, we do not know its shape a priori. We do know, however, that it will have mass points according to Proposition 6. Also, we know that the same growth model without aggregate shocks, i.e., where $z^{ag}$ is fixed at either 0 or 1, has an endogenous cross-sectional distribution, which stays constant as time goes on. This case is easy to compute because $K' = K$ in the agents' optimization problem (7). Therefore, one just has to solve the individual optimization problem at different values of aggregate capital $K$. In a second step, given these optimal responses, one can compute the stationary distribution as a fixed point of the distribution's law of motion (6). Naturally, this cross-sectional distribution will have features similar to the distribution of the model with aggregate shocks. Hence, we fix $\xi^0$ as the cross-sectional distribution of the unemployed in the model without aggregate shocks averaged over the two cases of keeping $z^{ag}$ fixed as 0 or 1. Analogously, $\xi^1$ is fixed as the distribution of the employed. I compute these stationary distributions using histograms. They are displayed in Figure 2. The distributions exhibit several mass points measured as the local extrema in the histogram representation. One can observe that the distribution of the unemployed is shifted to the left compared to the distribution of the employed as they generally have lower capital savings. Note that the capital constraint at the beginning of the period is binding for a small fraction of employed agents as well because these are agents who were unemployed in the previous period. But the constraint will not be binding at the end of the period because all employed agents optimally choose positive capital savings. To obtain an accurate approximation of these stationary distributions, one should choose a reasonably small bin size for the histogram.

3.2.4. *Generation of the Corresponding Orthogonal Polynomials*

As our basic random variables are represented by histograms, they are essentially discrete distributions where the end points of the bins $\{\xi_n^0\}_{n=1}^N$ have probability $\{p_n^0\}_{n=1}^N$. Generally, orthogonal polynomials w.r.t. a discrete distribution with finite support are considered discrete as well in the sense that their maximal degree is $N$. Furthermore, the highest-order polynomial $\Phi_N^0$ has the points $\{\xi_n^0\}_{n=1}^N$ as roots.

In Zheng et al. (2015), different methods for generating polynomials corresponding to discrete distributions are compared. Of their suggested methods, we use the

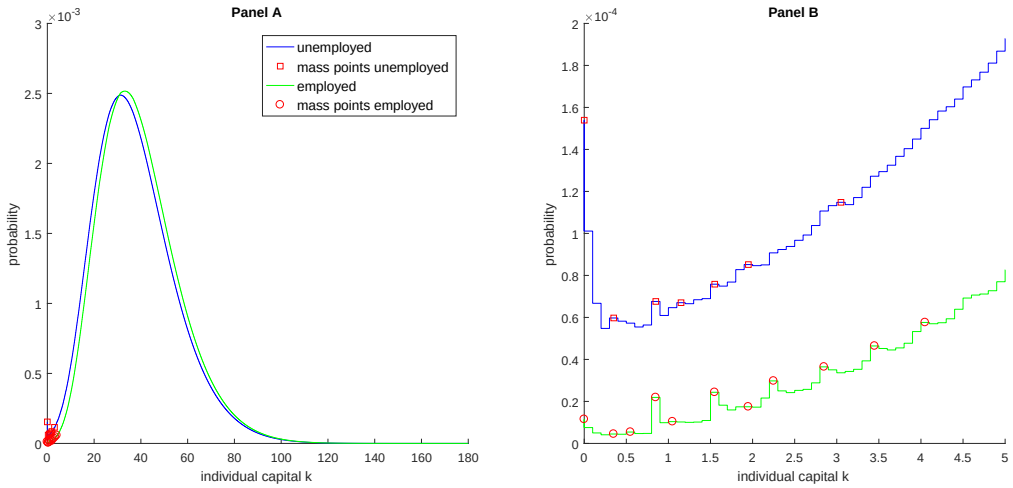


FIGURE 2. **Stationary cross-sectional capital distribution for the growth model without aggregate shocks.** Panel A shows the histogram representation with bin size 0.1 of the stationary distribution conditional on the employment status of the agents. A mass point is identified as a bin whose probability is higher than the ones of its direct neighbors, but the global maxima are excluded. Panel B zooms into the left tail of the distribution.Note that the distribution displayed here is the beginning-of-period distribution $\tilde{\mu}'$ from (5), i.e., before the optimal savings are chosen.

Stieltjes method, which performs well in terms of precision. It directly computes the parameters $\theta_i$ and $\omega_i$ in (11) using the standard inner product of $L^2$ and is explained in detail in Gautschi (1982). The constant parameters are given by

$$\theta_i = \frac{\langle \Phi_i, \xi \Phi_i \rangle}{\langle \Phi_i, \Phi_i \rangle}, \; i \in \{0, 1, \ldots\}$$
$$\omega_i = \frac{\langle \Phi_i, \Phi_i \rangle}{\langle \Phi_{i-1}, \Phi_{i-1} \rangle}, \; i \in \{1, 2, \ldots\}$$

with $\langle ., . \rangle$ denoting the standard inner product of $L^2$ w.r.t. the corresponding basic distribution. As these distributions are represented as discrete distributions, the inner product is a sum rather than an integral. The definitions of these parameters follow from inserting the three-term recurrence relation (11) into the orthogonality condition (10). With the parameters defined as above, the orthogonal polynomials are easily constructed using (11). They are displayed in Figure 3. As usual, the number of roots of each polynomial corresponds to its degree. Note that each first-order polynomial has its root at the mean of the distribution of the corresponding basic random variable.

With the basic random variables defined and the corresponding polynomials generated, the polynomial chaos expansion is fully defined up to order $N$. Any

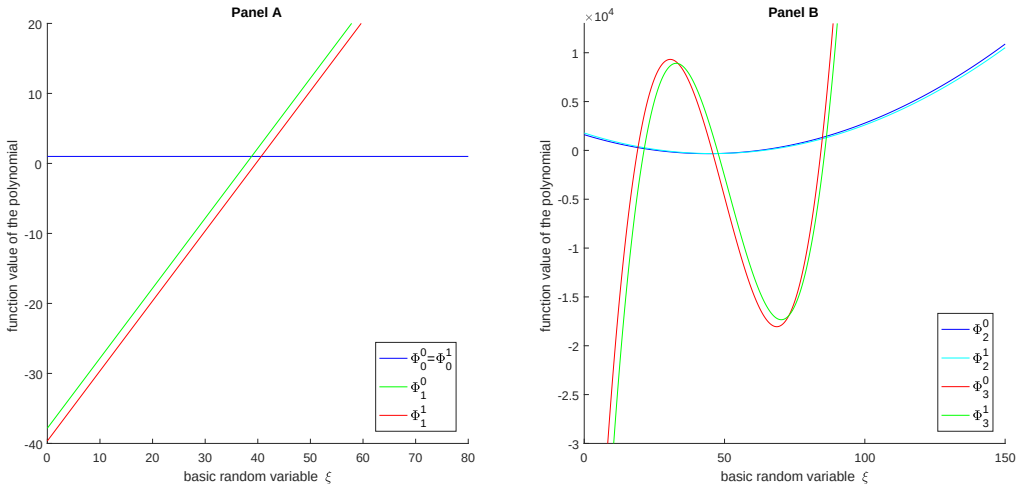


FIGURE 3. **Orthogonal polynomials corresponding to the stationary distribution for the model without aggregate shocks.** Panel A shows the polynomials $\Phi_i^{z^{id}}$ corresponding to the distribution of the unemployed ($z^{id} = 0$) and the employed ($z^{id} = 1$), respectively, up to order $i \leq 1$ and Panel B displays the polynomials $\Phi_i^{z^{id}}$ of order $i = 2$ and $i = 3$.

square integrable distribution measurable w.r.t. the basic random variables can now be projected. The polynomials with different degrees have different effects in this projection as can be seen in Figure 4. In this figure, we consider a polynomial chaos expansion with fixed projection coefficients $\{\varphi_i\}_{i=0}^{\infty}$, which is truncated at different orders. The zeroth-order polynomial results in a mass point at $\varphi_0$, which, due to its definition, is the mean of the projected distribution. Adding the first-order polynomial simply stretches or compresses the distribution of the basic random variable depending on the projection coefficient. Adding the second-order polynomial modifies the skewness of the basic random variable whereas the third-order polynomial adjusts the kurtosis. Higher orders further refine the tails.

## 4. THEORETICAL CONVERGENCE RESULTS

With the methodology laid out in the previous section, we now show convergence of the proposed accelerated proximal point algorithm, which uses polynomial chaos to discretize the cross-sectional distribution. This is done in two steps. Firstly, the convergence for the theoretical resolvent, i.e., the iteration on the policy functions $c' = h_1(z', k, \{\varphi_i^0\}_{i=0}^{\infty}, \{\varphi_i^1\}_{i=0}^{\infty})$ and $k' = h_2(z', k, \{\varphi_i^0\}_{i=0}^{\infty}, \{\varphi_i^1\}_{i=0}^{\infty})$ is shown. In the second step, we show that also the discretized resolvent, i.e., the iteration on the discretized policy functions $c' = h_1^D(z', k, \{\varphi_i^0\}_{i=0}^{M}, \{\varphi_i^1\}_{i=0}^{M})$ and $k' = h_2^D(z', k, \{\varphi_i^0\}_{i=0}^{M}, \{\varphi_i^1\}_{i=0}^{M})$, converges. The superscript $D$ denotes the finite-element discretization w.r.t. the arguments $k$ and $\varphi_i^j$, $i \in \{0, 1, \ldots\}$, $j \in \{0, 1\}$,

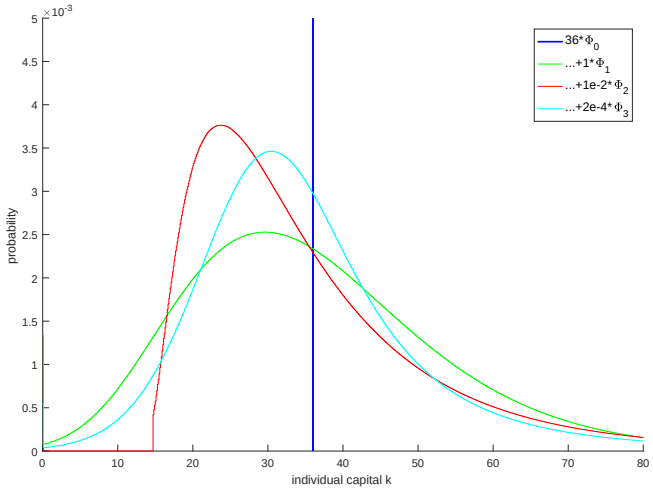


FIGURE 4. **Example distributions resulting from truncated polynomial chaos expansions.** The graph displays the histogram representations with bin size 0.1 of distributions, which result from the polynomial chaos series truncated at different orders ranging from order 0 to 3. The basic random variable used is the stationary cross-sectional distribution of the unemployed in the growth model without aggregate shocks as in Figure 2. The projection coefficients for this example are fixed as $[\varphi_0, \ldots, \varphi_3] = [36, 1, 0.01, 0.0002]$.

whereas $M \leq N$ denotes the order, at which the polynomial chaos is truncated.

### 4.1. *Convergence of the Theoretical Solution Operator to the Recursive Equilibrium*

According to the theory on the proximal point algorithm (i.e., Rockafellar, 1970, 1976a,b; Güler, 1992; Salzo and Villa, 2012) summarized in Section 3.1, it suffices to show that the Lagrangian of the agents' optimization problem (7) satisfies the conditions of Corollary 4. This guarantees the nonexpansiveness of the corresponding resolvent and hence, convergence of the proximal point algorithm.

THEOREM 7 (Convergence) *Consider the growth model from Section 2. Consider the function space $\mathcal{H}$ defined in Proposition 10 for the consumption and capital savings choice. The Lagrangian $L : \mathcal{H} \times \prod_{i=1}^{3} L^2(\mathcal{Z}^{id} \times \mathbb{R}, \mathcal{B}(\mathcal{Z}^{id} \times \mathbb{R}), \mu) \to [-\infty, \infty]$ of the agents' optimization problem (7) in the growth model given by*

$$
\begin{aligned}
(14) \qquad L\left(h_1, h_2, y_1, y_2, y_3\right) = & -u(h_1) \\
& - \sum_{z'' \in \mathcal{Z}} p^{z''|z'} \beta u\left(I\left(z'', h_2, K'\right) + \left[1-\rho\right] h_2 - h_2'\right) \\
& + y_1\left(I\left(z', k, K\right) + \left[1-\rho\right] k - h_1 - h_2\right) - y_2 h_1 - y_3 h_2
\end{aligned}
$$

*satisfies the conditions of Corollary 4 and therefore, Algorithm 1 and Algorithm 2 converge to a recursive equilibrium of the growth model.*

### 4.2. *Convergence of the Discretized Solution Operator*

When implementing the proximal point algorithm, we need to use approximations of two different categories. Firstly, we approximate the minimizer of the resolvent in each iteration of the proximal point algorithm since this optimization problem doesn't have a closed-form solution. Secondly, we have to approximate the policy functions with a finite-dimensional representation. Let me explain the first approximation category. We already pointed out in Section 3.1 that the resolvent can be approximated without affecting convergence if the approximation satisfies Definition 5. This can be achieved as follows.

PROPOSITION 8 (Implementation of the resolvent approximation) *Computing the next iterate $h^{n+1}$ in Algorithm 1, line 5, or Algorithm 2, line 8, as a solution to the formula*

$$X\left(z',k,K\right)\left\|\nabla L^{A}\left(h_{1},h_{2},y_{1}^{n},y_{2}^{n},y_{3}^{n};h^{n}\right)\right\|_{1}\leq\frac{\epsilon^{2}}{2\lambda} \tag{15}$$

*for any $(z',k,K)\in\mathcal{Z}\times\mathbb{R}_{\geq0}^{2}$ results in an approximation with $\epsilon$-precision of the resolvent according to Definition 5. Note that $X$ is the total wealth, i.e., productive income plus savings, of the agent with start capital $k$*

$$X\left(z',k,K\right)=I\left(z',k,K\right)+\left[1-\rho\right]k$$

*with the income $I$ as in (3). Furthermore, $L^{A}$ is the augmented Lagrangian as defined in (8) and $\nabla$ denotes its gradient w.r.t. $(h_1,h_2)$.*

REMARK Equation (15) is easily implemented by any root solver using a tolerance level of $\epsilon^2/(2\lambda)$.

Now, let us investigate the second category of approximation: We have to ensure convergence when using approximations for the functions of optimal consumption $h_1(z',k,\{\varphi_i^0\}_{i=0}^{\infty},\{\varphi_i^1\}_{i=0}^{\infty})$ and capital savings $h_2(z',k,\{\varphi_i^0\}_{i=0}^{\infty},\{\varphi_i^1\}_{i=0}^{\infty})$ in each dimension. Approximation of these policies occurs in two steps. Firstly, we truncate the polynomial chaos expansion, and secondly, we discretize all dimensions and apply the finite element method with first-order Lagrange elements, which amounts to linear interpolation. We denote the truncated policy

by $h^M = h_1(z', k, \{\varphi_i^0\}_{i=0}^M, \{\varphi_i^1\}_{i=0}^M)$. Its interpolant, denoted by $h^{M,D}$, is defined on a tensor product of finite grids of the state space elements

$$D = \Big\{ \Big(k_{i_0}, \varphi^0_{1,i_1}, \ldots, \varphi^0_{M,i_M}, \varphi^1_{1,i_{M+1}}, \ldots, \varphi^1_{M,i_{2M}}\Big) \Big| \\ i_m = 1, \ldots, I_m < \infty \, \forall \, m = 0, \ldots, 2M \Big\}.$$

It is clear that such an interpolant stays within the admissible set of the policies $\mathcal{H}$ defined in Proposition 10. Therefore, convergence follows from a vanishing approximation error. The total policy function approximation error is composed of two parts corresponding to the truncation and interpolation error

$$\|h - h^{M,D}\|_{L^2} \leq \|h - h^M\|_{L^2} + \|h^M - h^{M,D}\|_{L^2}.$$

The following theorem derives bounds on these two parts of the error. The bound on the second part is a well established result from the theory on finite elements (see e.g., Brenner and Scott, 2007) whereas the bound on the first part is more involved. It follows the methodology of the error analysis in Babuška et al. (2007).

THEOREM 9 (Error bounds of the approximation) *Consider the growth model from Section 2 with the function space $\mathcal{H}$ defined in Proposition 10 for the consumption and capital savings choice. Consider Algorithm 1 or Algorithm 2 with polynomial chaos extension as in Section 3.2, i.e., using the basic random variables $\xi^j$ and the corresponding orthogonal polynomials $\Phi^j$ to project any square-integrable $\kappa^j \sim \mu^j$ with $\kappa^j = \sum_{i=0}^N \phi_i^j \Phi_i^j(\xi^j)$, $j \in \{0, 1\}$. Note that the polynomial chaos expansion is real analytic, i.e., it holds that*

$$\left\| \frac{\partial^p}{[\partial \xi^j]^p} \sum_{i=0}^{N} \phi_i^j \Phi_i^j \left(\xi^j\right) \right\| \leq \mathbf{c}^{\Phi,j} p!, \; j \in \{0,1\}, \; p \in \{1, 2, \ldots\}, \tag{16}$$

*where $\mathbf{c}^{\Phi,j}$ is a constant. Assume that, for any fixed exogenous shock, start capital and individual capital distribution $(z', k, \mu)$, the initial guess of the consumption policy $h_1^0$, the savings policy $h_2^0$ and the Lagrange multipliers $(y_1^0, y_2^0, y_3^0)$ for the proximal point algorithm are real analytic and hence, satisfy*

$$\left\| \frac{\partial^p}{[\partial \xi^j]^p} f \right\| \leq \mathbf{c}^{f,j} p!, \; j \in \{0,1\}, \; p \in \{1, 2, \ldots\}, \tag{17}$$

*for some constant $\mathbf{c}^{f,j}$ where $f$ is a handle for $h_1^0, h_2^0, y_1^0, y_2^0$ and $y_3^0$. Consider the*

*following subsets of the complex plane*

$$\Sigma\left(\tau^j, \Gamma^j\right) = \left\{ x \in \mathbb{C} \,\middle|\, \inf_{\xi^j \in \Gamma^j} |x - \xi^j| \leq \tau^j \right\},\ j \in \{0, 1\},$$

*where* $\Gamma^j$ *is the range of* $\xi^j$ *and* $0 < \tau^j < 1/\|\mathbf{A}\|_\infty < \infty$ *with the matrix* $\mathbf{A}$ *given in (26) in the proof. Then, the approximation error bound resulting from truncating the polynomial chaos expansion at order* $M$ *and using linear interpolation on a rectangular tensor-product grid*

$$\begin{aligned} D = \Big\{ \Big(k_{i_0}, \varphi^0_{0,i_1}, \ldots, \varphi^0_{M,i_M}, \varphi^1_{0,i_{M+2}}, \ldots, \varphi^1_{M,i_{2(M+1)}}\Big) \Big| \, k_{i_n} < k_{i_{n+1}}, \\ \varphi^j_{m,i_n} < \varphi^j_{m,i_{n+1}} \, \forall\, i_n \in \{1, \ldots, d_n\},\, m \in \{1, \ldots, M\}, j \in \{0, 1\} \Big\} \end{aligned}$$

*with maximum mesh-size* $s$ *are given by*

$$\begin{aligned} \|h - h^{M,D}\|_{L^2} \leq \sum_{j=0}^{1} \mathbf{b}_j \frac{2}{\eta^j - 1} e^{-M \log(\eta^j)} \max_{f \in \{h^0_1, h^0_2, y^0_1, y^0_2, y^0_3, \Phi\}} \left|\mathbf{c}^{\mathbf{f},\mathbf{j}}\right| \frac{1}{1 - \tau^j \|A\|_\infty} \\ + \mathbf{b}_2 s^2 \left( \sum_{i=0}^{2(M+1)} \left\| \frac{\partial^2 h^M}{[\partial D_i]^2} \right\|^2_{L^2} \right)^{\frac{1}{2}}, \end{aligned}$$

*where* $\mathbf{b}_j$, $j \in \{0, 1, 2\}$, *are constants and*

$$\eta^j = \frac{2\tau^j}{|\Gamma^j|} + \sqrt{1 + \frac{4(\tau^j)^2}{|\Gamma^j|^2}} > 1,\ j \in \{0, 1\}.$$

REMARK The theorem implies that the error from the truncation of the polynomial chaos expansion decreases exponentially with the order of the expansion. Furthermore, the error from the interpolation decreases proportionately to the step size of the discretization.

## 5. NUMERICAL RESULTS

We compute the recursive equilibrium solution of Algorithm 2 truncating the polynomial chaos expansion at different orders using Matlab R2016b.[10] We also compute the solutions of existing algorithms for comparison. We choose the algorithm by Krusell and Smith (1998) as it is the most prominent existing method. We use its Matlab implementation by Maliar et al. (2010). Furthermore, there has been an effort to improve on this original algorithm with a special issue of the Jour-

[10]The computations were performed on the Baobab cluster at the University of Geneva.

nal of Economic Dynamics and Control in January 2010. From these more recent methods, we use the backward induction algorithm by Reiter (2010a) and the explicit aggregation algorithm by den Haan and Rendahl (2010), both implemented in Matlab, since they perform best in the comparison by den Haan (2010).

To ensure comparability, we run all these methods using the same grid for individual capital and the same termination criterion 5e-5. Additionally, I configure the discretizations of the cross-sectional distribution so that they are as close as possible. The Krusell-Smith and the Reiter method use total aggregate capital whereas the den Haan-Rendahl algorithm uses the aggregate capital of the unemployed and employed. In our algorithm, the aggregate capital of the unemployed and employed is equivalent to the projection coefficients $\phi_0^0$ and $\phi_0^1$ of the polynomial chaos expansion. We use 3 grid points for aggregate capital per idiosyncratic shock such that the latter two algorithms have $3 \times 3$ grid points in aggregate capital. For the former two algorithms, we compute the total aggregate capital from each combination of these $3 \times 3$ grid points weighted by the employment rate, which results in 9 grid points. Keep in mind that the proximal point algorithm has additional dimensions to discretize the cross-sectional distribution depending on the order of truncation. The different methods are summarized in Table 1. Note that the proximal point algorithm is implemented with parallelized Matlab

| Algorithm | # Grid Points for $\mu$ | # CPUs | Compute Time |
|---|---|---|---|
| Krusell-Smith (K-S) | 9 | 4 | 38s |
| Reiter (R) | 9 | 4 | 2m 41s |
| den Haan-Rendahl (D-R) | 9 | 4 | 13s |
| PPA M=0 (PPA0) | 9 | 16 | 1m 15s |
| PPA M=1 (PPA1) | 36 | 16 | 3m 12s |
| PPA M=2 (PPA2) | 144 | 16 | 15m 36s |
| PPA M=3 (PPA3) | 576 | 16 | 2h 12m 44s |

TABLE 1. **Summary of the algorithms to be compared.** In the first column, $M$ denotes the order of truncation of the proximal point algorithm. The abbreviation in the parenthesis is the algorithm identifier used in the comparison analysis, which follows. The second column displays the total number of grid points to discretize the cross-sectional distribution. Note that the methods of discretizing the distribution vary across algorithms.

code run on a HPC cluster whereas the other three algorithms are implemented with serial code run on a desktop computer. This is the reason for the differences in the number of CPUs used. The compute time is higher for the proximal point algorithm. This is mainly due to the fact that it solves a full optimization problem in each iteration to ensure convergence. Also, it uses a nonparametric law of mo-

tion, which adds to the complexity. I argue that the goal of being more accurate and having a theoretically sound algorithm justifies the increased compute time. These features are especially important for solving models with a higher degree of nonlinearity, which is ultimately the goal of our algorithm.

In the following, I investigate whether the proximal point algorithm really yields higher precision than the existing methods to compensate its slower performance. Furthermore, we examine the order of truncation of the polynomial chaos expansion resulting in sufficient precision and which economic implications the proximal point algorithm yields.

### 5.1. *Precision of the Proximal Point Algorithm versus Existing Algorithms*

One way of comparing these sets of numerical solutions is to analyze their Euler equation errors. There have been two different Euler equation error tests put forward in the literature (see e.g., den Haan, 2010), the standard Euler equation error test and the dynamic Euler equation error test. The standard Euler equation errors are calculated by comparing the numerical solution for optimal consumption $c$ against the explicitly calculated conditional expectation in the Euler equation denoted by $\tilde{c}$. It is the absolute percentage error

$$\epsilon^{SEee} = \frac{|c - \tilde{c}|}{\tilde{c}}.$$

In contrast to the standard Euler equation error, the dynamic equivalent denoted by $\epsilon^{DEee}$ is computed for several consecutive periods. This test is more stringent as the numerical solution and the explicit conditional expectation usually diverge with more periods. We compute the standard and the dynamic Euler equation error for a random sample of aggregate shocks over N periods for the different numerical solutions from Table 1. We choose the number of periods such that the power of the subjective discount factor $\beta^t$, $t = 0, \ldots, N-1$, is always above machine precision. Our configuration leads to $N = 3587$. Otherwise, we would only add noise to the infinite sum of utility. Note that we compute the standard Euler Equation error test also over multiple periods but it is reset every period and hence, does not accumulate. The errors' summary statistics are displayed in Table 2, Panel A. One can see that the proximal point algorithm irrespective of the truncation order improves the median error by one order of magnitude and the maximum error by two orders. When looking at the dynamic error, the den Haan-Rendahl algorithm produces a median error of the same order of magnitude. The summary statistics can be deceiving in that respect. The improvement becomes

clearer when displaying the full error distribution in terms of boxplots in Figure 5. One can see that the existing algorithms produce much wider error distributions

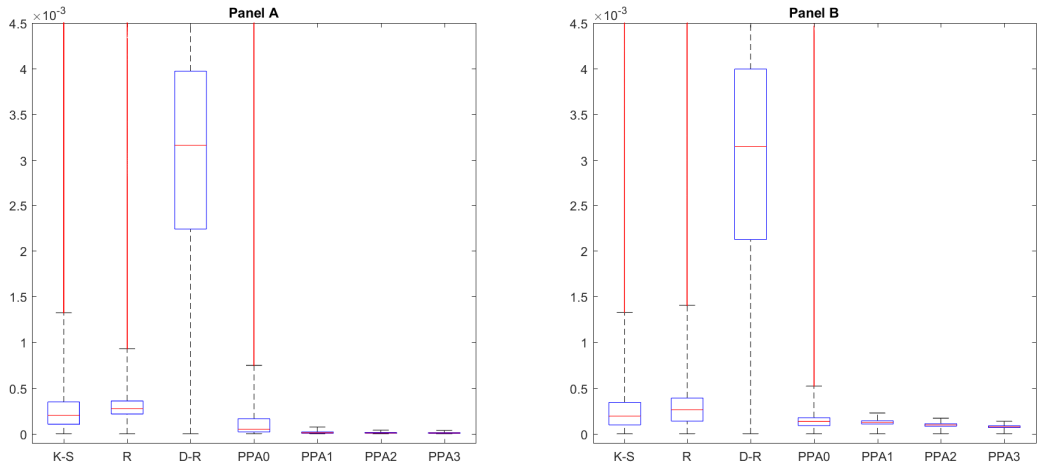


FIGURE 5. **Boxplots of the Euler equation error distributions of individual capital.** Panel A shows the standard Euler equation error and Panel B displays the dynamic Euler equation error for the numerical solutions from Table 1. The error is computed over all grid points in the dimension of individual capital and the idiosyncratic shock for a random sample of aggregate shocks over $N = 3587$ periods. The initial cross-sectional distribution is the same for all algorithms. The red lines mark the medians whereas the blue boxes denote the $25^{\text{th}}$ to $75^{\text{th}}$ percentiles. The whiskers indicate the range of the distribution and the red dots outside are outliers.

for both the standard and the dynamic Euler equation error. It is interesting to observe that the Reiter algorithm, although improving on the extreme points of the error distribution, does not lead to any significant improvement compared to the Krusell-Smith algorithm. The same is true for the den Haan-Rendahl method, which performs even worse. It seems that the reason why they performed well in the comparison by den Haan (2010) is that they use considerably more grid points whereas here we deliberately run all methods on the same discretization.

| | K-S | R | D-R | PPA1 | PPA2 | PPA3 | PPA4 |
|---|---|---|---|---|---|---|---|
| Panel A: Individual consumption | | | | | | | |
| $\epsilon^{SEee}$ Median | 1.74e-04 | 3.04e-04 | -1.67e-04 | -3.92e-05 | -1.26e-05 | -1.42e-05 | -1.55e-05 |
| $\epsilon^{SEee}$ Min | -6.24e-03 | -2.61e-03 | -7.05e-03 | -2.21e-03 | -2.22e-03 | -2.21e-03 | -2.21e-03 |
| $\epsilon^{SEee}$ Max | 1.67e-02 | 2.08e-03 | 6.44e-03 | 1.38e-05 | 2.65e-05 | 3.17e-05 | 3.54e-05 |
| $\epsilon^{DEee}$ Median | 1.72e-04 | 3.18e-04 | 8.25e-05 | 7.48e-05 | 6.91e-05 | 8.67e-05 | 8.60e-05 |
| $\epsilon^{DEee}$ Min | -6.22e-03 | -2.61e-03 | -6.95e-03 | -2.15e-03 | -2.17e-03 | -2.15e-03 | -2.15e-03 |
| $\epsilon^{DEee}$ Max | 1.67e-02 | 2.36e-03 | 7.33e-03 | 2.25e-04 | 2.34e-04 | 2.21e-04 | 2.23e-04 |
| Panel B: Aggregate consumption | | | | | | | |
| $\epsilon^{SEee}$ Median | 5.23e-05 | 2.99e-04 | -1.27e-03 | -3.27e-05 | -4.87e-06 | -1.16e-05 | -1.42e-05 |
| $\epsilon^{SEee}$ Min | -5.08e-05 | -6.73e-04 | -2.45e-03 | -6.37e-05 | -5.16e-05 | -5.67e-05 | -5.69e-05 |
| $\epsilon^{SEee}$ Max | 1.44e-04 | 5.08e-04 | 2.16e-03 | -5.01e-06 | 2.19e-05 | 1.57e-05 | 9.79e-06 |
| $\epsilon^{DEee}$ Median | 1.33e-06 | 2.64e-04 | 1.89e-04 | 1.72e-04 | 1.60e-04 | 1.70e-04 | 1.70e-04 |
| $\epsilon^{DEee}$ Min | -1.03e-04 | -7.53e-04 | -3.89e-03 | -3.74e-05 | -1.26e-06 | -4.86e-06 | -8.38e-06 |
| $\epsilon^{DEee}$ Max | 1.41e-04 | 1.13e-03 | 2.75e-03 | 2.80e-04 | 2.51e-04 | 2.65e-04 | 2.68e-04 |

TABLE 2. **Euler equation errors for the numerical solutions from Table 1.** This table displays the summary statistics of the standard Euler equation error $\epsilon^{SEee}$ and of the dynamic Euler equation error $\epsilon^{DEee}$ . Panel A displays the error for individual consumption where it is computed over all grid points in the dimension of individual capital and the idiosyncratic shock for a random sample of aggregate shocks over $N = 3587$ periods. Panel B displays the error for aggregate consumption, i.e., the percentage deviation of the aggregate consumption implied by the algorithms from the aggregate consumption implied from the Euler equation. The initial cross-sectional distribution is the same for all algorithms.

In comparison, all proximal point solutions produce much narrower error bands. As one would expect, they become wider for the dynamic Euler equation error, but are still much narrower than the Krusell-Smith error. This is mainly due to the better anticipation of the cross-sectional distribution's law of motion in the proximal point algorithms.

One can argue that the Euler equation error computed from individual consumption can result in a distorted picture because the number of agents grouped by different start capitals are weighed equally. However, there will be fewer agents with high start capital than agents with medium start capitals depending on the cross-sectional distribution. A remedy is to compute the Euler equation errors w.r.t. aggregate consumption, i.e., taking the relative difference of the mean of the cross-sectional distribution of individual consumption implied by the algorithms and the mean implied by the Euler equation. The summary statistics of these errors are displayed in Table 2, Panel B, and the corresponding boxplots are shown in Figure 6.

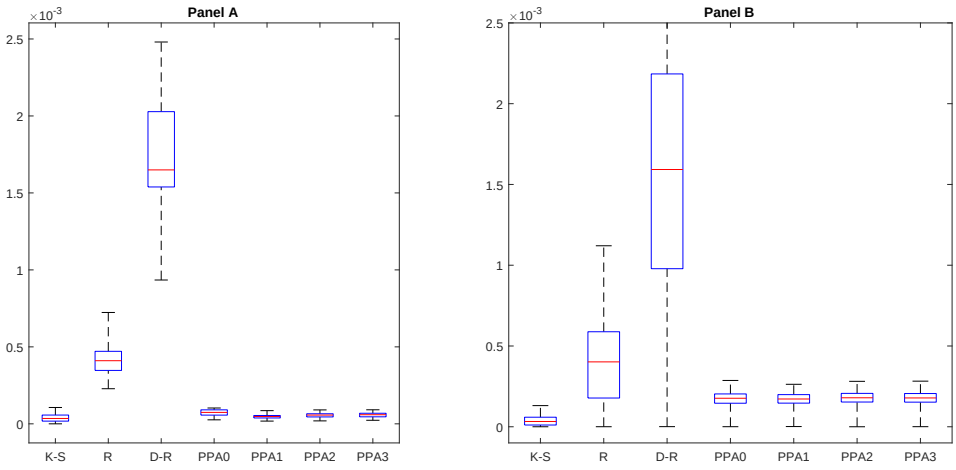


FIGURE 6. **Boxplots of the Euler equation error distributions of aggregate capital.** Panel A shows the standard Euler equation error and Panel B displays the dynamic Euler equation error for the numerical solutions from Table 1. The error is computed from aggregate consumption implied by the algorithms versus the Euler equation implied aggregate consumption for a random sample of aggregate shocks over $N = 3587$ periods. The initial cross-sectional distribution of start capital is the same for all algorithms. The red lines mark the medians whereas the blue boxes denote the $25^{\text{th}}$ to $75^{\text{th}}$ percentiles. The whiskers indicate the range of the distribution and the red dots outside are outliers.

Interestingly, the Krusell-Smith algorithm's performance is now closer to the proximal point algorithm than it is to the Reiter and den Haan-Rendahl algorithms, which perform much worse. It means that its large Euler equation errors

for individual capital occur in regions of the cross-sectional distribution with little mass. All in all, we see that the proximal point algorithm yields much higher precision in terms of Euler equation errors both for the individual agents as well as for aggregate variables.

### 5.2. *Order of Truncation*

An important question is which order of truncation of the polynomial chaos expansion will yield sufficient precision for our algorithm. To investigate that, we look again at the Euler equation errors for individual and aggregate consumption in Figure 7. One can see that there is a significant shift of the whole error

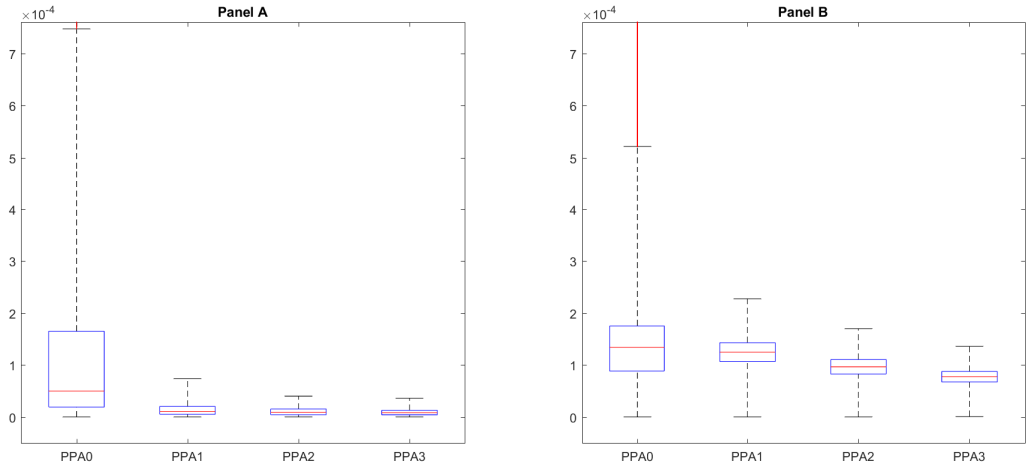


FIGURE 7. **Boxplots of the Euler equation error distributions for different orders of truncation for proximal point algorithm.** Panel A shows the standard Euler equation error for individual capital whereas Panel B displays the dynamic Euler equation error for individual capital for the different proximal point algorithms in Table 1. The errors are computed for a random sample of aggregate shocks over $N = 3587$ periods. The initial cross-sectional distribution of start capital is the same for all algorithms. The red lines mark the medians whereas the blue boxes denote the $25^{\text{th}}$ to $75^{\text{th}}$ percentiles. The whiskers indicate the range of the distribution and the red dots outside are outliers.

distribution towards zero from order zero to one. With higher orders than one, the error distribution for individual capital stays almost the same. It narrows further to a smaller extend for the dynamic error on individual consumption but does not change substantially. Therefore, a polynomial chaos expansion of solely first order seems to be enough to produce a satisfactory precision for the growth model at hand. Recall that order zero implies that the optimal policies depend on the aggregate capital of the employed and unemployed. Order one and higher,

however, imply a dependance on the full approximated distribution. Therefore, to approximate the rational expectations equilibrium of the growth model sufficiently, the agents need to know more than the aggregate capital. However, a crude approximation of the cross-sectional distribution seems to be enough.

### 5.3. *Economic Implications*

Let me now compare the economic implications of the different numerical solutions. As the Euler equation errors for the Reiter and den Haan-Rendahl algorithm were much worse, especially in the aggregates, I will compare the proximal point algorithm only to the Krusell-Smith algorithm. The largest conceptual difference between these two types of algorithms is that the Krusell-Smith method assumes bounded rationality in terms of a "rule of thumb", i.e., a parametric law of motion for the aggregate variables depending on a finite number of moments. Our algorithm, however, just uses the nonparametric law of motion of the aggregate variables.

To compare the implications of these conceptual differences, we look at the cross-sectional distributions, which the different numerical solutions produce on average. We cannot compute the full stationary state distribution for this model though, since this is a distribution of distributions $\mathcal{P}(z', k, \mu)$. However, we can look at the expected conditional cross-sectional distribution $\mathbb{E}^{\mu}(\mathcal{P}(z', k|\mu))$, which is essentially the average stationary cross-sectional distribution. It is computed as a fixed point of the cross-sectional distribution's law of motion and displayed in Figure 8. Similar to the Euler equation errors, one can see a significant change of the distribution of the proximal point algorithm with order zero compared to all higher orders producing distributions, which are indistinguishable. This confirms that higher orders of truncation of the polynomial chaos expansion than one do not have a high impact on the approximate solution of the growth model but it is crucial to use an approximation of the whole distribution rather than only the mean. Economically, this means that even for this simple growth model, there is no aggregation in the approximate fully rational expectations equilibrium. This stands in stark contrast to the aggregation result by Krusell and Smith (1998) for the approximate bounded rational expectations equilibrium. They find that the solutions do not differ if one adds more moments to their parametric law of motion.

A second economic implication is that the distribution by the Krusell-Smith algorithm has a much thicker tail to the right than the distributions of the proximal point algorithms. The corresponding Lorenz curves also confirm that the inequality

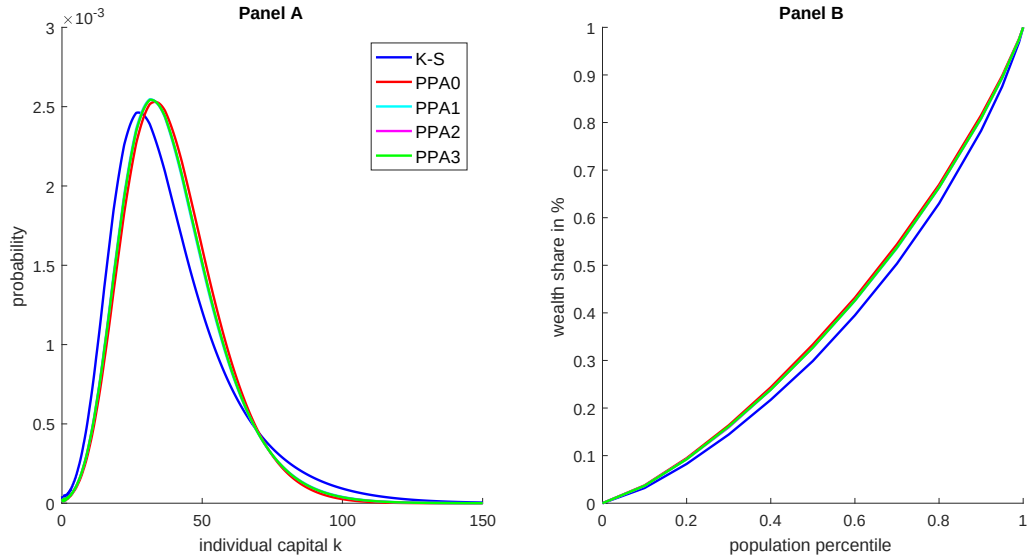


FIGURE 8. **Average cross-sectional distributions produced by by the Krusell-Smith algorithm and the proximal point algorithms from Table 1.** This graph displays the average stationary cross-sectional distribution, i.e., the expectation of the stationary distribution of the state space $(z', k, \mu)$ conditional on $\mu$. Panel A shows the p.d.f.s with a histogram approximation with bin size 0.1 whereas Panel B displays the corresponding Lorenz curves.

in the economies resulting from the proximal point algorithms is less than in the Krusell-Smith economy. This implies that the bounded rationality assumption introduces more inequality.

## 6. CONCLUSIONS

In this paper, I develop a novel global solution algorithm for DSGE models with ex-post heterogeneity and aggregate risk, which is purely based on projection methods. What sets this algorithm apart from existing methods is that, rather than approximating the law of motion of aggregate variables with a more or less parametric formula, it approximates the cross-sectional distribution of individual variables. I use generalized polynomial chaos expansions to do so. This projection technique essentially extends orthogonal polynomial projection from spaces of smooth functions to the space of square-integrable random variables.

Furthermore, the algorithm in this paper extends the proximal point algorithm for convex saddle point problems in a way, which preserves its convergence properties. To go beyond the convergence of the iteration on the theoretical operator, I also derive approximation error bounds concerning the truncation of the polynomial chaos expansion and the interpolation of the optimal policies. This ensures that the approximate solution is indeed close to the rational expectations equilib-

rium.

By taking this conceptually quite different approach with a sound theoretical underpinning, we are able to increase the precision of the solution in terms of its Euler equation errors by a significant amount. Additionally, we obtain the surprising result that, even for the comparatively simple growth model from Krusell and Smith (1998), we do not obtain aggregation for our approximate fully rational expectations equilibrium. The agents need to know more than the aggregate capital in order to have a sufficiently precise approximation of the optimal policies. This contrasts the aggregation result by Krusell and Smith (1998) for their approximate bounded rational expectations equilibrium.

Overall, my approach provides a new tool to analyze numerical solutions of DSGE models with ex-post heterogeneity and nonlinearities. This type of models is crucial for analyzing wealth effects. An interesting application are macro-finance models, which investigate systemic risk in financial markets and its effect on the real economy. Future research should, hence, be aimed at extending the algorithm in this paper to accommodate more complex models, e.g., models with occasionally binding constraints or a higher number of variables. In terms of the former, one needs to ensure convexity of the overall problem, which actually means that nonconvexity in some parts of the problem might be feasible. In terms of the latter, one can explore the integration of sparse grids (see e.g., Brumm and Scheidegger, 2016).

## APPENDIX A: PROOFS

### A.1. *Proof of Proposition 6*

PROOF: Let us denote the support of the marginal distribution w.r.t. $k$ of the cross-sectional distribution by $\operatorname{supp}\mu^k$. The minimum value of $k$, which has positive probability, is denoted by $\underline{k} = \min_k \operatorname{supp}\mu^k$. First, let us show that the constraint has positive probability $\delta \in \operatorname{supp}\mu^k$. Because of $p^{\hat{z}|\hat{z}} > 0$, eventually we have $\hat{z}^{ag}$ in the previous and the current period. Suppose that the start capital, at which the constraint starts binding, is not in the support $k^* < \underline{k} \leq \hat{k}$. Applying the optimal capital savings function, we obtain that $\underline{k}' = h_2(\hat{z}, \underline{k}, \mu) \leq \underline{k}$. By induction, this contradicts $k^* \notin \operatorname{supp}\mu^k$. Now let us show that there is a mass point at $\delta$. Assume that $\delta < k^* = \underline{k}$. Because $h_2$ is continuous and strictly increasing to the right of its kink, there exists an interval $[k^*, \bar{k}]$ with $\bar{k} := \max\{k \geq \delta \,|\, h_2(\hat{z}, k, \mu) = k^*\} > k^*$ and positive measure $\mu^k([k^*, \bar{k}]) > 0$. Due to $p^{\hat{z}|\hat{z}} > 0$, a strictly positive part of this mass will stay at $\hat{z}$ and have future value $\delta$. Hence, $\mu^k(\delta) > 0$ and $\underline{k} = \delta$. This yields the mass point at the constraint for the cross-sectional distribution. Using

the same reasoning, one can easily see that this mass point at zero propagates to higher levels of individual capital at $\bar{z} \in \mathcal{Z}$. *Q.E.D.*

## A.2. *Proof of Theorem 7*

In order to proof Theorem 7, we need the following preliminary result.

PROPOSITION 10 (Admissible Set of the Growth Model) *Consider the growth model from Section 2. Define a subspace $\mathcal{H}$ of the intersection of the square-integrable functions w.r.t. $\mu \in L^2$, i.e., $L^2\left(\mathcal{Z}^{id} \times \mathbb{R}, \mathcal{B}(\mathcal{Z}^{id} \times \mathbb{R}), \mu\right)$, and the functions with bounded first and second variation*[11] *such that for any element $h = (h_1, h_2) \in \mathcal{H}$, the following inequalities are satisfied almost surely for any $z, z' \in \mathcal{Z}$ and $(z^{id}, \kappa) \sim \mu$*

*(i) Nonnegative consumption: $h_1(z', \kappa, \mu) \geq 0$*

*(ii) Limited capital savings: $h_2(z', \kappa, \mu) \leq \mathbf{I}(z', \kappa) + [1 - \rho]\kappa$*

*(iii) Lower bound on the average second variation of capital savings:*

$$\sum_{z' \in \mathcal{Z}} \mathbf{P}(z', z, \kappa)\, \delta^2 h_2(z', \kappa, \mu; \tilde{\kappa}) \geq \sum_{z' \in \mathcal{Z}} \mathbf{P}(z', z, \kappa) \left[ \delta^2 \mathbf{I}(z', \kappa; \tilde{\kappa}) - \gamma \frac{[\delta \mathbf{C}(z', \kappa; \tilde{\kappa})]^2}{\mathbf{C}(z', \kappa)} \right]$$

*where $\mathbf{P}$ denotes the probability operator, $\mathbf{C}$ denotes the consumption operator and $\mathbf{I}$ denotes the income operator defined as follows*

$$\begin{aligned}
\mathbf{P}(z', z, \kappa) &= \frac{p^{z'|z} \mathbf{C}(z', \kappa)^{-\gamma}}{\sum_{z' \in \mathcal{Z}} p^{z'|z} \mathbf{C}(z', \kappa)^{-\gamma}} \\
\mathbf{C}(z', \kappa) &= \mathbf{I}(z', \kappa) + [1 - \rho]\kappa - h_2(z', \kappa, \mu) \\
\mathbf{I}(z', \kappa) &= \alpha \tilde{a}(z^{ag'}) \left[ \frac{\langle \kappa, \mathbf{1} \rangle}{\pi p^{e'}} \right]^{\alpha - 1} \kappa + [1 - \alpha] \tilde{a}(z^{ag'}) \tilde{\nu}(z') \left[ \frac{\langle \kappa, \mathbf{1} \rangle}{\pi p^{e'}} \right]^{\alpha}
\end{aligned}$$

*with*

$$\begin{aligned}
\tilde{a}(z^{ag'}) &= 1 + z^{ag'} a - [1 - z^{ag'}] a \\
\tilde{\nu}(z') &= \nu + z^{id'} \left[ 1 - \frac{\nu}{p^{e'}} \right].
\end{aligned}$$

*Then, $\mathcal{H}$ is a Hilbert space.*

PROOF: It is well known that the subspace of functions with bounded variation

[11] $n^{\text{th}}$ variation (see e.g., Zeidler, 1986b): Let $\mathcal{E}$ be a Hilbert space. The $n^{\text{th}}$ variation of an operator $\mathbf{T} : \mathcal{E} \to \mathcal{E}$ at a point $e \in \mathcal{E}$ in the direction $\tilde{e} \in \mathcal{E}$ is defined by $\delta^n \mathbf{T}(e; \tilde{e}) = d^n / dt^n \mathbf{T}(e + t\tilde{e})|_{t=0}$.

within $L^2$ is complete and hence, a Hilbert space itself. With conditions $(i)-(iii)$, we take yet another subset of functions with bounded variations. It is easy to see that any limiting element $h^*$ of a Cauchy sequence $h^n \in \mathcal{H}$, $n \in \{1, 2, \ldots\}$, satisfies conditions $(i)-(iii)$ as well. The subspace $\mathcal{H}$ is therefore complete and a Hilbert space itself. *Q.E.D.*

REMARK Note that condition $(iii)$ implies that the expected second variation of capital savings in a nonnegative direction $\tilde{\kappa} \geq 0$ is greater equal a nonpositive threshold if $K \geq ([2-\alpha]\pi p^{e'})/([1-\alpha]\nu)$. In our calibration, this translates into $K \geq 18.2222$, which seems to be a reasonable minimum value for aggregate capital. Due to this fact, additionally to convex capital savings functions, the subspace also allows for capital savings functions with concave sections.

PROOF OF THEOREM 7: It suffices to show that the Lagrangian (14) of the agents' optimization problem (7) satisfies the conditions of Corollary 4. This guarantees the nonexpansiveness of the corresponding resolvent and therefore implies convergence of the (accelerated) proximal point algorithm.

**Saddle function:** Let us start by specifying the Hilbert spaces $\mathcal{C} \times \mathcal{D}$ the Lagrangian is defined on. $L$ depends on the optimal controls $h = (h_1, h_2)$. Hence, we define the first Hilbert space by $\mathcal{H}$ as given in Proposition 10. The Lagrange multipliers lie in the corresponding dual space, which implies $\mathcal{D} \subseteq L^2(\mathcal{Z}^{id} \times \mathbb{R}, \mathcal{B}(\mathcal{Z}^{id} \times \mathbb{R}), \mu)$, such that $y_2, y_3 \geq 0$ for any exogenous shock and start capital.

Now it remains to show that the Lagrangian is convex in the optimal controls and concave in the Lagrange multipliers. The latter is trivial as the Lagrangian is linear in the multipliers. The former means that the Hessian (in terms of second variations) w.r.t. $h_1$ and $h_2$ is positive semidefinite (see e.g., Zeidler, 1986b, Corollary 42.8). As the cross-variation is zero and the second variation of the Lagrangian w.r.t. $h_1$ is nonnegative, we need to show that the second variation of the operator

$$\mathbf{U}(h_2) = \sum_{z'' \in \mathcal{Z}} p^{z''|z'} \beta u \left( \mathbf{I}\left(z'', h_2\right) + [1-\rho]\, h_2 - h_2' \right)$$

is nonpositive. Defining $\mathbf{C}\left(z'', h_2\right) = \mathbf{I}\left(z'', h_2\right) + [1-\rho]h_2 - h_2'$, we have

$$\begin{aligned} \delta^2 \mathbf{U}\left(h_2; \tilde{h}\right) = \sum_{z'' \in \mathcal{Z}} p^{z''|z'} \beta \Big\{ & u_{cc}\left(\mathbf{C}\left(z'', h_2\right)\right) \left[\delta \mathbf{C}\left(z'', h_2; \tilde{h}\right)\right]^2 \\ & + u_c\left(\mathbf{C}\left(z'', h_2\right)\right) \delta^2 \mathbf{C}\left(z'', h_2; \tilde{h}\right) \Big\} \end{aligned}$$

and therefore convexity is achieved by a next-period consumption choice with second variation bounded from above

$$\sum_{z''\in\mathcal{Z}} p^{z''|z'}\beta\mathbf{C}\left(z'',h_2\right)^{-\gamma}\delta^2\mathbf{C}\left(z'',h_2;\tilde{h}\right)\leq\gamma\sum_{z''\in\mathcal{Z}} p^{z''|z'}\beta\mathbf{C}\left(z'',h_2\right)^{-\gamma}\frac{\left[\delta\mathbf{C}\left(z'',h_2;\tilde{h}\right)\right]^2}{\mathbf{C}\left(z'',h_2\right)}.$$

This condition is equivalent to our condition $(iii)$ of bounded second variation for $h\in\mathcal{H}$ and is therefore satisfied. Applying the resolvent corresponding to the Lagrangian amounts to finding the minimum of the augmented Lagrangian (8). Any such minimum stays in $\mathcal{H}$ (see e.g., Bauschke and Combettes, 2011, Proposition 23.2), which ensures convexity throughout the algorithm.

**Properness:** We proceed in two steps. First we show that there exist Lagrange multipliers $y=(y_1,y_2,y_3)\in\mathcal{D}$ such that $L(h,y)>-\infty$ for all $h\in\mathcal{H}$. Let $y=0$, then $L(h,y)>-\infty$ by definition because any $h\in\mathcal{H}$ results in a number on the real line including $+\infty$. Secondly, there exists a capital savings and consumption choice $h=(h_1,h_2)\in\mathcal{H}$ such that $L(h,y)<\infty$ for all $y\in\mathcal{D}$. For any $\mu$ with aggregate capital $K>0$, the productive income for any agent is positive $I>0$. Fix a constant $0<\epsilon<I$ and set $h_1=\epsilon$ for any exogenous shock and start capital. Set capital savings according to the budget constraint. Then, the value of $y_1$ does not change the value of the Lagrangian. The nonnegativity of the other two multipliers ensures $L(h,y)<\infty$.

**Semicontinuity:** What is missing to conclude, is the continuity property of the Lagrangian in the policies $h\in\mathcal{H}$ as well as in the Lagrange multipliers $y\in\mathcal{D}$, which simply follows from the definition of the Lagrangian. *Q.E.D.*

## A.3. *Proof of Proposition 8*

PROOF: Because $\max_{h\in\mathcal{H}}(h^{n+1}-h)=h^{n+1}\leq X(z',k,K)$, Equation (15) implies that

$$\langle h^{n+1}-h,\nabla\tilde{L}\left(h^{n+1},y^n;h^n\right)-v\rangle\leq\frac{\epsilon^2}{2\lambda},\ \forall h\in\mathcal{H},$$

where

$$\begin{aligned}\tilde{L}\left(h^{n+1},y^n;h^n\right)=&L^A\left(h^{n+1},y^n;h^n\right)-\frac{1}{2\lambda}\left(h_1-h_1^n\right)^2-\frac{1}{2\lambda}\left(h_2-h_2^n\right)^2\\ v=&\frac{1}{\lambda}\left(h^n-h^{n+1}\right).\end{aligned}$$

Adding a zero and applying the definition of the gradient then implies

$$\begin{aligned}
&\left[\tilde{L}\left(h^{n+1}, y^n; h^n\right) - \tilde{L}\left(h, y^n; h^n\right)\right] - \left[\tilde{L}\left(h^{n+1}, y^n; h^n\right) - \tilde{L}\left(h, y^n; h^n\right)\right] \\
&+ \langle h - h^{n+1}, v - \nabla \tilde{L}\left(h^{n+1}, y^n; h^n\right)\rangle \leq \frac{\epsilon^2}{2\lambda} \\
\Rightarrow &\begin{cases} \tilde{L}\left(h^{n+1}, y^n; h^n\right) - \tilde{L}\left(h, y^n; h^n\right) + \langle h - h^{n+1}, v\rangle \leq \frac{\epsilon^2}{2\lambda} \\ \tilde{L}\left(h^{n+1}, y^n; h^n\right) - \tilde{L}\left(h, y^n; h^n\right) + \langle h - h^{n+1}, \nabla \tilde{L}\left(h^{n+1}, y^n; h^n\right)\rangle \leq 0 \end{cases}
\end{aligned}$$

for all $h \in \mathcal{H}$. Note that $\tilde{L}(., y^n; h^n) = L(., y^{n+1})$ with $L$ as in (14). Therefore, we have that $v \in \mathbf{T}_L^{\epsilon^2/(2\lambda)}(h^{n+1})$, which leads to $h^{n+1} \in \left(\mathbf{Id} + \lambda \mathbf{T}_L^{\epsilon^2/(2\lambda)}\right)^{-1}(h^n)$ and concludes the proof. *Q.E.D.*

## A.4. *Proof of Theorem 9*

In order to proof Theorem 9, we first need to establish that the optimal policies $h_2$ and $h_1$ as computed in the proximal point algorithm are analytic in the basic random variables $\xi^0$ and $\xi^1$. It is clear from equation (12) that the optimal policies depend on the basic random variables through the projection coefficients $\varphi_i^j$, $i \in \{0, 1, \ldots\}$, $j \in \{0, 1\}$.

PROPOSITION 11 (Analytic policies) *Under the assumptions of Theorem 9, all iterates of the consumption policy, the savings policy and the Lagrange multipliers as functions of $\xi^j$ admit analytic extensions in the complex plane, namely in the region $\Sigma(\tau^j, \Gamma^j)$ given in Theorem 9. Furthermore, it holds that*

$$\max_{x \in \Sigma(\tau^j, \Gamma^j)} |f(x)| \leq \max_f \left|\mathbf{c}^{\mathbf{f},\mathbf{j}}\right| \frac{A}{A - \tau}$$

*where $f$ is a handle for the policies and Lagrange multipliers.*

PROOF: The proof proceeds in two steps. First, we establish that all iterates of the policies and hence, Lagrange multipliers are real analytic functions of the basic random variables. Secondly, we construct the complex analytic extension.

**Real analytic:** Equation (15) implies that the $(n+1)$-th iterate of the consumption policy $h_1^{n+1}$ and the savings policy$h_2^{n+1}$ in the proximal point algorithm

solve the following first-order conditions

$$X\frac{\partial L^A}{\partial h_1^{n+1}} = \mathbf{e}_1 \tag{18}$$

$$X\frac{\partial L^A}{\partial h_2^{n+1}} = \mathbf{e}_2 \tag{19}$$

with constants $\|\mathbf{e}_1\|, \|\mathbf{e}_2\| \leq \frac{\epsilon^2}{2\lambda}$ for any fixed exogenous shock and start capital $(z', k)$ . Now let us take the derivatives of the first-order conditions (18) and (19) w.r.t. $\xi^0$ and $\xi^1$. It is obvious that $X$, $\mathbf{e}_1$ and $\mathbf{e}_2$ do not depend on the basic random variables. The partial derivatives of the augmented Lagrangian, however, do because the optimal policies and hence, also the Lagrange multipliers depend on $\xi^0$ and $\xi^1$. Additionally, the next-period aggregate capital $K'$ depends on the basic random variables, which is due to (6). Using polynomial chaos as in (13), the next-period aggregate capital can be written as

$$K' = \sum_{g,j=0}^{1} \frac{p^{(z^{ag'},g)|(z^{ag},j)}}{p^{z^{ag'}|z^{ag}}} \int_{-\infty}^{\infty} h_2^{n+1}\left(z^{ag'}, g, \sum_{i=0}^{\infty}\varphi_i^j\Phi_i^j\left(\xi^j\right), \{\varphi_{i_0}^0, \varphi_{i_1}^1\}_{i_0,i_1=0}^{\infty}\right) dF^j\left(\xi^j\right).$$

The derivatives of the first-order conditions w.r.t. $\xi^0$ and $\xi^1$ hence satisfy

$$\begin{aligned}
\underbrace{\frac{\partial^2 L^A}{[\partial h_1^{n+1}]^2}}_{=:L_1^A}\frac{\partial h_1^{n+1}}{\partial \xi^j} + \lambda\frac{\partial h_2^{n+1}}{\partial \xi^j} &= \frac{\partial}{\partial \xi^j}\underbrace{\left(\frac{1}{\lambda}h_1^n + y_1^n + \left\{h_1^{n+1}\leq y_2^n/\lambda\right\}y_2^n\right)}_{=:g_1} \\
\lambda\frac{\partial h_1^{n+1}}{\partial \xi^j} + \underbrace{\frac{\partial^2 L^A}{[\partial h_2^{n+1}]^2}}_{=:L_2^A}\frac{\partial h_2^{n+1}}{\partial \xi^j} &= \frac{\partial}{\partial \xi^j}\underbrace{\left(\frac{1}{\lambda}h_2^n + y_1^n + \left\{h_2^{n+1}\leq y_3^n/\lambda\right\}y_3^n\right)}_{=:g_2} \\
&\quad - \underbrace{\frac{\partial^2 L^A}{\partial h_2^{n+1}\partial K'}}_{=:L_{K'}^A}\frac{\partial K'}{\partial \xi^j},
\end{aligned}$$

where $j \in \{0, 1\}$. This system of equations is easily solved:

$$\begin{aligned}
\frac{\partial h_1^{n+1}}{\partial \xi^j} &= \frac{1}{L_1^A L_2^A - \lambda^2}\left(L_2^A\frac{\partial g_1}{\partial \xi^j} - \lambda\frac{\partial g_2}{\partial \xi^j} + \lambda L_{K'}^A\frac{\partial K'}{\partial \xi^j}\right) \\
\frac{\partial h_2^{n+1}}{\partial \xi^j} &= \underbrace{\frac{1}{L_1^A L_2^A - \lambda^2}}_{=:d}\left(L_1^A\frac{\partial g_2}{\partial \xi^j} - L_1^A L_{K'}^A\frac{\partial K'}{\partial \xi^j} - \lambda\frac{\partial g_1}{\partial \xi^j}\right).
\end{aligned}$$

Note that $\partial K'/\partial \xi^j$ is well defined since our admissible set from Proposition 10 requires $h_2$ to have bounded first variation and the polynomial chaos expansion

$\sum_{i=0}^{\infty} \varphi_i^j \Phi_i^j$ is real analytic. Now, applying conditions (16) and (17) results in

$$
\begin{aligned}
(20) \qquad \left\| \frac{\partial h_1^{n+1}}{\partial \xi^j} \right\| \;\leq\; & \frac{1}{d} \left\{ \frac{L_2^A}{\lambda} \mathbf{c}^{h_1^n,j} + \mathbf{c}^{h_2^n,j} + \left(L_2^A + \lambda\right) \mathbf{c}^{y_1^n,j} + L_2^A \mathbf{c}^{y_2^n,j} + \lambda \mathbf{c}^{y_3^n,j} \right. \\
& \left. + \lambda L_{K'}^A \mathbf{c}^{\Phi,j} \right\} 1! \\
(21) \qquad \left\| \frac{\partial h_2^{n+1}}{\partial \xi^j} \right\| \;\leq\; & \frac{1}{d} \left\{ \mathbf{c}^{h_1^n,j} + \frac{L_1^A}{\lambda} \mathbf{c}^{h_2^n,j} + \left(L_1^A + \lambda\right) \mathbf{c}^{y_1^n,j} + \lambda \mathbf{c}^{y_2^n,j} + L_1^A \mathbf{c}^{y_3^n,j} \right. \\
& \left. + L_1^A L_{K'}^A \mathbf{c}^{\Phi,j} \right\} 1!
\end{aligned}
$$

Furthermore, by taking the derivative of the Lagrange multipliers, defined in line 5 of Algorithm 1 or line 8 of Algorithm 2, and applying condition (17) together with (20) and (21), we obtain

$$
\begin{aligned}
(22) \left\| \frac{\partial y_1^{n+1}}{\partial \xi^j} \right\| \;\leq\; & \frac{\lambda}{d} \left\{ \frac{L_2^A + \lambda}{\lambda} \mathbf{c}^{h_1^n,j} + \frac{L_1^A + \lambda}{\lambda} \mathbf{c}^{h_2^n,j} + \left( \frac{d}{\lambda} + L_1^A + L_2^A + 2\lambda \right) \mathbf{c}^{y_1^n,j} \right. \\
& \left. + \left(L_2^A + \lambda\right) \mathbf{c}^{y_2^n,j} + \left(L_1^A + \lambda\right) \mathbf{c}^{y_3^n,j} + \left(L_1^A + \lambda\right) L_{K'}^A \mathbf{c}^{\Phi,j} \right\} 1! \\
(23) \left\| \frac{\partial y_2^{n+1}}{\partial \xi^j} \right\| \;\leq\; & \frac{\lambda}{d} \left\{ \frac{L_2^A}{\lambda} \mathbf{c}^{h_1^n,j} + \mathbf{c}^{h_2^n,j} + \left(L_2^A + \lambda\right) \mathbf{c}^{y_1^n,j} + \left( \frac{d}{\lambda} + L_2^A \right) \mathbf{c}^{y_2^n,j} \right. \\
& \left. + \lambda \mathbf{c}^{y_3^n,j} + \lambda L_{K'}^A \mathbf{c}^{\Phi,j} \right\} 1! \\
(24) \left\| \frac{\partial y_3^{n+1}}{\partial \xi^j} \right\| \;\leq\; & \frac{\lambda}{d} \left\{ \mathbf{c}^{h_1^n,j} + \frac{L_1^A}{\lambda} \mathbf{c}^{h_2^n,j} + \left(L_1^A + \lambda\right) \mathbf{c}^{y_1^n,j} + \lambda \mathbf{c}^{y_2^n,j} + \left( \frac{d}{\lambda} + L_1^A \right) \mathbf{c}^{y_3^n,j} \right. \\
& \left. + L_1^A L_{K'}^A \mathbf{c}^{\Phi,j} \right\} 1!.
\end{aligned}
$$

The equations for higher-order derivatives are analogous. Therefore, as condition (17) is equivalent with $f$ being analytic in $\xi^0$ and $\xi^1$, the iterates in the proximal point algorithm stay analytic in $\xi^0$ and $\xi^1$ on the whole domain except at the points $h_1^{n+1} = y_2^n/\lambda$ and $h_2^{n+1} = y_3^n/\lambda$ due to the inequality constraints. As we have flexibility on the error $(\mathbf{e}_1, \mathbf{e}_2)$ in the first-order conditions (18) and (19), however, we can always achieve $h_1^{n+1} \neq y_2^n/\lambda$ and $h_2^{n+1} \neq y_3^n/\lambda$ so that there are no singularities. That all iterates are real analytic follows by induction.

**Complex continuation:** We define the following power series for the $(n+1)$-th iterate on the complex plane

$$
f(x) = \sum_{p=0}^{\infty} \frac{(x - \xi^j)^p}{p!} \frac{\partial^p}{[\partial \xi^j]^p} f\left(\xi^j\right),
$$

where $f = [h_1 \quad h_2 \quad y_1 \quad y_2 \quad y_3 \quad \kappa^j]^T$, where $\kappa^j$ denotes the polynomial chaos expansion of the cross-sectional distribution. Note that we dropped the iteration count superscript for notational ease here. Taking norms and exploiting equations

(20)-(24) leads to

$$|f_i(x)| \leq \sum_{p=0}^{\infty} \left\| \left(x - \xi^j\right) \mathbf{Id} \right\|_\infty^p \|\mathbf{A}\|_\infty^p \max_{i=1,\ldots,5} \left|\mathbf{c}^{\mathbf{f_i,j}}\right|$$

$$\leq \max_{i=1,\ldots,6} \left|\mathbf{c}^{\mathbf{f_i,j}}\right| \sum_{p=0}^{\infty} \left|x - \xi^j\right|^p \|\mathbf{A}\|_\infty^p, \; i \in \{1,\ldots,5\}, \tag{25}$$

where **Id** denotes the identity matrix and the matrix **A** directly results from equations (20)-(24) and (16)

(26)

$$\mathbf{A} = \frac{\lambda}{d} \begin{bmatrix} \frac{L_2^A}{\lambda^2} & \frac{1}{\lambda} & \frac{L_2^A+\lambda}{\lambda} & \frac{L_2^A}{\lambda} & 1 & L_{K'}^A \\ \frac{1}{\lambda} & \frac{L_1^A}{\lambda^2} & \frac{L_1^A+\lambda}{\lambda} & 1 & \frac{L_1^A}{\lambda} & \frac{L_1^A L_{K'}^A}{\lambda} \\ \frac{L_2^A+\lambda}{\lambda} & \frac{L_1^A+\lambda}{\lambda} & \frac{d}{\lambda} + L_1^A + L_2^A + 2\lambda & L_2^A + \lambda & L_1^A + \lambda & \left(L_1^A + \lambda\right) L_{K'}^A \\ \frac{L_2^A}{\lambda} & 1 & L_2^A + \lambda & \frac{d}{\lambda} + L_2^A & \lambda & \lambda L_{K'}^A \\ 1 & \frac{L_1^A}{\lambda} & L_1^A + \lambda & \lambda & \frac{d}{\lambda} + L_1^A & L_1^A L_{K'}^A \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that all entries of matrix **A** are bounded due to the assumptions in Proposition 10. The power series in (25) converges for $|x - \xi^j| \leq \tau^j < 1/\|\mathbf{A}\|_\infty$ such that

$$|f_i(x)| \leq \max_{i=1,\ldots,6} \left|\mathbf{c}^{\mathbf{f_i,j}}\right| \frac{1}{1 - \tau\|\mathbf{A}\|_\infty}.$$

Therefore, by continuation the iterates can be extended analytically in the whole region $\Sigma(\tau^j, \Gamma^j)$, which concludes the proof. *Q.E.D.*

Remark Note that we follow the proof of Theorem 4.1 in Babuška et al. (2007) with bounded range of the basic random variables for our proof of Theorem 9. We use bounded range since we choose a histogram approximation of the basic random variables. For other types of approximation, one might need to modify the error bound estimates to accommodate an unbounded range. We refer to Babuška et al. (2007) for that case.

Proof of Theorem 9: The last term of the bound is the interpolation error from tensor-product finite elements of order 1 on a rectangular discretization $D$. It is well established (see e.g., Brenner and Scott, 2007, Theorem 4.6.14). The error bound due to truncation of the polynomial chaos expansion is a little more involved. Due to the fact that the continuous functions of the basic random vari-

ables are a subset of the square-integrable functions, i.e., $C^0(\Gamma^j) \subset L^2(\Gamma^j)$, we have that the truncation error is bounded by the best approximation error (see Babuška et al., 2007, Lemma 4.3)

$$\left\|h - h^M\right\|_{L^2} \leq \mathbf{b} \inf_{w \in \mathcal{H}^M} \|h - w\|_{C^0},$$

where constant $\mathbf{b}$ is independent of the order of truncation $M$. Given that $h$ admits an analytic extension on the complex plane $\Sigma(\tau^j, \Gamma^j)$, the best approximation error is bounded by (see Babuška et al., 2007, Lemma 4.4)

$$\inf_{w \in \mathcal{H}^M} \|h - w\|_{C^0(\Gamma^j)} \leq \frac{2}{\eta^j - 1} e^{-M \log(\eta^j)} \max_{x \in \Sigma(\tau^j, \Gamma^j)} |h(x)|, \; j \in \{0, 1\},$$

where

$$\eta^j = \frac{2\tau^j}{|\Gamma^j|} + \sqrt{1 + \frac{4(\tau^j)^2}{|\Gamma^j|^2}} > 1.$$

Combining this with Proposition 11 and keeping in mind that we truncate once for the unemployed distribution and once for the employed distribution leads to the truncation error bound. *Q.E.D.*

## REFERENCES


Yves Achdou, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. Heterogeneous agent models in continuous time. *Working Paper*, 2014.

S. Rao Aiyagari. Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics*, 109(3):659–684, 1994.

Yann Algan, Olivier Allais, and Wouter J. Den Haan. Solving heterogeneous-agent models with parameterized cross-sectional distributions. *Journal of Economic Dynamics and Control*, 32 (3):875–908, 2008.

Yann Algan, Olivier Allais, and Wouter J. Den Haan. Solving the incomplete markets model with aggregate uncertainty using parameterized cross-sectional distributions. *Journal of Economic Dynamics and Control*, 34(1):59–68, 2010.

Ivo Babuška, Fabio Nobile, and Raúl Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 45(3):1005–1034, 2007.

Heinz H. Bauschke and Patrick L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics. Springer, 2011.

Truman Bewley. The permanent income hypothesis: A theoretical formulation. *Journal of Economic Theory*, 16(2):252–292, 1977.

Susanne Brenner and Ridgway Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer, 2007.

Johannes Brumm and Simon Scheidegger. Using adaptive sparse grids to solve high-dimensional dynamic models. *Working Paper, Available at SSRN 2349281*, 2016.

Patrick Cheridito and Juan Sagredo. Existence of sequential competitive equilibrium in Krusell-Smith type economies. *Working Paper, Available at SSRN 2794629*, 2016a.

Patrick Cheridito and Juan Sagredo. Comment on 'Competitive equilibria of economies with a continuum of consumers and aggregate shocks' by J. Miao. *Working Paper, Available at SSRN 2815971*, 2016b.

David Childers. On the solution and application of rational expectations models with function-valued states. *Working Paper*, 2015.

Wouter J. den Haan. Comparison of solutions to the incomplete markets model with aggregate uncertainty. *Journal of Economic Dynamics and Control*, 34(1):4–27, 2010.

Wouter J. den Haan and Pontus Rendahl. Solving the incomplete markets model with aggregate uncertainty using explicit aggregation. *Journal of Economic Dynamics and Control*, 34(1): 69–78, 2010.

Wouter J. den Haan, Kenneth L. Judd, and Michel Juillard. Computational suite of models with heterogeneous agents: Incomplete markets and aggregate uncertainty. *Journal of Economic Dynamics and Control*, 34(1):1–3, 2010.

Wouter J. den Haan, Kenneth L. Judd, and Michel Juillard. Computational suite of models with heterogeneous agents ii: Multi-country real business cycle models. *Journal of Economic Dynamics and Control*, 35(2):175–177, 2011.

Darrell Duffie and Wayne Shafer. Equilibrium in incomplete markets: I: A basic model of generic existence. *Journal of Mathematical Economics*, 14(3):285–300, 1985.

Darrell Duffie and Wayne Shafer. Equilibrium in incomplete markets: Ii: Generic existence in stochastic economies. *Journal of Mathematical Economics*, 15(3):199–216, 1986.

Darrell Duffie, John Geanakoplos, Andreu Mas-Colell, and Andrew McLennan. Stationary markov equilibria. *Econometrica*, pages 745–781, 1994.

Oliver G. Ernst, Antje Mugler, Hans-Jörg Starkloff, and Elisabeth Ullmann. On the convergence of generalized polynomial chaos expansions. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(02):317–339, 2012.

Walter Gautschi. On generating orthogonal polynomials. *SIAM Journal on Scientific and Statistical Computing*, 3(3):289–317, 1982.

Osman Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29(2):403–419, 1991.

Osman Güler. New proximal point algorithms for convex minimization. *SIAM Journal on Optimization*, 2(4):649–664, 1992.

Hugo A. Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60 (5):1127–1150, 1992.

Greg Kaplan, Benjamin Moll, and Giovanni L. Violante. Monetary policy according to HANK. *NBER Working Paper No. 21897*, 2016.

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

Felix Kubler and Karl Schmedders. Recursive equilibria in economies with incomplete markets. *Macroeconomic dynamics*, 6(02):284–306, 2002.

Lilia Maliar, Serguei Maliar, and Fernando Valli. Solving the incomplete markets model with aggregate uncertainty using the Krusell-Smith algorithm. *Journal of Economic Dynamics and*

*Control*, 34(1):42–49, 2010.

Thomas Mertens and Kenneth Judd. Equilibrium existence and approximation for incomplete market models with substantial heterogeneity. *Working Paper*, 2013.

Jianjun Miao. Competitive equilibria of economies with a continuum of consumers and aggregate shocks. *Journal of Economic Theory*, 128(1):274–298, 2006.

George J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Mathematical Journal*, 29(3):341–346, 1962.

Robert R. Phelps. Lectures on maximal monotone operators. *Extracta Mathematicae*, 12(3): 193–230, 1997.

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

Michael Reiter. Solving the incomplete markets model with aggregate uncertainty by backward induction. *Journal of Economic Dynamics and Control*, 34(1):28–35, 2010a.

Michael Reiter. Approximate and almost-exact aggregation in dynamic stochastic heterogeneous-agent models. *Economics Series Working Paper 258*, 2010b.

José-Víctor Ríos-Rull. Computation of equilibria in heterogeneous agent models. *Federal Reserve Bank of Minneapolis Staff Report 231*, pages 238–264, 1997.

R. Tyrrell Rockafellar. *Monotone operators associated with saddle-functions and minimax problems*, volume 18 of *Proceedings of Symposia in Pure Mathematics: Nonlinear Functional Analysis, Part 1*, pages 248–257. American Mathematical Society, 1970.

R. Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976a.

R. Tyrrell Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1(2):97–116, 1976b.

Saverio Salzo and Silvia Villa. Inexact and accelerated proximal point algorithms. *Journal of Convex Analysis*, 19(4):1167–1192, 2012.

Thomas Winberry. A toolbox for solving and estimating heterogeneous agent macro models. *Working Paper*, 2016.

Dongbin Xiu and George E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.

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

Eberhard Zeidler. *Nonlinear Functional Analysis and its Applications I. (Fixed Point Theorems).* Springer, 1986a.

Eberhard Zeidler. *Nonlinear Functional Analysis and its Applications III. (Variational Methods and Optimization).* Springer, 1986b.

Mengdi Zheng, Xiaoliang Wan, and George E. Karniadakis. Adaptive multi-element polynomial chaos with discrete measure: Algorithms and application to SPDEs. *Applied Numerical Mathematics*, 90:91–110, 2015.