

Job Market Paper

APPROXIMATING EQUILIBRIA WITH EX-POST HETEROGENEITY AND AGGREGATE RISK

ELISABETH PRÖHL

University of Geneva and Swiss Finance Institute

October 30, 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 have assumed bounded rationality in terms of a parametric law of motion of aggregate variables in order to reduce dimensionality. In this paper, we remove that assumption and compute a fully rational equilibrium dependent 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. We establish existence of the computed recursive equilibrium and provide theoretical convergence results. Economically, we find that idiosyncratic risk does not aggregate in our fully rational approximate equilibrium, which contrasts the well-known approximate aggregation result for the bounded rational approximate equilibrium by Krusell and Smith (1998).

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

JEL CLASSIFICATION: C62, 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, Victor Zhorin, and participants at the 6th Lindau Nobel Laureate Meeting, the NBER Summer Institute 2017, WEHIA 2017, the 2017 Workshop on Recent Developments in Numerical Methods with Applications in Statistics and Finance, SNDE 2017, the RGS Doctoral Conference 2017, ESEM 2016, EcoMod 2016, CEF 2016, ESNASM 2016, the MFM Summer Camp 2016, the 12th 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 "Discretizing the Infinite-Dimensional Space of Distributions to Approximate Markov Equilibria with Ex-Post Heterogeneity and Aggregate Risk".

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. Other examples of idiosyncratic risks are firm-specific productivity shocks in models of firm exit and entry as in [Hopenhayn \(1992\)](#). More recently, [Khan and Thomas \(2008\)](#) consider heterogeneity to produce a consistent investment rate distribution, whereas, [Kaplan et al. \(2017\)](#) achieve a realistic wealth distribution with heterogeneous households trading in two assets. In finance, [Constantinides and Duffie \(1996\)](#) show that risk premia increase when heterogeneity is introduced. [Storesletten et al. \(2007\)](#) find a moderate effect on the Sharpe ratio, but a significant negative impact on intergenerational risk sharing. Overall, there is plenty of evidence that idiosyncratic risks have a sizable impact on the economy.

Many of these models, however, do not feature aggregate risk to avoid the difficulties it poses for solving the model. 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 the realization of aggregate shocks. The aggregate variables evolve, in turn, depending on how the cross-sectional distribution changes. [Storesletten et al. \(2007\)](#) and [Khan and Thomas \(2008\)](#) do include aggregate risk, they resort to the Krusell-Smith algorithm to solve their models.

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 for 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: Solving for the optimal policies given a guess of parameters of the aggregate variables' law of motion, and secondly, estimating 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 moments higher 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 for 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 twofold. The first contribution is of a theoretical nature concerning the existence of a solution. Even though existence of a sequential equilibrium has been shown in [Cheridito and Sagredo \(2016a\)](#), the existence of simple recursive equilibria is still not fully established for the growth model with aggregate risk and a continuum of agents. [Miao \(2006\)](#) contains an argument for recursive equilibria which depend on the exogenous shocks, endogenous variables, cross-sectional distribution and the value function. [Brumm et al. \(2017\)](#) show existence of a simpler recursive equilibrium which does not require the value function. Their proof, however, does not apply to a continuum of agents with unbounded utility functions. This paper closes that gap by using an alternative fixed point argument. The existing literature relies on the compactness of the state space which requires bounded utility functions, whereas, this paper exploits the monotonicity properties of the model. To do so, the policy function is interpreted as a square-integrable function and hence, as a random variable itself rather than a real function.

Secondly, we construct a global solution algorithm for DSGE models with heterogeneous agents and aggregate risk, which does not assume bounded rationality and for which convergence is proven. Rather than assuming a parametric law of motion for the aggregate variables, we discretize the space of cross-sectional distributions. We use a projection method which extends the polynomial projection of

real functions to a projection of square-integrable random variables and hence, can be interpreted as a probabilistic polynomial projection. This technique is called generalized polynomial chaos. It has several advantages over standard polynomial projection using, for instance, Chebyshev polynomials. First of all, polynomial chaos does not require smooth functions and can therefore handle distributions with mass points. This is very relevant for our example of the Aiyagari-Bewley growth model which features mass points in the cross-sectional distribution due to the hard borrowing constraint. A further complication is that the location of these mass points is also endogenous such that one cannot simply treat them separately. Another advantage is that this probabilistic polynomial projection converges very fast. We find that a projection on polynomials up to the first order is enough to obtain a satisfactory precision of the solution. Hence, we only need two dimensions to sufficiently approximate the cross-sectional distribution in the growth model. Furthermore, by approximating the full distribution, 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.

When comparing the results of our algorithm to existing methods for the benchmark Aiyagari-Bewley growth model with aggregate risk, we find a significant improvement in precision for individual policies in terms of Euler equation errors as well as errors in the prediction of the law of motion of aggregate variables. Furthermore, there is a significant improvement in precision when truncating the polynomial chaos expansion in our algorithm at order two, rather than at order zero. Note that the latter leads to optimal policies, which depend solely 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 in this fully rational equilibrium. We do find a reason though, why [Krusell and Smith \(1998\)](#) obtained their approximate aggregation result. When computing the expected ergodic cross-sectional distributions, we see that they do not change significantly when the order of truncation increases. In another calibration of the model with high unemployment benefit, approximate aggregation does not persist. The Krusell-Smith algorithm produces an expected ergodic cross-sectional distribution with much fatter tails than our theoretically founded algorithm in this case. This suggests that bounded rationality can, in fact, be considered as one driver of inequality. A second interesting economic result arises when comparing the expected ergodic distributions resulting from our algorithm for the case of low and high unemployment benefit, respectively. We

find that the case of more redistribution, i.e. increased risk sharing, leads to fatter tails in the expected ergodic distribution. This means that more risk sharing leads to more systemic risk which is a similar result to the volatility paradox in [Brunnermeier and Sannikov \(2014\)](#).

This paper is related to several strands of literature. First of all, it is related to 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 first been 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 a sequential equilibrium. The question of existence of a recursive equilibrium with a simple state space is treated in [Brumm et al. \(2017\)](#), but only for finitely many agents.

Secondly, it is related to 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 aforementioned seminal algorithm by [Krusell and Smith \(1998\)](#) belong to the latter group. The algorithm by [Krusell and Smith \(1998\)](#) was also the subject of a special issue of the Journal of Economic Dynamics and Control in January 2010. This special issue presented various alternative algorithms, and compared 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 \(2010a\)](#) parameterize 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 rationality. 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 law of motion approach of [Algan et al. \(2008\)](#) with the perturbation of [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 is often only linear, or at most quadratic. Therefore, any higher-order nonlinear effects of aggregate shocks like risk are not accounted for. Secondly, as for all perturbation methods, the solutions are only accurate for small aggregate shocks. Crises scenarios consisting of a large aggregate shock or a long series of aggregate shocks in one direction cannot be analyzed with confidence.

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. [?](#) put forward a very interesting application of this methodology to monetary policy questions. However, their models incorporate only idiosyncratic shocks without 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 theoretical argument and algorithm. In Section 3, we introduce the existence argument. It is followed by a section on the corresponding iterative procedure for which we also show convergence. Section 5 explains the technique we use to discretize the space of distributions, and thus, enables use to keep track of the cross-sectional distribution. We also derive approximation error bounds. In Section 6, 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 and an extension of our iterative solution procedure.

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 $(z_t^{ag}, z_t^{id})_{t \geq 0}$ 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

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

where μ_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

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

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 μ_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 μ_{-1} ¹ with non-negative support, each agent adheres to a budget constraint, which equates individual consumption and current

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

capital stock to productive income and saved capital stock

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

The time line underlying this equation is clarified in Figure 1.² The parameters in

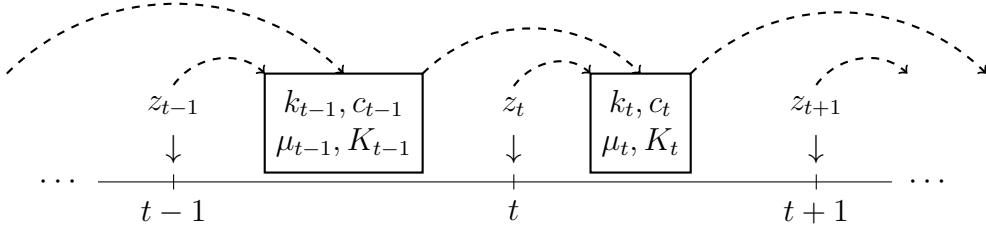


FIGURE 1. **Time line of events.** Before period t , the agent observes how much capital she saved in the previous period k_{t-1} and what the cross-sectional distribution of individual capital savings μ_{t-1} and hence, aggregate capital K_{t-1} is. At period t , the agent first observes the exogenous 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

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

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 proportional unemployment benefit νW 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

²Note 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.

derived from a Cobb-Douglas production function for the consumption good

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

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 π .

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 μ_{-1} with non-negative support, the individual optimization problem reads

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

where the productive income I is defined as in (3). To ensure that the Euler equation is sufficient for optimality, we assume that the following transversality condition holds.

ASSUMPTION 1 (Transversality condition) We suppose that the following condition is satisfied for all agents and any history of shocks in $[0, T]$

$$\lim_{T \rightarrow \infty} \mathbb{E} [\beta^{T+1} (1 - \rho + R(z_{T+1}^{ag}, K_T)) c_{T+1}^{-\gamma}] k_T = 0.$$

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. In this paper, we consider a particular competitive equilibrium of recursive form. 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 2 (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 μ_{-1} with non-negative support is called recursive if there exist functions $h_i : \mathcal{Z} \times \mathbb{R} \times \mathcal{P}(\mathcal{Z}^{id} \times \mathbb{R}) \rightarrow \mathbb{R}$, $i \in \{1, 2\}$, 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 μ 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 3 (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 μ , 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 μ to μ' . Given a fixed distribution μ 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 $\mu \rightarrow \tilde{\mu}' \rightarrow \mu'$. 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

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

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

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

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 establish existence of the recursive equilibrium and an iterative procedure to compute the equilibrium.

3. EXISTENCE OF RECURSIVE EQUILIBRIA FOR THE GROWTH MODEL

Existence of a solution to the Aiyagari-Bewley growth model with aggregate risk as proposed by [Krusell and Smith \(1998\)](#) has long been an open research question. It has first been examined by [Miao \(2006\)](#) whose idea is to show existence of a fixed point of the value function which directly depends on the cross-sectional distributions. A flaw in the theoretical argument that weak convergence of measures does not imply convergence of moments, however, has been pointed out in [Cheridito and Sagredo \(2016b\)](#). The same authors then derive existence of a sequential equilibrium in [Cheridito and Sagredo \(2016a\)](#) by first proving existence of a value function which depends on a series of fixed aggregates and in a second step proving existence of a fixed-point in aggregates which are consistent with the agent's optimal choices. Both works build on fixed-point theory relying on compactness. Therefore, the existence of sequential equilibria has only be shown for bounded utility functions. Furthermore, even though [Miao \(2006\)](#) extends his results to recursive equilibria depending on the current state of exogenous and endogenous variables, the distribution and the current value function, [Cheridito and Sagredo \(2016a\)](#) do not show existence of recursive equilibria. Another work which shows existence of simple recursive equilibria is [Brumm et al. \(2017\)](#). However, they restrict the growth model to bounded utility and finitely many agents. The existence of recursive equilibria in the form of [Krusell and Smith \(1998\)](#) given in Definition 2 is still an open question.

In this section, I extend the existence results to recursive equilibria according to Definition 2 for the growth model from Section 2 which has unbounded utility function and a continuum of agents. As is shown in [Stokey et al. \(1989\)](#), the extension to unbounded utility functions in the case of CRRA is typically done via constant returns to scale. However, due to the idiosyncratic shocks, there is a

disjunction between individual capital and the rental rate on capital as a function of aggregate capital within the budget constraint. Furthermore, each agent in the continuum has zero weight and cannot influence aggregates. Therefore, it can happen that individual capital grows substantially for an agent, but aggregate capital does not and the rental rate therefore, does not reduce. In terms of Stokey et al. (1989), this model thus falls into the category of unbounded returns. To establish existence, we have to rely on arguments of monotonicity because compactness cannot be proven.

In the following, an existence results is derived from the monotonicity properties of the individual optimization problem. The methodology proposed herein builds on the observation that the optimal policy function for capital savings denoted by h solves the Euler equation, which is equivalent to the first-order condition of the following constrained optimization problem

$$(7) \quad \begin{aligned} \min_{h \in \mathcal{H}} & -u(I(z', k, K) + [1 - \rho]k - h) \\ & - \sum_{z'' \in \mathcal{Z}} p^{z''|z'} \beta u(I(z'', h, K') + [1 - \rho]h - h') \\ \text{s.t. } & 0 \leq h. \end{aligned}$$

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

The admissible set \mathcal{H} has to be defined carefully to ensure that the optimization problem is well-posed and convex. The key idea is that instead of viewing policy functions as continuous functions on the real line, we rather view them as square-integrable Borel-measurable functions and hence, treat the policy as a random variable.

PROPOSITION 4 (Admissible Set of the Growth Model) *Consider the growth model from Section 2. Define the subspace \mathcal{H} as the intersection of the square-integrable functions w.r.t. μ , i.e., $L^2(\mathcal{Z}^{id} \times \mathbb{R}, \mathcal{B}(\mathcal{Z}^{id} \times \mathbb{R}), \mu)$, and the functions with bounded first and second variation³ BV^2 such that for any element $h \in \mathcal{H}$, the following inequalities are satisfied a.s. for the random variable $\kappa = h(z', k, \mu)$ with $(z^{id'}, k) \sim \mu$ and a.e. $z, z' \in \mathcal{Z}$*

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

³ n^{th} variation (see e.g., Zeidler, 1986b): Let \mathcal{E} be a Hilbert space. The n^{th} variation of an operator $\mathbf{T} : \mathcal{E} \rightarrow \mathcal{E}$ at a point $e \in \mathcal{E}$ in the direction $\tilde{e} \in \mathcal{E}$ is defined by $\delta^n \mathbf{T}(e; \tilde{e}) = d^n/dt^n \mathbf{T}(e + t\tilde{e})|_{t=0}$.

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

$$\sum_{z' \in \mathcal{Z}} \mathbf{P}(z', z, \kappa) \delta^2 h(z', \kappa, \mu; \tilde{\kappa}) \geq \sum_{z' \in \mathcal{Z}} \mathbf{P}(z', z, \kappa) [\delta^2 \mathbf{I}(z', \kappa; \tilde{\kappa}) - \gamma \left(\mathbf{C}(z, k)^{-(\gamma+1)} + \frac{[\delta \mathbf{C}(z', \kappa; \tilde{\kappa})]^2}{\mathbf{C}(z', \kappa)} \right)]$$

where \mathbf{P} denotes the probability operator, \mathbf{C} denotes the consumption operator⁴ and \mathbf{I} denotes the income operator defined as follows

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

with the random productivity and wage level given as

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

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

The admissible set essentially includes all functions $h(z', k, \mu)$ which are convex and continuous in k and are bounded from above by the agents' cash-at-hand, i.e. income plus savings from the previous period.

The standard way of solving a constrained convex 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 is equivalent to finding a root of the first-order condition of the optimization problem. Due to the transversality condition and the monotonicity properties of the optimization problem (7), any root of its first-order condition results in a recursive equilibrium. The existence of such a root is derived with the following theorem.

THEOREM 5 (Existence of a recursive equilibrium) *Consider the growth model from Section 2 together with Assumptions 1 and 3. Define the admissible set \mathcal{H} as in Proposition 4. If the model parameters satisfy that*

$$(8) \quad \beta(1 - \rho)^{1-\gamma} < 1,$$

⁴ Note that the consumption operator contains the composition $h \circ h$ which is Borel-measurable due to h being a Borel-measurable function.

then there exists a function $h \in \mathcal{H}$, $h : \mathcal{Z} \times \mathbb{R} \times \mathcal{P}(\mathcal{Z}^{id} \times \mathbb{R}) \rightarrow \mathbb{R}$, which minimizes the optimization problem (7) and hence, constitutes the optimal savings choice of a recursive equilibrium according to Definition 2.

The proof of this existence theorem relies on the monotonicity properties of our equilibrium problem. This approach is inspired by series of papers by Rockafellar (1969, 1970, 1976a,b). Rather than using fixed-point theory on compact spaces, it relies on results from convex analysis and monotone operator theory. Due to the convexity of the objective and the constraints in the policy, the corresponding first-order condition as a functional of the policy is a maximal monotone operator. This property is comparable to a monotone function which spans the whole image space. There exists a root of such an operator if one can find a candidate policy at which the first-order condition has a negative value and another policy at which it has a positive value. The mathematical details are explained in Appendix A.2, which also contains the proof.

Note that due to the convex nature of the optimization problem (7), it is straightforward to investigate the uniqueness of an equilibrium.

PROPOSITION 6 (Condition for uniqueness) *If the optimal savings policy, i.e., the minimizer of the optimization problem (7), lies in the interior of the admissible set $h \in \text{int}(\mathcal{H})$, then this is the unique equilibrium.*

It is well known that optima of strictly convex objectives are unique (see e.g. Bauschke and Combettes, 2017, Corollary 11.9), which is the case when condition (ii) of Proposition 4 is satisfied strictly. This is true for interior policies. Note also that the boundary of the admissible set cannot contain any local optima in that case as any local minimum coincides with the global minimum for convex optimization problems.

Due to the fact that we cannot rely on compactness for this type of model, one can also not rely on the contraction property to design a convergent iterative procedure. Hence, we cannot compute the equilibrium using value function iteration. However, the monotonicity approach leads to another convergent iterative procedure. This procedure is explained subsequently.

4. A CONVERGENT ITERATIVE SCHEME

We know from the previous section that an equilibrium is the optimal policy at which the first-order condition of the optimization problem (7) equals zero. For illustrative purposes, let us denote the first-order condition by an operator $\mathbf{T}(h)$

which takes the optimal policy h as an argument. Hence, we obtain

$$\mathbf{T}(h) = 0 \Leftrightarrow \mathbf{T}(h) + h = h \Leftrightarrow (\mathbf{T} + \mathbf{Id})(h) = h \Leftrightarrow h = (\mathbf{T} + \mathbf{Id})^{-1}(h),$$

where \mathbf{Id} is the identity operator. The last equality contains the resolvent of the first-order equation $(\mathbf{T} + \mathbf{Id})^{-1}$. This operator has a very desirable property. Due to the maximal monotonicity of the first-order condition, its resolvent is Lipschitz continuous with coefficient one and it can be proven that iterating on the resolvent as in

$$h^{n+1} = (\mathbf{T} + \mathbf{Id})^{-1}(h^n),$$

where n is the iteration count, converges to the optimal policy, i.e. the root of the first-order condition. This iterative procedure results in the proximal point algorithm. 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 $F : \mathcal{E} \rightarrow [-\infty, \infty]$. It is well known that its subdifferential ∂F is maximal monotone (see e.g., [Bauschke and Combettes, 2017](#), Theorem 20.48). 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 ,

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

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

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

The latter is the update of the proximal point algorithm.⁵

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

⁵ 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 $0 < \lambda^n \nearrow \lambda^\infty \leq \infty$.

the update of the proximal point algorithm for the Lagrangian of our agents in the growth model in the following.

4.1. 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 first-order condition associated with the Lagrangian of (7). Hence, each iteration on the resolvent updates the agents' optimal choices for individual capital h as well as the Lagrange multiplier y for the inequality constraint 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^{n+1} , is the minimizer of the Lagrangian augmented by terms featuring the n -th iterate. The augmented Lagrangian is a function $L^A : \mathcal{H} \times L^2(\mathcal{Z}^{id} \times \mathbb{R}, \mathcal{B}(\mathcal{Z}^{id} \times \mathbb{R}), \mu) \rightarrow [-\infty, \infty]$ given by

$$(9) \quad \begin{aligned} L^A(h, y; h^n) = & -u(I(z', k, K) + [1 - \rho]k - h) \\ & - \sum_{z'' \in \mathcal{Z}} p^{z''|z'} \beta u(I(z'', h, K') + [1 - \rho]h - h') \\ & + \frac{1}{2\lambda} (h - h^n)^2 \\ & + \begin{cases} -yh + \frac{\lambda}{2}h^2 & , h \leq \frac{y}{\lambda} \\ -\frac{1}{2\lambda}y^2 & , h > \frac{y}{\lambda} \end{cases}, \end{aligned}$$

where $\lambda > 0$ is the step size parameter of the proximal point algorithm. Note that the next-period optimal capital savings policy is naturally given by the composition $h' = h^n \circ h$. The first two lines of the Lagrangian features the objective of (7). The third line consists of the objective's proximal point augmentation, which transforms the first-order condition into its resolvent. The last line corresponds to the inequality constraint. It also consists of the Lagrange term and the augmentation, but it is 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 an optimum of the Lagrangian even if the update of the optimal consumption and individual capital in line 5 is only approximate. This is important as the minimizer of the augmented Lagrangian is often not known in closed form, but it can be approximated with standard nonlinear solvers.

Algorithm 1 Proximal point algorithm for the growth model

▷ *A Initialization*

- 1: Set $n = 0$. Initialize the agents' choices individual capital and the Lagrange multiplier $H^n = (h^n, y^n)$.
- 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

$$h^{n+1} \approx \arg \min_{h \in \mathcal{H}} L^A(h, y^n; h^n)$$

$$y^{n+1} = \max(0, y^n - \lambda h^{n+1})$$

where L^A is defined as in (9).

- 6: Compute the distance $d = \|H^{n+1} - H^n\|$.
- 7: Set $n = n + 1$.
- 8: **end while**

4.2. Convergence of the Proximal Point Algorithm

We make the form of approximation of the policy update precise to show convergence in the following theorem.

THEOREM 7 (Convergence of Algorithm 1) *Consider the growth model from Section 2 together with Assumptions 1 and 3. Define the admissible set of policies \mathcal{H} as in Proposition 4. When computing the next iterate h^{n+1} in Algorithm 1, line 5, as a solution to the formula*

$$(10) \quad X(z', k, K) \|\partial L^A(h, y^n; h^n)\|_1 \leq \frac{\epsilon^2}{2\lambda}$$

for any $(z', k, K) \in \mathcal{Z} \times \mathbb{R}_{\geq 0}^2$, where X is the cash-at-hand, i.e., productive income plus savings, of the agent with start capital k

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

with the income I as in (3), the augmented Lagrangian L^A as in (9) and ∂ denoting the subdifferential w.r.t. h , then, Algorithm 1 converges to a recursive equilibrium of the growth model.

REMARK (i) Equation (10) is easily implemented by any root solver using a tolerance level of $\epsilon^2/(2\lambda)$.
(ii) Note that the augmentations in the augmented Lagrangian (9) go to zero

when the proximal point algorithm converges. Hence, the optimal policy to which the Algorithm 1 converges, minimizes the standard Lagrangian of the individual optimization problem (7) which is equivalent to the equilibrium's Euler equation.

The mathematical details are explained in Appendix A.3, which also contains the proof. The convergence rate of Algorithm 1 is $O(n^{-1})$ as is shown by Güler (1991). 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). I explain the acceleration in Appendix B.

We have to discretize the policy function in order to implement the algorithm. The difficulty herein lies in the cross-sectional distribution as an element of the state space because the space of distributions is infinite-dimensional. A solution to this problem is discussed in the next section.

5. DISCRETIZING THE SPACE OF DISTRIBUTIONS

The recursive equilibrium, we want to solve for, depends on the cross-sectional distribution $h(z', k, \mu)$ as an element of the state space. Therefore, we need to discretize the space of distributions. The existing literature often resorts to using a finite number of moments to characterize the cross-sectional distribution. However, even though there is a one-to-one correspondence between a distribution and its moment-generating function, this function does not exist for all distributions. Hence, any moment-based method cannot span the full space of square-integrable distributions and for some models, especially the ones producing fat-tailed cross-sectional distributions, such an approximation method is bound to fail. Another option would be a histogram representation or a spline interpolation of the distribution, but the discretized state space becomes very large very quickly in this case. Unfortunately, we have to rule out projection on orthogonal polynomials, which is widely used in computational economics, as well because a prerequisite is a sufficiently smooth distribution. Due to the occasionally binding borrowing constraint, however, the cross-sectional distribution exhibits mass points at the constraint and elsewhere. This fact is illustrated in the following proposition.

PROPOSITION 8 (A condition for mass points⁶) *Consider a recursive equilibrium as in Definition 2 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 $h_2(\hat{z}, \hat{k}, \mu) \leq k$ for all*

⁶The proof can be found in Appendix A.4.

$k \in [\delta, \hat{k}]$. Furthermore, assume that the optimal capital savings function has a kink at $k^* := \max\{k \geq \delta \mid h_2(\hat{z}, k, \mu) = \delta\} > \delta$, 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 δ . 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 μ . Hence, standard orthogonal polynomial projection methods do not work well here as one would need a very large number of projection points and would be confronted with the curse of dimensionality.

There is an efficient way of approximating distributions circumventing the previously mentioned problems. Instead of polynomial projection on the real line, we will use polynomial projection in the space of square integrable random variables. One could interpret this as probabilistic rather than deterministic polynomial projection. This technique is called polynomial chaos and is well known in the physics and engineering literature. It is a method, which projects a square-integrable distribution onto orthogonal polynomials which have random variables rather than the real line as arguments. The advantage of this approach is that it spans the whole space of square-integrable random variables and hence, one can be sure that one can approximate any cross-sectional distribution sufficiently well. This includes discrete distributions and mixtures of discrete and continuous distributions. Furthermore, when the basic random variables and their corresponding family of polynomials are chosen carefully, the speed of convergence easily outperforms standard polynomial projection. Hence, a low order of polynomials is enough to obtain a good approximation of the cross-sectional distribution. In the growth model, we find that order two to three is sufficient, i.e. we can substitute μ with two to three projection coefficients. In the following, I will summarize the method of polynomial chaos in general and how this technique is applied to our growth model. Subsequently, I derive approximation error bounds which yield convergence of this method.

5.1. 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

$$(11) \quad \kappa = \sum_{i=0}^{\infty} \varphi_i \Phi_i(\xi),$$

where φ_i are constant projection coefficients.

It is important to note that there is an explicit 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 $i, j \in \{0, 1, \dots\}$, it reads

$$(12) \quad \langle \Phi_i, \Phi_j \rangle = \int_{-\infty}^{\infty} \Phi_i(\xi) \Phi_j(\xi) dF(\xi) = \frac{\delta_{ij}}{a_i^2},$$

where δ_{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](#))

$$(13) \quad \Phi_{i+1}(\xi) = (\xi - \theta_i) \Phi_i(\xi) - \omega_i \Phi_{i-1}(\xi), \quad i \in \{0, 1, \dots\},$$

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 $i \in \{0, 1, \dots\}$. If κ is not a direct function of the basic random variable ξ , one uses the fact that both c.d.f.s $\mu, F \sim \mathcal{U}[0, 1]$ are uniform to compute the

coefficients

$$(14) \quad \varphi_i = \frac{1}{\langle \Phi_i, \Phi_i \rangle} \int \kappa \Phi_i (F^{-1} \circ \mu(\kappa)) d\mu(\kappa) \quad \forall i \in \{0, 1, \dots\},$$

where μ^{-1} is the generalized inverse distribution function of κ . 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 truncate the series of projection coefficients later on.

For practical reasons, it is important to note that polynomial chaos is easily extended to multivariate distributions by the tensor product rule. To approximate a joint distribution of n random variables denoted by $\kappa \sim \mu$, we simply need to fix n independent basic random variables ξ^1, \dots, ξ^n and determine their corresponding univariate orthogonal polynomials $\Phi^{\xi^1}, \dots, \Phi^{\xi^n}$ separately. Then, the polynomial chaos expansion for the multivariate distribution equals

$$(15) \quad \kappa = \sum_{i=0}^{\infty} \varphi_i \Phi_i (\xi^1, \dots, \xi^n) = \sum_{i=0}^{\infty} \varphi_i \sum_{\substack{0 \leq i_1, \dots, i_n \leq i, \\ i_1 + \dots + i_n = i}} \Phi_{i_1}^{\xi^1} (\xi^1) \cdot \dots \cdot \Phi_{i_n}^{\xi^n} (\xi^n).$$

The projection coefficient is then computed as in (14) for $\xi = (\xi^1, \dots, \xi^n)$ which reduces to the sum of a composition of integrals due to the independence of the basic random variables.

Let me now summarize which steps are necessary to approximate the distribution space with polynomial chaos expansion.

Before starting the solution algorithm:

1. Determine how many basic random variables are necessary.
2. Fix the distribution of each basic random variable.⁷
3. For each basic random variable, generate its corresponding orthogonal polynomials using the orthogonality condition (12) and the three-term recurrence relation (13).
4. Compute the multivariate orthogonal polynomials by multiplying the univariate polynomials according to (15).

During the solution algorithm:

⁷ Any such distribution has to possess finite moments of all orders and be determinate in the Hamburger sense (see [Ernst et al., 2012](#)). A distribution is determinate in the Hamburger sense if it uniquely solves the Hamburger moment problem or in other words if it is uniquely determined by the sequence of its moments.

5. Represent any endogenous distribution by projecting it onto the predetermined polynomial chaos expansion according to (14).

In the following, I will explain how to tackle steps 1 to 3 for the growth model in more detail.

5.2. Applying Polynomial Chaos to our Growth Model

We want to approximate the cross-sectional distribution of idiosyncratic variables $\kappa = (z^{id}, k) \sim \mu$, i.e. a bivariate distribution. Hence, we have to define two independent basic random variables $\xi^z \sim F^z$ and $\xi^k \sim F^k$ and their corresponding univariate orthogonal polynomials Φ^z and Φ^k . Then, the polynomial chaos expansion reads

$$\kappa = \sum_{i=0}^{\infty} \varphi_i \Phi_i (\xi^z, \xi^k).$$

This expression can be rewritten in terms of individual capital conditional on employment status $k|z^{id} \sim \mu_{z^{id}}$

$$(16) \quad k|z^{id} = \begin{cases} \frac{1}{1-p^e} \sum_{i=0}^{\infty} \varphi_i \Phi_i \left(\xi^z |_{\{\xi^z \leq [F^z]^{-1}(1-p^e)\}}, \xi^k \right) & , z^{id} = 0 \\ \frac{1}{p^e} \sum_{i=0}^{\infty} \varphi_i \Phi_i \left(\xi^z |_{\{\xi^z > [F^z]^{-1}(1-p^e)\}}, \xi^k \right) & , z^{id} = 1, \end{cases}$$

which yields the following definition of the projection coefficient for a cross-sectional distribution μ

$$\varphi_i(\mu) = \left\{ [1 - p^e] \int k \Phi_i ([F^k]^{-1} \circ \mu_{z^{id}=0}(k)) d\mu_{z^{id}=0}(k) dF^z(\xi^z) + p^e \int k \Phi_i ([F^k]^{-1} \circ \mu_{z^{id}=1}(k)) d\mu_{z^{id}=1}(k) dF^z(\xi^z) \right\} \frac{1}{\langle \Phi_i, \Phi_i \rangle}.$$

Accordingly, the law of motion of the projection coefficients is $\varphi'_i = \varphi_i(\mu')$ where, according to (6),

$$(17) \quad \mu' (z^{id'}, k) = \sum_{z^{id} \in \mathcal{Z}^{id}} \frac{p^{z'|z}}{p^{z^{ag'}|z^{ag}}} p^{z^{id}|z^{ag}} \mu_{z^{id}} (h (z', k|z^{id}, \{\varphi_i\}_{i=0}^{\infty}) \leq k)$$

with $k|z^{id}$ as in (16).

5.2.1. A Specific Choice of the Basic Random Variables

We have to determine two independent basic random variable for the growth model ξ^k and ξ^z . It was illustrated in [Xiu and Karniadakis \(2002\)](#) that the speed of convergence of the polynomial chaos expansion significantly improves if the distribution of the basic random variables is not too far from the distribution we want to approximate. Let me first explain, how we fix the basic random variable corresponding to individual capital and subsequently, the one for the exogenous idiosyncratic shock.

As 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 8](#). 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 stationary cross-sectional distribution. 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). I do this using histograms. Naturally, the stationary cross-sectional distributions will have features similar to the distribution of the model with aggregate shocks. Hence, we fix ξ^k as the cross-sectional distribution of individual capital in the model without aggregate shocks averaged over the two cases of keeping z^{ag} fixed at 0 and 1. The distribution of the basic random variable ξ^k is displayed in [Figure 2](#). The distribution exhibits several mass points measured as the local extrema in the histogram representation. Note that the capital constraint at the end of the period is binding only for a small fraction of unemployed agents because all employed agents optimally choose positive capital savings. To obtain an accurate approximation of the stationary distributions in the model without aggregate risk, one should choose a reasonably small bin size for the histogram.

We now move on to fix a distribution for the basic random variable corresponding to the idiosyncratic employment shock. This distribution needs to accommodate both the idiosyncratic shock in the good economic state as well as in the bad aggregate state. Hence, we define it as a discrete distribution with three states

$$F(\xi^z) = \begin{cases} p^{z^{id}=0|z^{ag}=1} & , \xi^z = 1 \\ p^{z^{id}=0|z^{ag}=0} & , \xi^z = 2 \\ 1 & , \xi^z = 3. \end{cases}$$

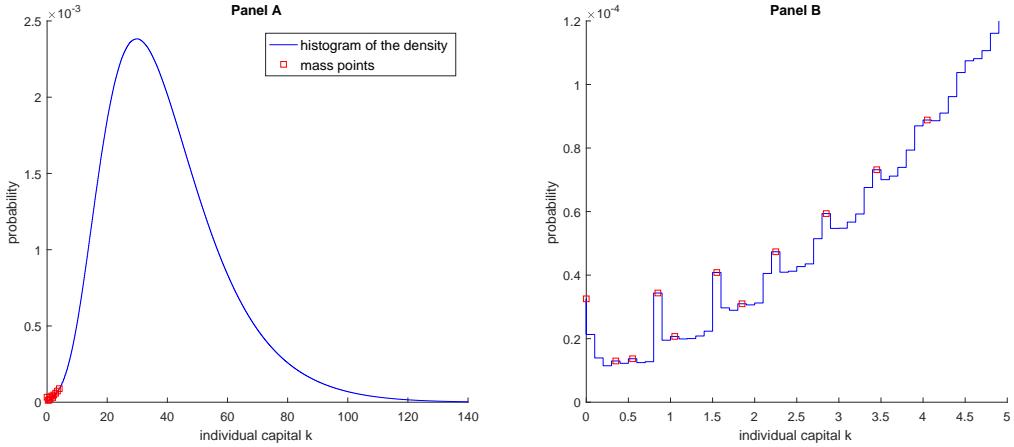


FIGURE 2. Distribution for the basic random variable ξ^k . Panel A shows a histogram representation with bin size 0.1. A mass point is identified as a bin whose probability is higher than the ones of its direct neighbors, but the global maximum is excluded. Panel B zooms into the left tail of the distribution. Note that the distribution displayed here is the average of the stationary end-of-period distributions from (6) of the model without aggregate shocks.

It follows that $z^{id}|\{z^{ag} = 0\} = \mathbb{1}_{\{\xi^z > 2\}}$, whereas, $z^{id}|\{z^{ag} = 1\} = \mathbb{1}_{\{\xi^z > 1\}}$. Having chosen the two basic random variables, we explain how to generate their orthogonal polynomials next.

5.2.2. Generating the Corresponding Orthogonal Polynomials

As our basic random variable ξ^k is represented by a histogram, both basic random variables have discrete distributions where the end points of the bins $\{\xi_n\}_{n=1}^N$ have probability $\{p_n\}_{n=1}^N$. 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 Φ_N has the points $\{\xi_n\}_{n=1}^N$ 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 Stieltjes method, which performs well in terms of precision. It directly computes the parameters θ_i and ω_i in (13) using the standard inner product of L^2 and is explained in detail in [Gautschi \(1982\)](#). The constant parameters are given by

$$\begin{aligned}\theta_i &= \frac{\langle \Phi_i, \xi \Phi_i \rangle}{\langle \Phi_i, \Phi_i \rangle}, \quad i \in \{0, 1, \dots\} \\ \omega_i &= \frac{\langle \Phi_i, \Phi_i \rangle}{\langle \Phi_{i-1}, \Phi_{i-1} \rangle}, \quad i \in \{1, 2, \dots\}\end{aligned}$$

with $\langle \cdot, \cdot \rangle$ 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 (13) into the orthogonality condition (12). With the parameters defined as above, the orthogonal polynomials are easily constructed using (13).

The orthogonal polynomials of ξ^k and ξ^z are displayed in Figure 3. As usual,

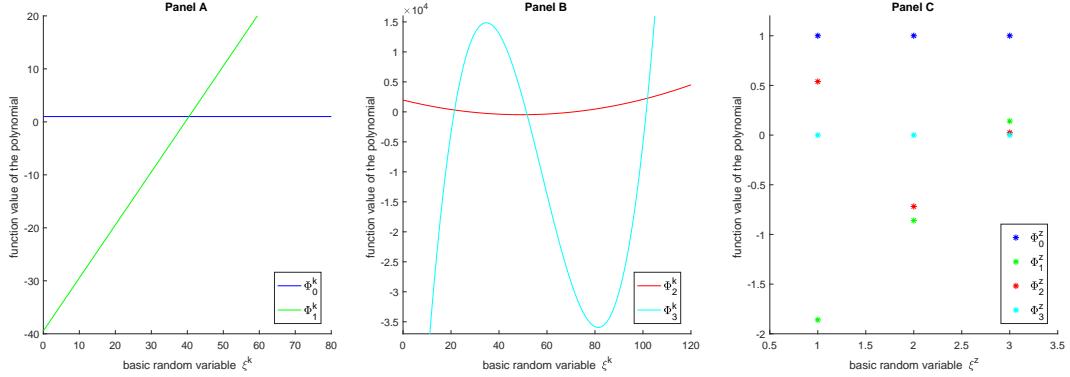


FIGURE 3. Orthogonal polynomials corresponding to the basic random variables of the growth model. Panel A shows the polynomials Φ_i^k corresponding to individual capital up to order $i \leq 1$, whereas, Panel B displays the polynomials Φ_i^k of order $i = 2$ and $i = 3$. Panel C plots the polynomials Φ_i^z corresponding to the idiosyncratic shock up to order $i \leq 3$.

the number of roots of each polynomial corresponds to its degree. Each first-order polynomial has its root at the mean of the distribution of the corresponding basic random variable. The polynomials of ξ^z are plotted only at its three states $\{1, 2, 3\}$. This graph confirms that the polynomial with maximal degree, which is the third-order polynomial Φ_3^z in this case, has its roots at the states of the corresponding distribution, i.e. at $\{1, 2, 3\}$.

With the basic random variables defined and the corresponding polynomials generated, the polynomial chaos expansion is fully determined. Any 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 $\sum_{i=0}^{\infty} \varphi_i \Phi_i(\xi^k)$ with fixed projection coefficients $\{\varphi_i\}_{i=0}^{\infty}$. The expansion is truncated at increasing order. The zeroth-order polynomial results in a mass point at φ_0 , which, due to its definition, is the mean of the projected distribution. The first-order polynomial term simply stretches or compresses the distribution of the basic random variable depending on its projection coefficient. Summing the zeroth- and

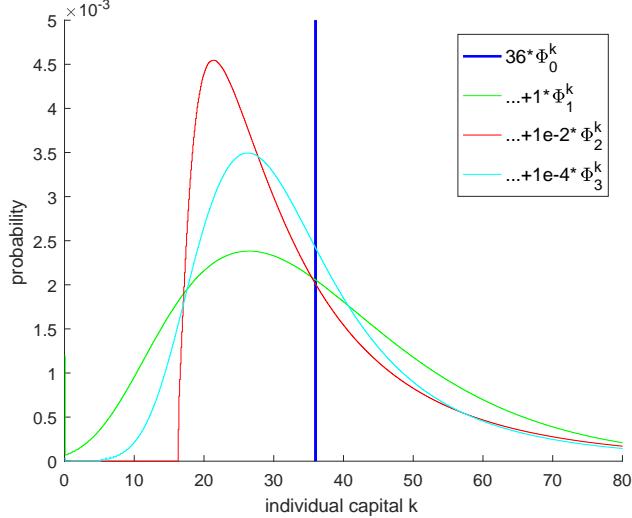


FIGURE 4. Example distributions resulting from truncated polynomial chaos expansions. The graph displays the histogram representations with bin size 0.1 of distributions resulting from the polynomial chaos expansion truncated at different orders ranging from order 0 to 3. The basic random variable used is ξ^k distributed as in Figure 2. The projection coefficients for this example are fixed as $[\varphi_0, \dots, \varphi_3] = [36, 1, 0.01, 0.0002]$.

first-order term simply centers the stretched/compressed distribution of the basic random variable around the mean of the projected distribution. Further adding the second-order polynomial modifies the skewness of the polynomial chaos expansion, whereas, the third-order polynomial adjusts the kurtosis. Higher orders further refine the tails. Hence, the polynomial chaos expansion gets closer to the projected distribution the higher the order of truncation.

5.3. Convergence of the Discretized Policy

When using polynomial chaos expansions for the cross-sectional distribution, we obtain the policy $h(z', k, \mu) = h(z', k, \{\varphi_i\}_{i=0}^\infty)$ where ξ^z and ξ^k are fixed. Note that this is not an approximation because we can replicate any square-integrable distribution perfectly with a polynomial chaos expansion. Approximation of this policy 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 $h^M = h(z', k, \{\varphi_i\}_{i=0}^M)$. Its interpolant, denoted by $h^{M,D}$, is defined on a tensor product of finite grids of the state space elements

$$D = \{(k_{i_0}, \varphi_{1,i_1}, \dots, \varphi_{M,i_M}) \mid i_m = 1, \dots, I_m < \infty \forall m = 0, \dots, M\}.$$

The question is whether the algorithm converges for such a discretized policy. This is shown in the following.

It is clear that the interpolant $h^{M,D}$ stays within the admissible set of the policies \mathcal{H} defined in Proposition 4. 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. To derive it, I follow 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 4. Consider Algorithm 1 with polynomial chaos extension as in Section 5, i.e., using the basic random variables ξ^z and ξ^k and the corresponding orthogonal polynomials Φ to project any square-integrable cross-sectional distribution $\kappa \sim \mu$ with $\kappa = \sum_{j=0}^M \varphi_j \Phi_j(\xi^z, \xi^k)$. Assume that, for any fixed exogenous shock, start capital and individual capital distribution (z', k, μ) , the initial guess of the savings policy h^0 and the Lagrange multiplier y^0 for the proximal point algorithm are real analytic in the basic random variables and hence, satisfy*

$$(18) \quad \left\| \frac{\partial^p}{[\partial \xi^j]^p} f \right\| \leq \mathbf{c}_f^p p!, \quad p \in \{1, 2, \dots\}, \quad j \in \{z, k\},$$

for some constants \mathbf{c}_f where f is a handle for h^0 and y^0 . Furthermore, assume that the initial policy guess h^0 is real analytic in start capital. Consider the following subsets of the complex plane

$$\Sigma(\tau_{n+1}^i, \Gamma^i) = \left\{ x \in \mathbb{C} \mid \inf_{\xi^i \in \Gamma^i} |x - \xi^i| \leq \tau_{n+1}^i \right\}, \quad i \in \{z, k\},$$

where Γ^i is the range of ξ^i and $0 < \tau_{n+1}^i < \frac{\min(1, \mathbf{L}_{n+1})}{2A_{n+1,i}^1} < \infty$. \mathbf{L}_{n+1} is the value of the second derivative of the Lagrangian L^A as in (9) evaluated at the $(n+1)$ -th policy iterate and $A_{n+1,i}^1$ is given in (25) in the proof. Then, the approximation error bound for the $(n+1)$ -th policy iterate resulting from truncating the polynomial chaos expansion at order M and using linear interpolation on a rectangular tensor-

product grid

$$D = \left\{ (k_{i_0}, \varphi_{0,i_1}, \dots, \varphi_{M,i_M}) \mid k_{i_n} < k_{i_{n+1}}, \varphi_{m,i_n} < \varphi_{m,i_{n+1}} \right. \\ \left. \forall i_n \in \{1, \dots, d_n\}, m \in \{1, \dots, M\} \right\}$$

with maximum mesh-size s is given by

$$\|h^{n+1} - h^{M,D}\|_{L^2} \leq \sum_{i \in \{z, k\}} \mathbf{b}_i \frac{2}{\eta^i - 1} e^{-M \log(\eta^i)} \frac{\min(1, \mathbf{L}_{n+1})}{\min(1, \mathbf{L}_{n+1}) - 2\tau_{n+1}^i A_{n+1,i}^1} \\ + \mathbf{b}_d s^2 \left(\sum_{j=0}^{M+1} \left\| \frac{\partial^2 h^M}{[\partial D_j]^2} \right\|_{L^2}^2 \right)^{\frac{1}{2}}, \quad (19)$$

where \mathbf{b}_i , $i \in \{z, k, d\}$, are constants and

$$\eta^i = \frac{2\tau_{n+1}^i}{|\Gamma^i|} + \sqrt{1 + \frac{4(\tau_{n+1}^i)^2}{|\Gamma^i|^2}} > 1, \quad i \in \{z, k\}.$$

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.

6. NUMERICAL RESULTS

We compute the recursive equilibrium solution of the accelerated proximal point algorithm explained in Appendix B truncating the polynomial chaos expansion at different orders using Matlab R2016b.⁸ Furthermore, to demonstrate that the polynomial chaos expansion can be combined with other solution methods, we also compute the recursive solution via standard policy function iteration, which is possible when the utility function is bounded from above. We also compute the solutions of existing algorithms for comparison. We choose the algorithm by Krusell and Smith (1998) since 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 in a special issue of the Journal 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, as they perform best in the comparison by den Haan (2010).

⁸The computations were performed on the Baobab cluster at the University of Geneva.

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 proximal point algorithm as well as the policy function iteration, the total aggregate capital is equivalent to the projection coefficient ϕ_0 of the polynomial chaos expansion. We use 4 grid points for aggregate capital for the Krusell-Smith, the Reiter and our algorithms. Note that the Hahn-Rendahl algorithm then has 4×4 grid points in aggregate capital because it differentiates unemployed and employed aggregate capital. 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 and the policy function iteration

Algorithm	# Grid Points for $z' \times k \times \mu$	# CPUs	Compute Time
Krusell-Smith (K-S)	$4 \times 80 \times 04$	4	42s
Reiter (R)	$4 \times 80 \times 04$	4	2m 58s
den Haan-Rendahl (D-R)	$4 \times 80 \times 16$	4	23s
Proximal Point Algorithm M=0 (PPA0)	$4 \times 80 \times 04$	20	1m 6s
Proximal Point Algorithm M=1 (PPA1)	$4 \times 80 \times 12$	20	-
Proximal Point Algorithm M=2 (PPA2)	$4 \times 80 \times 24$	20	30m 12s
Proximal Point Algorithm M=3 (PPA3)	$4 \times 80 \times 48$	20	1h 33m 19s
Policy Function Iteration M=0 (PFI0)	$4 \times 80 \times 04$	20	7s
Policy Function Iteration M=1 (PFI1)	$4 \times 80 \times 12$	20	21s
Policy Function Iteration M=2 (PFI2)	$4 \times 80 \times 24$	20	1m 40s
Policy Function Iteration M=3 (PFI3)	$4 \times 80 \times 48$	20	4m 33s

TABLE 1. **Summary of the algorithms to be compared.** In the first column, M denotes the order of truncation of the polynomial chaos expansion. 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 policy function. Note that the methods of discretizing the distribution μ vary across algorithms. If the distribution is discretized with several parameters, the number of grid points is the product of the number of grid points in each parameter.

is implemented with parallelized Matlab 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 proximal point algorithm truncated at first order does not converge⁹ which is why

⁹Note that we showed convergence for the proximal point algorithm using the full polynomial

there is no computation time. 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 for a model with unbounded utility function. I argue that the goal of being more accurate and having a theoretically sound algorithm justifies the increased compute time. However, if one can ensure that policy function iteration converges, then this should be the method of choice. If convergence of a faster ad hoc method is not clear, one can use the proximal point algorithm as a benchmark.

In the following, we investigate whether the algorithms using polynomial chaos expansions really yield higher precision than the existing methods. Furthermore, we examine the order of truncation of the polynomial chaos expansion resulting in sufficient precision and which economic implications the polynomial chaos algorithms yield.

6.1. 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

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

In contrast to the standard Euler equation error, the dynamic equivalent denoted by ϵ^{DE} is computed for several consecutive periods. This test is more stringent as the numerical solution and the explicit conditional expectation usually diverges 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 β^t , $t = 0, \dots, N - 1$, is always above machine precision. Our configuration leads to $N = 3587$. Otherwise, we would only add noise to the infinite sum of the utility. Note that we compute the standard Euler Equation error test also over multiple periods, but it is reset every period

chaos expansion, i.e. without truncation. When truncating at very low orders, one can therefore run into stability issues.

and hence, does not accumulate. The errors' summary statistics are displayed in Table 2. Note that the den Haan-Rendahl algorithm does not seem to work in our configuration. The other solution algorithms produce more or less the same numbers in terms of the mean and median, although Reiter and the algorithms based on polynomial chaos improve the maximum error.

The advantage of the polynomial chaos algorithms becomes clearer when displaying the full error distribution in terms of boxplots in Figure 5. One can see

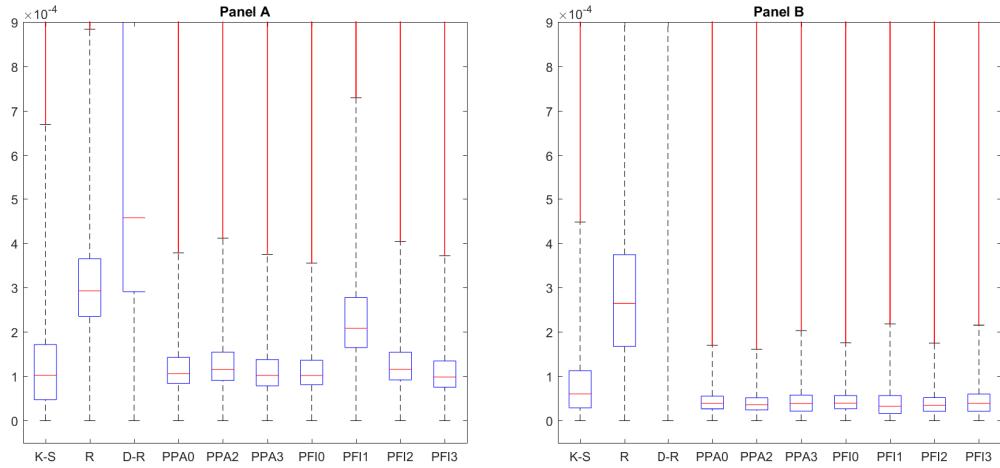


FIGURE 5. Boxplots of the Euler equation error distributions of individual capital. Panel A shows the standard Euler equation error and Panel B displays the dynamic Euler equation error for the numerical solutions from Table 1. The error is computed over a finer grid in the dimension of individual capital than the grid on which the solutions are computed. We compute the error 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 25th to 75th percentiles. The whiskers indicate the range of the distribution and the red dots outside are outliers.

that the existing algorithms produce much wider error distributions 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 improvement compared to the Krusell-Smith algorithm. The same is true for the den Haan-Rendahl method, which performs much 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. In comparison, all polynomial chaos based solutions produce much narrower error bands. This is mainly due to the better anticipation of the cross-sectional distribution's law of motion in the polynomial chaos algorithms.

	SE Mean	SE Median	SE Min	SE Max	DE Mean	DE Median	DE Min	DE Max
K-S	1.5450e-04	1.0250e-04	0	8.3959e-03	1.2220e-04	6.0565e-05	0	8.3199e-03
R	3.1636e-04	2.9273e-04	6.5846e-10	2.9434e-03	2.8791e-04	2.6446e-04	1.4075e-10	2.9706e-03
D-R	1.1744e+15	4.5746e-04	0	5.4910e+15	1.2826e+13	8.9241e-03	1.8271e-09	5.4904e+15
PPA0	1.3101e-04	1.0592e-04	4.4331e-09	2.9837e-03	5.3277e-05	3.9124e-05	1.4229e-12	2.8710e-03
PPA1	2.1844e-04	1.8168e-04	2.0443e-09	3.1966e-03	6.0037e-05	4.3256e-05	5.3937e-12	2.9824e-03
PPA2	1.4074e-04	1.1530e-04	4.1057e-08	3.0028e-03	4.9667e-05	3.6176e-05	1.2937e-11	2.8671e-03
PPA3	1.2726e-04	1.0238e-04	1.1388e-08	2.9874e-03	5.1997e-05	3.8513e-05	1.2379e-12	2.8646e-03
PFI0	1.2737e-04	1.0184e-04	2.2559e-08	2.9819e-03	5.3903e-05	3.9725e-05	1.4888e-11	2.8711e-03
PFI1	2.4063e-04	2.0824e-04	1.6600e-10	3.2041e-03	5.1694e-05	3.2844e-05	4.4933e-13	2.9808e-03
PFI2	1.4204e-04	1.1568e-04	4.1952e-08	3.0059e-03	4.8581e-05	3.4518e-05	8.9723e-12	2.8704e-03
PFI3	1.2480e-04	9.8397e-05	2.7090e-08	2.9837e-03	5.3215e-05	3.8933e-05	1.4970e-11	2.8646e-03

TABLE 2. Euler equation errors for the numerical solutions from Table 1. This table displays the summary statistics of the standard Euler equation error ϵ^{SE} and of the dynamic Euler equation error ϵ^{DE} . It is computed over a finer grid in the dimension of individual capital than the grid on which the solution is computed. It is computed for a random sample of aggregate shocks over $N = 3587$ periods. The initial cross-sectional distribution is the same for all algorithms.

Since consumption is concave and increasing in individual capital, the interpolation error when not taking the absolute value is expected to be negative due to the underestimation of the true consumption away from the interpolation points. Hence, by looking at the error without absolute values in Figure 6, we can identify any systematic biases. We see that the standard errors of the algorithms based

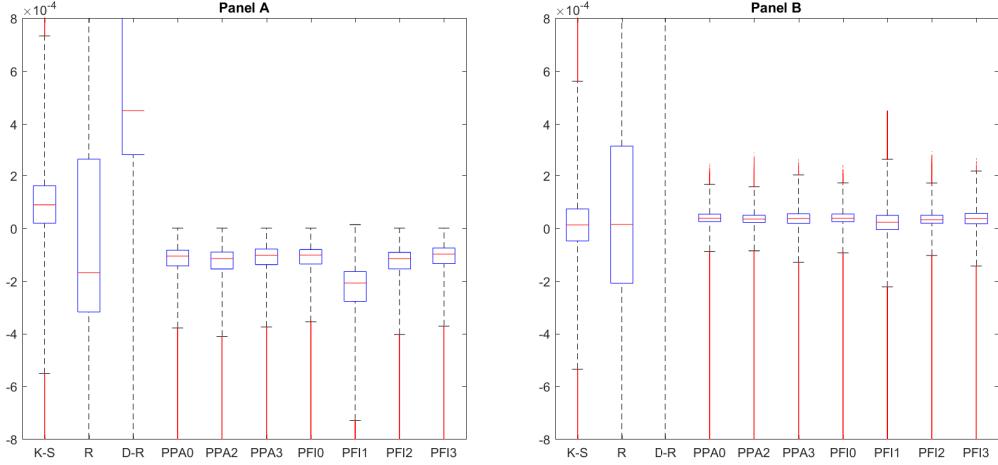


FIGURE 6. Boxplots of the Euler equation error distributions of individual capital without absolute value. Panel A shows the standard Euler equation error and Panel B displays the dynamic Euler equation error for the numerical solutions from Table 1. The error is computed over a finer grid in the dimension of individual capital than the grid on which the solutions are computed. We compute the error 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 25th to 75th percentiles. The whiskers indicate the range of the distribution and the red dots outside are outliers.

on polynomial chaos are indeed negative which indicates that the approximation at the interpolation points is good. However, the existing algorithms have a systematic positive bias. The dynamic error for the polynomial chaos based algorithms shows positive bias as well. This is due to the equilibrium effect that less consumption today increases consumption tomorrow. However, not taking the absolute error still does not help in distinguishing between the different truncation orders.

It seems counter-intuitive that the error distributions for the algorithms based on polynomial chaos do not differ much. One would expect that they decrease with increasing order. To understand why, recall that the approximation error bound in (19) consists of two terms. A decrease in error due to the increase in the truncation order of the polynomial chaos might be offset by an increase in

the interpolation error. Note that even though we use the same interpolation grid for all algorithms, the interpolation error may still differ as it depends on the curvature of the solution. The higher the curvature, the higher the interpolation error. Therefore, we need to disentangle the truncation error from the interpolation error. This is possible by comparing the prediction of the next-period aggregate capital by the algorithm K with the true next-period aggregate capital \tilde{K} . We display the law of motion error, computed by

$$\epsilon^{LoM} = \frac{K - \tilde{K}}{\tilde{K}},$$

in Figure 7. We exclude the den Haan-Rendahl algorithm because it already per-

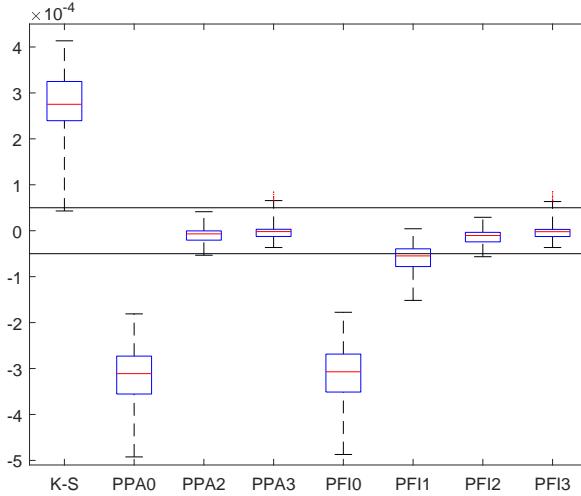


FIGURE 7. Boxplots of the law of motion error distributions of aggregate capital. This figure shows the low of motion error ϵ^{LoM} produced by the Krusell-Smith algorithm and the polynomial chaos algorithms from Table 1. The red lines mark the medians, whereas, the blue boxes denote the 25th to 75th percentiles. The whiskers indicate the range of the distribution and the red dots outside are outliers.

formed worse than the others in terms of the Euler equation errors. We also exclude the law of motion error for the Reiter algorithm because it was not possible to extract the prediction of next-period aggregate capital from Reiter's Matlab implementation. The figure shows that the low of motion error indeed becomes smaller when the truncation order is increased. In particular, we observe an exponential decrease as predicted by the error bound. Therefore, We conclude that the Euler equation error is indeed dominated by the interpolation error which offsets the decrease of the low of motion error when the curvature of the solution increases with the truncation order. Furthermore, the law of motion error is in the region of

the truncation criterion when truncating at second or higher order. This indicates that a polynomial chaos expansion up to second order yields sufficient precision.

Overall, the error analysis showed that the polynomial chaos based algorithms outperforms the existing algorithms and that the polynomial chaos expansion up to second order suffices to approximate the growth model. Recall that order zero implies that the optimal policies depend solely on aggregate capital. Order one and higher, however, imply a dependence on the full approximated distribution. Hence, 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.

6.2. Economic Implications and Reasons for Approximate Aggregation

Let me now compare the economic implications of the different numerical solutions. As the Euler equation errors for the den Haan-Rendahl algorithm were large, I will compare the polynomial chaos algorithms only to the Krusell-Smith and the Reiter algorithm. The largest conceptual difference between these 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. The polynomial chaos based algorithms, however, use the nonparametric law of motion of the aggregate variables stemming from the cross-sectional distribution. The Reiter algorithm lies conceptually in between the former two. It maps the set of moments to a parameterized cross-sectional distribution rather than a rule of thumb to compute the prediction for aggregate variables.

To compare the implications of these conceptual differences, we look at the stationary cross-sectional distributions. 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 consider 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. We can see that the distributions of the proximal point algorithm with different orders of truncation are almost indistinguishable. They are further to the right than the distribution of the Krusell-Smith algorithm. As expected the distribution of the Reiter algorithm is in between the former two. Furthermore, the distribution by the Krusell-Smith algorithm has a much thicker tail to the right than the other distributions. The corresponding Lorenz curves also confirm that the inequality in the economies resulting from the

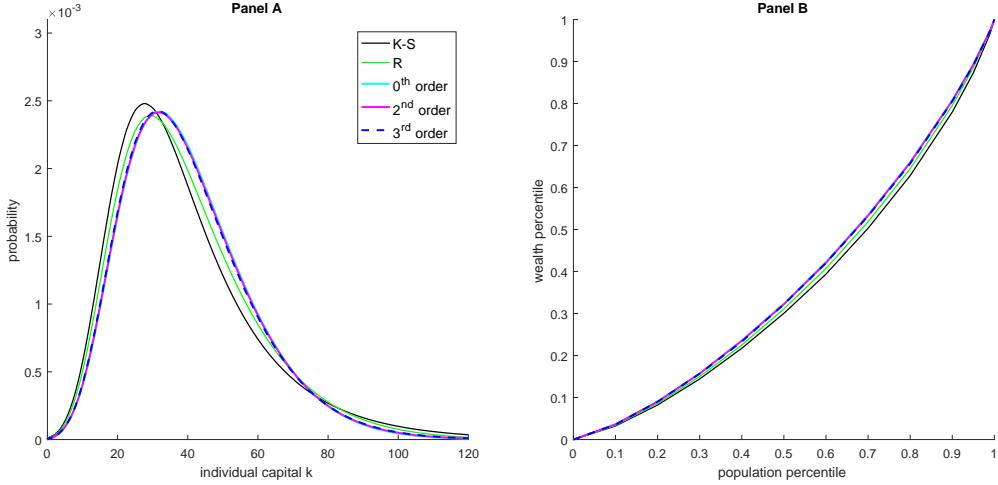


FIGURE 8. Average cross-sectional distributions produced by the Krusell-Smith algorithm and the polynomial chaos algorithms from Table 1. This graph displays the average stationary cross-sectional distribution, i.e., the expectation of the stationary distribution of the state space (z', k, μ) conditional on μ . 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. Only the distributions of the proximal point algorithms are displayed because they are identical with the ones produced by value function iteration.

other algorithms is slightly less than in the Krusell-Smith economy. It seems that the bounded rationality assumption introduces more inequality. It might therefore be a valid modeling tool to produce more realistic wealth distributions.

The fact that the distributions of the different proximal point algorithms are so close explains the approximate aggregation result from Krusell and Smith (1998). In terms of stationary distributions, higher orders do not matter in this calibration of the growth model. This is confirmed when computing the changes in the stationary distributions for increasing truncation order, as displayed in Figure 9. The change from order zero to two is already smaller than the termination threshold $5e - 5$ of the algorithm. This means that the solution of this calibration of the growth model is not sensitive to errors in the law of motion of aggregate capital. Note that the expected ergodic distribution is used in Krusell and Smith (1998) to update the parameter estimate for their law of motion. Since this ergodic distribution does not change substantially with a different law of motion, they did not find any changes in the solution when using more than one moment.

Hence, we conclude that approximate aggregation does hold in this calibration of the model. An interesting question, hence, remains. By how much would the solutions differ in a model where approximate aggregation does not hold. To

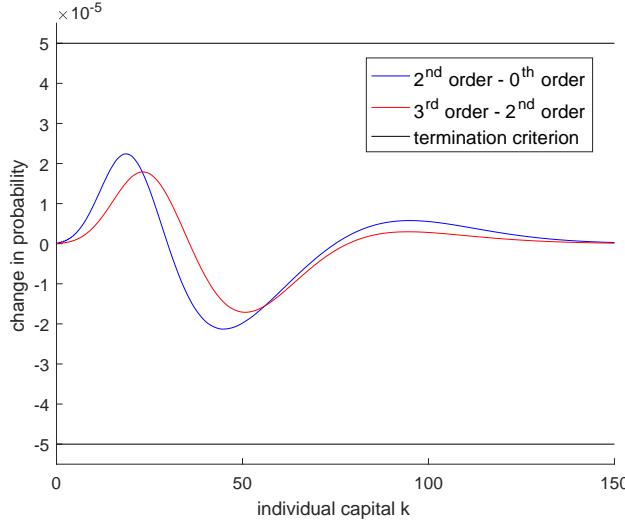


FIGURE 9. Changes in the average cross-sectional distributions produced by the proximal point algorithms from Table 1. This graph displays the changes in the average stationary cross-sectional distribution, i.e., the expectation of the expected ergodic distribution of the state space (z', k, μ) conditional on μ , when the order of truncation of the polynomial chaos expansion is increased. It shows the differences in p.d.f.s with a histogram approximation with bin size 0.1.

investigate this, we do not even have to change the model. In the following section, I show that approximate aggregation fails for a different calibration of the benchmark growth model.

6.3. Economic Implications when there is no Approximate Aggregation

In the previous calibration taken from [den Haan et al. \(2010\)](#), the unemployment benefit is set to 15% of gross wage. In the following, I set it to 65%. Therefore, there is more redistribution in the new calibration and hence, better risk sharing. As in the previous calibration, the law of motion error decreases in an exponential manner with increasing truncation order as can be seen in Figure 10. The error is overall smaller than in the first calibration. This makes sense since fewer agents will hit the borrowing constraint meaning that the kink in the policy has less impact. However, as we can see from the Euler equation errors in Figure 11, the curvature of the policy away from the kink increases with higher order of truncation. The Euler equation errors indicate that one should use truncation order 3 or higher. To investigate, whether order 3 or 4 yield sufficient precision, we also plot the changes in the stationary distributions in Figure 12. We can see that the changes between the first two distributions are substantial and we need order three. This also leads

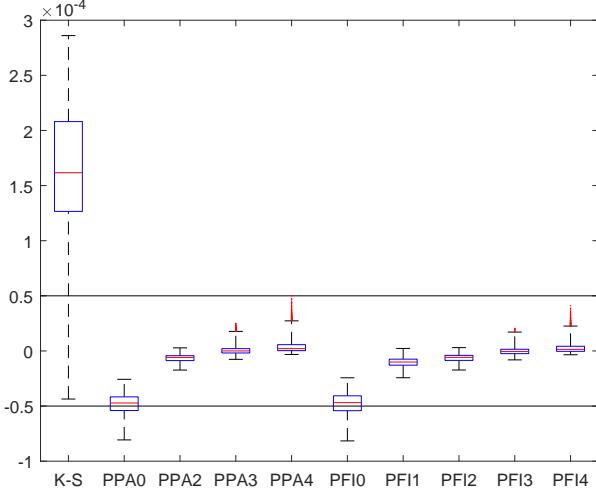


FIGURE 10. Boxplots of the law of motion error distributions of aggregate capital. This figure shows the law of motion error ϵ^{LoM} produced by the Krusell-Smith algorithm and the polynomial chaos algorithms from Table 1 for 65 % unemployment benefit rate. The red lines mark the medians, whereas, the blue boxes denote the 25th to 75th percentiles. The whiskers indicate the range of the distribution and the red dots outside are outliers.

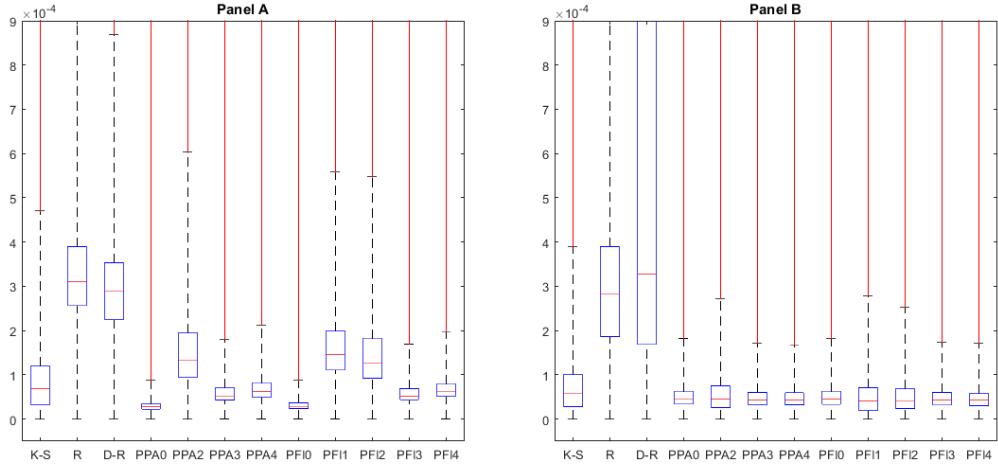


FIGURE 11. Boxplots of the Euler equation error distributions of individual capital. Panel A shows the standard Euler equation error and Panel B displays the dynamic Euler equation error for the numerical solutions from Table 1 for 65 % unemployment benefit rate. The error is computed over a finer grid in the dimension of individual capital than the grid on which the solutions are computed. We compute the error for a random sample of aggregate shocks over $N = 3587$ periods. The whiskers indicate the range of the distribution and the red dots outside are outliers.

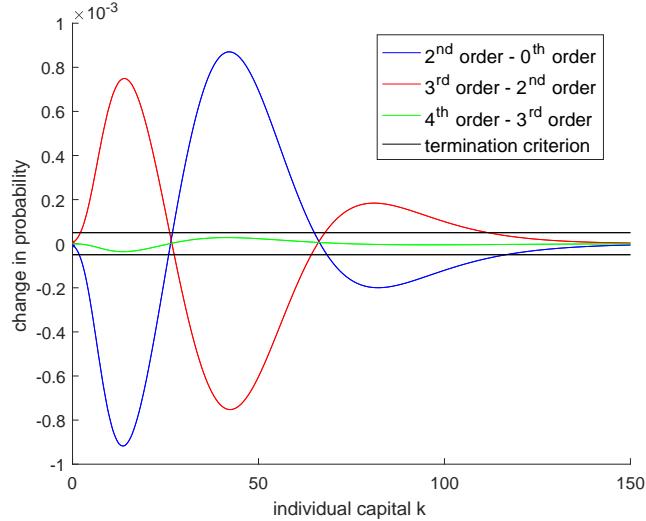


FIGURE 12. Changes in the average cross-sectional distributions produced by the proximal point algorithms from Table 1 for 65 % unemployment benefit rate. This graph displays the changes in the average stationary cross-sectional distribution, i.e., the expectation of the expected ergodic distribution of the state space (z', k, μ) conditional on μ , when the order of truncation of the polynomial chaos expansion is increased. It shows the differences in p.d.f.s with a histogram approximation with bin size 0.1.

to the conclusion that approximate aggregation does not hold in this calibration.

The failure of approximate aggregation leads to stark differences in the expected ergodic distributions as can be seen in Figure 13. The distribution resulting from the Krusell-Smith algorithm exhibits much fatter tails. Given it's methodology, it is not surprising that the distribution from Reiter's method is closer to ours than to Krusell-Smith's distribution. This graph confirms that bounded rationality can in fact be a substantial contributor to inequality.

Another interesting result emerges by comparing the expected ergodic distributions of the proximal point algorithm for the two different calibrations. As can be seen in Figure 14, the expected ergodic distribution in the case of a high unemployment benefit has fatter tails than the distribution in the case of a low benefit. This result even holds for the model without aggregate shocks. Therefore, albeit the better risk sharing by more redistribution, the volatility of the cross-sectional distribution, which can be interpreted as systemic risk, increases. This result mirrors the volatility paradox in Brunnermeier and Sannikov (2014).

7. CONCLUSIONS

This paper has two contributions. Firstly, I contribute to the theoretical underpinnings of DSGE models with ex-post heterogeneity and aggregate risk by

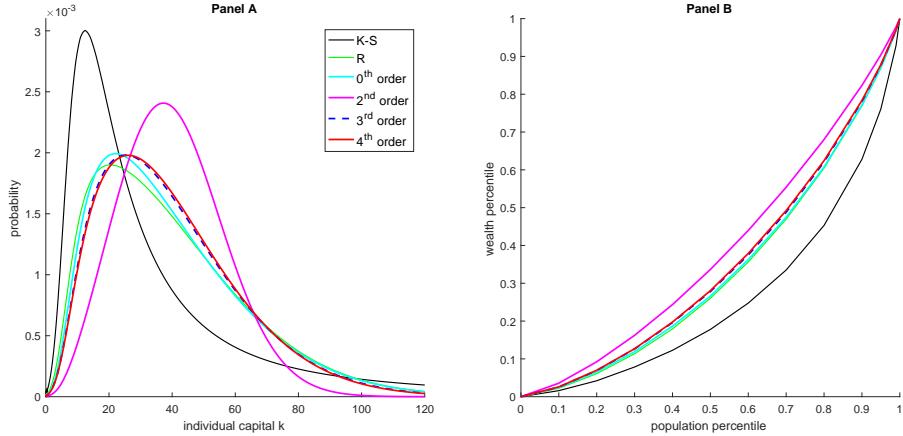


FIGURE 13. Average cross-sectional distributions produced by the Krusell-Smith algorithm and the polynomial chaos algorithms from Table 1. This graph displays the average stationary cross-sectional distribution, i.e., the expectation of the stationary distribution of the state space (z', k, μ) conditional on μ for 65 % unemployment benefit rate. 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. Only the distributions of the proximal point algorithms are displayed because they are identical with the ones produced by value function iteration.

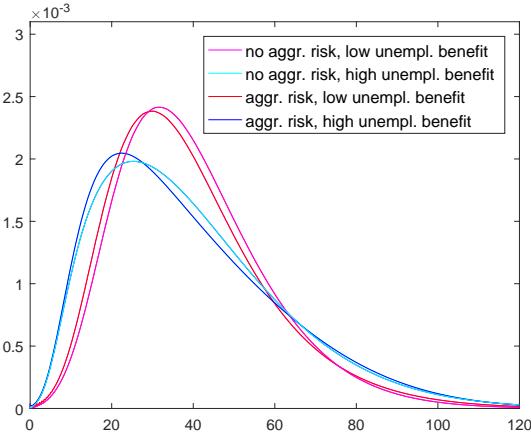


FIGURE 14. Average cross-sectional distributions produced by the polynomial chaos algorithm for low and high unemployment benefit. This graph displays the average stationary cross-sectional distribution, i.e., the expectation of the stationary distribution of the state space (z', k, μ) conditional on μ for 15 % and 65 % unemployment benefit rate, respectively. The proximal point algorithm for the former case is truncated at second order, whereas, the proximal point algorithm for the latter case is truncated at third order. It also displays the stationary distributions for the model without aggregate shocks for both unemployment benefit rates.

investigating the Aiyagari-Bewley economy with aggregate risk. Using a monotonicity argument, I show existence of a recursive equilibrium where the state space depends solely on the exogenous shocks, individual start capital and the cross-sectional capital distribution. Instead of relying on the compactness of the state space, which is usually done in existence arguments, I exploit the convexity of the model. An important element to this is to reinterpret the policy functions as random variables rather than real functions.

Secondly, I develop a novel global solution algorithm relying on the aforementioned theory, 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 a projection technique which extends orthogonal polynomial projection from spaces of smooth functions to the space of square-integrable random variables. This technique is known as generalized polynomial chaos and can be interpreted as a probabilistic polynomial projection method.

By taking this conceptually quite different approach with a sound theoretical underpinning, I am able to increase the precision of the solution by a significant amount, especially in terms of the law of motion of aggregate variables. However, this requires the agent to know more about the distribution than just its current mean, namely the polynomial chaos expansion of the distribution up to second order which is an approximation of the full cross-sectional distribution. We do find though, that the solutions in terms of their expected ergodic distribution do not change significantly when the truncation order increases. It means that the solutions are insensitive to changes in the law of motion. This observation mirrors the approximate aggregation result in [Krusell and Smith \(1998\)](#). In contrast, when increasing the unemployment benefit and thus, the amount of redistribution and risk sharing, approximate aggregation fails because the solution becomes very sensitive to changes in the law of motion. In this case, the Krusell-Smith algorithm produces substantially more inequality compared to our algorithm which approximates the fully rational equilibrium. This leads to the conclusion that bounded rationality can in fact contribute to the modeling of realistic inequality. A second important economic result is that more risk sharing leads to more volatility in the cross-sectional distribution, which can be interpreted as higher systemic risk. This essentially mirrors the volatility paradox described in [Brunnermeier and Sannikov \(2014\)](#).

Overall, my approach provides a new tool to analyze numerical solutions of

DSGE models with ex-post heterogeneity and nonlinearities. Even though the algorithm was illustrated with a simple benchmark model in this paper, it's true potential lies in more challenging models. Especially the two results that bounded rationality might be a useful tool to produce more inequality in wealth or in other words a fat-tailed distribution of asset allocations and that better risk sharing can have the same effect are well worth exploring in future research. The result related to bounded rationality may in fact give a micro-level explanation of noise trading models. The result on risk sharing yields a way to endogenize systemic risk in terms of tightening or loosening the financial constraints of heterogeneous agents. Future research should, hence, be also aimed at extending the algorithm in this paper to accommodate more complex models, e.g., models with a higher number of variables and occasionally binding constraints. In terms of the latter, 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 former, one can explore the integration of sparse grids (see e.g., [Brumm and Scheidegger, 2016](#)).

APPENDIX A: PROOFS

A.1. *Proof of Proposition 4*

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

A.2. *Proof of Theorem 5*

Before proving Theorem 5, let me recall some notions from monotone operator theory which are important to our equilibrium problem. As outlined in Section 3, we want to find a minimizer of a convex constraint optimization problem by finding a root of the first-order conditions of the Lagrangian. To do so, we will establish some important properties of the Lagrangian and its first-order conditions. Let me first define what a saddle function is in this context.

DEFINITION 10 (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 $L : \mathcal{C} \times \mathcal{D} \rightarrow [-\infty, \infty]$ 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

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

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 $0 \in \mathbf{T}_L(c^*, d^*)$.¹⁰

Note that if our Lagrangian satisfies all properties of a saddle function, then the first-order conditions coincide with the operator \mathbf{T}_L . This operator can be further characterized by the following Corollary.

COROLLARY 11 (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.¹¹*

Maximal monotonicity is the key property which will lead to existence of a root of the first-order conditions $0 \in \mathbf{T}_L(h)$ for an optimal savings choice h .

COROLLARY 12 (Rockafellar (1969)) *Let \mathcal{C} be a Hilbert spaces over \mathbb{R} , and let $\mathbf{T} : \mathcal{C} \rightarrow \mathcal{C}^*$ be a maximal monotone operator. Suppose that there exists a subset $B \subset \mathcal{C}$ such that $0 \in \text{int}(\text{conv}(\mathbf{T}(B)))$. Then, there exists a $c \in \mathcal{C}$ such that $0 \in \mathbf{T}(c)$.*

REMARK This corollary implies that we need to find a subset B such that the interior of the convex hull of its image contains zero. Hence, it is enough to find two elements c_- and c_+ such that the image $\mathbf{T}(c_-)$ is negative and the image $\mathbf{T}(c_+)$ is positive.

¹⁰ 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 $d \in \mathcal{D}$.

¹¹ Maximal monotonicity (see e.g., Phelps, 1997; Bauschke and Combettes, 2017): Let \mathcal{E} be a Hilbert space. An operator $\mathbf{T} : \mathcal{E} \rightarrow \mathcal{E}$ 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 $(\tilde{e}, \tilde{f}) \in G(\mathbf{T})$.

We now have all ingredients to prove Theorem 5. In a first step, I show that the Lagrangian of the agent's optimization problem is indeed a saddle function according to Definition 10 and satisfies the conditions of Corollary 11.

LEMMA 13 (Saddle function) *Consider the growth model from Section 2. Consider the function space \mathcal{H} defined in Proposition 4 for the consumption and capital savings choice. The Lagrangian $L : \mathcal{H} \times L^2(\mathcal{Z}^{id} \times \mathbb{R}, \mathcal{B}(\mathcal{Z}^{id} \times \mathbb{R}), \mu) \rightarrow [-\infty, \infty]$ of the agents' optimization problem (7) in the growth model given by*

$$(20) \quad L(h, y) = -u(I(z', k, K)) + [1 - \rho]k - h \\ - \sum_{z'' \in \mathcal{Z}} p^{z''|z'} \beta u(I(z'', h, K')) + [1 - \rho]h - h' - yh$$

satisfies the conditions of Corollary 11 and therefore, its first-order condition constitutes a maximal monotone operator.

PROOF OF LEMMA 13: We have to show that the Lagrangian (20) of the agents' optimization problem (7) satisfies the conditions of Corollary 11.

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 control h . Hence, we define the first Hilbert space by \mathcal{H} as given in Proposition 4. The Lagrange multiplier lies 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 $y \geq 0$ for any exogenous shock and start capital.

Now, it remains to show that the Lagrangian is convex in the optimal control and concave in the Lagrange multiplier. The latter is trivial as the Lagrangian is linear in the multiplier. The former means that the second variation of the Lagrangian w.r.t. h is nonnegative (see e.g., Zeidler, 1986b, Corollary 42.8). Defining $\mathbf{C}(z'', h) = \mathbf{I}(z'', h) + [1 - \rho]h - h'$, we have

$$-u_{cc}(\mathbf{C}(z', k)) - \sum_{z'' \in \mathcal{Z}} p^{z''|z'} \beta \left\{ u_{cc}(\mathbf{C}(z'', h)) [\delta \mathbf{C}(z'', h; \tilde{h})]^2 \right. \\ \left. + u_c(\mathbf{C}(z'', h)) \delta^2 \mathbf{C}(z'', h; \tilde{h}) \right\}$$

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

$$\sum_{z'' \in \mathcal{Z}} p^{z''|z'} \beta \mathbf{C}(z'', h)^{-\gamma} \delta^2 \mathbf{C}(z'', h; \tilde{h}) \leq \gamma (\mathbf{C}(z', k)^{-(\gamma+1)} \\ + \sum_{z'' \in \mathcal{Z}} p^{z''|z'} \beta \mathbf{C}(z'', h)^{-\gamma} \frac{[\delta \mathbf{C}(z'', h; \tilde{h})]^2}{\mathbf{C}(z'', h)}).$$

This condition is equivalent to our condition (*ii*) of bounded second variation for $h \in \mathcal{H}$ and is therefore satisfied.

Properness: We proceed in two steps. First we show that there exists a Lagrange multipliers $y \in \mathcal{D}$ 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 $h \in \mathcal{H}$ such that $L(h, y) < \infty$ for all $y \in \mathcal{D}$. For any μ with aggregate capital $K > 0$, the productive income for any agent is positive $I > 0$. Fix a constant $0 < \epsilon < I$ and set $h = I - \epsilon$ for any exogenous shock and start capital. The nonnegativity of the Lagrange multiplier ensures $L(h, y) < \infty$.

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

The second step to prove Theorem 5 now applies Corollary 12.

PROOF OF THEOREM 5: After having shown in Lemma 13 that the conditions of Corollary 11 hold, it remains to construct a subset $B \in \mathcal{H}$ such that Corollary 12 applies. This ensures that there exists a solution to the first-order condition which due to the convexity of the optimization problem and the transversality of the equilibrium problem coincides with a recursive equilibrium itself. We construct two savings policies h at which the Euler equation is positive and negative, respectively. The idea is to use the two polar strategies save everything/consume nothing and save nothing/consume everything.

Let me first define the candidate policy

$$h(z', k, \mu) = (1 - \epsilon)(I(z', k, K) + (1 - \rho)k),$$

where productive income I is as in (3). This implies that aggregate capital and current and next-period consumption are given by

$$\begin{aligned} K' &= (1 - \epsilon)(\mathbb{E}^\mu[I(z', k, K)] + (1 - \rho)K) \\ c(z', k, \mu) &= \epsilon(I(z', k, K) + (1 - \rho)k) \\ c'(z', k, \mu) &= \epsilon[(1 - \epsilon)^\alpha I(z'', h|_{\epsilon=0}, K'|_{\epsilon=0}) + (1 - \epsilon)(1 - \rho)h|_{\epsilon=0}]. \end{aligned}$$

The first-order condition is given by

$$\frac{\partial}{\partial c} u(c) - \beta \sum_{z'' \in \mathcal{Z}} p^{z''|z'} (1 - \rho + R(z'', K')) \frac{\partial}{\partial c} u(c').$$

Positive FOC: This outcome is equivalent to

$$(21) \quad 1 > \beta \sum_{z'' \in \mathcal{Z}} p^{z''|z'} (1 - \rho + R(z'', K')) \left(\frac{c}{c'}\right)^\gamma,$$

where

$$\frac{c}{c'} = \frac{I(z', k, K) + (1 - \rho)k}{(1 - \epsilon)^\alpha I(z'', h|_{\epsilon=0}, K'|_{\epsilon=0}) + (1 - \epsilon)(1 - \rho)h|_{\epsilon=0}}.$$

We let $\epsilon \rightarrow 0$ which is equivalent to the save everything/consume nothing strategy. Clearly, this strategy is admissible $h \in \mathcal{H}$. Also, it is easy to see that c/c' is an increasing function in individual capital in this case. We can compute its limit by applying l'Hôpital's rule

$$\lim_{k \rightarrow \infty} \frac{c}{c'} = \frac{1}{1 - \rho + R(z'', K'|_{\epsilon=0})}.$$

If $\gamma = 1$, the right side of (21) equals $\beta < 1$ which results in the positive value of the first-order condition. When $\gamma > 1$, the right hand side is an increasing function of K' . It goes to zero when $K' \rightarrow 0$ and to $\beta(1 - \rho)(1 - \gamma)$ when $K' \rightarrow \infty$. This also results in a positive value of the first-order condition by assumption.

Negative FOC: This outcome is equivalent to

$$1 < \beta \sum_{z'' \in \mathcal{Z}} p^{z''|z'} (1 - \rho + R(z'', K')) \left(\frac{c}{c'}\right)^\gamma.$$

We now let $\epsilon \rightarrow 1$ which corresponds to the save nothing/consume everything strategy. It is obvious that $c/c' \rightarrow +\infty$ in this case which gives us a negative value of the first-order condition. It remains to check that this strategy is admissible, i.e. $h \in \mathcal{H}$. Condition (i) of Proposition 4 is trivial, whereas, condition (ii) requires more care. Clearly, the left side of condition (ii) equals zero. Let us now analyze the first and second variation of income at $\kappa = (1 - \epsilon)[I(z', k, K) + (1 - \rho)k]$ w.l.o.g. in the direction of $\tilde{\kappa}^\epsilon = (1 - \epsilon)\tilde{\kappa}$

$$\begin{aligned} \delta \mathbf{I}(z', \kappa; \tilde{\kappa}^\epsilon) &= \frac{\partial}{\partial K'} I(z', \kappa) \cdot \langle \tilde{\kappa}^\epsilon, \mathbf{1} \rangle + R(z', \kappa) \tilde{\kappa}^\epsilon \\ &= (1 - \epsilon)^\alpha \delta \mathbf{I}(z', \kappa|_{\epsilon=0}; \tilde{\kappa}), \end{aligned}$$

i.e., we can pull out ϵ from the variation. It follows that the first and second variation of income converge to zero for $\epsilon \rightarrow 1$. This shows that $h \in \mathcal{H}$ and concludes the proof. *Q.E.D.*

A.3. Proof of Theorem 7

We showed in Lemma 13 in Appendix A.2 that the operator \mathbf{T}_L associated with the Lagrangian of the optimization problem (7), i.e. its first-order condition, is maximal monotone. The reason for the importance of this property is that the resolvent¹² of a maximal monotone operator is firmly nonexpansive.¹³ This fact is due to Minty (1962). 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, 2017, Remark 4.34 (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. Hence, iterating on the resolvent of a maximal monotone operator yields the proximal point algorithm.

In Rockafellar (1976a), this convergence argument is made precise for Lagrange problems like the one we consider. The author furthermore shows that the proximal point algorithm converges to an optimum of the Lagrangian even if the update of the optimal policy is only approximate. Salzo and Villa (2012) extend this result to different concepts of approximation. Let me define which kind of approximation applies in this work.

DEFINITION 14 (Resolvent approximation¹⁴) *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 ϵ -precision of this resolvent at $c \in \mathcal{C}$ is defined as $\tilde{c} \in (\mathbf{Id} + \lambda\mathbf{T}_L^{\epsilon^2/(2\lambda)})^{-1}(c)$ where*

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

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

PROOF OF THEOREM 7: Generally, convergence of the proximal point algorithm is well established as explained above. It remains to show that our specific ap-

¹² Resolvent (see e.g., Bauschke and Combettes, 2017): Let \mathcal{E} be a Hilbert space. The resolvent of an operator $\mathbf{T} : \mathcal{E} \rightarrow \mathcal{E}$ is the operator $(\mathbf{Id} + \mathbf{T})^{-1}$ where \mathbf{Id} is the identity operator.

¹³ Nonexpansiveness (see e.g., Bauschke and Combettes, 2017): Let \mathcal{E} be a Hilbert space. An operator $\mathbf{T} : \mathcal{E} \rightarrow \mathcal{E}$ 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 $\|\mathbf{T}(e) - \mathbf{T}(\tilde{e})\|^2 \leq \langle e - \tilde{e}, \mathbf{T}(e) - \mathbf{T}(\tilde{e}) \rangle$.

¹⁴ This definition corresponds to the type 2 approximation with ϵ -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.

proximate policy update satisfies Definition 14. As $\max_{h \in \mathcal{H}} (h^{n+1} - h) = h^{n+1} \leq X(z', k, K)$, Equation (10) implies that

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

where

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

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

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

for all $h \in \mathcal{H}$. Note that $\tilde{L}(\cdot, y^n; h^n) = L(\cdot, y^{n+1})$ with L as in (20). Therefore, we have that $v \in \mathbf{T}_L^{\epsilon^2/(2\lambda)}(h^{n+1})$, which leads to $h^{n+1} \in (\mathbf{Id} + \lambda \mathbf{T}_L^{\epsilon^2/(2\lambda)})^{-1}(h^n)$ and concludes the proof. $Q.E.D.$

A.4. Proof of Proposition 8

PROOF: Let us denote the support of the marginal distribution w.r.t. k of the cross-sectional distribution by $\text{supp } \mu^k$. The minimum value of k , which has positive probability, is denoted by $\underline{k} = \min_k \text{supp } \mu^k$. First, let us show that the constraint has positive probability $\delta \in \text{supp } \mu^k$. 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 $k^* \notin \text{supp } \mu^k$. Now let us show that there is a mass point at δ . 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 $\bar{k} := \max\{k \geq \delta \mid h_2(\hat{z}, k, \mu) = k^*\} > k^*$ 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 δ . 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.5. Proof of Theorem 9

In order to proof Theorem 9, we first need to establish that any iterate of the optimal policy h^{n+1} , $n \geq 0$, as computed in the proximal point algorithm is analytic in the basic random variables ξ^z and ξ^k , i.e. there exist constants $\mathbf{c}_{h^{n+1},j}$ such that

$$\|D_j^p h^{n+1}\| \leq \mathbf{c}_{h^{n+1},j}^p p!, \quad p \in \{1, 2, \dots\}, \quad j \in \{z, k\},$$

where the p -th derivative is denoted by $D_j^p = \partial/\partial\xi^j$. Before I show analyticity, let me specify how exactly the policy depends on the basic variables. In our discretization of the optimal savings policy, we impose a grid on the exogenous shocks, the individual start capital and the projection coefficients. Hence, at each grid point of $h(z', k, \{\varphi_i\}_{i=0}^M)$, the states z' , the individual capital and the projection coefficients are fixed. However, the policy implicitly depends on the basic random variables through the Euler equation because it contains the next-period aggregate capital

$$K' = \sum_{\xi^z=1}^3 \int_{-\infty}^{\infty} h\left(z^{ag'}, \mathbb{1}_{\{\xi^z > 2 - z^{ag'}\}}, \kappa, \{\varphi_i\}_{i=0}^M\right) dF^k(\xi^k) dF^z(\xi^z),$$

where, due to (16),

$$\kappa = \begin{cases} \frac{1}{1-p^{e'}} \sum_{i=0}^M \varphi_i \Phi_i(\xi^z, \xi^k) & , \xi^z \leq 2 - z^{ag'} \\ \frac{1}{p^{e'}} \sum_{i=0}^M \varphi_i \Phi_i(\xi^z, \xi^k) & , \xi^z > 2 - z^{ag'} \end{cases}$$

is a function of the basic random variables. Therefore, the savings policy is a function of ξ^z and ξ^k .

PROPOSITION 15 (Analytic policies) *Under the assumptions of Theorem 9, all iterates of the the savings policy and the Lagrange multipliers as functions of ξ^z and ξ^k admit analytic extensions in the complex plane, namely in the region $\Sigma(\tau_{n+1}^j, \Gamma^j)$, $j \in \{z, k\}$, given in Theorem 9. Furthermore, it holds that the $(n+1)$ -th policy iterate*

$$\max_{x \in \Sigma(\tau_{n+1}^j, \Gamma^j)} |h^{n+1}(x)| \leq \frac{\min(1, \mathbf{L}_{n+1})}{\min(1, \mathbf{L}_{n+1}) - 2\tau_{n+1}^j A_{n+1,j}^1}.$$

is bounded in the region $\Sigma(\tau_{n+1}^j, \Gamma^j)$, $j \in \{z, k\}$.

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

Real analytic: Equation (10) implies that the $(n + 1)$ -th iterate of the savings policy h^{n+1} in the proximal point algorithm solves the following first-order condition

$$(22) \quad X \frac{\partial}{\partial h^{n+1}} L^A(h^{n+1}, y^n; h^n) = \mathbf{e}$$

with constant $\|\mathbf{e}\| \leq \frac{\xi^2}{2\lambda}$ for any fixed exogenous shock and start capital (z', k) . Now, let us take the derivatives of the first-order condition (22) w.r.t. ξ^z and ξ^k . It is obvious that X and \mathbf{e} do not depend on the basic random variables. The partial derivative of the augmented Lagrangian, however, does due to its dependence on K' and because the optimal policies and hence, also the Lagrange multipliers depend on ξ^z and ξ^k as can be seen in

$$(23) \quad \begin{aligned} \frac{\partial}{\partial h^{n+1}} L^A(h^{n+1}, y^n; h^n) &= \frac{\partial}{\partial c} u(I(z', k, K) + [1 - \rho]k - h^{n+1}) \\ &\quad - \beta \sum_{z'' \in \mathcal{Z}} \left\{ p^{z''|z'} [1 - \rho + R(z'', K')] \right. \\ &\quad \left. \frac{\partial}{\partial c} u(I(z'', h^{n+1}, K') + [1 - \rho]h^{n+1} - h^{(n+1)'}) \right\} \\ &\quad + \frac{1}{\lambda} (h^{n+1} - h^n) - \mathbb{1}_{\{h^{n+1} \leq \frac{y^n}{\lambda}\}} (y^n - \lambda h^{n+1}). \end{aligned}$$

We now investigate the derivatives of (22) w.r.t. the basic random variable ξ^j , $j \in \{z, k\}$. Trivially,

$$D_j^p \left(\frac{\partial}{\partial h^{n+1}} L^A \right) = 0, \quad p \in \{1, 2, \dots\}.$$

It follows that

$$(24) \quad D_j^{p-1} \left(\frac{\partial^2 L^A}{\partial h^{n+1} \partial K'} D_j^1 K' \right) + D_j^{p-1} \left(\frac{\partial^2 L^A}{[\partial h^{n+1}]^2} D_j^1 h^{n+1} \right) = \frac{1}{\lambda} D_j^p h^n + \mathbb{1}_{\{h^{n+1} \leq \frac{y^n}{\lambda}\}} D_j^p y^n.$$

Let us first analyze the derivative of K' . It is easy to see from $D_j^1 K' = \partial/\partial \kappa K' D_j^1 \kappa$ that all derivatives of K' are composed of the derivatives of the optimal policy w.r.t. start capital and the derivatives of κ w.r.t. the basic random variables. The latter component is obviously analytic. Hence, K' is analytic in the basic random variables if the optimal policy is analytic in start capital. This fact is

easily established by induction when taking derivatives of the first-order condition w.r.t. k and taking into account that h^0 is analytic in k . We exploit the fact that products, sums and compositions of analytic functions are analytic. Hence, there is a $\mathbf{c}_{K',j}$ such that

$$\|D_j^p K'\| \leq \mathbf{c}_{K',j}^p p!, \quad p \in \{1, 2, \dots\}.$$

Furthermore, note that it follows from (23) that $\partial/\partial h^{n+1} L^A$ is analytic in K' and h^{n+1} .

We can now show analyticity of the optimal policy by induction in two dimensions: First increasing the iterate of the policy, then increasing the order of the derivative. Assume that all iterates h^j , $j \leq n$, are analytic and that the derivatives of h^{n+1} w.r.t. the basic random variables up to order $p - 1$ are bounded as required for analyticity. This implies that the derivatives w.r.t. K' and h^{n+1} of the first-order condition also satisfy the analyticity condition up to order $p - 1$ with coefficients $\mathbf{c}_{L_{hK'}^A, j}$ and $\mathbf{c}_{L_{hh}^A, j}$. W.l.o.g., choose $\mathbf{c}_{L_{hK'}^A, j} \geq \mathbf{c}_{L_{hh}^A, j}$. Applying the product rule, we rewrite (24) as

$$\begin{aligned} \frac{\partial^2 L^A}{[\partial h^{n+1}]^2} D_j^p h^{n+1} &= \frac{1}{\lambda} D_j^p h^n + \mathbb{1}_{\{h^{n+1} \leq \frac{y^n}{\lambda}\}} D_j^p y^n \\ &\quad - \sum_{l=0}^{p-1} \binom{p-1}{l} D_j^{p-l} K' D_j^l \left(\frac{\partial^2 L^A}{\partial h^{n+1} \partial K'} \right) \\ &\quad - \sum_{l=1}^{p-1} \binom{p-1}{l} D_j^{p-l} h^{n+1} D_j^l \left(\frac{\partial^2 L^A}{[\partial h^{n+1}]^2} \right). \end{aligned}$$

Dividing by $p!$, taking norms and denoting $R_{n+1,j}^p = \|D_j^p h^{n+1}\|/p!$ leads to

$$\begin{aligned} \underbrace{\frac{\partial^2 L^A}{[\partial h^{n+1}]^2} R_{n+1,j}^p}_{=: \mathbf{L}_{n+1}} &\leq \frac{1}{\lambda} \mathbf{c}_{h^n, j}^p + \mathbb{1}_{\{h^{n+1} \leq \frac{y^n}{\lambda}\}} \mathbf{c}_{y^n, j}^p + \sum_{l=0}^{p-1} \mathbf{c}_{K', j}^{p-l} \mathbf{c}_{L_{hK'}^A, j}^l \\ &\quad + \sum_{l=1}^{p-1} R_{n+1,j}^{p-l} \mathbf{c}_{L_{hh}^A, j}^l \\ (25) \quad &\leq \underbrace{\frac{1}{\lambda} \mathbf{c}_{h^n, j}^p + \mathbb{1}_{\{h^{n+1} \leq \frac{y^n}{\lambda}\}} \mathbf{c}_{y^n, j}^p + \max \left(2\mathbf{c}_{K', j}, 2\mathbf{c}_{L_{hK'}^A, j} \right)^p}_{=: A_{n+1,j}^p} \\ &\quad + \sum_{l=1}^{p-1} R_{n+1,j}^{p-l} \mathbf{c}_{L_{hh}^A, j}^l. \end{aligned}$$

Note that, due to convexity, $\mathbf{L}_{n+1} > 0$. Solving this recursion yields

$$R_{n+1,j}^p \leq \frac{A_{n+1,j}^p}{\mathbf{L}_{n+1}} + \sum_{l=0}^{p-1} \frac{A_{n+1,j}^{p-1-l}}{\mathbf{L}_{n+1}} 2^l \mathbf{c}_{L_{hh}^A}^{l+1} \leq 2^p \frac{A_{n+1,j}^p}{\mathbf{L}_{n+1}} \leq \left(\frac{2A_{n+1,j}^1}{\min(1, \mathbf{L}_{n+1})} \right)^p,$$

where $A_0 = \|\mathbf{e}\|/X$. Hence, we obtain a uniform bound for all derivatives of the optimal policy. Analyticity follows by induction.

Complex continuation: We define the following power series for the $(n+1)$ -th iterate in terms of the basic random variable ξ^j , $j \in \{z, k\}$, on the complex plane

$$h^{n+1}(x) = \sum_{p=0}^{\infty} \frac{(x - \xi^j)^p}{p!} D_j^p h^{n+1}(\xi).$$

Taking norms leads to

$$|h^{n+1}(x)| = \sum_{p=0}^{\infty} \left(|x - \xi^j| \frac{2A_{n+1,j}^1}{\min(1, \mathbf{L}_{n+1})} \right)^p.$$

This series converges for all $|x - \xi^j| \leq \tau_{n+1}^j < \frac{\min(1, \mathbf{L}_{n+1})}{2A_{n+1,j}^1}$ such that

$$|h^{n+1}(x)| \leq \frac{\min(1, \mathbf{L}_{n+1})}{\min(1, \mathbf{L}_{n+1}) - 2\tau_{n+1}^j A_{n+1,j}^1}.$$

Therefore, by continuation the iterates can be extended analytically in the whole region $\Sigma(\tau_{n+1}^j, \Gamma^j)$, 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 variables 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)

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

where constant \mathbf{b} is independent of the order of truncation M . Given that h admits an analytic extension on the complex plane $\Sigma(\tau_{n+1}^j, \Gamma^j)$, the best approximation error is bounded by (see Babuška et al., 2007, Lemma 4.4)

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

where

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

Combining this with Proposition 15 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.*

APPENDIX B: 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 choice h is then determined such that the following condition is satisfied

$$L^A(h^{n+1}, y^n; h^n) \leq \hat{\phi}^{n+1} = \min_h \phi^{n+1}(h),$$

where ϕ^{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 $q > 3/2$.

Algorithm 2 Accelerated 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 $H^n = (h^n, y^n)$. Set $\nu^n = h^n$.
- 2: Set the parameters $\lambda > 0$, $A^n > 0$ and $b \in [0, 2)$.
- 3: Set the resolvent approximation precision $\{\epsilon^n\}_{n=0}^\infty$.
- 4: Set the termination criterion small $\tau > 0$ and the initial distance larger $d > \tau$.

▷ *B Iterative procedure*

- 5: **while** $d > \tau$ **do**
- 6: Update $\alpha^n = \frac{1}{2} \left(\sqrt{(b\lambda A^n)^2 + 4b\lambda A^n} - b\lambda A^n \right)$.
- 7: Update $x^n = (1 - \alpha^n)h^n + \alpha^n \nu^n$.
- 8: Update H^{n+1} by

$$h^{n+1} \approx \arg \min_h L^A(h, y^n; x^n)$$

$$y^{n+1} = \max(0, y^n - \lambda h^{n+1})$$

where L^A is defined as in (9).

- 9: Update $A^{n+1} = (1 - \alpha^n)A^n$.
- 10: Update $\nu^{n+1} = \nu^n - \frac{\alpha^n}{(1-\alpha^n)\lambda A^n}(x^n - h^{n+1})$.
- 11: Compute the distance $d = \|H^{n+1} - H^n\|$.
- 12: Set $n = n + 1$.
- 13: **end while**

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, 2nd edition, 2017.
- 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.
- Johannes Brumm, Dominika Kryczka, and Felix Kubler. Recursive equilibria in dynamic economies with stochastic production. *forthcoming in Econometrica*, 2017.
- Markus K. Brunnermeier and Yuliy Sannikov. A macroeconomic model with a financial sector. *The American Economic Review*, 104(2):379–421, 2014.
- 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.
- George M. Constantinides and Darrell Duffie. Asset pricing with heterogeneous consumers. *Journal of Political Economy*, 104(2):219–240, 1996.
- 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.
- 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. *American Economic Review (forthcoming)*, 2017.
- Aubhik Khan and Julia K. Thomas. Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica*, 76(2):395–436, 2008.
- 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. Local boundedness of nonlinear, monotone operators. *The Michigan Mathematical Journal*, 16(4):397–407, 1969.
- 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.
- N. L. Stokey, R. E. Lucas, and E. C. Prescott. *Recursive Methods in Economic Dynamics*. Harvard University Press, 1989.
- Kjetil Storesletten, Christopher I. Telmer, and Amir Yaron. Asset pricing with idiosyncratic risk and overlapping generations. *Review of Economic Dynamics*, 10(4):519–548, October 2007.
- 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.