Quantitative Economics 11 (2020), 1253–1288 1759-7331/20201253

Solving discrete time heterogeneous agent models with aggregate risk and many idiosyncratic states by perturbation

Christian Bayer
Department of Economics, Universität Bonn, CEPR, and IZA

Ralph Luetticke
Department of Economics, University College London, CEPR, and CFM

This paper describes a method for solving heterogeneous agent models with aggregate risk and many idiosyncratic states formulated in discrete time. It extends the method proposed by Reiter (2009) and complements recent work by Ahn, Kaplan, Moll, Winberry, and Wolf (2017) on how to solve such models in continuous time. We suggest first solving for the stationary equilibrium of the model without aggregate risk. We then write the functionals that describe the dynamic equilibrium as sparse expansions around their stationary equilibrium counterparts. Finally, we use the perturbation method of Schmitt-Grohé and Uribe (2004) to approximate the aggregate dynamics of the model.

Keywords. Numerical methods, heterogeneous agent models, linearization, incomplete markets.

JEL classification. C63, E32.

1. Introduction

Models of heterogeneous agents have become widespread in macroeconomics, at least since Krusell and Smith (1997, 1998) developed the first widely applicable algorithm to solve them in an environment of aggregate risk. Yet, their use has been limited initially by the computational resources needed to solve these models. Over the last decade, substantial progress has been made in developing algorithms that can solve these models more efficiently.[1] One of the most popular and powerful of these methods was originally developed by Reiter (2002, 2009). This method extends perturbation methods to heterogeneous agent environments, that is, it builds on the methods often used to solve

Christian Bayer: christian.bayer@uni-bonn.de
Ralph Luetticke: r.luetticke@ucl.ac.uk

We would like to thank Benjamin Born and Philip Jung as well as conference and seminar participants at AMCM 2019 in Lillehammer, PASC 2019 in Zurich, Universitat Autonoma Barcelona, Universität Konstanz, and University of Oxford, for helpful comments and suggestions as well as Seungcheol Lee and Seungmoon Park for translating our codes into Python for the HARK toolkit (github.com/econ-ark/HARK). Christian Bayer gratefully acknowledges support through the ERC-CoG project Liquid-House-Cycle funded by the European Union's Horizon 2020 Program under grant agreement No. 724204.

[1] See, for example, the JEDC comparison project: Den Haan, Judd, and Juillard (2010), Den Haan (2010b), Reiter (2010b), Young (2010), Maliar, Maliar, and Valli (2010), Kim, Kollmann, and Kim (2010), Algan (2010), Den Haan and Rendahl (2010).

 https://doi.org/10.3982/QE1243

dynamic stochastic general equilibrium models with a representative agent (see, e.g., Schmitt-Grohé and Uribe (2004)). Our paper restates this procedure and additionally shows how the necessary dimensionality reduction of the heterogeneous agent model can be achieved in a new, intuitive way.

The extension of perturbation methods to heterogeneous agent models relies on writing the model in the form of a nonlinear difference equation that is function-valued instead of vector-valued (as in representative agent models). This equation is then (linearly) approximated around the stationary equilibrium of the heterogeneous agent model without aggregate risk. The (at least) two functionals that enter the difference equation are the distribution of agents over idiosyncratic states (e.g., the wealth distribution) and the function (value or policy function) that describes the optimal individual behavior. These functionals can be seen as replacements for the aggregate capital accumulation and consumption Euler equation in representative agent models. These replacements allow us to maintain all nonlinearity with respect to microeconomic shocks—yet obtaining a model that is linear in aggregate variables.

While all of this is straightforward in theory, the key practical issue is how to approximate the functionals involved because they need to be replaced by finite-dimensional objects for the actual computation of the model's dynamics. In particular, when the individual planning problem is rich insofar as it has many idiosyncratic states, this issue is severe. The curse of dimensionality implies that it is hard to come up with a small enough finite-dimensional representation of the distribution function and the value/policy function without having any a priori knowledge of their shape.

However, the solution of the stationary equilibrium provides us with such knowledge in most practical cases. Therefore, we propose a dimensionality reduction step *after* the stationary equilibrium of the economy (i.e., without aggregate risk) has been determined, but *before* perturbing the system. This dimensionality reduction is adaptive and takes into account the shape of the distribution and value function in the stationary equilibrium. As a result, the stationary equilibrium can be computed without taking into account that the goal is to solve for aggregate dynamics in the end.

In detail, we suggest using sparse expansions of value and distribution functions around their nonsparse stationary equilibrium counterparts. First, we write the value function in the stationary equilibrium as a sum of a full set of basis functions and determine the coefficients on these. We then allow only those coefficients of the basis functions to vary outside the stationary equilibrium that are large in the stationary equilibrium while we keep all small coefficients at their stationary equilibrium values. This is analogous to lossy video compression where the compressed video stream is coded by strongly compressing the difference to a lightly compressed reference frame. In fact, we borrow further from image compression by writing the value functions in the form of their discrete cosine transform (i.e., as Chebyshev polynomials on the Chebyshev nodes). Second, we split the high-dimensional distribution function into the histograms of its marginals and their (joint) copula. As a baseline, we suggest keeping the copula fixed at its stationary equilibrium value. This, as a second dimensionality reduction, picks up the idea of Krusell and Smith (1997, 1998) that not all moments of the wealth-income distribution are equally important for price formation and, therefore, relevant

for the equilibrium dynamics. The assumption of a fixed copula implies that the rank correlation among, say, wealth in various kinds of assets and income is time constant without imposing any restriction on changes in the shape of the marginal distributions. However, one can also treat the copula as time varying, applying the same dimensionality reduction for the copula as we do for the value/policy functions, that is, using the discrete cosine transforms.

Concretely, we show, both for an incomplete markets model with one asset and for a model with two assets, that the assumption of a fixed copula has little impact on the model dynamics but substantially speeds up the computation. The largest share of the computation time falls on the calculation of the stationary equilibrium followed by the calculation of the derivatives of the nonlinear difference equation. However, both can be sped up by parallelization. At any rate, the models we consider can all be solved on a standard desktop computer in a matter of seconds or minutes using our algorithm.

By reducing the dimensionality after the solution of the stationary equilibrium but before linearizing, our method differs from existing proposals. The original proposal by Reiter (2002) was to represent distribution functions by histograms without any dimensionality reduction and to write value functions (or other functionals describing the dynamic planning problem) as finite-dimensional parametric objects—for example, by using splines. However, when the individual planning problem is rich insofar as it has many idiosyncratic states, this procedure can become inaccurate and in many cases even infeasible to solve numerically. The first idea to tackle this issue was to be as sparse as possible in the parametric approximation of functions when solving for the stationary equilibrium (see, e.g., Reiter (2009)), for example, through sparse grid methods in the dynamic planning problem (see, e.g., Bungartz and Griebel (2004), Krueger and Kubler (2004)) and by using mixtures of parametric distributions as proposed by Winberry (2018). In other words, these methods rely on achieving dimensionality reduction ex ante, before solving for the stationary equilibrium, and hence impose a numerical constraint on this solution. The analogy of this is still in image compression, or the compression of a sequence of images picture-by-picture, which is in general inefficient for video compression because of many nonmoving parts. For a dynamic equilibrium model, this analogy carries over: Many aspects of value and policy functions do not change much with aggregate shocks, such that the stationary equilibrium functions are good "reference frames."

An alternative attack, also suggested by Reiter (2010a), is to use singular value decomposition for dimensionality reduction of the Jacobian of the system *after* linearizing the difference equation but before solving it. Ahn et al. (2017) developed this approach further in that they write the planning problem in continuous time and suggest using automatic differentiation in order to obtain a sparse Jacobian. This helps with both the memory requirements, and with the computing time for both the singular value decomposition and the solving of the difference equation itself. In addition, they suggest perturbing the deviations of value and distribution functions from their stationary equilibrium counterparts instead of perturbing the functions themselves. This allows for different parametric classes for deviations and stationary equilibrium functionals. As a result, it decouples the number of perturbed parameters from the number of parameters used

in the approximation of the functions in the stationary equilibrium (which can potentially be richer). Our approach shares the latter aspect with the approach of Ahn et al. (2017). Compared to their method, ours has the advantage of avoiding the calculation of a very large Jacobian because the dimensionality is reduced before this step. Thus, it can be applied to models formulated in discrete time, where the Jacobian would otherwise be too nonsparse to be efficiently stored in a PC's memory. Another advantage is that this allows us to calculate second-order (or higher) perturbations, because the number of (higher-order) derivatives to be calculated does not increase too fast. Concretely, we provide an example where it takes a few minutes to calculate a brute-force second-order perturbation solution to the Krusell and Smith (1998) model—brute-force in the sense that we do not exploit prior theoretical knowledge of some second-order derivatives being zero.

The remainder of the paper is organized as follows: Section 2 defines the generic model we aim to solve with our method and lays out the solution method itself. Section 3 provides the economic model of two application examples: first, a standard incomplete markets model with just a single asset, capital, as in Krusell and Smith (1998); second, an extension of that model, in which households have to choose between assets of different liquidity. They can hold a liquid nominal asset or illiquid capital. We add a nominal rigidity to this model, such that it is of the New-Keynesian flavor. For the first model variant, we can compare our solution to the original Krusell and Smith (1998) algorithm and to the standard Reiter (2009) approach. Our method is equally as precise as Reiter's standard approach but faster. It is faster and slightly less precise than the Krusell and Smith algorithm in our example. The second model variant is too rich in terms of aggregate states to be solved by Reiter's standard approach. Here, we only show that simulating the model along the lines of Den Haan's (2010a) test proves the method to be accurate, that business cycle properties of the model change little when perturbing a larger set of DCT-coefficients, and that the model produces realistic asset return premia and a reasonably good approximation of asset market clearing. Section 5 concludes. Example codes are provided in the Appendix (Bayer and Luetticke (2020)).

2. Method

We consider a generic economy with a continuum of heterogeneous agents and aggregate risk. We first define the objects we need to work with. Thereafter, we define a stationary equilibrium and a sequential equilibrium (with recursive individual planning) for this economy. Then we describe how the sequential equilibrium can be solved for locally and how a reduction of the state space can be achieved. Finally, we give an overview of the suggested algorithm.

2.1 *Prerequisites and notation*

Let $S_t \in \mathbb{R}^n$ denote the aggregate states in this economy other than the distribution of agents over their idiosyncratic states $s_{it} \in \mathbb{R}^m$ for individual i at time t. In a representative agent model, these S_t would be the only state variables. With heterogeneous agents,

the distribution *function* μ_t of agents over s_{it} is also part of the aggregate states of the economy but for notational purposes shall not be included in S_t.

Both S_t and s_{it} shall be partitioned into an exogenous stochastic and an endogenous deterministic component

$$S_t = \begin{bmatrix} X_t \\ D_t \end{bmatrix}, \qquad s_{it} = \begin{bmatrix} x_{it} \\ d_{it} \end{bmatrix}, \tag{1}$$

with length $n = n_x + n_d$ and $m = m_x + m_d$, respectively.

With stochastic elements in S_t and s_{it}, agents in the economy face both aggregate and idiosyncratic risk. We denote the stochastic elements of the aggregate and idiosyncratic state space by X_t and x_{it}, respectively. We assume that all stochastic variables follow a stationary Markov chain, such that

$$X_{t+1} = H^X(X_t) + \varepsilon_{t+1}, \qquad x_{it+1} = h^x(x_{it}) + \epsilon_{it+1}, \tag{2}$$

and the innovations ε_{t+1}, ϵ_{it+1} have variances $\omega\Omega$ and $\sigma\Sigma$ for the aggregate and idiosyncratic variables, respectively.

The remaining idiosyncratic state variables d_{it} are chosen by households in order to maximize their utility. This choice shall be described by the generic planning problem

$$v(x_{it}, d_{it}, S_t, \mu_t) = \max_{d_{it+1}} u(x_{it}, d_{it}, d_{it+1}; P_t) + \beta \mathbb{E} v(x_{it+1}, d_{it+1}, S_{t+1}, \mu_{t+1}), \tag{3}$$

subject to $d_{it+1} \in \Gamma(x_{it}, d_{it}, P_t)$ where Γ is a budget set and $P_t = P(X_t, D_t, \mu_t)$ is a pricing kernel.[2] Prices may result from market clearing in the sense introduced below, but may also be directly determined by the aggregate state or the distribution, such as, for example, interest rates set by the central bank or the wage rate as a function of the aggregate amount of capital. The further aggregate states move for simplicity according to some given law of motion $D_{t+1} = H^D(X_t, D_t, \mu_t)$.[3] Note that this does not preclude prices from also depending on choices for state variables D_{t+1} made at time t because we can write these as functions of states in t.

It will come in handy later to simplify notation for the Bellman equation by observing that, from the individual's point of view, aggregates and distributions only matter through prices. These, in turn, we can summarize by adding a time index t to the value functions. Dropping the indexes to the *idiosyncratic* states and using $'$ to denote the next period variables, we can write the individual planning problem recursively as

$$v_t(x, d) = \max_{d' \in \Gamma_t(x,d)} u_t\big(x, d, d'\big) + \beta \mathbb{E} v_{t+1}\big(x', d'\big), \tag{4}$$

where the time index here stands for conditioning the individual planning problem and the pricing kernel on all state variables of time t. Individual policy functions h_t^d can be defined accordingly.

[2]Potentially, prices indirectly enter into the utility function because they may change the mapping of states to consumption.

[3]The law of motion H^D can be the outcome of some other aggregate planning problem as well. Importantly, it is neither stochastic nor influenced by a single individual decision.

To close the model, we need a description of market clearing. We define an excess demand function $\Phi_t(h_t^d, \mu_t)$ that maps the idiosyncratic policies and the distribution, as well as prices and aggregate states (captured by the time index), into a real vector. Typically, we have as many prices as idiosyncratic endogenous states, given that we assumed an exogenous law of motion for aggregate states, that is, $\Phi_t(h_t^d, \mu_t) \in \mathbb{R}^{m_d}$.

For example, in an economy as in Krusell and Smith (1998), that is, with capital and aggregate productivity risk, Φ is given by the difference between the marginal product of capital and the rate of return on capital. In a bond economy with only IOUs, in contrast, we would have $\Phi = \int h_t^d(s)\, d\mu_t$ (a time constant Φ), and in an economy with government bonds this would be $\Phi_t = \int h_t^d(s)\, d\mu_t - B_t$, where B_t is the amount of government bonds issued and circulating in t (such that Φ changes in aggregates).

2.2 *Stationary equilibrium and approximate solution*

Since the method developed by Reiter (2009) approximates the aggregate dynamics around the stationary equilibrium, we first consider an economy without aggregate risk, that is, where $\omega = 0$. For such an economy, prices, distributions, and hence value functions do not change over time, and we can define a stationary equilibrium generically as follows.

DEFINITION 1. A *stationary equilibrium* is a value function $\overline{v}$, a distribution function $\overline{\mu}$, a policy function $\overline{h}^d(s)$, and prices $\overline{P}$ such that:

1. The individual policy $\overline{h}^d(s)$ is the maximizer of the Bellman equation (3) given $\overline{P}$,
$$\overline{h}^d(x, d) = \arg \max_{d' \in \Gamma_{\overline{P}}(x,d)} u\big(x, d, d'\big) + \beta \mathbb{E}\overline{v}\big(x', d'\big). \tag{5}$$
2. The value function solves the Bellman equation (3) given the individual policy $\overline{h}^d(s)$.
3. Markets clear, that is, $\Phi(\overline{h}^d, \overline{\mu}) = 0$.
4. The distribution $\overline{\mu}$ is the stationary distribution of the Markov chain induced by $\overline{h}(s, \epsilon) := \left[\begin{smallmatrix} h^x(s)+\epsilon \\ \overline{h}^d(s) \end{smallmatrix}\right]$.

To solve for the equilibrium, it is necessary to approximate the model. Typically, the model is solved for a (full tensor) grid of points in $\mathbb{R}^m$ replacing the functionals by some parametric approximation. A common approach is, for example, to replace the value functions with splines with the nodes of the spline being equal to the grid points. When first-order conditions are sufficient and the problem is differentiable, we can replace the Bellman equation with an Euler equation to describe the planning problem. Since the techniques to find the equilibrium value functions are standard, we refer only to the literature here (see, e.g., Carroll (2006), Hintermaier and Koeniger (2010)).

Similarly, the distribution is often approximated by a step function (the density being replaced by a point mass) on the grid or by a piecewise linear function (the density

function being a step function).[4] Since policy functions map potentially into nongrid points, a standard technique is to introduce some trembling to the policy function such that policies fall on neighboring grid points with such probabilities that the off-grid policy equals the expected value of the tremble; see Young (2010).[5]

Under these assumptions, the dynamics of the wealth distribution can be described by the point-mass, in short a histogram, $d\mu$, replacing the density, and a transition matrix $\Pi_{\overline{h}}$ induced by the policy function $\overline{h}$. In the stationary economy,

$$d\overline{\mu} = d\overline{\mu}\Pi_{\overline{h}} \tag{6}$$

needs to hold. This is the discrete time analogue to the Kolmogorov forward/Fokker–Planck equation in continuous time systems. For a given transition probability matrix, that is, for a given policy function, the stationary distribution can then be calculated efficiently by determining the eigenvector of $\Pi_{\overline{h}}$ to the eigenvalue 1. Similarly, if we assume that the value function is replaced by a linear interpolant, we obtain the result that the solution to the Bellman equation is given by a finite vector of values, with a slight abuse of notation also denoted by $\overline{v}$, which needs to satisfy

$$\overline{v} = u_{\overline{h}^d} + \beta\Pi_{\overline{h}}\overline{v}, \tag{7}$$

where $u_{\overline{h}^d}$ is the period payoff under the optimal policy.[6]

In the following, we assume that the stationary equilibrium is solved for in this way on a full tensor grid, because these methods are readily available and easy to implement, and their application is, in most cases, not constrained by memory availability even on desktop computers. However, the method laid out below extends readily to the case where the stationary equilibrium is solved for by sparse grid methods; see Bungartz and Griebel (2004) or Krueger and Kubler (2004).

2.3 *Sequential equilibrium with recursive individual planning*

If there is uncertainty regarding the aggregate states, value functions written as functions of idiosyncratic states are no longer time constant in equilibrium. The same holds true for the distribution functions. Instead, if the model is stationary, value functions and distributions will converge to a sequence that fulfills the following equilibrium conditions.[7]

[4]We follow Young (2010) in using the point-mass approach throughout and understand the word "histogram" as a synonym for point-mass distributions.

[5]If one wants to read this in a strict way, then we assume that the individual planner can choose only mixed strategies over two neighboring grid points and that the current payoffs depend on the two grid points and the relative probability weights chosen. Then the solution with linear interpolation is an exact solution to the described surrogate planning problem.

[6]If first-order conditions are sufficient such that, say, a standard consumption Euler equation holds, we can also work with

$$\overline{u}'_{\overline{h}^d} = \beta(1+r)\Pi_{\overline{h}}\overline{u}'_{\overline{h}^d},$$

instead of (7), where $\overline{u}'_{\overline{h}}$ is the marginal utility of consumption under the optimal policy.

[7]Note that we write the problem still in recursive form from a household's point of view.

DEFINITION 2. A *sequential competitive equilibrium with recursive individual planning* is a sequence of value functions ν_t, a sequence of distribution functions μ_t, a sequence of policy functions $h_t^d(s)$, a sequence of aggregate states S_t, and a sequence of prices P_t such that at each point in time t:

1. The individual policy is the maximizer of the Bellman equation (3) given the prices P_t,
$$h_t^d(x,d) = \arg\max_{d' \in \Gamma(x,d;P_t)} u\big(x,d,d';P_t\big) + \beta \mathbb{E}\nu_{t+1}\big(x',d'\big). \tag{8}$$
2. The value function solves the Bellman equation (3) given the individual policy h_t^d and the expected continuation value ν_{t+1}.
3. Markets clear, that is, $\Phi_t(h_t^d, \mu_t, P_t, S_t) = 0$.
4. The distribution μ_{t+1} is induced by $h_t(s,\epsilon) := \left[\begin{smallmatrix} h^x(s)+\epsilon_t \\ h_t^d(s) \end{smallmatrix}\right]$ and the distribution μ_t.
5. The sequence of aggregate states is induced by $\left[\begin{smallmatrix} X_{t+1} \\ D_{t+1} \end{smallmatrix}\right] = \left[\begin{smallmatrix} H^X(X_t,D_t)+\varepsilon_{t+1} \\ H^D(X_t,D_t,\mu_t) \end{smallmatrix}\right]$

Again, we need to approximate the functions involved in the model in a suitable way to solve the model. For that purpose, we replace the distribution function by a histogram and add trembles to the policy. Finally, we write the value function as a linear interpolant. This implies that the discrete time Fokker–Planck equation (6) takes the form

$$d\mu_{t+1} = d\mu_t \Pi_{h_t}, \tag{9}$$

which makes clear its forward equation character. Further, note that due to the continuum-of-agents assumption, there is no randomness in the transition other than through aggregate states and, therefore, shocks changing h_t. The Bellman equation (7) now takes the form

$$\nu_t = u_{h_t^d} + \beta \Pi_{h_t} \nu_{t+1}, \tag{10}$$

where $u_{h_t^d}$ is the period payoff under the optimal policy at time t.

Combining these equilibrium conditions, we can summarize the sequential equilibrium conditions by the nonlinear difference equation given by

$$F(d\mu_t, S_t, d\mu_{t+1}, S_{t+1}, \nu_t, P_t, \nu_{t+1}, P_{t+1}, \varepsilon_{t+1}) = \begin{bmatrix} d\mu_{t+1} - d\mu_t \Pi_{h_t} \\ X_{t+1} - H^X(X_t, D_t) + \varepsilon_{t+1} \\ D_{t+1} - H^D(X_t, D_t, d\mu_t) \\ \nu_t - (u_{h_t^d} + \beta \Pi_{h_t} \nu_{t+1}) \\ \Phi_t\big(h_t^d, d\mu_t\big) \\ \varepsilon_{t+1} \end{bmatrix} \tag{11}$$

s.t.

$$h_t^d(s) = \arg\max_{d' \in \Gamma(x,d;P_t)} u\big(x,d,d';P_t\big) + \beta \mathbb{E}\nu_{t+1}\big(x',d'\big). \tag{12}$$

A sequential equilibrium now fulfills

$$\mathbb{E}_t F(d\mu_t, S_t, d\mu_{t+1}, S_{t+1}, \nu_t, P_t, \nu_{t+1}, P_{t+1}, \varepsilon_{t+1}) = 0. \tag{13}$$

For notational simplicity, it is useful to define, $\hat{S}_t := [d\mu_t \; X_t \; D_t]'$ as all the aggregate states of this system, including the distribution, and $\hat{C}_t := [\nu_t \; P_t]'$ as all the controls of the system, that is, prices and value functions; to be more precise, their function values at the grid points (nodes). Again if we are working with first-order conditions, value functions might be replaced with marginal utilities.

2.4 *Approximating the sequential equilibrium around the stationary equilibrium*

There are various ways to solve the nonlinear difference equation, $\mathbb{E}_t F = 0$, by perturbation methods. Here, we follow Klein (2000) and Schmitt-Grohé and Uribe (2004), who show how to solve the system (11) by first- and second-order perturbation. These methods can be readily applied here as well, choosing the stationary equilibrium solution as the point around which to perturb the system, as in Reiter (2002).

For expositional purposes, we focus on first-order perturbation here. This means that it is necessary to calculate the Jacobian matrix of the system (dropping ε_{t+1}), $J = [F_{\hat{S}} \; F_{\hat{S}'} \; F_{\hat{C}} \; F_{\hat{C}'}]$, and solve the linearized difference equation by relating its solution to the generalized eigenvalue problem

$$\underbrace{\begin{bmatrix} F_{\hat{S}'} & F_{\hat{C}'} \end{bmatrix}}_{A:=} Z\Lambda = \underbrace{-\begin{bmatrix} F_{\hat{S}} & F_{\hat{C}} \end{bmatrix}}_{B:=} Z, \tag{14}$$

with Z being the matrix of eigenvectors and Λ the diagonal matrix of eigenvalues. Splitting the eigenvalues such that Λ_1 contains the eigenvalues in the unit circle, we can write $\Lambda = \left[\begin{smallmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{smallmatrix}\right]$ and $Z = \left[\begin{smallmatrix} Z_{11} & Z_{21} \\ Z_{12} & Z_{22} \end{smallmatrix}\right]$. If a local equilibrium exists and is unique, the number of eigenvalues in the unit-circle is equal to the number of state variables and the linearized law of motion for state variables is given by $O := Z_{11}\Lambda_1 Z_{11}^{-1}$, while states map to controls through $G := Z_{12}Z_{11}^{-1}$. For details, we refer to Schmitt-Grohé and Uribe (2004). The fact that the distribution function over idiosyncratic states is part of the aggregate state vector and that the value functions (or marginal utilities) are part of the aggregate vector of controls does not change the solution in principle.

In practice, however, solving the generalized eigenvalue problem (or equivalently making a qz-decomposition of A, B) becomes easily numerically infeasible because the number of state variables (and controls) becomes very large, and thus A and B are large matrices. If the idiosyncratic state-space is high dimensional, both value functions and distribution functions are objects hard to approximate. A simple tensor grid to describe the value function or histogram has easily a large number points, even if it has a small number of points in each dimension of heterogeneity among households. Consider, for example, a household planning problem with two assets and idiosyncratic income. Even if we use only 9 points for the income grid and 50 points for each of the two asset grids, then both $d\mu$ and ν are vectors with a length of 22,500 entries, and with this resolution, the precision is at the lower bound of what one would like to have. This creates various

numerical problems. First, one needs to calculate many derivatives numerically. In our example, both A and B would be more than $45{,}000 \times 45{,}000$ entries large. While this calculating of the Jacobian is time consuming, the numerical complexity is only quadratic in the number of grid points. On top, modern automatic differentiation can speed this up. Still, the matrix to be stored remains large; each has more than 7 GB in our example if stored as a full double precision matrix.[8] Second, the qz-decomposition and the calculation of generalized eigenvalues become very time-consuming (cubic in the number of grid points).

The literature has suggested ways to deal with the issue. First, Reiter (2009) suggested replacing the value function with splines in order to decrease the number of nodes needed to describe the value function. Building on this suggestion, Winberry (2018) suggested using parametric families for the distribution functions to reduce the number of parameters that describe the distributions at each point in time. A downside of these two approaches is that they might impose tight restrictions on the value function and distribution in the stationary equilibrium.[9] What is more, they no longer allow us to represent the Bellman equation and the distribution dynamics by conveniently linear systems. For this reason, Ahn et al. (2017) suggested working in continuous time, which increases the sparsity of the Jacobians. Then they suggest, following the original paper by Reiter (2009), using singular-value decomposition of the Jacobians to project the state space of the model into a lower dimensional space without losing much of the dynamics of the system. Similar to what we suggest next, Ahn et al. (2017) linearized around the stationary equilibrium value and distribution functions without imposing any a priori restrictions on the functional forms.

2.5 *State-space reduction: Fixed copula, compressed value function*

We suggest reducing the dimensionality of the dynamic system before calculating the Jacobian, but after solving for the stationary equilibrium, which we then can use as a "reference frame." This allows us to solve the model in discrete time, where the Jacobian of the full system is much less sparse than in continuous time.

We achieve dimensionality reduction of the control space by writing the node values (on the tensor grid) of the value functions as some form of sparse expansions around their stationary equilibrium values:

$$\hat{\nu}_t(s) = g_\nu(s; \theta_t, \overline{\nu}), \tag{15}$$

where the length of the time-varying parameter vector θ_t is much smaller than the size of the tensor grid for s. We keep the setup with the surrogate planning problem that uses

[8]Clearly, many of the derivatives are (close to) zero, and thus storing the matrices as sparse matrices further helps. In fact, this is one of the main advantages of writing the model in continuous time, because then Π_h is very sparse and it is easy to see how this translates into a very sparse Jacobian; see Ahn et al. (2017).

[9]Another approach in the literature is to assume a finite but potentially large number of agents; see for example, Mertens and Judd (2018). Ragot (2018) provided an overview.

a linear interpolant outside the tensor grid for s to calculate the value function for nonnode values. This avoids oscillating behavior, which g_ν might show outside the nodes, and is computationally convenient. However, it is not central to our algorithm.

Yet, we have not specified how to select g_ν. One particularly useful way to construct g_ν is through (inverse) discrete cosine transformation of the stationary equilibrium value function. The discrete cosine transformation of a data array yields the coefficients of the fitted (multidimensional) Chebyshev polynomial, where the polynomial is constructed such that the tensor grid for s is mapped to the Chebyshev knots.[10] Importantly, the absolute value of the coefficients has an interpretation in terms of the contribution of the corresponding polynomial to the R^2 statistics in terms of fitting the data. This allows us to order and select the polynomial terms based on their importance.

To discuss this procedure in detail, with a slight abuse of notation, let $\overline{\nu}$ be the array of the value function values at the nodes of the full tensor grid in the stationary equilibrium. Further, let $\overline{\Theta} = \text{dct}(\overline{\nu})$ be its discrete cosine transform. The inverse cosine transformation of $\overline{\Theta}$ again produces $\overline{\nu}$. What is key for our procedure later on is that the larger (in absolute value) a coefficient $\overline{\Theta}(i)$ is, the more important is its corresponding Chebyshev polynomial for fitting $\overline{\nu}$; see Hu and Yu (1998). Therefore, it is useful to define $\mathcal{I}$ as the index set of some α% largest elements of $\overline{\Theta}$ (or equivalently the set that explains γ% of the total Euclidean norm of $\overline{\Theta}$) and define the sparse coefficient vector

$$\tilde{\Theta} = \begin{cases} \overline{\Theta}(i) & \forall i \in \mathcal{I}, \\ 0 & \text{else} \end{cases}$$

as the vector that shrinks all coefficients outside this set to zero. Then the inverse discrete cosine transformation of $\text{idct}(\tilde{\Theta})$ is the closest one to $\overline{\nu}$ in a least squares sense among all potential inverse discrete cosine transforms of arrays of the same level of sparseness. One can roughly read the suggested procedures by Reiter (2009) and Winberry (2018) as being sparse in this sense when calculating the stationary equilibrium, and then perturb *all* the coefficients that are used in calculating the stationary equilibrium.

Our approach by contrast does not try to be particularly sparse in calculating the stationary equilibrium, but can reach a higher degree of sparseness when calculating the dynamics. This is achieved by using all coefficients $\overline{\Theta}$ as a "reference frame" for calculating $g^\nu(s)$, by defining

$$\hat{\Theta}(\theta_t) = \begin{cases} \overline{\Theta}(i) + \theta_t(i) & \forall i \in \mathcal{I}, \\ \overline{\Theta}(i) & \text{else} \end{cases}$$

and $g^\nu(s)$ as its inverse discrete cosine transform $\text{idct}[\hat{\Theta}(\theta_t)]$, for a sparse vector θ_t. Importantly, for $\theta_t = 0$ it follows that $g^\nu = \overline{\nu}$ and our method thus fully recovers the stationary equilibrium value function at the same precision as is used in the computation of the

[10] See Ahmed, Natarajan, and Rao (1974) for the seminal contribution.

stationary equilibrium, that is, without creating any approximation error irrespective of the degree of sparseness that is used in the calculation of the model dynamics.[11]

This leaves us with the need to reduce the dimensionality of the distribution function. For this purpose, we split the distribution into a copula Ξ_t and marginal distributions $\{\mu_{1t}(s), \ldots, \mu_{mt}(s)\}$:

$$\mu_t(s) = \Xi_t\big\{\mu_{1t}(s), \ldots, \mu_{mt}(s)\big\}. \tag{16}$$

Again, as with the value functions, we can treat the copula as an interpolant defined on the grid of steady-state marginal distributions, and also approximate Ξ_t as a sparse expansion around the steady-state copula $\overline{\Xi}$, as we just did for the value function. The most extreme variant of this is to treat the copula as time fixed. We show in later sections that this works extremely well in practice, and hence we focus on this variant in what follows. We provide an extension treating the copula as time-varying in Appendix A.1.

The finding that the assumption of a fixed copula may work well follows from the insight by Krusell and Smith (1998) that not all moments of the cross-sectional distribution μ_t have a strong impact on the distribution of prices that economic agents need to forecast. In fact, for this reason Reiter (2009) proposes reducing the dimensionality of the state space by projecting the histogram of the joint distribution on a lower dimensional object that is perturbed instead. The projection can be done in such a way that, for example, a list of moments of the distribution is preserved. Yet, if one uses this approach, the distribution function will in general not maintain the shape it has in the stationary equilibrium. With our method by contrast, it maintains its shape. Perturbing only the marginals can be expected to be locally exact if the rank-correlation structure has no significant impact on equilibrium prices or is relatively constant; see Bayer, Luetticke, Pham-Dao, and Tjaden (2019) or Luetticke (2018) for examples.

Under this approach, the dynamic system F replaces value functions and distributions by the parameters $\theta_t, d\mu_{1t}, \ldots, d\mu_{mt}$, where the $d\mu$-terms are the histograms of the marginal distributions. Since the system has more equations than unknowns now, we need to reduce the dimensionality of F, too. This can be done by projecting the differences back to a lower dimensional space. For example, for the distribution functions this can be done by comparing only the marginal distributions. For the value functions, one can focus on the coefficients of the discrete cosine transformation of the error terms on the value functions at all nodes $\nu_t - (u_{h_t^d} + \beta\Pi_{h_t}\nu_{t+1})$ in the index set $\mathcal{I}$.

One advantage of reducing the state space *before* calculating the Jacobian of the difference equation through fixing the copula and "compressing" the value function instead of reducing it *after* calculating the Jacobian (as in Reiter (2009), Ahn et al. (2017)) is that it reduces substantially the time needed for calculating derivatives and avoids the potentially large memory requirements to store them that arise in discrete time models. In addition, it avoids the singular-value decomposition altogether.

[11] Since the degree of sparseness and the index set $\mathcal{I}$ are chosen heuristically, the researcher should check the robustness of her findings to the choice of the degree of sparseness. Yet, for the state-space reduction based on singular-value decompositions of the Jacobian as in Ahn et al. (2017), one also needs to decide on the minimal singular value that is retained.

Its disadvantage is that it is not guaranteed that the coefficients of the expansion around the stationary equilibrium value function that are shrunk to zero are unimportant for the shape of the value function outside the stationary equilibrium. They are only unimportant in the stationary equilibrium (and hence would have been left out in procedures that reduce the dimensionality entirely ex ante). Yet, whether the latter leads to low-quality approximations can be checked through simulating the model along the lines of the tests suggested by Den Haan (2010a).

2.6 *The algorithm in a nutshell*

To give a practical guide on the implementation, we finally provide a summary of the proposed algorithm. Concrete implementations can differ in particular in how the dynamic programming problem is solved. In particular, we provide the algorithm here on the basis of value function iteration, for simplicity and generality. In practice, another recursive method such as an endogenous grid method might well be preferable.

For our algorithm, define grids $\mathfrak{s}^j = \{d_1^j \dots d_{n_j}^j\}$ for each $j = 1 \dots m_d$ of the idiosyncratic endogenous state variables d^j, with n_j being the number of grid points used for variable j (note that different from the section before, here we explicitly split up the endogenous state variables in their m_d-dimensions). In addition to the endogenous idiosyncratic states, there is the exogenous stochastic one, x, which evolves on the grid $\mathfrak{s}^0 = \{x_1 \dots x_{n_0}\}$, which together with the transition matrix Π_x defines a discrete Markov chain for this state variable (collapsing all idiosyncratic exogenous states m_x into one). Let $\mathcal{S} \bigotimes_{j=0\dots m_d} \mathfrak{s}^j$ be the tensor product (mesh) of these $m_d + 1$ grids, and let $\mathcal{IS}$ be the corresponding tensor product (mesh) of the indexes. This mesh has in total $J = \prod_{j=0}^{m_d} n_j$ grid points.

We define $\mathcal{V}$ as the m_d+1-dimensional array that stores the values of a value function at each point of the mesh $\mathcal{S}$. We define $\hat{\nu}[(x, d^1 \dots d^{m_d})|\Pi_x \mathcal{V}]$ as the linear interpolant defined by the mesh $\mathcal{S}$ and node values $\Pi_x \mathcal{V}$, where $\Pi_x \mathcal{V}$ is the matrix product of Π_x and $\mathcal{V}$ reshaped accordingly. With $d\mu \in \mathbb{R}^{n_0 \times n_1 \times \dots n_{m_d}}$, we denote the histogram of the distribution of agents over all states $s \in \mathcal{S}$ in array form; $d\mu$ is the same, but vectorized (stacked). Let X be the (exogenous) aggregate state of the economy with $\overline{X}$ its steady-state value.

PREREQUISITES 1. 1. Define for a given price system P a mapping $T(\mathcal{V}|P) : \mathbb{R}^J \to \mathbb{R}^J$ such that

$$\forall s = \left(x, d^1 \dots d^{m_d}\right) \in \mathcal{S}:$$
$$T(\mathcal{V}|P)(s) := \max_{(d^{1'} \dots d^{m_d\prime}) \in \Gamma(s,P)} u\left(s, d^{1'} \dots d^{m_d\prime}\right) + \beta \hat{\nu}\left[\left(x, d^{1'} \dots d^{m_d\prime}\right)|\Pi_x \mathcal{V}\right].$$

In words, this mapping is one iteration of the value function. Define $h^d(\mathcal{V}|P) : \mathbb{R}^J \to \mathbb{R}^{J/n_0}$ as the corresponding policy function (the arg max).

2. Define a mapping $\Pi = \Pi(\mathcal{V}_P) : \mathbb{R}^J \to \mathbb{R}^{J\times J}$ such that

$$\forall k = \left(k^0 \dots k^{m_d}\right), l = \left(l^0 \dots l^{m_d}\right) \in \mathcal{IS} : \Pi(\mathcal{V}_P)(k,l) = \Pi_x\left(k^0, l^0\right) \prod_{j=1}^{m_d} \Pi_{d^j}(k,l),$$

where Π_{d^j} are the coefficients to represent the policy $h_P^d(x) = (h_1^d(x) \dots h_{m_d}^d(x))$ as convex combinations of the nearest neighbors on the index mesh $\mathcal{IS}$, that is,

$$\Pi_{d^j}(k,l) = \begin{cases} 0 & \text{if } h_j^d(k) \notin \left[d_{l-1}^j, d_{l+1}^j\right], \\ 1 - \dfrac{h_j^d(k) - d_l^j}{d_{l+1}^j - d_l^j} & \text{if } d_{l+1}^j \geq h_j^d > d_l^j, \\ \dfrac{h_j^d - d_{l-1}^j}{d_l^j - d_{l-1}^j} & \text{if } d_l^j \geq h_j^d \geq d_{l-1}^j. \end{cases} \tag{17}$$

3. The discrete cosine transformation of an array A along a dimension j is given by pre-multiplying a transformation matrix C_j to array A along the j-dimension. This is done by permuting the array such that dimension j becomes the first one and reshaping the array to matrix form. The result of this matrix multiplication has to be reshaped back to its array form, permuting the now first dimension back to the jth position. The inverse is defined analogously through premultiplication of $C_j^{-1} = C_j'$. The matrix C_j is constructed as

$$C_j(k,l) = \begin{cases} \dfrac{1}{\sqrt{n_j}} & \forall l = 1 \dots n_j, k = 1, \\ \sqrt{2/n_j} \cos\left(\pi \dfrac{(l-1/2)(k-1)}{n_j}\right) & \forall l = 1 \dots n_j, k = 2 \dots n_j. \end{cases} \tag{18}$$

Algorithm 1. 1. *Finding the stationary equilibrium*

(a) For a given price system P, iterate $T^{(n)} = \underbrace{T(T(\dots T(\mathcal{V}^{(0)}|P)|P)|P)}_{n \text{ times}}$ until convergence to obtain an equilibrium value function $\mathcal{V}_P$ as the limit $n \to \infty$.

(b) Calculate the equilibrium distribution $d\mu_P$ by solving $d\mu_P = d\mu_P \Pi(\mathcal{V}_P)$.

(c) Calculate excess demand Φ as a function $\Phi(h_P^d, d\mu_P)$.

(d) Search over prices, repeating (a) to (c) until $\Phi(h_P^d, d\mu_P) = 0$. The prices that set excess demand to zero are in the following denoted as $\overline{P}$ with $\overline{h}^d$ and $\overline{\mathcal{V}}$ being the corresponding policy and value functions and $\overline{d\mu}$ the equilibrium histogram.

2. *Dimensionality reduction*

(a) Define the joint distribution function $\overline{\mu}(s) = \sum_{x \leq s} \overline{d\mu}(x)$. Define $\overline{\mu}^j \in [0,1]^{n_j}$, $j = 0 \dots m_d$ as the $m_d + 1$ vectors of the marginal distributions corresponding to the n_j points on the $\mathfrak{s}_j$-grids. Generate the fixed copula $\overline{\Xi}(\mu^0, \dots, \mu^{m_d}|\overline{\mu}) : [0,1]^{m_d+1} \to [0,1]$ as an interpolant of $\overline{\mu}$ on the tensor product $\bigotimes_{j=0}^{m_d} \overline{\mu}^j$.

(b) Calculate the discrete cosine transformation of $\overline{\mathcal{V}}$ along all $m_d + 1$ dimensions. This yields coefficients $\overline{\Theta}$. Find the minimal index set $\mathcal{I}$, such that $\frac{\sum_{i\in\mathcal{I}}\overline{\Theta}(i)^2}{\sum_i \overline{\Theta}(i)^2} > 1 - \epsilon$ (by sorting the coefficients and retaining only the largest ones).

(c) Define a sparse vector that has $\#\mathcal{I}$ nonzero entries, and hence is effectively much shorter than $\overline{\Theta} \in \mathbb{R}^J$. In the following, when we speak of perturbing θ_t, we mean perturbing its nonzero entries. This vector is used to assign values to those coefficients of the discrete cosine transformation of $\overline{\mathcal{V}}$ that were found to be different from zero, and hence important. In other words, it assigns a value to each coefficient in the index set $\mathcal{I}$, such that we obtain the full set of coefficients, $\hat{\Theta}(\theta|\Theta, \mathcal{I}) \in \mathbb{R}^J$, which is given by

$$\hat{\Theta} = \begin{cases} \overline{\Theta}(i) + \theta(i) & \text{if } i \in \mathcal{I}, \\ \overline{\Theta}(i) & \text{if } i \notin \mathcal{I}. \end{cases}$$

The mapping of this array $\hat{\Theta}$ to the value function values $\hat{\mathcal{V}}(\theta)$ is obtained through an inverse cosine transformation.

3. *Linearization*

(a) Define the following objects:

- the difference between the value function implied from one backward iteration based on its value at time $t+1$ and the value function for time t as implied by θ_t. We apply the discrete cosine transformation to the value functions and evaluate on all points in $\mathcal{S}$,

$$\Delta_\nu(\theta_t, \theta_{t+1}, P_t) := \theta_t - \mathrm{dct}\{T[\hat{\mathcal{V}}(\theta_{t+1})|P_t]\} \in \mathbb{R}^J. \tag{19}$$

The shorter vector Δ_ν^* selects out of Δ_ν only those elements that correspond to the index set $\mathcal{I}$.

- for all variables $j = 0 \ldots m_d$ the difference between the marginal distribution for time $t+1$ obtained from iterating forward once (using the optimal policies) the distribution implied by $(\mu_t^j)_{j=0\ldots m_d}$ and the copula $\overline{\Xi}$,

$$\Delta_\mu^*[\{\mu_t^j\}_{j=0\ldots m_d}, \{\mu_{t+1}^j\}_{j=0\ldots m_d}, P_t, \theta_{t+1}] \in \mathbb{R}^{\sum_{(m_d+1)} n_j}. \tag{20}$$

- the excess demand function

$$\Phi(\{\mu_t^j\}_{j=0\ldots m_d}, \theta_{t+1}, P_t, S_t, S_{t+1}) := \Phi[d\overline{\Xi}(\{\mu_t^j\}_{j=0\ldots m_d}), h^d_{P_t, \hat{\mathcal{V}}(\theta_{t+1})}, S_t, S_{t+1}]. \tag{21}$$

(b) Use these differences (19)–(21) to define a function

$$F(\{\mu_t^j\}_{j=0\ldots m_d}, S_t, S_{t+1}, \{\mu_{t+1}^j\}_{j=0\ldots m_d}, \theta_t, P_t, \theta_{t+1}, P_{t+1}|\overline{\Xi}, \overline{\mathcal{V}}, \mathcal{I})$$

that describes the economy as a system of nonlinear difference equations

$$F = \begin{bmatrix} \Delta_{\nu}^{*}(\theta_t, \theta_{t+1}, P_t) \\ \Delta_{\mu}^{*}\left[\{\mu_t^j\}_{j=0\ldots m_d}, \{\mu_{t+1}^j\}_{j=0\ldots m_d}, P_t, \theta_{t+1}\right] \\ S_{t+1} - H(S_t) \\ \Phi\left(\{\mu_t^j\}_{j=0\ldots m_d}, \theta_{t+1}, P_t, S_t, S_{t+1}\right) \end{bmatrix}. \tag{22}$$

(c) Calculate the Jacobian of F. Define A, B as defined in the text before and as in Schmitt-Grohé and Uribe (2004).

(d) Calculate the qz decomposition and solve for the linearized dynamics using the algorithm provided by Schmitt-Grohé and Uribe (2004).

3. Examples

In the following, we discuss two examples to illustrate our modification of Reiter's method to solve general equilibrium models with heterogeneous agents and aggregate risk. Both examples share the same model of consumption-savings choice in which households face uninsurable income risk and use assets to self-insure. We then specify two variants of the model: one without nominal frictions and only one asset, that is, the setup of Krusell and Smith (1998); second, a setup with two assets of different liquidity and a nominal rigidity. The first example can be solved using the original Krusell and Smith algorithm and the Reiter algorithm without state-space reduction. For the second example, state-space reduction is necessary to render the computation feasible. Details on the numerical precision of the various algorithms are provided in Section 4.

3.1 *Household sector*

There is a continuum of ex ante identical households of measure one, indexed by i. Households are infinitely lived, have time-separable preferences with time-discount factor β, and derive felicity from consumption c_{it} and leisure. Households have Greenwood–Hercowitz–Huffman (GHH) preferences, and maximize the discounted sum of felicity:[12]

$$E_0 \max_{\{c_{it}, n_{it}, \Delta k_{it}\}} \sum_{t=0}^{\infty} \beta^t u\left[c_{it} - G(h_{it}, n_{it})\right].$$

The maximization is subject to the budget constraints described further below. The felicity function u exhibits a constant relative risk aversion (CRRA) with risk aversion parameter $\xi > 0$,

$$u(x_{it}) = \frac{1}{1-\xi} x_{it}^{1-\xi},$$

[12]The assumption of GHH preferences simplifies the numerical analysis of the stationary equilibrium substantially but is not necessary for our implementation of Reiter's method.

where $x_{it} = c_{it} - G(h_{it}, n_{it})$ is household i's composite demand for goods consumption c_{it} and leisure and G measures the disutility from work. Goods consumption bundles varieties j of differentiated goods according to a Dixit–Stiglitz aggregator:

$$c_{it} = \left(\int c_{ijt}^{\frac{\eta-1}{\eta}} \, dj \right)^{\frac{\eta}{\eta-1}}.$$

Each of these differentiated goods is offered at price p_{jt}, so that for the aggregate price level, $P_t = (\int p_{jt}^{1-\eta} \, dj)^{\frac{1}{1-\eta}}$, the demand for each of the varieties is given by

$$c_{ijt} = \left(\frac{p_{jt}}{P_t} \right)^{-\eta} c_{it}.$$

The disutility of work, $G(h_{it}, n_{it})$, determines a household's labor supply given the aggregate wage rate, w_t, and a labor income tax, τ, through the first-order condition:

$$\frac{\partial G(h_{it}, n_{it})}{\partial n_{it}} = (1 - \tau) w_t h_{it}.$$

Assuming that G has a constant elasticity w.r.t. n, $\frac{\partial G(h_{it}, n_{it})}{\partial n_{it}} = (1 + \gamma) \frac{G(h_{it}, n_{it})}{n_{it}}$ with $\gamma > 0$, we can simplify the expression for the composite consumption good x_{it} making use of the first-order condition (3.1):

$$x_{it} = c_{it} - G(h_{it}, n_{it}) = c_{it} - \frac{(1 - \tau) w_t h_{it} n_{it}}{1 + \gamma}.$$

When the Frisch elasticity of labor supply is constant, the disutility of labor is always a constant fraction of labor income. Therefore, in both the budget constraint of the household and its felicity function, only after-tax income enters, and neither hours worked nor productivity appears separately.

This implies that we can assume $G(h_{it}, n_{it}) = h_{it} \frac{n_{it}^{1+\gamma}}{1+\gamma}$ without further loss of generality as long as we treat the empirical distribution of income as a calibration target. This functional form simplifies the household problem as h_{it} drops out from the first-order condition and all households supply the same number of hours $n_{it} = N(w_t)$. Total effective labor input, $\int n_{it} h_{it} \, di$, is hence also equal to $N(w_t)$ because $\int h_{it} \, di = 1$.

A household's labor income $w_t h_{it} n_{it}$ is composed of the aggregate wage rate, w_t, the household's hours worked, n_{it}, and its idiosyncratic labor productivity, h_{it}. Productivity evolves according to a log-AR(1) process and a fixed probability of transition to a high income state in which $h_{it} = 0$ but households receive a share of pure rents, that is, they become entrepreneurs:

$$h_{it} = \begin{cases} \exp\left(\rho_h \log h_{it-1} + \epsilon_{it}^h\right) & \text{with probability } 1 - \zeta \text{ if } h_{it-1} \neq 0, \\ 1 & \text{with probability } \iota \text{ if } h_{it-1} = 0, \\ 0 & \text{otherwise,} \end{cases}$$

with shocks to productivity ϵ_{it}^h being normally distributed.

With probability ζ households become entrepreneurs ($h = 0$). With probability ι an entrepreneur returns to the labor force with median productivity. An entrepreneurial household obtains a fixed share of the pure rents, Π_t, in the economy (from monopolistic competition and creation of capital). We assume that the claim to the pure rent cannot be traded as an asset. The idea here is that a household that becomes an entrepreneur develops a variety only it can produce out of intermediate goods and it loses this capacity (because its variety is replaced by another household's drastic innovation) when returning to the labor force.

3.2 *Price setting*

These entrepreneur households, that is, the final-goods producers, differentiate the intermediate good and set prices. We assume price adjustment costs à la Rotemberg (1982). For tractability, we assume that the actual price setting is delegated to a mass-zero group of households (managers) that are risk neutral and compensated by a share in profits. They do not participate in any asset market. Under this assumption, managers maximize the present value of real profits given the demand for good j,

$$y_{jt} = (p_{jt}/P_t)^{-\eta} Y_t,$$

and quadratic costs of price adjustment, that is, they maximize

$$E_0 \sum_{t=0}^{\infty} \beta^t Y_t \left\{ \left(\frac{p_{jt}}{P_t} - \mathrm{MC}_t \right) \left(\frac{p_{jt}}{P_t} \right)^{-\eta} - \frac{\eta}{2\kappa} \left(\log \frac{p_{jt}}{p_{jt-1}} \right)^2 \right\},$$

with a time-constant discount factor. From the corresponding first-order condition for price setting, it is straightforward to derive the Phillips curve:

$$\log(\pi_t) = \beta E_t \left[\log(\pi_{t+1}) \frac{Y_{t+1}}{Y_t} \right] + \kappa \left(\mathrm{MC}_t - \frac{\eta - 1}{\eta} \right), \tag{23}$$

where π_t is the gross inflation rate, $\pi_t := \frac{P_t}{P_{t-1}}$, and MC_t is the real marginal costs. The price adjustment then creates real costs $\frac{\eta}{2\kappa} Y_t \log(\pi_t)^2$.

Since managers are a mass-zero group in the economy, their consumption does not show up in any resource constraint and all profits—net of price adjustment costs—go to the entrepreneur households (whose $h = 0$). In the case of the two-asset economy, these households also obtain profit income from adjusting the aggregate capital stock. They can transform I_t consumption goods into ΔK_{t+1} new capital goods (and back) according to the transformation function:[13]

$$I_t = \frac{\phi}{2} (\Delta K_{t+1}/K_t)^2 K_t + \Delta K_{t+1}.$$

[13]We assume that capital goods producers are each small, and thus ignore their externality on the future cost of capital goods production.

Since they are facing perfect competition in this market, entrepreneurs will adjust the stock of capital until the following first-order condition holds:

$$q_t = 1 + \phi \Delta K_{t+1}/K_t,$$

where q_t is the price of capital.[14]

3.3 *Intermediate-goods producers*

Intermediate goods are produced with a constant returns to scale production function:

$$Y_t = A_t N_t^{\alpha} K_t^{(1-\alpha)},$$

where $K_t = E(k_{it})$ is the aggregate capital supply, $N_t = E(h)[(1-\tau)w_t]^{\frac{1}{\gamma}}$ is the aggregate labor supply, and A_t is total factor productivity.

Let MC_t be the relative price at which the intermediate good is sold to entrepreneurs. The intermediate-good producer maximizes profits,

$$\mathrm{MC}_t Y_t - w_t N_t - (r_t + \delta)K_t = \mathrm{MC}_t A_t N_t^{\alpha} K_t^{(1-\alpha)} - w_t N_t - (r_t + \delta)K_t,$$

but it operates in perfectly competitive markets, such that the real wage and the user costs of capital are given by the marginal products of labor and capital:

$$w_t = \alpha A_t \mathrm{MC}_t (K_t/N_t)^{1-\alpha}, \qquad r_t + \delta = (1-\alpha) A_t \mathrm{MC}_t (N_t/K_t)^{\alpha}.$$

3.4 *Model variants*

To close the model, we still need to define which assets households can trade. As stated before, we consider two model variants. First, we have a variant of the original Krusell and Smith (1998) economy where only capital is traded, which is a perfectly liquid asset. This variant serves to benchmark our solution strategy against other discrete time methods. Second, we use the economy as in Bayer et al. (2019) and Luetticke (2018) with a liquid nominal asset and illiquid capital. This economy cannot be solved without state-space reduction and serves as an application example for those cases.

3.4.1 *A neoclassical economy with one asset: The Krusell–Smith setup* Our model nests the Krusell and Smith (1998) economy. In that economy, households save only in capital that is perfectly liquid. There are no entrepreneurs ($\zeta = 0$), labor supply is constant, competition is perfect, and price adjustment is costless ($\eta, \kappa \to \infty$, $\frac{\eta}{\kappa} \to 0$). In addition, there is no capital adjustment cost, $\phi = 0$, such that $q_t = 1$. Taxes τ are zero, too.

Therefore, households optimize subject to this budget constraint:

$$c_{it} + k_{it+1} = k_{it}(1 + r_t) + w_t h_{it} N,$$
$$k_{it+1} \geq 0,$$

where r_t is the real return on capital.

[14]We assume for simplicity that all depreciation is replaced immediately through maintenance investment that transforms consumption goods into replacement investment one-for-one.

Substituting the expression $c_{it} = x_{it} + \frac{w_t h_{it} N}{1+\gamma}$ for consumption, we obtain

$$x_{it} + k_{it+1} = k_{it}(1 + r_t) + \left(\frac{\gamma}{1+\gamma} w_t h_{it} N\right),$$
$$k_{it+1} \geq 0.$$

With this setup, one Bellman equation characterizes the dynamic planning problem of a household:

$$V(k, h; \mu, A) = \max_{k'} u\big[x\big(k, k', h\big)\big] + \beta V\big(k', h'; \mu', A'\big),$$

where μ is the wealth-income distribution and A is aggregate productivity as the only other state variable. Capital and labor market clearing are the only equilibrium conditions (there is classical dichotomy and the nominal side is not determined):

$$w_t = \alpha A_t (K_t/N)^{1-\alpha}, \qquad r_t + \delta = (1-\alpha) A_t (N/K_t)^{\alpha}.$$

3.4.2 *New-Keynesian variant with liquid and illiquid assets* The second model variant introduces a nominal rigidity, such that the Phillips curve (23) is not vertical, and a nominal bond that pays R_t, and makes capital illiquid, such that the two assets are not close substitutes. Illiquidity is modeled as follows: Only a randomly selected fraction of households, ν, participates in the market for capital each period and can thus actively sell or buy capital. All other households obtain dividends, but may only adjust their holdings of nominal bonds. Holdings of bonds have to be above an exogenous debt limit $\underline{B}$, and holdings of capital have to be nonnegative.

Therefore, households optimize subject to their budget constraint:

$$c_{it} + b_{it+1} + q_t k_{it+1} = b_{it} \frac{R\big(b_{it}, R_t^b\big)}{\pi_t} + (q_t + r_t) k_{it} + (1-\tau)(w_t h_{it} N_t + \mathbb{I}_{h_{it}=0} \Pi_t),$$
$$k_{it+1} \geq 0, b_{it+1} \geq \underline{B},$$

where b_{it} is real bond holdings, $\underline{B}$ is an exogenous borrowing constraint, k_{it} is the amount of illiquid assets, q_t is the price of these assets, r_t is their dividend, $\pi_t = \frac{P_t - P_{t-1}}{P_{t-1}}$ is realized inflation, and R is the nominal interest rate on bonds, which depends on the portfolio position of the household and the central bank's interest rate R_t^b, which is set one period before. All households that do not participate in the capital market ($k_{it+1} = k_{it}$) still obtain dividends and can adjust their bond holdings. Depreciated capital has to be replaced for maintenance, such that the dividend, r_t, is the net return on capital.

We assume that there is a wasted intermediation cost, $\overline{R}$, when households resort to unsecured borrowing and specify:

$$R\big(b_{it}, R_t^b\big) = \begin{cases} R_t^b & \text{if } b_{it} \geq 0, \\ R_t^b + \overline{R} & \text{if } b_{it} < 0. \end{cases}$$

This assumption creates a mass of households with zero unsecured credit but with the possibility to borrow, though at a penalty rate.

Substituting the expression $c_{it} = x_{it} + \frac{(1-\tau)w_t h_{it} N_t}{1+\gamma}$ for consumption, we obtain

$$x_{it} + b_{it+1} + q_t k_{it+1} = b_{it}\frac{R(b_{it}, R_t^b)}{\pi_t} + (q_t + r_t)k_{it} + (1-\tau)\left(\frac{\gamma}{1+\gamma}w_t h_{it} N_t + \mathbb{I}_{h_{it}=0}\Pi_t\right),$$
$$k_{it+1} \geq 0, b_{it+1} \geq \underline{B}.$$

With this setup, two Bellman equations characterize the dynamic planning problem of a household: V_a in the case where the household can adjust its capital holdings and V_n otherwise:

$$V_a(b,k,h;\mu,R^b,A) = \max_{k',b'_a} u\big[x(b,b'_a,k,k',h)\big] + \beta\big[\nu EV^a(b'_a,k',h';\mu',R^{b'},A')$$
$$+ (1-\nu)EV^n(b'_a,k',h';\mu',R^{b'},A')\big]$$
$$V_n(b,k,h;\mu,R^b,A) = \max_{b'_n} u\big[x(b,b'_n,k,k,h)\big] + \beta\big[\nu EV^a(b'_n,k,h';\mu',R^{b'},A')$$
$$+ (1-\nu)EV^n(b'_n,k,h';\mu',R^{b'},A')\big]$$

Since we allow for a nominal rigidity, the equilibrium is only determined when a monetary and a fiscal policy are specified. Monetary policy controls the nominal interest rate on liquid assets, while fiscal policy determines the amount of government bonds by controlling fiscal deficits through the adjustment of expenditures. We assume that the monetary and fiscal authorities operate independently and their behavior is described by simple rules.

We assume that monetary policy sets the nominal interest rate on bonds following a Taylor-type 1993 rule with interest rate smoothing:

$$\frac{R_{t+1}^b}{\overline{R}^b} = \left(\frac{R_t^b}{\overline{R}^b}\right)^{\rho_R}\left(\frac{\pi_t}{\overline{\pi}}\right)^{(1-\rho_R)\theta_\pi}.$$

The coefficient $\overline{R}^b \geq 0$ determines the nominal interest rate in the steady state. The coefficient $\theta_\pi \geq 0$ governs the extent to which the central bank attempts to stabilize inflation around its steady-state value: the larger θ_π the stronger is the reaction of the central bank to deviations from the inflation target. When $\theta_\pi \to \infty$, inflation is perfectly stabilized at its steady-state value. $\rho_R \geq 0$ captures interest rate smoothing.

We assume that the government issues bonds according to the rule (cf. Woodford (1995)):

$$\frac{B_{t+1}}{B_t} = \left(\frac{B_t}{\overline{B}}\right)^{\rho_B}\left(\frac{\pi_t}{\overline{\pi}}\right)^{-\gamma_\pi}\left(\frac{\mathcal{T}_t}{\overline{\mathcal{T}}}\right)^{-\gamma_\mathcal{T}},$$

using tax revenues $\mathcal{T}_t = \tau(w_t N_t + \Pi_t)$ to finance government consumption, G_t, and interest on debt. In other words, the government seeks to stabilize debt in the long run and

output in the short run. The coefficient ρ_B captures whether and how fast the government seeks to repay its outstanding obligations B_t. For $\rho_B < 1$, the government actively stabilizes real government debt, and for $\rho_B = 1$ the government rolls over all outstanding debt. The coefficients γ_π, $\gamma_\mathcal{T}$ capture the cyclicality of debt issuance: for $\gamma_\pi = \gamma_\mathcal{T} = 0$, new debt does not actively react to tax revenues and inflation, but only to the value of outstanding debt. For $\gamma_\pi > 0 > \gamma_\mathcal{T}$, debt is countercyclical; for $\gamma_\pi < 0 < \gamma_\mathcal{T}$, it is procyclical.

In equilibrium, we need both factor markets to clear, such that

$$w_t = \alpha \mathrm{MC}_t A_t (K_t/N_t)^{1-\alpha}, \qquad r_t + \delta = (1-\alpha) A_t \mathrm{MC}_t (N_t/K_t)^{\alpha},$$

and we also need asset markets to clear. This requires first

$$B_{t+1} = B^d\big(\mu_t; R_t^b, A_t; q_t, \pi_t; V_{a,t+1}, V_{n,t+1}\big) := E\big[\nu b_a^* + (1-\nu) b_n^*\big], \tag{24}$$

where b_a^*, b_n^* are bond demand functions of adjusters and nonadjusters. They are functions of the states $(b, k, h; R_t^b, A_t)$, of current prices q_t, π_t, and of expectations of future prices summarized in the marginal value functions $V_{a,t+1}, V_{n,t+1}$. Expectations in the right-hand side expression are taken w.r.t. the distribution $\mu_t(b, k, h)$. Equilibrium requires the total *net* amount of bonds the household sector demands, B^d, to equal the supply of government bonds. In gross terms, there are more liquid assets in circulation as some households borrow up to $\underline{B}$.

Second, the asset market for capital has to clear. This requires that

$$\begin{aligned} q_t &= 1 + \phi \frac{K_{t+1} - K_t}{K_t}, \\ K_{t+1} &= K^d\big(\mu_t; R_t^b, A_t; q_t, \pi_t; V_{a,t+1}, V_{n,t+1}\big) := E\big[\nu k^* + (1-\nu) k\big]. \end{aligned} \tag{25}$$

Again expectations are taken w.r.t. the distribution $\mu_t(b, k, h)$.

4. Numerical performance

In the following, we first demonstrate the performance and accuracy of our method by comparing it to the Krusell and Smith (1998) algorithm for the standard Krusell and Smith (K-S) model, as described in Section 3.4.1. We then show the scalability of our method by solving heterogeneous agent New-Keynesian (HANK) models with higher dimensional heterogeneity, providing accuracy measures for the variant described in Section 3.4.2. Finally, we also show that our approach also practically renders second-order approximatableions feasible. All codes are available on the authors' websites and as a replication file inside the Supplementary Material (Bayer and Luetticke (2020)).

4.1 *Comparison to Krusell and Smith (1998)*

To compare the performance and accuracy of our method, we solve Krusell and Smith's (1998) model with the standard parameterization of the JEDC comparison project (cf.

Den Haan, Judd, and Juillard (2010)).[15] A period in the model is a quarter, the discount factor is $\beta = 0.99$, the coefficient of relative risk aversion is $\xi = 1$, and the rate of depreciation equals 2.5% per quarter.[16] Idiosyncratic and aggregate productivity risk both follow two-state Markov chains. We solve the household problem on 100 grid points for idiosyncratic capital. The grid for the aggregate capital stock has 3 points for the Krusell–Smith algorithm and covers the unconditional ±3 STD interval from the linearized solution.

4.1.1 *Numerical quality* Figure 1 shows simulations of the K-S model for three different solution methods: (1) perturbation with state-space reduction via the fixed copula assumption and policy function compression (25 coefficients of the discrete cosine transformation conserve 99.99% of the energy), (2) perturbation with a full policy function and histogram on the tensor product of the income and capital grid as in Reiter (2002), and (3) the original Krusell and Smith algorithm.[17] The response of aggregate capital to TFP shocks is virtually the same in all three simulations. Table 1 confirms this. The mean absolute error between the time series from the two linearization methods and the K-S algorithm is 0.03%. What is more, the linearization methods with and without state and control space reduction yield basically the same simulation for the aggregate stock of capital with a maximum absolute error of 0.001%.

To further evaluate the accuracy of our solution method, we use the error metric suggested by Den Haan (2010a), comparing the simulation from the linearized solution of the model to one in which we solve for the equilibrium interest rate every period and

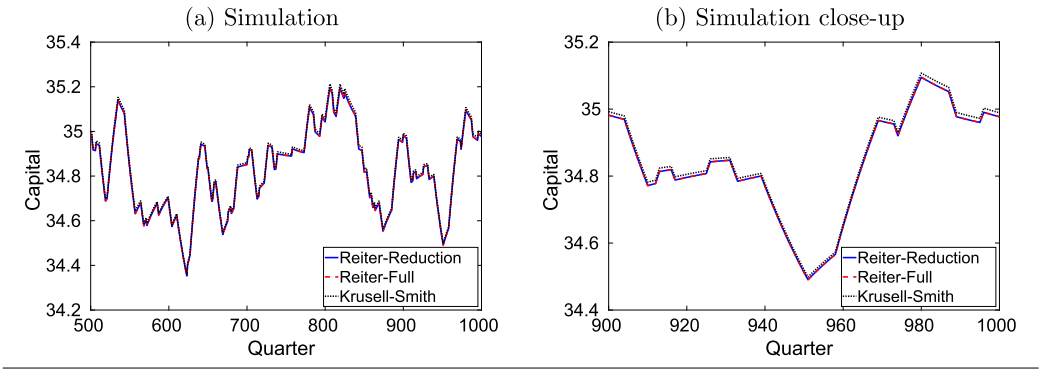

FIGURE 1. Simulations of Krusell & Smith model. *Notes*: Both panels show simulations of the Krusell and Smith (1998) model with TFP shocks solved with (1) the Reiter method with our proposed state-space reduction, (2) the original Reiter method without state-space reduction, and (3) the original Krusell and Smith algorithm; simulated for 1000 periods. The draws for the productivity process are kept constant across solution methods.

[15]Setting $\eta \to \infty$ and $\kappa \to \infty$, that is, no markups and flexible prices, yields the standard neoclassical incomplete markets model.

[16]See Appendix B, Table 11 for the calibration.

[17]The simulations start from the steady state without aggregate risk, which is the same for all three methods. For all statistics, we simulate the model for 1000 periods.

TABLE 1. Simulation errors relative to Krusell and Smith algorithm.

	Absolute difference (in %) of log capital stocks K_t between simulations		
	Reiter-Reduction vs. K-S	Reiter-Full vs. K-S	R.-Reduction vs. R.-Full
Mean	0.0324	0.0324	0.0003
Max	0.0670	0.0662	0.0012

Note: Differences in percent between simulations of aggregate capital for the Krusell and Smith (1998) model solved with (1) the Reiter method with our proposed state-space reduction, (2) the original Reiter method without state-space reduction, (3) the original Krusell and Smith algorithm. The first two columns show the performance of (1) and (2) relative to (3), and the last column shows the performance of (1) relative to (2) for 1000 periods. The draws for the productivity process are kept constant across solution methods.

TABLE 2. Den Haan errors.

	Absolute error (in %) for log capital K_t		
	Reiter-Reduction	Reiter-Full	K-S
Mean	0.0100	0.0102	0.0051
Max	0.0191	0.0193	0.0131

Note: Differences in percent between the simulation of the linearized solutions of the model and simulations in which we solve for the intratemporal equilibrium prices in every period and track the full histogram over time for $t = \{1, \ldots, 1000\}$; see Den Haan (2010a).

TABLE 3. Run time for Krusell and Smith model.

	Stationary equilibrium	Krusell and Smith	Reiter-Reduction	Reiter-Full
in seconds	7.05	91.61	0.38	1.19

Note: Run time in seconds on a Dell laptop with an Intel i7-7500U CPU at 2.70 GHz at 4. Model calibration and number of grid points as in Den Haan, Judd, and Juillard (2010). Code in Matlab.

track the full histogram over time. The mean absolute error is 0.01% and the maximum error is 0.019%; see Table 2. The K-S algorithm, which is the most accurate algorithm in Den Haan, Judd, and Juillard (2010), here is also most precise with a mean absolute error of 0.005%. In Appendix A.2, we show that this result is not specific to the parameterization (as can be expected for a first-order perturbation, solution quality deteriorates with the variance of shocks).

Finally, Table 3 shows the run times of all three methods and the steady state separately. The Reiter method with state and control space reduction only takes 0.4 seconds. This makes it more than 240 times faster than the Krusell and Smith algorithm. Without reduction, the run time increases by a factor of 3. Even when the time to compute the stationary equilibrium is taken into account, our linearization method is 13 times faster than the Krusell–Smith algorithm. The main advantage of linearization with state and control space reduction, however, lies in its capacity to solve models with many idiosyncratic states fast and precisely as the next section shows. Before going there, we will provide a short illustration of our dimension reduction procedure in the Krusell–Smith economy.

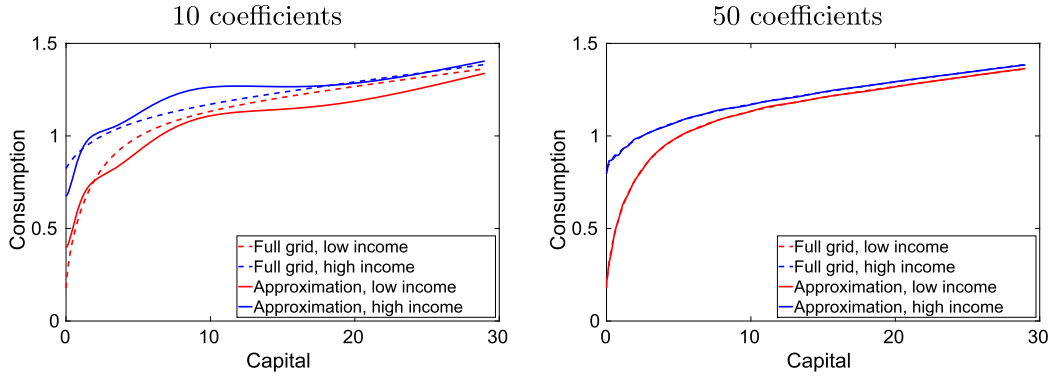

FIGURE 2. Stationary equilibrium consumption policies by sparseness of θ.

4.1.2 *Details on using the DCT for dimensionality reduction* The small size of the Krusell and Smith example allows us to discuss the advantages of our dimensionality reduction procedure by displaying the implied approximations and approximation errors for different levels of state-space reductions for the policy functions (since we solve with EGM). Here, we apply a much rougher approximation than in the previous subsection to show where the potential of strong dimensionality reduction comes from and compare this to the alternative of selecting the perturbed coefficients as those of a complete polynomial of a given order (a nonadaptive "sparse" type of approximation).

We present in the following the solution of the model in terms of policy functions and impulse responses based on retaining 10, 50, and all 200 coefficients of the discrete cosine transform of the policy function. First, we compare the policy function in the stationary equilibrium with the policy function that would have been obtained by solving the stationary equilibrium with the sparse Chebyshev polynomial, that is, actually shrinking the remaining smaller coefficients to zero already in the stationary equilibrium solution. The comparison can be seen in Figure 2. The approximation with 10 coefficients is fairly rough and unsatisfactory in quality as a description of the stationary equilibrium policy. It shows excessive fluctuation and oscillation. With 50 out of 200 coefficients, the approximation becomes much better, but small oscillations and approximation errors remain. Applying the method of Reiter (2009) or Winberry (2018), one might accept the sparse Chebyshev polynomial with 50 coefficients as an ex ante dimensionality reduction.

A low number of coefficients, however, has hardly any impact on the response of individual policies to a TFP shock; see Figure 3. The figure shows how consumption policies change (according to our solution) for different levels of sparseness of θ, that is, for a different number of retained coefficients. The reason for this is that the shock mostly produces a level shift for consumption together with a small change in the steepness of the consumption policy in wealth and income. Using the stationary equilibrium values of the small coefficients, changes in the large coefficients of the discrete cosine transform of the consumption policy can capture these shifts well. In other words, the stationary equilibrium policies provide a good "reference frame" that we can exploit for

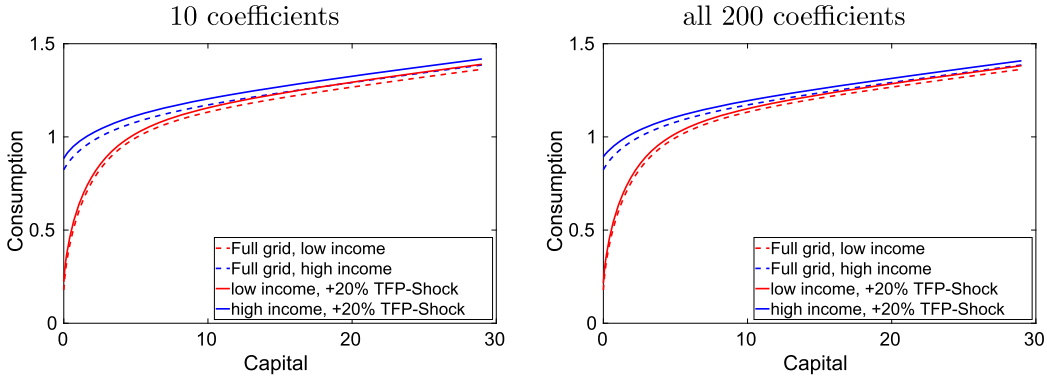

FIGURE 3. Change in consumption policies after a 20% TFP shock by sparseness of θ.

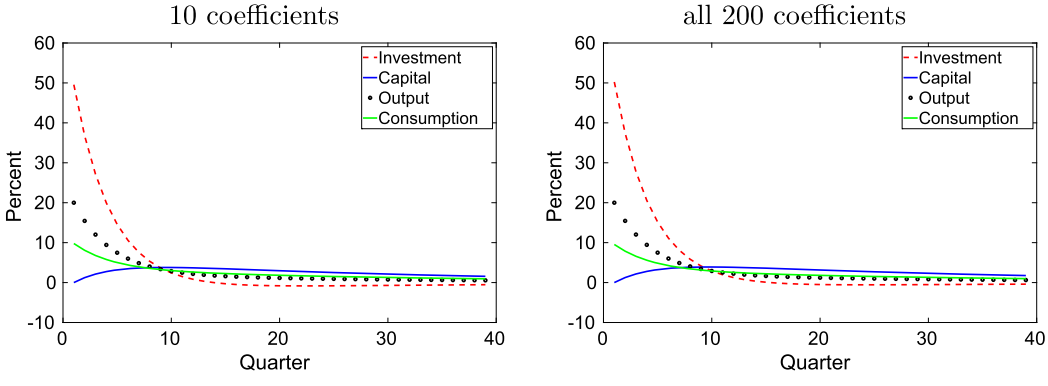

FIGURE 4. Aggregate response after a 20% TFP shock by sparseness of θ.

our solution. Not very surprisingly, with these small differences in individual policies, the aggregate responses look also indistinguishable; see Figure 4.

As we argued before, finding which of the coefficients are perturbed of a Chebyshev polynomial representation of the value/policy functions is in principle not an ex ante well-defined problem. Retaining those coefficients that are large when representing the *stationary equilibrium* value/policy functions is only a heuristic. An alternative (heuristic) would be to retain those coefficients that correspond to the complete (instead of full) polynomial. In practice, this means that we retain those coefficients that correspond to polynomial terms (over the two dimensions) that have a sum of exponents of at most some number N.

We compare this choice in Table 4 to our suggested choice of finding the coefficients to retain, that is, by perturbing only those coefficients that are large in the stationary equilibrium solution. Despite the fact that the complete polynomial choice has a somewhat stronger theoretical underpinning (being a Taylor expansion), in our practical example it performs substantially worse—especially when the number of retained coefficients becomes small. For less than 35 retained coefficients, the selection based on forming a complete polynomial of given order yields such a bad approximation that we get a

TABLE 4. Comparison of DCT-based coefficient selection to a nonadaptive rule.

Degree of polynomial, N	50	40	30	20
Number of coefficients	101	81	61	41
Selection of coefficients	Max absolute difference of log capital stocks ($\times 1e^{-8}$)			
(a) Complete polynomial	0.08	0.80	6.24	37.37
(b) DCT	0.10	0.43	0.07	0.46
	Mean absolute difference of log capital stocks ($\times 1e^{-8}$)			
(a) Complete polynomial	0.02	0.25	1.95	11.59
(b) DCT	0.03	0.13	0.02	0.13

Note: Relative differences between the simulated capital stock for 1000 periods obtained from using all coefficients of the policy function and either (a) the reduction Reiter method where coefficients are retained that form the complete polynomial of at most order N or (b) the reduction Reiter method with our proposed DCT-based selection of coefficients that retains the same number of coefficients as in (a).

violation of the Blanchard–Kahn condition and the model fails to solve. The DCT-based selection allows us to still solve for much fewer retained coefficients with relatively high precision.

The reason for the superior performance of the adaptive DCT-based method is that across different income states, the policy functions are relatively similar in the stationary equilibrium (think: one is an affine transformation of the other); the DCT method detects this, and this remains true even when prices change after a shock.

4.1.3 *Details on using the copula for dimensionality reduction* To understand how restrictive the assumption of a fixed copula is, we compare the model-implied distributions over time for the solution that does no reduction (Reiter–Full) and our method, which fixes the copula. We further consider an in-between case where we treat the copula as a functional that we represent through its DCT, perturbing only its most important coefficients. Details about the implementation can be found in Appendix A.1.

Figure 5 shows the result of this exercise. For the top row, we simulate the model using TFP shocks (as described before) as the driving force. As all households are similarly affected by the TFP shocks, there is no strong a priori reason for the copula to vary much over the cycle—of course, the marginals vary and so does the entire joint distribution. Indeed, we find that the approximation error measured in terms of the Jensen–Shannon distance (left column)[18] between the joint distribution (of assets and income) in the Reiter solution with and without the fixed copula assumption is an order of magnitude

[18]The Jensen–Shannon distance (JSD) is a metric for distribution functions. It is the square root of a symmetricized Kullback–Leibler divergence, where for two distribution functions f_1, f_2 over discrete support X the JSD is defined as

$$\mathrm{JSD}(f_1, f_2) = \sqrt{\frac{1}{2}\sum_{x\in X} f_1(x)\log\left[\frac{2f_1(x)}{f_1(x)+f_2(x)}\right] + f_2(x)\left[\log\frac{2f_2(x)}{f_1(x)+f_2(x)}\right]}. \tag{26}$$

To put the Jensen–Shannon distance in perspective, it is useful to think of comparing two normal distributions with unit variance that differ in means. The distance in that case is half the absolute difference of the means.

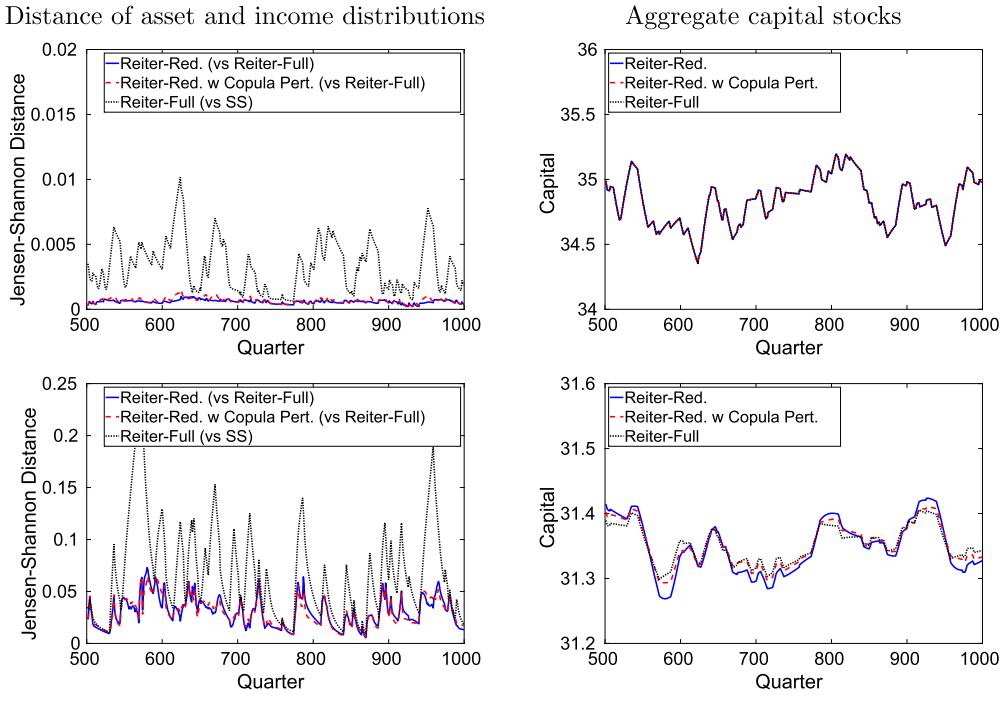

FIGURE 5. Distance between the distribution with and without fixed copula assumption. *Notes*: The top row compares simulated solutions of the Krusell and Smith model for a series of TFP shocks; the bottom row does the same for a series of income uncertainty shocks. The left column shows the Jensen–Shannon distance between the distribution of capital and income between the one implied by the full-grid Reiter method and by our reduction method, which treats the copula as fixed or perturbs only a few coefficients of the polynomial approximation for the copula obtained through a DCT. The right column compares the model solutions through the lens of the aggregate capital stock.

smaller than the distance between either solution and the stationary equilibrium distribution. The distance between the distributions is, at 0.0005, negligibly small. There is virtually no difference in the capital stock series (right column), as we know from the results in the previous section.

To consider a case where the copula varies more, we simulate the model with shocks to idiosyncratic income uncertainty as a driver of the cycle (see the next section as well). These shocks affect the joint distribution of assets and income directly, so that the fixed-copula assumption has more potential to introduce approximation errors. The bottom row of Figure 5 shows the results of this exercise. Now, the distance of the simulated distributions to the steady-state one is much larger and the difference between the distribution from the full Reiter solution and the one with a fixed copula attains a significant order of magnitude. We also find some difference in the fluctuations of the capital stock that the model implies—a model where the fluctuations in capital are small, as there is little aggregate feedback. However, perturbing the most important 41 coefficients (out of

possible 2100) of the DCT of the copula virtually eliminates the already small difference to the full Reiter solution.

4.2 *Second-order approximation*

Given that our approach keeps the number of derivatives to be calculated relatively low, it is possible to solve the model by second-order perturbation using the method of Schmitt-Grohé and Uribe (2004). This requires first obtaining the first-order solution (using the described qz-decomposition technique), then calculating second-order derivatives of F and finally solving a system of linear equations. For the Krusell and Smith model, this requires to calculate roughly 88 times the number of derivatives as for the first-order perturbation (in total 30,450). Of course, the calculation of derivatives can be parallelized, which allows to speed up higher order approximations substantially on computers with more cores.[19]

The left panel of Figure 6 presents the IRF of capital to a large TFP shock ($10\sigma_S$) for both the first-order and the second-order approximation of the K-S model. The right panel displays the ergodic distribution of capital for the same model in the first-order approximation (stationary equilibrium) and second-order approximation (average capital distribution over simulations).

We view this primarily as a proof-of-concept. For practical applications, one will need to further decrease the number of derivatives to be calculated by exploiting the economic structure of the problem, where, for example, the law of motion for the distributions is linear in the distribution. In Appendix A.3, we provide further details along this line.

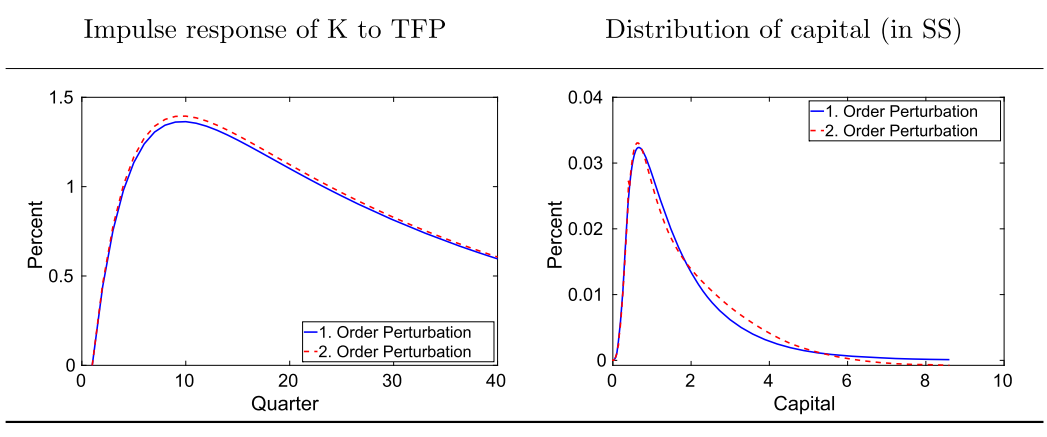

FIGURE 6. Second-order perturbation of Krusell–Smith model. *Notes*: Comparison of Reiter-reduction solution with first-order and second-order perturbation (for tenfold standard deviation of TFP shocks). Left panel shows the impulse response of capital. Right panel shows the steady-state marginal distribution of capital (as a multiple of steady-state aggregate capital).

[19]Note that solving the higher-order system itself requires limited additional computations, because the time consuming qz-decomposition does not change size when moving from first to higher order.

TABLE 5. Run times for two-asset model.

	Running times*	
	Stationary equilibrium	Reiter-Reduction
In seconds	1311	326

*On a Dell laptop with an Intel i7-7500U CPU @ 2.70 GHz, 4 cores. Code in Julia.

4.3 *Two-asset model*

The true advantage of the state and control space reduction through separating marginals and copula and compressing the latter alongside the value functions lies in tackling the curse of dimensionality, and thus, making it possible to solve models with high dimensional heterogeneity. In the following, we provide accuracy statistics and computational time for our model with a portfolio choice between liquid and illiquid assets as set out in Section 3, in particular Section 3.4.2. This model features heterogeneity with respect to three dimensions: (1) liquid asset holdings, (2) illiquid asset holdings, and (3) idiosyncratic productivity. We solve the household problem on 100 grid points for both asset choices and 12 grid points for productivity. With 120,000 states and 240,000 controls (for the two value functions), it is infeasible to solve for the aggregate dynamics of the model on the full histogram. The fixed copula approximation reduces the number of states to 236. Maintaining only the coefficients of the discrete cosine transform of the value functions with the cumulative highest 99.9999% energy reduces the number of controls to 1427.[20] This all together makes it possible to solve the model on a laptop computer in, as the top panel of Table 5 shows, only 5 minutes (plus an additional 22 minutes for the stationary equilibrium).

4.3.1 *Frictionless version* We first solve the model for the same calibration as the Krusell and Smith model in the previous section.[21] Table 6 shows the error metric suggested by Den Haan (2010a) for the capital stock implied by the two-asset model in response to TFP shocks. The maximum absolute error is 0.12% and the mean absolute error is 0.05%, which are comparable to the errors in Table 2 for the single-asset model. The errors for equilibrium bonds are slightly larger when measured relative to bonds themselves. Bonds are only 10% of the capital stock in the steady state so that, relative to capital or output, the errors are comparable to the errors for capital.

To assess how sensitive this result is to our baseline numerical specification, we also consider a specification that perturbs the copula, too, and a specification that retains more coefficients of the DCT, allowing the energy to drop only by 10^{-7} instead of 10^{-6}. Retaining more DCT coefficients or perturbing the copula has little impact on the approximation in terms of the den Haan test; see again Table 6. The quality of the approximation for capital slightly increases with more DCT coefficients. Surprisingly, perturbing also the copula worsens the approximation quality marginally.

[20]With the richer model, some of the histogram entries contain very little mass and numerical derivatives become less precise. In particular, this issue is important when treating the copula as time varying. For this reason, we solve the two-asset model in Julia where we can use a package for automatic differentiation.

[21]Appendix B, Table 12 provides the full calibration.

TABLE 6. Accuracy for two-asset model.

	Absolute error (in %)*			
	Mean, for ...		Max, for ...	
	capital K_t	bonds B_t	capital K_t	bonds B_t
Baseline	0.033	0.081	0.092	0.617
Retain *more* DCT coefficients	0.031	0.080	0.087	0.610
Baseline + perturb copula	0.043	0.080	0.151	0.777

*Differences in percent between the simulation of the linearized solution of the model and a simulation in which we solve for the intratemporal equilibrium prices in every period and track the full histogram over time for $t = \{1, \ldots, 1000\}$; see Den Haan (2010a).

TABLE 7. Business cycle statistics for the two-asset model.

	$\sigma(Y_t)$	$\sigma(C_t)$	$\sigma(I_t)$
Baseline	1.40	1.39	4.49
Retain *more* DCT coefficients	1.42	1.42	4.49
Baseline + perturb copula	1.31	1.28	4.56

Note: The table displays the standard deviations of output Y_t, consumption C_t, and investment I_t, alongside the Sharpe-ratio for the excess return of illiquid assets for the two-asset model with only a TFP shock as described in the main text. Baseline refers to the numerical baseline specification, where we retain coefficients of the DCTs to maintain 99.9999% of the "energy" in the policy functions. In the second row, we increase this to 99.99999%. The third row refers to a setup, where we perturb also 50 coefficients of the DCT representation of the copula.

4.3.2 *Business cycles, asset prices* The calibration of the model as in the previous subsection is meant to be as comparable as possible to the Krusell and Smith setup with one asset—there are no further frictions besides market incompleteness. A practical application of the method to a business cycle model, however, typically will feature nominal and real frictions. Appendix B, Table 13 summarizes such a calibration that we use in the following.

Table 7 shows that for this business cycle calibration with TFP, monetary, and uncertainty shocks also the business cycle statistics do vary relatively little, when we change the numerical specification. This resembles the results displayed in Figure 4 for the KS model. Again, we consider a specification that perturbs the copula, too, and a specification that retains more coefficients of the DCT, allowing the energy to drop only by 10^{-7} instead of 10^{-6}. Perturbing the copula slightly lowers output volatility and increases the volatility of investment somewhat.

In addition to the business cycle statistics, we also look at how the model performs in terms of asset prices; see Table 8. We report the average and maximum absolute deviation from asset market clearing that the linearized solution produces, that is, we evaluate the difference in asset supply and demand for both B and K given prices and the wealth distribution that we get from the simulation of the linearized solution. We find that deviations from exact market clearing for the two assets are small and of similar order of magnitude as the Den Haan statistics in Table 6.

What is more, we calculate the Sharpe ratio for the model across the various numerical variants. Since the model produces a steady state return difference, an illiquidity

TABLE 8. Asset market clearing for the two-asset model.

	Deviation from Market Clearing in %				
	Mean-Absolute		Max-Absolute		
	on K_t	on B_t	on K_t	on B_t	Sharpe-Ratio
Baseline	0.03%	0.05%	0.29%	0.65%	0.70
Retain *more* DCT coefficients	0.03%	0.05%	0.28%	0.64%	0.67
Baseline + perturb copula	0.01%	0.02%	0.22%	0.59%	0.68

Note: The table displays the deviation from market clearing in the asset markets obtained by comparing the model solution for bonds B and capital K (the bond supply and capital demand by firms) to values implied by the solution for the wealth distribution. Baseline refers to the numerical baseline specification, where we retain coefficients of the DCTs to maintain 99.9999% of the "energy" in the policy functions. In the second row, we increase this to 99.99999%. The third row refers to a setup, where we perturb also 50 coefficients of the DCT representation of the copula.

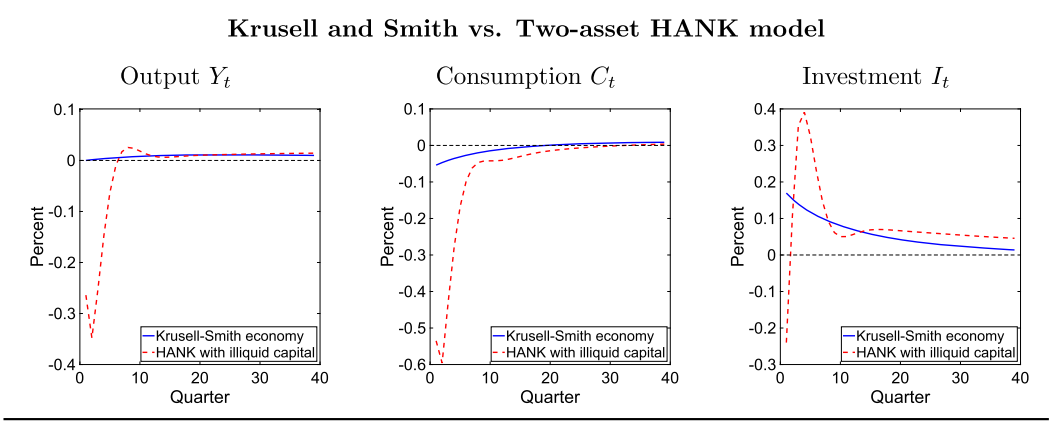

FIGURE 7. Aggregate response to idiosyncratic uncertainty shock. *Notes*: Response to a 54% increase in uncertainty (measured by STD) of idiosyncratic income.

premium for capital, the model gets a long way in terms of being close to the observed Sharpe-ratios of Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019) that range between 0.6 for housing and 0.25 for equities. The model has somewhat to stable asset returns if anything.

This illiquidity premium is closely tied to the concept of "wealthy hand-to-mouth" households (Kaplan and Violante (2014)) and the latter implies different investment behavior (see Bayer et al. (2019) and Luetticke (2018)). Figure 7 shows the effect of higher uncertainty about idiosyncratic productivity in the Krusell and Smith model and the two-asset HANK model. Consumption falls in both models as households increase their precautionary savings in response to higher uncertainty. In the Krusell and Smith model, higher savings translate one-for-one into capital, which leads to an economic expansion. In the two-asset model, by contrast, households prefer to hold more liquid portfolios. They sell illiquid capital to save more in liquid assets. Higher uncertainty therefore causes a simultaneous fall in consumption, investment, and output. The recessionary effect is further amplified through sticky prices, which makes the economy demand-

driven in the short run. See Bayer et al. (2019) for a more detailed discussion of the portfolio rebalancing channel of uncertainty.

5. Conclusion

In this paper, we propose an extension of Reiter's method to solve heterogeneous agent models with aggregate risk by perturbation. The proposed method relies on reducing the state space after solving for the stationary equilibrium but before linearizing the nonlinear difference equation that characterizes the equilibrium dynamics. The state-space reduction is achieved by "lossy compression" of the value functions, which are control variables of the system, and by approximating the dynamics of the multidimensional distribution of individual characteristics by a distribution with an (almost) fixed copula and varying marginals. Both steps effectively reduce the problem that high-dimensional idiosyncratic state spaces pose and allow us to efficiently and precisely solve for the equilibrium dynamics of heterogeneous agent economies as we have shown in two examples.

Dealing with the curse of dimensionality is essential because it allows us to analyze business cycle models with rich heterogeneity. Examples that go beyond what we show here are models where aging adds another dimension to the household problem or where a richer household portfolio needs to be modeled, for example, when households own liquid assets, own houses, and write mortgages at the same time. To all these setups, the proposed method lends itself well to solving for equilibrium dynamics.

References

Ahmed, N., T. Natarajan, and K. R. Rao (1974), "Discrete cosine transform." *IEEE Transactions on Computers*, C-23 (1), 90–93. [1263]

Ahn, S., G. Kaplan, B. Moll, T. Winberry, and C. Wolf (2017), "When inequality matters for macro and macro matters for inequality." In *NBER Macroeconomics Annual*, Vol. 32, National Bureau of Economic Research. [1253, 1255, 1256, 1262, 1264]

Algan, Y. (2010), "Solving the incomplete markets model with aggregate uncertainty using parameterized cross-sectional distributions." *Journal of Economic Dynamics and Control*, 34 (1), 59–68. Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty. [1253]

Bayer, C., R. Luetticke, L. Pham-Dao, and V. Tjaden (2019), "Precautionary savings, illiquid assets, and the aggregate consequences of shocks to household income risk." *Econometrica*, 87 (1), 255–290. [1264, 1271, 1284, 1285]

Bayer, C. and R. Luetticke (2020), "Supplement to 'Solving discrete time heterogeneous agent models with aggregate risk and many idiosyncratic states by perturbation'." *Quantitative Economics Supplemental Material*, 11, https://doi.org/10.3982/QE1243. [1256, 1274]

Bungartz, H.-J. and M. Griebel (2004), "Sparse grids." *Acta Numerica*, 13, 147–269. [1255, 1259]

Carroll, C. D. (2006), "The method of endogenous gridpoints for solving dynamic stochastic optimization problems." *Economics Letters*, 91 (3), 312–320. [1258]

Den Haan, W. J. (2010a), "Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents." *Journal of Economic Dynamics and Control*, 34 (1), 79–99. [1256, 1265, 1275, 1276, 1282, 1283]

Den Haan, W. J. (2010b), "Comparison of solutions to the incomplete markets model with aggregate uncertainty." *Journal of Economic Dynamics and Control*, 34 (1), 4–27. Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty. [1253]

Den Haan, W. J., K. L. Judd, and M. Juillard (2010), "Computational suite of models with heterogeneous agents: Incomplete markets and aggregate uncertainty." *Journal of Economic Dynamics and Control*, 34 (1), 1–3. [1253, 1275, 1276]

Den Haan, W. J. and P. Rendahl (2010), "Solving the incomplete markets model with aggregate uncertainty using explicit aggregation." *Journal of Economic Dynamics and Control*, 34 (1), 69–78. Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty. [1253]

Hintermaier, T. and W. Koeniger (2010), "The method of endogenous gridpoints with occasionally binding constraints among endogenous variables." *Journal of Economic Dynamics and Control*, 34 (10), 2074–2088. [1258]

Hu, X.-Y. and X.-H. Yu (1998), "The Chebyshev polynomial fitting properties of discrete cosine transform." *Signal Processing: Image Communication*, 13 (1), 15–20. [1263]

Jordà, Ò., K. Knoll, D. Kuvshinov, M. Schularick, and A. M. Taylor (2019), "The rate of return on everything, 1870–2015." *The Quarterly Journal of Economics*, 134 (3), 1225–1298. [1284]

Kaplan, G. and G. L. Violante (2014), "A model of the consumption response to fiscal stimulus payments." *Econometrica*, 82, 1199–1239. [1284]

Kim, S. H., R. Kollmann, and J. Kim (2010), "Solving the incomplete market model with aggregate uncertainty using a perturbation method." *Journal of Economic Dynamics and Control*, 34 (1), 50–58. Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty. [1253]

Klein, P. (2000), "Using the generalized Schur form to solve a multivariate linear rational expectations model." *Journal of Economic Dynamics and Control*, 24 (10), 1405–1423. [1261]

Krueger, D. and F. Kubler (2004), "Computing equilibrium in OLG models with stochastic production." *Journal of Economic Dynamics and Control*, 28 (7), 1411–1436. [1255, 1259]

Krusell, P. and A. A. Smith (1997), "Income and wealth heterogeneity, portfolio choice, and equilibrium asset returns." *Macroeconomic Dynamics*, 1 (2), 387–422. [1253, 1254]

Krusell, P. and A. A. Smith (1998), "Income and wealth heterogeneity in the macroeconomy." *Journal of Political Economy*, 106 (5), 867–896. [1253, 1254, 1256, 1258, 1264, 1268, 1271, 1274, 1275, 1276, 1281]

Luetticke, R. (2018), "Transmission of monetary policy with heterogeneity in household portfolios." CFM Discussion Paper Series (CFM-DP2018-19). [1264, 1271, 1284]

Maliar, L., S. Maliar, and F. Valli (2010), "Solving the incomplete markets model with aggregate uncertainty using the Krusell–Smith algorithm." *Journal of Economic Dynamics and Control*, 34 (1), 42–49. Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty. [1253]

Mertens, T. M. and K. L. Judd (2018), "Solving an incomplete markets model with a large cross-section of agents." *Journal of Economic Dynamics and Control*, 91 (C), 349–368. [1262]

Ragot, X. (2018), "Heterogeneous agents in the macroeconomy: Reduced-heterogeneity representations." In *Handbook of Computational Economics*, Vol. 4, 215–253, Elsevier. [1262]

Reiter, M. (2002), "Recursive computation of heterogeneous agent models." Report, Universitat Pompeu Fabra. [1253, 1255, 1261, 1275]

Reiter, M. (2009), "Solving heterogeneous-agent models by projection and perturbation." *Journal of Economic Dynamics and Control*, 33 (3), 649–665. [1253, 1255, 1256, 1258, 1262, 1263, 1264, 1268, 1277, 1285]

Reiter, M. (2010a), "Approximate and almost-exact aggregation in dynamic stochastic heterogeneous-agent models." Economics Series 258, Institute for Advanced Studies. [1255]

Reiter, M. (2010b), "Solving the incomplete markets model with aggregate uncertainty by backward induction." *Journal of Economic Dynamics and Control*, 34 (1), 28–35. Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty. [1253]

Rotemberg, J. J. (1982), "Sticky prices in the United States." *Journal of Political Economy*, 1187–1211. [1270]

Schmitt-Grohé, S. and M. Uribe (2004), "Solving dynamic general equilibrium models using a second-order approximation to the policy function." *Journal of Economic Dynamics and Control*, 28 (4), 755–775. [1253, 1254, 1261, 1268, 1281]

Taylor, J. B. (1993), "Discretion versus policy rules in practice." *Carnegie–Rochester Conference Series on Public Policy*, 39, 195–214. [1273]

Winberry, T. (2018), "A method for solving and estimating heterogeneous agent macro models." *Quantitative Economics*, 9 (3), 1123–1151. [1255, 1262, 1263, 1277]

Woodford, M. (1995), "Price-level determinacy without control of a monetary aggregate." *Carnegie–Rochester Conference Series on Public Policy*, 43 (Supplement C), 1–46. [1273]

Young, E. R. (2010), "Solving the incomplete markets model with aggregate uncertainty using the Krusell Smith algorithm and non-stochastic simulations." *Journal of Economic Dynamics and Control*, 34 (1), 36–41. Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty. [1253, 1259]

Co-editor Kjetil Storesletten handled this manuscript.

Manuscript received 15 November, 2018; final version accepted 10 July, 2020; available online 5 August, 2020.