Solving Heterogeneous Agent Models with Non-convex Optimization Problems: Linearization and Beyond

Michael Reiter, Institute for Advanced Studies, Vienna and NYU Abu Dhabi

February 15, 2019

PRELIMINARY VERSION!

Abstract

This paper presents methods for heterogeneous agent models where agents solve non-convex optimization problems. It shows how to apply the linearization approach of Reiter (2009) to non-convex models, and develops a theory of state and value function reduction to handle models with very large state spaces. It shows the potential problems of the linearization approach and ways to diagnose them. To overcome these problems, global nonlinear solution algorithms are presented, based on temporary equilibrium concepts. The methods are applied to models with heterogeneous households and indivisible labor, as well as to a model of heterogeneous firms with lumpy investment.

JEL classification: C63, C68, E21
Keywords: heterogeneous agents; model reduction; linearization; temporary equilibrium

Address of the author:

Michael Reiter
Institute for Advanced Studies
Josefstädter Strasse 39
A-1080 Vienna
Austria

e-mail: michael.reiter@ihs.ac.at

1 Introduction

The solution of heterogeneous agent models with incomplete markets poses important technical challenges, mainly because the underlying state space is very high-dimensional. Krusell and Smith (1998) use a very low-dimensional approximation to the aggregate law of motion, which is highly successful in their specific application, and this method has been the workhorse in this field for 20 years. In Reiter (2009) I take the opposite approach: the solution is linearized in aggregate variables (while fully nonlinear in the solution of the individual problem), which allows to keep track of very high-dimensional approximation of the cross-sectional distribution. I the current paper I push this agenda further, making three contributions. First and most importantly, I present a theory of state and value function reduction which allows to reduce the model dimension by several orders of magnitude, with almost no loss in accuracy. This greatly expands the set of models that can be solved by linearization, and will facilitate the estimation of heterogeneous agent models. This method is rather complicated, but it is automated in a toolkit that is programmed in Julia and uses a special syntax for heterogeneous agent models, similar to what Dynare does for representative agent models. Second, I show how to linearize models where the optimization problems of economic agents is non-convex, due to one or more choice variables being discrete. This case is very common in state-of-the art heterogeneous agent models. Discrete choice generally leads to discontinuities in the continuous policy functions and to irregular cross-sectional distributions. To correctly identify extensive-margin effects, my method identifies the switch points between different discrete choices, and monitors the change of switch points during the simulation. Third, I illustrate limitations of linearization, how to diagnose them and ways to solve them. It is clear that certain classes of models require solution methods that are nonlinear also in aggregate variables. Examples are models with occasionally binding constraints at the aggregate level, such as the zero lower bound on the nominal interest rate, and some types of models of portfolio choice. But there are less obvious reasons why linearization may fail to give a reliable approximation. Models with non-convex optimization problems often generate cross-sectional distributions with rapidly changing densities, which lead to asymmetries in aggregate impulse responses. It is not difficult to diagnose these asymmetries, but to overcome the problem it is necessary to compute global nonlinear approximations. I will present a global collocation method that is relatively easy to compute if the linearized solution is available.

The model reduction presented here is using the state reduction in my working paper Reiter (2010a), and completes the theory by adding optimal value function reduction. The approach in Reiter (2010a), has been applied in McKay and Reis (2016) and Reiter, Sveen, and Weinke (2013). Ahn, Kaplan, Moll, Winberry, and Wolf (2018) developed a similar linearization technique for continuous time models, using a smooth approximation of value function. The approach also has similarities to Kubler and Scheidegger (2018), who reduce the state space to a very small dimension, using the concept of "self-justified equilibria",

and give it a bounded-rationality interpretation. In contrast, my approach tries to give an approximation that is as close as possible to the rational expectations equilibrium under full information.

There has been growing interest recently in linearization and perturbation approaches to heterogeneous agent models. Boppart, Krusell, and Mitman (2018) show that there is an alternative and simpler approach to compute linearized solution, because the simulation of a linearized model is just a linear superposition of impulse responses to "MIT shocks". Bhandari, Evans, Golosov, and Sargent (2018) derive a new perturbation method that can be applied to certain problems of optimal policy with heterogeneous agents. Winberry (2018) uses a low-dimensional smooth approximation of cross-sectional distribution, to allow for higher-order perturbation solutions. Mertens and Judd (2017) perform the perturbation around deterministic steady state without aggregate and with idiosyncratic shocks. Using methods of functional analysis, Childers (2016) gives a theoretical foundation of finite-dimensional approximations to models with infinite-dimensional state space.

2 Example Models

2.1 Example 1: the Model of Chang and Kim (2007)

This model is very similar to the well known model in Krusell and Smith (1998), except for introducing indivisible labor. Ex-ante identical households face shocks to idiosyncratic labor productivity, which follows a Markov process. Labor markets work frictionlessly, but labor is indivisible: a household can work either zero hours or a fixed number of hours.

Technology is standard. There is a representative firm with production function:

$$Y_t = F(L_t, K_t, \lambda_t) = \lambda_t L_t^{\alpha} K_t^{1-\alpha} \tag{1}$$

where λ_t is Markov with transition probability distribution π_λ. Marginal productivity determines factor prices.

The household value function is given by

$$V(a, x; \lambda, \phi) = \max_{a' \in \mathcal{A}, h \in \{0, \bar{h}\}} \left\{ u(c, h) + \beta \, \mathrm{E} \, V(a', x'; \lambda', \phi') \right\} \tag{2}$$

s.t.

$$c = w(\lambda, \phi) x h + (1 + r(\lambda, \phi)) a) - a' \tag{3}$$

$$a' \geq \bar{a} \tag{4}$$

$$\phi' = T(\lambda, \phi) \tag{5}$$

where x is exogenous individual productivity, λ is aggregate TFP, and ϕ is the cross-sectional distribution of agents over (a, x). In this model, the aggregate number of hours is determined

exclusively by the extensive margin, namely the fraction of households who decide to work in a given period. The effective labor supply also depends on the productivity level of the working households. The labor supply reaction is determined by two things: the change in the threshold level of capital where households switch from working to non-working, and the mass of households close to this threshold. A main computational challenge is therefore to pin down the labor supply response.

To get a first idea of the properties of the model, let us look at the steady-state results presented in Figure 1). The upper left panel shows the cumulative density function of household assets, summed up over all productivity levels. Since aggregate capital is 11.2, we see that very few households hold more than six times the average amount of assets. The upper right panel shows the density of assets for three different productivity levels (levels 9–11 out of 17). The asset distribution has clear spikes, which occur at approximately the same places for the three productivity levels. This is because households change frequently between adjacent productivity levels. The spikes are a consequence of the discontinuity of household saving, which is apparent in the lower left panel of the graph. It shows household savings (change in assets) for the middle levels of household productivity (7–11 out of 17 possible levels). Net saving is first positive, but drops sharply at the threshold level where households stop working. Since households tend to smooth consumption, the reduction in labor earnings is mostly reflected in savings. At a certain level of assets, households switch from employment to non-employment. Since they want to smooth consumption, the loss of labor income requires a reduction in savings. Notice that, in the absence of labor market frictions, households can frequently switch in and out of employment, The lower right panel shows some snippets of the cross-sectional distribution of assets in the steady state, for the same five levels of individual labor productivity. The x-axis indicates the histogram bin in the capital distribution, relative to the bin in which the labor market participation threshold lies. The y-axis draws the fraction of households in this bin. This means, for each level of productivity, the value at x=0 gives the number of households in the capital bin in which the participation threshold lies. Since changes in aggregate labor supply result from changes in the participation threshold, the density in these bins is crucial for determining aggregate labor supply. The sharp change in the density around the participation threshold makes it difficult to pin down the extensive margin effect of labor supply, and is likely to induce asymmetries in aggregate labor responses. We will analyze this in detail in Section 5.1 This is because households that switch from working to non-working also reduce their saving level.

The main objective of Chang and Kim (2007) is to study the observed "labor wedge", defined as the discrepancy between wage and observed MRS between consumption and leisure:

$$wedge = w - \frac{U_L}{U_C} \tag{6}$$

where we use a standard utility function $U(C, L)$. Why are there systematic fluctuations in the wedge over the business cycle? The claim is that this is caused by the combination of

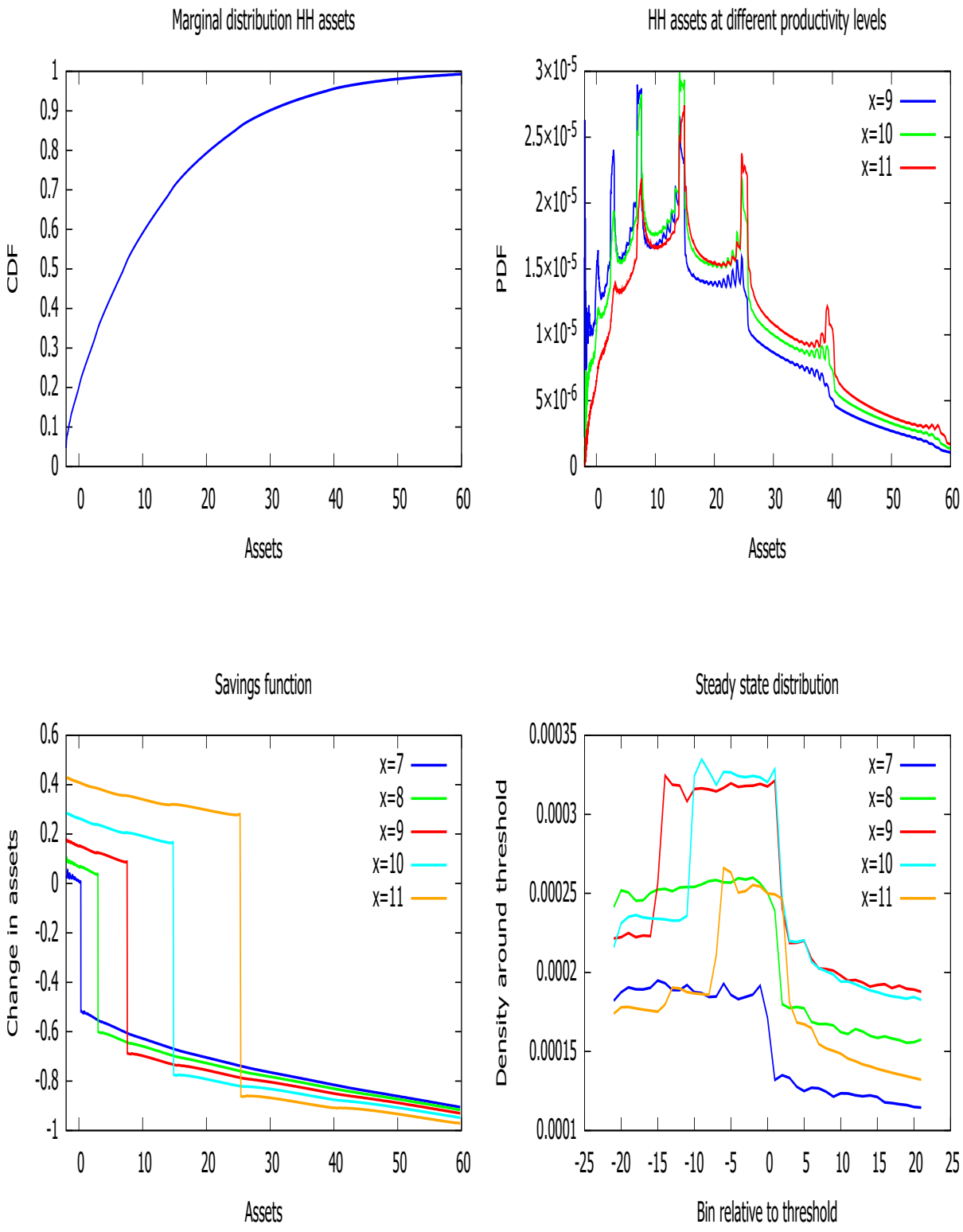


Figure 1: Chang/Kim model, steady state results

indivisible labor in combination with aggregation over heterogeneous households.

2.2 Example 2: The model of Khan and Thomas (2008)

This is a model of heterogeneous firms with lumpy investment. For the details of the model and the calibration, I refer the reader to the original paper.

2.3 A general model

For notational simplicity, we assume there is only one type of heterogeneous agent (the households in the Chang/Kim model, or the firms in the uncertainty model). Generalizing this to the case of several ex-ante heterogeneous agents is conceptually straightforward. The problem of this agent is described as a dynamic programming problem. The rest of the model consists of a finite set of dynamic equations, just as in a standard DSGE model.

The decision problem of the heterogeneous agent is assumed to be of the following form. For the agent there are four types of state variables:

1. One continuous individual endogenous state variable (such as capital), denoted by k.

2. One discrete individual endogenous state variable (such as employment states, size of owned house out of a finite choice of house sizes, etc.), denoted by e.

3. One discrete individual exogenous state variable (such as labor productivity) denoted by z.

4. The set of aggregate states, denoted by Ω, which consists of the predetermined states $\mathbf{D}$ and the current states Z. $\mathbf{D}$ contains the cross-sectional distribution of individual states at the end of the last period, and potentially some other predetermined states. Z typically consists of the current values of the exogenous driving processes. The agent takes the transition law of the aggregates state Ω as given.

The agent has two control variables:

1. The continuous control variable k', which for convenience we assume is equal to the end-of-period continuous state.

2. The discrete control variable d. We assume that this variable determines the end-of-period discrete state by the transition law

$$e' = T(e, d) \tag{7}$$

   It also influences current utility. For example, in the Chang/Kim model there is no endogenous discrete state, but the discrete choice (working or not working) affects current utility, conditional on the end-of-period capital of the household.

The transition law of the individual exogenous state is of course independent of the actions of the individual agent, but it can depend on the current aggregate state. We assume it is characterized by a finite Markov chain with transition probabilities $\pi(z, z'; \Omega)$. At this state we are not making any assumptions about the transition law of the aggregate state Ω.

This setup is somewhat more general than it may appear. The agent may have more continuous control variables, which one can handle if the other continuous controls are related to the end-of-period state by static optimality conditions, for example an optimal labor supply condition conditional on saving. A discrete state variable is often a discrete approximation to a continuous state variable, where we assume it is not necessary to have a very fine approximation. Several discrete variables can of course always combined into one discrete variable by forming a Cartesian product. We assume for notational simplicity that the end-of-period endogenous states are also the beginning-of-period states in the following period. One could easily generalize this by introducing some further exogenous noise into the individual transition laws.

Discrete choice renders the decision problem non-convex. At points in the state space where the discrete choices changes, the continuous choice generally also jumps. In the numerical solution, we will use a discrete grid also for the continuous state, but we will identify the value of the continuous choice where the discrete choice jumps (the "switch points"). The change of the switch points in reaction to changes in the aggregate environment defines the extensive-margin reaction of the agents.

At an aggregate state Ω, we define the individual policy function $\mathcal{A}(\Omega)$ as comprising the two decision functions $k'(k, e, z; \Omega)$ and $d(k, e, z; \Omega)$. Apart from the dynamic optimization problem of agents, the model is characterized by a transition equation for the cross-sectional distribution:

$$\mathbf{D}' = \mathcal{T}(\mathbf{D}, Z, p, \mathcal{A}) \tag{8}$$

and equilibrium conditions for endogenous aggregate variables as a function of state and decisions:

$$0 = \mathcal{E}(p, \mathbf{D}, Z, \mathcal{A}) \tag{9}$$

3 Linear Approximation and State Reduction

The main purpose of this section is to develop a model reduction technique which makes it possible to compute the linearized solution of Reiter (2009) for models with even larger state spaces, where standard methods for linear rational expectation systems are infeasible. In this case, the linearized solution requires the following steps:

1. A finite-dimensional approximation of the model, cf. Section 3.1.

2. Model reduction (Sections 3.2–3.5).

3. Simulating the model (Section 3.6).

4. Accuracy check (Section 3.7).

5. Balanced reduction (Section 3.8).

3.1 Finite approximation of HA model

The finite approximation (discretization) of a HA model was explained in Reiter (2009) (cf. also Costain and Nakov (2011)) and is now standard. Here I will mainly explain the elements that refer to the value function approximation, because this is the part that is most affected by the non-convexity of the agents' decision problem.

3.1.1 Discretization of the value function

The problem of the heterogeneous agents is solved as a discrete dynamic programming problem. For this, choose a grid G_k of the continuous individual state variable with n_k points, denoted by $\bar{k}_j$ for $j = 1, \ldots, n_k$. From this, form a grid of individual states G_x of $n_k \cdot n_e \cdot n_z$ points by a Cartesian product between the capital grid and the sets of discrete individual states:

$$G_x = G_k \times \{e_1, \ldots, e_{n_e}\} \times \{z_1, \ldots, z_{n_z}\} \tag{10}$$

Denote the elements of this grid as $\bar{x}_i$ for $i = 1, \ldots, n_k \cdot n_e \cdot n_z$. Between grid points, the value function will be interpolated in k by a quadratic interpolation method that preserves the contraction property of the Bellman iteration (for details, cf. Appendix A).

This grid will be used in both the linearized and the nonlinear solution.

3.1.2 Finite approximation of the cross-sectional distribution

The cross-sectional distribution of individual states will also be approximated on a finite grid. There are two simple ways to approximate continuous distribution on a finite grid. The first one is to model them as a combination of point masses, more concretely as the fraction of agents at each point of a fixed grid, which was used for example in Young (2010). The second one is to model the distribution as a histogram, assuming a constant density within a histogram bin, which means *between* two grid points (Reiter 2009). For a problem with both continuous and discrete decisions, the histogram approach is the appropriate one, for the following reason. The model solution determines threshold points, where the discrete choice changes. The extensive margin effect in a model will be determined by the change in the threshold points, as well as the density of agents at the threshold. Since the thresholds will generally not lie on a predetermined grid, the approximation as point masses cannot capture this effect. In the histogram approach, the relevant density is the density in the histogram bin in which the threshold lies.

For the histogram of the cross-sectional distribution one can use the same grid as for the value function approximation, but in may in general be useful to allow for a different choice of grid points, to model the cross-sectional dynamics with better precision. We therefore introduce a different grid G_k^D of the continuous individual state variable with n_D points, denoted by $\bar{k}_j^D$ for $j = 1, \ldots, n_D$. From this, we form a grid of individual states G_x^D of $n_D \cdot n_e \cdot n_z$ points.

3.1.3 The discretized model

The model contains three types of variables and equations.

1. The value function at each of the points in the grid G_k. This gives $n_k \cdot n_e \cdot n_z$ points in each period.

   The corresponding equation for each of these variables is the Bellman equation (11) at this grid point.

2. The mass of agents in each histogram bin defined by the grid G_k^D. This gives $(n_k - 1) \cdot n_e \cdot n_z$ points in each period.

   The corresponding equation is the transition dynamics at this point, which depends on the individual policy function. Details are given in Appendix B.

3. Aggregate variables such as GDP, TFP etc., and their corresponding equations. This is the same as in standard DSGE models.

All these variables at time t are collected into the vector θ_t. Since this vector can be huge, some form of model reduction may be needed, which is the topic of Sections 3.2–3.5. Notice that the optimal decisions k and e are not part of θ_t. Given current states and next period's value function, they are implicitly given by the optimality conditions.

3.1.4 Differentiating the value function

The solution of the agent problem is characterized by the Bellman equation

$$V\left(k, e, z, \Omega, Z, p\right) = \max_{d, k'} \tilde{V}\left(k, e, z, \Omega, Z, d, k', p\right) \tag{11}$$

where the current value function conditional on current equilibrium variables p and on current actions d, k' is defined as

$$\tilde{V}\left(k, e, z, \Omega, Z, d, k', p\right) \equiv U\left(k, e, z, d, k'; p, \Omega\right) + \beta \sum_{z'} \pi(z, z'; \Omega)\, \mathrm{E}_{\Omega'} \left\{V\left(k', T(e, d), z', \Omega'\right)\right\} \tag{12}$$

Writing $k'()$ and $d()$ for the optimal decision, we can write the Bellman equation as

$$V\left(k, e, z, \Omega, Z\right) = U\left(k, e, z, d(), k'(); p, \Omega\right) + \beta \sum_{z'} \pi(z, z'; \Omega)\, \mathrm{E}_{\Omega'} \left\{V\left(k'(), T(e, d()), z', \Omega'\right)\right\} \tag{13}$$

For the linearized solution, we have to differentiate (13) with respect to all the variables of the model. Define by ω any of the variables with respect to which we differentiate. Then we get from (13) that

$$\begin{aligned}\frac{\partial V\left(k,e,z,\Omega,Z\right)}{\partial\omega} &= \frac{\partial U\left(k,e,z,d(),k'();p,\Omega\right)}{\partial k'}\frac{\partial k'()}{\partial\omega} + \frac{\partial U\left(k,e,z,d(),k'();p,\Omega\right)}{\partial\omega} \\ &\quad + \beta\Big[\frac{\partial\sum_{z'}\pi(z,z';\Omega)}{\partial\omega}\,\mathrm{E}_{\Omega'}\left\{V\left(k'(),T(e,d()),z',\Omega'\right)\right\} \\ &\quad + \sum_{z'}\pi(z,z';\Omega)\,\mathrm{E}_{\Omega'}\left\{\frac{\partial V\left(k'(),T(e,d()),z',\Omega'\right)}{\partial\omega} + \frac{\partial V\left(k'(),T(e,d()),z',\Omega'\right)}{\partial k'}\frac{\partial k'}{\partial\omega}\right\}\Big] \quad (14)\end{aligned}$$

The problem is that $\frac{\partial k'}{\partial\omega}$ is not known. However, since we assume that k' is bound constrained, there are two possible cases. Either the optimal k' is interior, in which case the envelope theorem applies and (14) simplies to

$$\begin{aligned}\frac{\partial V\left(k,e,z,\Omega,Z\right)}{\partial\omega} &= \frac{\partial U\left(k,e,z,d(),k'();p,\Omega\right)}{\partial\omega} \\ &\quad + \beta\Big[\frac{\partial\sum_{z'}\pi(z,z';\Omega)}{\partial\omega}\,\mathrm{E}_{\Omega'}\left\{V\left(k'(),T(e,d()),z',\Omega'\right)\right\} \\ &\quad + \sum_{z'}\pi(z,z';\Omega)\,\mathrm{E}_{\Omega'}\left\{\frac{\partial V\left(k'(),T(e,d()),z',\Omega'\right)}{\partial\omega}\right\}\Big] \quad (15)\end{aligned}$$

Or k' is at a constraint, and which case $\frac{\partial k'}{\partial\omega}$ is simply the derivative of the constraint.

For given equilibrium vector p and discrete states e,z a threshold point k^* where the continuous choice switches from k'_- to k'_+ is characterized by

$$\begin{aligned}&\max_{k'_-} U\left(k^*,e,z,d_-,k'_-;p,\Omega\right) + \beta\sum_{z'}\pi(z,z';\Omega)\,\mathrm{E}_{\Omega'}\left\{V\left(k'_-,T(e,d_-),z',\Omega'\right)\right\} = \\ &\qquad \max_{k'_+} U\left(k^*,e,z,d_+,k'_+;p,\Omega\right) + \beta\sum_{z'}\pi(z,z';\Omega)\,\mathrm{E}_{\Omega'}\left\{V\left(k'_+,T(e,d_-),z',\Omega'\right)\right\} \quad (16)\end{aligned}$$

The extensive margin effect of any shock or change in parameters depends largely on how it affects this thresholds, and therefore it is essential that we compute the derivative of the threshold w.r.t. any other variables. It is given by

$$\frac{\partial k^*}{\partial\omega} = -\frac{\frac{\partial\tilde{V}(X_+)}{\partial\Omega}\frac{\partial\Omega|}{\partial\omega} + \frac{\partial\tilde{V}(X_+)}{\partial p}\frac{\partial p}{\partial\omega} + \frac{\partial\tilde{V}(X_+)}{\partial k'_+}\frac{\partial k'_+}{\partial\omega} - \frac{\partial\tilde{V}(X_-)}{\partial\Omega}\frac{\partial\Omega|}{\partial\omega} - \frac{\partial\tilde{V}(X_-)}{\partial p}\frac{\partial p}{\partial\omega} - \frac{\partial\tilde{V}(X_-)}{\partial k'_-}\frac{\partial k'_-}{\partial\omega}}{\frac{\partial\tilde{V}(X_+)}{\partial k} - \frac{\partial\tilde{V}(X_-)}{\partial k}} \quad (17)$$

where $X_- \equiv (k,e,z,\Omega,d_-,k'_-,p)$ and $X_+ \equiv (k,e,z,\Omega,d_+,k'_+,p)$. Following the discussion of how to differentiate the value function, we set $\frac{\partial k'_-}{\partial\omega}$ and $\frac{\partial k'_+}{\partial\omega}$ in (17) as zero in case they are an interior solution, or equal to the derivative of the relevant bound constraint in case they are constrained.

3.2 Model reduction: general outline

Differentiating the equations of the model outlined in Section 3.1.3 we obtain a system of linear rational expectations equations

$$\Lambda\theta_{t-1} + \Gamma\theta_t + \mathrm{E}_t\,\Phi\theta_{t+1} + \Psi\epsilon_t = 0 \tag{18}$$

with a very large vector of variables θ. We want to reduce the dimension of the model without any significant loss in accuracy. The engineering literature (Antoulas 2005) shows how to do this if the model is already given in the VAR form $\theta_t = A\theta_{t-1} + B\epsilon_t$ (or in state space form). We cannot apply this directly because we first have to solve the model in order to know the dynamics.

The first step is to partition the equation system (18) as

$$\begin{bmatrix} \Lambda_{ss} & 0 & 0 \\ \Lambda_{ys} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} s_{t-1} \\ y_{t-1} \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} \Gamma_{ss} & \Gamma_{sy} & \Gamma_{sv} \\ \Gamma_{ys} & \Gamma_{yy} & \Gamma_{yv} \\ 0 & \Gamma_{vy} & \Gamma_{vv} \end{bmatrix} \begin{bmatrix} s_t \\ y_t \\ v_t \end{bmatrix} + \\ \mathrm{E}_t \begin{bmatrix} 0 & 0 & \Phi_{sv} \\ 0 & 0 & \Phi_{yv} \\ 0 & 0 & \Phi_{vv} \end{bmatrix} \begin{bmatrix} s_{t+1} \\ y_{t+1} \\ v_{t+1} \end{bmatrix} + \begin{bmatrix} \Psi_s \\ \Psi_y \\ 0 \end{bmatrix} \epsilon_t = 0 \tag{19}$$

The variable vector θ partitioned as (s, y, v) such that only v appears with time index $t+1$, only s appears with time index $t-1$, and only y, not s enters equations for v. We assume that Γss, Γyy, and Γvv are regular. We further assume that Γss and Γvv are very sparse (often just the identity matrix) and therefore easy to invert.

In practice, we take as s all the variables that appear with time index $t-1$, and as v all the variables that appear with time index $t+1$. These two groups must not overlap. y are all the other variables. We split the equation systems into equations so as to satisfy the constraints implicit in (19).

With the above assumptions, we can rewrite (19) as

$$\begin{bmatrix} \Gamma_{ss}^{-1}\Lambda_{ss} \\ \Lambda_{ys} \\ 0 \end{bmatrix} s_{t-1} + \begin{bmatrix} I & \Gamma_{ss}^{-1}\Gamma_{sy} & \Gamma_{ss}^{-1}\Gamma_{sv} \\ \Gamma_{ys} & \Gamma_{yy} & \Gamma_{yv} \\ 0 & \Gamma_{vv}^{-1}\Gamma_{vy} & I \end{bmatrix} \begin{bmatrix} s_t \\ y_t \\ v_t \end{bmatrix} + \\ \mathrm{E}_t \begin{bmatrix} \Gamma_{ss}^{-1}\Phi_{sv} \\ \Phi_{yv} \\ \Gamma_{vv}^{-1}\Phi_{vv} \end{bmatrix} v_{t+1} + \begin{bmatrix} \Gamma_{ss}^{-1}\Psi_s \\ \Psi_y \\ 0 \end{bmatrix} \epsilon_t = 0 \tag{20}$$

Because both s and v are very large vectors, the task is to reduce the dimension of the model. This has two components:

1. State reduction: choose an $n_m \times n_s$ matrix $\bar{M}$ with $n_m < n_s$ and define

$$m_t = \bar{M} s_t \tag{21}$$

Interpretation: m_t denotes the statistics ("m" is a memo of "moments) of the cross-sectional distribution that agents based their decision on (bounded rationality). We assume that those statistics are linear functions of the distribution.

The matrix must be such that there exist $\tilde{\Lambda}_{ys}$ and $\tilde{\Gamma}_{ys}$ with

$$\Lambda_{ys} = \tilde{\Lambda}_{ys}\bar{M}, \qquad \Gamma_{ys} = \tilde{\Gamma}_{ys}\bar{M} \tag{22}$$

(22) is satisfied if the rows of Λ_{ys} and Γ_{ys} are spanned by the rows of $\bar{M}$:

$$\begin{bmatrix}\Lambda'_{ys} & \Gamma'_{ys}\end{bmatrix} \in span(\bar{M}') \tag{23}$$

2. Value reduction: choose a matrix $\bar{V}$ that spans the space in which the value function is assumed to live:

$$v_t = \bar{V} f_t \tag{24}$$

$dimen(f) << dimen(v)$. W.l.o.g. we can choose $\bar{V}$ as orthonormal so that $\bar{V}'\bar{V} = I$.

Using (21), (22) and (24), and premultiplying the first block of equations in (20) by $\bar{M}$, (20) can be written as and the second block by $\bar{V}'$, (20) can be written as

$$\begin{bmatrix}\bar{M}\Gamma_{ss}^{-1}\Lambda_{ss} \\ \tilde{\Lambda}_{ys}\bar{M} \\ 0\end{bmatrix} s_{t-1} + \begin{bmatrix}I & \bar{M}\Gamma_{ss}^{-1}\Gamma_{sy} & \bar{M}\Gamma_{ss}^{-1}\Gamma_{sv}\bar{V} \\ \tilde{\Gamma}_{ys} & \Gamma_{yy} & \Gamma_{yv}\bar{V} \\ 0 & \bar{V}'\Gamma_{vv}^{-1}\Gamma_{vy} & I\end{bmatrix} \begin{bmatrix}m_t \\ y_t \\ f_t\end{bmatrix} +$$

$$\mathrm{E}_t \begin{bmatrix}\bar{M}\Gamma_{ss}^{-1}\Phi_{sv}\bar{V} \\ \Phi_{yv}\bar{V} \\ \bar{V}'\Gamma_{vv}^{-1}\Phi_{vv}\bar{V}\end{bmatrix} f_{t+1} + \begin{bmatrix}\bar{M}\Gamma_{ss}^{-1}\Psi_s \\ \Psi_y \\ 0\end{bmatrix} \epsilon_t = 0 \tag{25}$$

Notice the asymmetry between state reduction versus value reduction: we have assumed we know a good approximation to the subspace in which the value function lives. I will show in Section 3.3 how to compute such a subspace by iterating forward on the Bellman equation. For the cross-sectional distribution, in contrast, we only assume that the statistics m_t contain the most relevant information about the distribution. There is no easy way to specify an approximate subspace, in particular because the cross-sectional distribution can take a quite erratic shape when individual policy functions are discontinuous, as they usually are with non-convex decision problems. Because of this problem, the equation system (25) still contains the full state vector s. The problem is the term $\bar{M}\Gamma_{ss}^{-1}\Lambda_{ss}s_{t-1}$. One can approach this problem in two ways. The simpler approach, which can be applied for any matrix $\bar{M}$, is to make some plausible choice of this subspace. Concretely, we set $s_t = \bar{P}m_{t-1}$ for some matrix $\bar{P}$ so that $\bar{M}\Gamma_{ss}^{-1}\Lambda_{ss}s_{t-1}$ in (25) is replaced by $\bar{M}\Gamma_{ss}^{-1}\Lambda_{ss}\bar{P}m_{t-1}$. The logic behind this approach is the following. If it is true that $\bar{M}s_t$ contains the relevant information about s_t, the exact choice of s_t should not matter much. This should at least be good enough for a first solution of the

model. One can then simulate the model to find a more appropriate PD and solve the model again. This is explained in Section 3.4

The alternative approach, which can lead to an high-precision solution, is to find a suitable matrix $\bar{M}$ such that there exists a matrix $\hat{A}$ with $\bar{M}\Gamma_{ss}^{-1}\Lambda_{ss} = \hat{A}\bar{M}$. In this case, $\bar{M}\Gamma_{ss}^{-1}\Lambda_{ss}s_{t-1}$ in (25) becomes $\hat{A}m_{t-1}$. This is explained in Section 3.5.

3.3 Almost Exact Value Function Reduction

The first step is to find lower-dimensional basis $\bar{V}$ of the space in which the value function v must lie. This can be done with almost zero loss of accuracy in the following way:[1] Setting $\tilde{\Gamma}_{vy} = -\Gamma_{vv}^{-1}\Gamma_{vy}$ and $\tilde{\Phi}_{vv}v_{t+1} = -\Gamma_{vv}^{-1}\Phi_{vv}$, we can write the third line of equations in (20) as

$$v_t = \mathrm{E}_t\left(\tilde{\Gamma}_{vy}y_t + \tilde{\Phi}_{vv}v_{t+1}\right) \tag{26}$$

Iterating gives

$$\begin{aligned} v_t &= \mathrm{E}_t\left[\tilde{\Gamma}_{vy}y_t + \tilde{\Phi}_{vv}\tilde{\Gamma}_{vy}y_{t+1} + \tilde{\Phi}_{vv}^{y}\tilde{\Gamma}_{vy}y_{t+y} + \ldots\right] \\ &= \mathrm{E}_t\sum_{i=0}^{\infty}\tilde{\Phi}_{vv}^{i}\tilde{\Gamma}_{vy}y_{t+i} \end{aligned} \tag{27}$$

At this stage, we do not know $\mathrm{E}_t\, y_{t+i}$, but (27) implies that v_t is spanned by the columns of the $\tilde{\Phi}_{vv}^{i}\tilde{\Gamma}_{vy}$. This information is useless if the rank of the $\tilde{\Phi}_{vv}^{i}\tilde{\Gamma}_{vy}$ taken together equals the dimension of v. It turns out, however, that the numerical rank of the $\tilde{\Phi}_{vv}^{i}\tilde{\Gamma}_{vy}$ as determined by the finite machine precision is much smaller than the dimension of v. An essential condition for this is that Γ_{vy} has small rank, which says again that there is only a small set of equilibrium variables y that enter the agents' utility functions.

For the basis $\bar{V}$ of the space in which the value function v must lie, we take an orthonormal basis of $\cup_{i=0}^{k}\tilde{\Phi}_{vv}^{i}\tilde{\Gamma}_{vy}$ for some finite k. Truncating k implies no significant loss of accuracy because of discounting.

3.4 Simple state aggregation: proxy distributions

We choose a matrix $\bar{P}$ which selects to any vector of statistics m_t one specific distribution (called "proxy distribution", (Reiter 2010b)) $s_t^{PD} = \bar{P}m_t$ which has those statistics. This means we require

$$\bar{M}\bar{P} = I_{n_m} \tag{28}$$

The interpretation of this approach is that the state equations in (19) are satisfied not at all possible state vectors s, but only at those that are "typical" distributions in the sense $s = \bar{P}m$ for any m. One general way to choose the proxy distribution is to use steady state information:

[1] In Reiter (2010a) I was using an iterative algorithm to determine $\bar{V}$, for which there is no convergence proof. The procedure described here avoids this problem and is faster.

given any moment matrix $\bar{M}$, one can choose the corresponding proxy distributions as the distributions that are closest to steady subject to the moment constraint:

$$\max_{\tilde{s}} \frac{1}{2}\tilde{s}'\Omega^{-1}\tilde{s} \qquad s.t. \qquad \bar{M}\tilde{s} = \tilde{m} \tag{29}$$

The solution to (29) is given by $\tilde{s} = \Omega\bar{M}'(\bar{M}\Omega\bar{M}')^{-1}\tilde{m}$, which means

$$\bar{P} = \Omega\bar{M}'(\bar{M}\Omega\bar{M}')^{-1} \tag{30}$$

For an illustrative example (which does not satisfy (22)), choose $\bar{M}$ such that it selects adjacent bins:

$$\bar{M} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & \dots & 0 & 0 & 0 \\ & \dots & & & \dots & & \dots & \dots & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 1 & 1 \end{pmatrix}$$

With $\Omega = I$, (30) gives

$$\bar{P} = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & \dots & 0 & 0 & 0 \\ & \dots & & & \dots & & \dots & \dots & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1/3 & 1/3 & 1/3 \end{pmatrix}$$

The proxy distribution assumes equal distribution within bins. Alternatively, if Ω is the diagonal matrix with elements equal to the steady state distribution s^*, the proxy distribution is proportional to the steady state distribution within bins:

$$\bar{P} = \begin{pmatrix} \frac{s_1^*}{\sum_{i=1}^3 s_i^*} & \frac{s_2^*}{\sum_{i=1}^3 s_i^*} & \frac{s_3^*}{\sum_{i=1}^3 s_i^*} & \dots & 0 & 0 & 0 & & & \\ 0 & 0 & 0 & \frac{s_4^*}{\sum_{i=4}^6 s_i^*} & \frac{s_5^*}{\sum_{i=4}^6 s_i^*} & \frac{s_6^*}{\sum_{i=4}^6 s_i^*} & \dots & 0 & 0 & 0 \\ & \dots & & & \dots & & \dots & \dots & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & \frac{s_{n-2}^*}{\sum_{i=n-2}^n s_i^*} & \frac{s_{n-1}^*}{\sum_{i=n-2}^n s_i^*} & \frac{s_n^*}{\sum_{i=n-2}^n s_i^*} \end{pmatrix}$$

This approximation is good enough for many models, but the following section describes a much more precise, if computationally more involved, approach.

3.5 Almost-exact state aggregation

As we have discussed at the end of Section 3.2, our aim is to find a selection matrix $\bar{M}$ such that (22) is satisfied, and there exists a matrix $\hat{A}$ with $\bar{M}\Gamma_{ss}^{-1}\Lambda_{ss} = \hat{A}\bar{M}$. To do this, we follow what in Reiter (2010a) is called the "Conditional Expectations Approach". So assume we want to predict the endogenous variables y_{t+i} for $i = 0, \dots, \infty$. in a linear model where y is related to the states s by

$$y_t = Cs_t \tag{31}$$

and we have the state transition equation

$$\mathrm{E}_t\, s_{t+i} = A^i s_t \tag{32}$$

where the sysem matrix A has dimension $n \times n$, and C has dimension $m \times n$. Obviously, we need the following linear combinations of s:

$$Cs_t, \qquad CAs_t, \qquad CA^2 s_t, \qquad \dots \tag{33}$$

Stack them into

$$Q = \begin{bmatrix} C \\ CA \\ CA^2 \\ \cdots \\ CA^{n-1} \end{bmatrix} \tag{34}$$

The $m \cdot n \times n$ matrix Q is called "observability matrix". This approach is only useful if the rank of Q is substantially lower than n. Define $k \leq n$ as the rank of Q. The SVD of Q can be written as

$$Q = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} = U_1 S V_1' \tag{35}$$

$$S \equiv diag(\sigma_1, \dots, \sigma_k) \tag{36}$$

where U_1 and V_1 have dimension $m \cdot n \times k$, $U_1'U_1 = V_1'V_1 = I_k$.

From the Cayley-Hamilton theorem, there exists a Λ such that

$$QA = \Lambda Q \tag{37}$$

Using (35) in (37) we get $U_1 S V_1' A = \Lambda U_1 S V_1'$. Premultiplying by $S^{-1}U_1'$ we get $V_1'A = S^{-1}U_1'\Lambda U_1 S V_1'$ (notice that S is invertible). Setting $\bar{M} = V_1'$ we get

1. $\bar{M}\bar{M}' = I_k$
2. $\bar{M}$ can be interchanged with A:
$$\bar{M}A = \hat{A}\bar{M} \tag{38}$$
with $\hat{A} = S^{-1}U_1'\Lambda U_1 S$ being a $k \times k$-matrix. From 1. it follows that $\hat{A} = \bar{M}A\bar{M}'$.
3. We have $C' \in span(\bar{M}')$:
$$C = \begin{bmatrix} I & 0 & \dots & 0 \end{bmatrix} Q = \left(\begin{bmatrix} I & 0 & \dots & 0 \end{bmatrix} U_1 S \right) \bar{M} \tag{39}$$

We can summarize the above discussion in the following

Proposition 1. *Assume that Q has rank $k << n$ when choosing $A = \Gamma_{ss}^{-1}\Lambda_{ss}$ and C such that*

$$\begin{bmatrix}\Lambda'_{ys} & \Gamma'_{ys}\end{bmatrix} \in span(C) \tag{40}$$

Then there is a $k \times n$-matrix $\bar{M}$ such that

- $\Lambda_{ys} = \tilde{\Lambda}_{ys}\bar{M}, \qquad \Gamma_{ys} = \tilde{\Gamma}_{ys}\bar{M}$
- *There exists an $n_m \times n_m$-matrix $\hat{A}$ such that*

$$\bar{M}\Gamma_{ss}^{-1}\Lambda_{ss} = \hat{A}\bar{M} \tag{41}$$

Using (41), we can replace $\bar{M}\Gamma_{ss}^{-1}\Lambda_{ss}s_t$ in (25) by $\hat{A}m_t$ and obtain

$$\begin{bmatrix}\hat{A}\\ \tilde{\Lambda}_{ys}\\ 0\end{bmatrix} m_{t-1} + \begin{bmatrix}I & \bar{M}\Gamma_{ss}^{-1}\Gamma_{sy} & \bar{M}\Gamma_{ss}^{-1}\Gamma_{sv}\bar{V}\\ \tilde{\Gamma}_{ys} & \Gamma_{yy} & \Gamma_{yv}\bar{V}\\ 0 & \bar{V}'\Gamma_{vv}^{-1}\Gamma_{vy} & I\end{bmatrix}\begin{bmatrix}m_t\\ y_t\\ f_t\end{bmatrix} +$$
$$\mathrm{E}_t\begin{bmatrix}\bar{M}\Gamma_{ss}^{-1}\Phi_{sv}\bar{V}\\ \Phi_{yv}\bar{V}\\ \bar{V}'\Gamma_{vv}^{-1}\Phi_{vv}\bar{V}\end{bmatrix} f_{t+1} + \begin{bmatrix}\bar{M}\Gamma_{ss}^{-1}\Psi_s\\ \Psi_y\\ 0\end{bmatrix}\epsilon_t = 0 \tag{42}$$

The commutation property (41) says that, knowing the statistics m_t, the exact distribution does not matter for the solution. It does not say that the distributions live in the space spanned by $\hat{A}$. In the terminology of Section 3.4, it does not necessarily give a good proxy distribution.

3.6 Simulation and Error Analysis

The reduced model (42) can be solved and simulated for (m_t, y_t, f_t) by standard methods. Having a simulation (m_t, y_t, f_t), can we recover the full state vector s_t and the value function v_t? Since we have established in Section 3.3 that v_t lives in the space spanned by $\bar{V}$ and parameterized by f_t, we get v_t directly by

$$v_t = \bar{V}f_t \tag{43}$$

Similarly, we could obtain the distribution by $s_t = \bar{P}m_t$, but this cannot be expected to yield a precise solution: If we follow the proxy distribution approach of Section 3.4, the $\bar{P}$ given by (30) is a rather arbitrary approximation. Following Section 3.5, we have mentioned above that the property (38) does not imply that $\bar{P} = \bar{M}'$ is a good proxy distribution. It only says that $(\bar{M}'m_t)$ is just as good for predicting the future as is s_t.[2] To avoid the arbitrariness of the proxy distribution, we do not not use $s_t = \bar{P}m_t$, but rather use the following, computationally more involved procedure. In any period t of the simulation

[2]Numerically, (38) is not very precisely satisfied anyway.

1. get y_t from the reduced simulation (y is part of the reduced model)j
2. compute $v_t = \bar{V} f_t$ from the reduced simulation
3. get $\mathrm{E}_t\, v_{t+1} = \bar{V}\, \mathrm{E}_t\, f_{t+1}$ from the reduced simulation
4. compute s_t by solving the first block of equations in (20).

To simulate a path for the s_t, we obviously have to specify an initial state s_0, for example the deterministic steady state.

3.7 Measuring accuracy

We have to deal with three types of approximation error:

1. The error from discretization. This arises already in steady state. We can check for it by varying the number of grid points etc.
2. The error from linearization. We will analyze it later, after obtaining nonlinear solutions. Alternatively, one can check for it by computing perfect foresight solutions after shocks of different size (Boppart, Krusell, and Mitman 2018).
3. The error from aggregation. Minimizing this error is the purpose of the aggregation procedure of Sections 3.3–3.5. By simulating the model as described in (cf. Section 3.6), we can compute the aggregation error by analyzing the residuals along any impulse response function. For any given initial state s_{t-1}, if $\epsilon_t = \epsilon_{t+1} = 0$ it follows from (18) and linearity that
$$\Lambda\theta_{t-1} + \Gamma\theta_t + \Phi\theta_{t+1} = 0 \tag{44}$$
We compute an IR function for $t = 1 : T$, starting from any s_0. For $t = 2 : T-1$, we compute the residual
$$Res(t) = \Lambda\theta_{t-1} + \Gamma\theta_t + \Phi\theta_{t+1} \tag{45}$$
To interpret the residual, we must scale it properly. Define
$$RScal(i,t) \equiv \frac{Res(i,t)}{\sum_j (|\Lambda_{i,j}\theta_{j,t-1}| + |\Gamma_{i,j}\theta_{j,t}| + |\Phi_{i,j}\theta_{j,t+1}|)} \tag{46}$$
as the residual of equation i normalized by the sum of the absolute entries in this equation.

With optimal state and value function reduction, the aggregation error turns out to be extremely small. For example, in the case of the Chang/Kim model, the maximum scaled error is of the order 10^{-10}, and the mean scaled error is of the order 10^{-11}.

3.8 Reducing even further: Balanced Reduction

Notice that the state and value function reduction of Sections 3.3–3.5 has not involved the matrix Ψ, which determines how the i.i.d. shocks affect the economy. The usual assumption is that the shocks enter the equations in a particular way, having a direct impact only on a small number of exogenous driving forces that follow AR processes. Then we would expect that the economy lives in a lower-dimensional subspace than it would do with a more general impact of shocks, which means that further state reductions should be possible. Having solved the reduced model (42), we get the solution in state-space form

$$\begin{aligned} m_t &= Am_{t-1} + B\epsilon_t \\ y_t &= Cm_t \end{aligned} \tag{47}$$

This model class can be further analyzed and reduced by the methods developed in the engineering literature, called "balanced reduction" (Antoulas 2005). Define the matrices R, $\mathcal{P}$, Q, $\mathcal{Q}$, U, S, V, $\tilde{H}$ as follows:

$$RR' = \mathcal{P} \equiv \mathcal{L}\left(A, B, \Sigma_\epsilon\right) \tag{48}$$

$$QQ' = \mathcal{Q} \equiv \mathcal{L}\left(A', C'\right) \tag{49}$$

$$USV' = R'Q \tag{50}$$

$$\tilde{H} = S^{-1/2}V'Q' \tag{51}$$

where $\mathcal{L}\left(A, B, \Sigma_\epsilon\right)$ is defined as the matrix Σ that solves $\Sigma = A\Sigma A' + B\Sigma_\epsilon$. (48) syas that R is the Cholesky factor of the covariance matrix $\mathcal{P}$, and Q in (49) is the Cholesky factor of the observability Gramian $\mathcal{Q}$, while U, S and V are the SVD of the matrix $R'Q$ with $U'U = I$, $V'V = I$ and S diagonal with decreasing entries. We take S as the square matrix containing only the non-zero singular values (and drop the columns of U and rows of V corresponding to the zero singular values), so that S is invertible by construction.

Now consider the variable transformation $\hat{m} = \tilde{H}m$, $\hat{A} = \tilde{H}A\tilde{H}^{-1}$, $\hat{B} = \tilde{H}B$, $\hat{C} = C\tilde{H}^{-1}$. Using that $\tilde{H}^{-1} = RUS^{-1/2}$, straightforward algebra shows that

$$\mathcal{L}\left(\hat{A}, \hat{B}, \Sigma_\epsilon\right) = \tilde{H}\mathcal{L}\left(A, B, \Sigma_\epsilon\right)\tilde{H}' = \mathcal{L}\left(\hat{A}', \hat{C}'\right) = (\tilde{H}')^{-1}\mathcal{L}\left(A', C'\right)\tilde{H}^{-1} = S. \tag{52}$$

Equ. (52) is a remarkable result. It shows that in the new vector $\hat{m}$ the variables are ordered such that $\hat{m}_i$ has both the i-th highest variance, and makes the i-th highest contribution to future values of y. For the reduced model, we pick the first k components of $\hat{m}$, or the first k rows of $\tilde{H}$, such that the diagonal elements $S_{i,i}$ are negligible for $i > k$:

$$H = \tilde{H}_{1:k,:} \tag{53}$$

Properties of balanced reduction

Is the (doubly) reduced model

$$\begin{aligned} \hat{m}_t &= \hat{A}\hat{m}_{t-1} + \hat{B}\epsilon_t \\ \hat{y}_t &= \hat{C}\hat{m}_t \end{aligned} \tag{54}$$

an optimal approximation to the (reduced) model (47) in any sense? With the choices of H that we have discussed, it is not a strictly optimal. Nevertheless, balanced reduction has a strong performance guarantee (cf. Antoulas (2005, Theorem 7.10), Antoulas (1999, Section 2.6)):

$$distance(ExactModel, ReducedModel) \leq 2(\sigma_{k+1} + \ldots + \sigma_n) \tag{55}$$

Here, the σ's are the singular values in (50) (called "Hankel singular values") that were omitted in the construction of H in (53). The distance measure in (55) is the Hankel norm, which is defined as the maximum distance in the future response

$$\sqrt{\sum_{i=0}^{\infty} ||y_{t+i} - \hat{y}_{t+i}||^2} \tag{56}$$

to any sequence of past shocks ϵ_{t-i} with unit length:

$$\sqrt{\sum_{i=0}^{\infty} ||\epsilon_{t-i}||^2} = 1 \tag{57}$$

In particular, the difference in the usual impulses responses between exact and the reduced models cannot be bigger than the bound (55). This explains why balanced reduction is the standard aggregation method in the control literature.

There exist even better, but more complicated approximations than balanced reduction. The theoretical lower bound on the distance between the two models is σ_{k+1}. This bound can actually be attained (Antoulas 1999, Sections 2.6.1,3.2). For us, it seems not worthwhile to investigate more complicated methods for the linearized model, because our main concern is whether the reduced state from the linear model is still suitable for the nonlinear model.

4 Computing the Nonlinear Solution

4.1 Approximating the value function

The household value function is a function of the individual state (k, e, z) and the aggregate state $\Omega = (\mathbf{D}, Z)$. To make the computation feasible, we assume that the value function can be well approximated as a function of the reduced information set M rather than the full state vector Ω:

$$V(k, e, z, \Omega, Z) \approx \hat{V}(k, e, z, M, Z) \tag{58}$$

where

$$M = H\mathbf{D} \tag{59}$$

for some given matrix H. For ease of notation, I assume that the full Z is in the information set, which is typically the case.

At each point $(k, e, z)_i$ in the individual grid G_x, with $i = 1, \ldots, n_k \cdot n_e \cdot n_z$, we approximate this function as a linear combination of n_p known basis functions $\phi_j(M, Z)$,

$$\hat{V}((k, e, z)_i, M, Z) = \sum_{j=1}^{n_p} \phi_j(M, Z)\, \gamma_{j,i} \tag{60}$$

Notice that we have separate coefficients $\gamma_{j,i}$ for each individual state $(k, e, z)_i$. Obviously, we can write the approximated value function as a function of the original state by $V((k, e, z)_i, \mathbf{D}, Z) = \hat{V}((k, e, z)_i, H\mathbf{D}, Z)$.

Having computed $\hat{V}$ at each grid point $(k, e, z)_i$, we can again interpolate quadratically in the individual state k. In this way, we can compute an interpolated value function at any state $(k, e, z, \mathbf{D}, Z$. Denote the interpolated value function as $V^I(k, e, z, \mathbf{D}, Z; \gamma)$. It is parameterized by the vector of coefficients γ.

For the approximated value function, we can define expected continuation values based on two different information sets. The expected continuation value conditional on *end-of-period* information V^{EC} is given by

$$V^{EC}(k', e', z, \Omega, Z; M') = \sum_{z'} \pi(z, z'; \Omega)\, \mathrm{E}_{Z'}\, V^I\left(k', e', z', M'Z'; \gamma\right) \tag{61}$$

If we make a finite approximation of the law of motion of aggregate Z, we can write it as

$$V^{EC}(k', e', z, \Omega, Z; M') = \sum_{z'} \pi(z, z'; \Omega) \sum_{Z'} \pi_Z(Z, Z') V^I\left(k', e', z', M'Z'; \gamma\right) \tag{62}$$

Notice that the continuation value depends on end-of-period moments M', which are determined in equilibrium. Since $\mathbf{D}'$ is predetermined (does not depend on next period's shocks), we only need to know the finite number of statistics $M' = H\mathbf{D}'$ in order to compute the expected continuation value.

Conditional on M', which is exogenous to the individual agent, the optimization problem can be compactly written as

$$\max_{d, k'} U\left(k, e, z, d, k'; p, \Omega\right) + \beta V^{EC}(k', T(e, d), z', \Omega, Z; M') \tag{63}$$

4.2 Temporary Equilibrium

Given any continuation value function $V^{EC}(k', e', z, \Omega, Z; M')$, a temporary equilibrium $(p, m', \mathcal{A})$ at any aggregate state $\mathbf{D}, Z$ is defined as consisting of

- a set of equilibrium values p

- a set of end-of-period statistics m'
- a policy function $\mathcal{A}$

such that

- policy functions are optimal, which means that they satisfy (63).
- prices satisfy the equilibrium conditions

$$\mathcal{E}(p, \mathbf{D}, Z, \mathcal{A}) = 0 \tag{64}$$

- end-of-period statistics satisfy the transition law

$$m' = H_D \mathcal{T}(\mathbf{D}, Z, p, \mathcal{A}) \tag{65}$$

This concept of temporary equilibrium requires to solve for the set of statistics M', which might be impractical if this set is large. We therefore also define a *partial temporary equilibrium*, where we replace the consistency condition (65) by $m' = \mathcal{T}_0(\mathbf{D}', Z)$ with an arbitrarily given transition law $\mathcal{T}_0(\mathbf{D}', Z)$. Below, we will use the transition law obtained from the linearized solution. In this case, the fixed point problem is reduced to finding the equilibrium variables p.

4.3 Backward iteration algorithm

I now describe an algorithm to solve for a global nonlinear approximation of the solution. The task is to find the parameters $\gamma_{j,i}$ that define the value function at each grid point as a function of the reduced aggregate state. This requires that we have chosen a set of states M and a set of basis functions $\phi_j(M, Z)$, as described above. The backward iteration is carried out on a finite set of aggregate grid points Ω_l for $l = 1 \ldots, n_A$. For the results reported in Section 5), I have used a grid obtained from an earlier simulation of the model, similar to the approach in (Judd, Maliar, and Maliar 2012).

In the following, I describe here a "conceptual" version of the algorithm, which solves for temporary equilibrium in each step and at each grid point. This is very time consuming, but should be the best guarantee for convergence. A practical algorithm makes choices about what to update in which step, so as to increase speed while still achieving convergence, but this is an implementation detail.

1. Denote the maximum iteration count by T. Initialize the value function $V_{T+1}\left(\bar{x}_i; \bar{\Omega}_j\right)$ for $i = 1, \ldots, n_k \cdot n_e \cdot n_z$ and $j = 1, \ldots, n_A$ by the value function obtained in an earlier solution (for example the linearized solution).
2. For $t = T : -1 : 1$ do

(a) Compute the coefficients of the polynomial approximation separately for each $k, e, zj \in G_x$ by the linear projection $\gamma_{i,t+1} = B \backslash \vec{V}_{i,t+1}$ where

$$B = \begin{pmatrix} \phi_1(\bar{M}_1) & \cdots & \phi_{n_p}(\bar{M}_1) \\ \vdots & \vdots & \vdots \\ \phi_1(\bar{M}_{n_M}) & \cdots & \phi_{n_p}(\bar{M}_{n_M}) \end{pmatrix}, \qquad \vec{V}_{i,t+1} = \begin{pmatrix} V_{t+1}(\bar{x}_i; \bar{\Omega}_1) \\ \vdots \\ V_{t+1}(\bar{x}_i; \bar{\Omega}_{n_A}) \end{pmatrix} \tag{66}$$

(b) For each aggregate grid point M_j, $j = 1, \ldots, n_A$ do

i. Guess (p, m')

ii. Define an expected continuation value at grid points as

$$\tilde{V}(k_i, e_i, z_i) = \sum_{z'} \pi(z_i, z'; \Omega) \sum_{Z'} \pi_Z(Z, Z') \sum_{j=1}^{n_p} \phi_i\left(m'\right) \gamma_{j,t+1}\left(k_i, e_i, z'\right)) \tag{67}$$

iii. Compute policy $\mathcal{A}$ that solves (63) with continuation value function

$$[d(x_i), k'(x_i)] = \underset{d,k'}{\operatorname{argmax}} \left\{ U\left(x_i, d, k'; p, M_j\right) + \beta \tilde{V}_{\mathcal{I}}\left(k'; T(e_i, d), z_i\right) \right\} \tag{68}$$

where x_i stands for k_i, e_i, z_i.

iv. Check whether $(p, m', \mathcal{A})$ satisfy the conditions for a temporary equilibrium (63), (64) and (65). If not, update (p, m') until convergence is achieved.

Finding the equilibrium (p, m') describes an $n_e + n_M$-dimensional fixed point problem which can be solved, for example, by quasi-Newton methods such as Broyden's algorithm.

(c) Update the value function for all $i = 1, \ldots, n_k n_e n_z$ by

$$V_t(x_i, M_j) = U\left(x_i, d(x_i), k'(x_i), p^*\right) + \beta \tilde{V}_{\mathcal{I}}\left(k'(x_i); (., T(e_i, d), z_i)\right) \tag{69}$$

After convergence, the value function should approximately satisfy the recursive relationship

$$\begin{aligned} V_t(x_i, M_j) = U\left(x_i, d(x_i, M_j), k'(x_i, M_j)), p(M_j)\right) \\ + \beta \tilde{V}_{\mathcal{I}}\left(k'(x_i, M_j); \left(., T(e_i, d(x_i, M_j)), z_i; m'(M_j)\right)\right) \end{aligned} \tag{70}$$

4.4 A set of nonlinear solution methods

With the methods described above, we can compute nonlinear solutions in a variety of ways. The three important distinctions are:

1. How is the continuation value function approximated

   (a) which states?

(b) which functional form?

I will consider polynomials up to degree 3 in the main aggregates (aggregate wealth, lagged exogenous variables, current shock), and in addition linear terms in either the states from balanced reduction (memo "B") or quantiles of the asset distribution (memo "Q").

2. Which variables are solved for in the temporary equilibrium. The two exreme cases are

   (a) we only solve for market clearing prices (memo "PE" for "price equilibrium") the end-of-period states that determine the expected continuation value are taken from some perceived aggregate law of motion

   (b) both prices and all the end-of-period states that determine the continuation value function are solved for in equilibrium; this is the default, if not otherwise noted.

3. How is the continuation value function computed?

   (a) by collocation (default)

   (b) taken from the linearized solution (memo "LIN")

   (c) by the Krusell/Smith algorithm (memo "KS")

For example, "P2B4" stands for the solution where prices and moments are solved for, with a value function that is quadratic in the main states, and has linear terms in first four balanced reduction terms (these are states 3–6, because the first two states are already closely spanned by the main states) and was computed on a collocation grid.

The simplest nonlinear solution is "PELIN": the continuation value function is taken from the linearized solution, and in the simulation step, we only solve for equilibrium prices. Despite the linear continuation value, this solution already captures important aspects of nonlinearity. First, the current policy function reacts nonlinearly to current prices. Second, the partial temporary equilibrium reflects the current cross-sectional distribution at any point in the simulation, Asymmetries in the impulse responses due to changes in the distribution are therefore reflected in this solution. As a minimum, this can serve as a check for the validity of linear approximations in models that are too big to be solved fully nonlinearly.

To capture the effect of uncertainty on behavior in an adequate way, in particular precautionary motives, one should at least obtain a "P2" solution, with a quadratic approximation in current states and shocks. To compute such a solution, one can take the following steps:

1. Solve the model by linearization.

2. Compute the continuation value function (62) from the linearized model. Use it to simulate a long time series of the model with method "PELIN". Choose a set of aggregate grid points from this simulation for the collocaiton grid (Judd, Maliar, and Maliar 2012).

3. Approximate the value function as a polynomial in the following variables:

   - the lagged state of the exogenous driving processes (such as TFP)
   - the current shocks of these processes
   - the aggregate variables that directly affect prices In the Chang/Kim model, only aggregate capital).

   Solve the model by backward iteration on the collocation grid.

Of course, one can then move on to more states, higher order approximation, or iterate the simulation step to obtain an update of the collocation grid.

4.5 Measuring accuracy of nonlinear solutions

One can measure accuracy by looking at the Bellman equation residual along a simulation path. At any aggregate state that we visit in the simulation, we first solve for the temporary equilibrium, and then compute the value function at this state by Equ. (69), as we do in the update step of the backward iteration algorithm. Denote this value function by V_t. Then we compute the approximated value function $\hat{V}_t$ given by (60). Denote by $V_t[i]$ the value at point i of the value function grid (for ease of notation, I suppress the arguments of the grid of discrete variables). The difference

$$100 \left| \delta V_t[i] = \hat{V}_t[i] - V_t[i] \right| \tag{71}$$

is the Bellman equation error, similar to the Euler residual in in models where the policy function is approximated. We need to scale this error so that it can be interpreted. What matters is not really the level but the slope of the value function. The scaled error of marginal value

$$\delta V_t'[i] = 100 \left| \frac{(\hat{V}_t[i+1] - V_t[i+1]) - (\hat{V}_t[i] - V_t[i])}{V_t[i+1] - V_t[i]} \right| \tag{72}$$

With log utility, this has the interpretation of a relative Euler approximation error in consumption. We measure the average and the maximum error over the whole simulation.

4.6 Collocation method vs. Krusell/Smith method

To compare the collocation method with the Krusell/Smith method, I implement Krusell/Smith so as to maximize the commonalities between the two methods. The household problem is solved on the same collocation grid with both methods. The difference is in the consistency criterion. In the collocation method, I solve for the temporary equilibrium at each grid, and consistency means that the value function at all grid points has converged in the time iteration. In the Krusell/Smith method, prices and end-of-period moments in the backward iteration are taken from a parameterized law of motion. Consistency means that this law of

motion is consistent with the OLS estimate from a simulation. Since simulation is relatively slow with nonlinear models, solving for the temporary equilibrium on the grid is faster than the repeated simulation necessary for the convergence of the Krusell-Smith algorithm. I use the same functional form for the perceived law of motion as for the value function approximation. A theoretical advantage of collocation is that it is not even necessary to parameterize the law of motion, since the relevant information about end-of-period states is solved for at grid point. Nevertheless, in the examples below, collocation and Krusell/Smith give results that are almost identical. This is not surprising: in a linear Gaussian model, collocation and Krusell/Smith are equivalent if the collocation grid is chosen to be the conditional expectation of the full state vector with respect to the reduced state vector (cf. Appendix C.1.3).

4.7 Acceleration steps

In infinite-horizon dynamic programming problems, acceleration steps (also called "Howard's improvement algorithm) are essential for computational efficiency. It means that after each optimization step, one updates the value function many times without changing the policy function. For dynamic programming, convergence of the backward iteration can be proved. A similar trick works for the collocation method, although there is no convergence proof. After a step in the time iteration with temporary equilibrium and update of the agents' policy function, the value function is updated many times, keeping equilibrium variables and individual policy functions constant. In this way, one has to solve for temporary equilibrium only about 20 times in the backward iteration process.

5 Numerical Results

5.1 The model of Chang and Kim

I have solved this model using the same parameter values as in the original paper. For the numerical solution, I have approximated the stochastic process of individual productivity by a 17-state Markov chain. I have chosen a grid of 1000 histogram bins for the capital distribution, and a grid of 400 points for the value function, for each level of productivity. With these choices, the linearized model has somewhat more than 23800 variables. Optimal state reduction reduces the number of states from around 17000 to 389. The value function parameters were reduced from 6800 to 253. As mentioned in Section 3.7, the maximum aggregation error of the linearized solution is of the order 10^{-10}.

Although the model is easy to write down, the numerical solution still poses substantial problems. Takahashi (2014) argues that the solution presented in the original paper (Chang and Kim 2007) is numerically imprecise, and presents a corrected version of the solution. Table 1 presents the key results of the model. The first columns reports the data, taken from Chang and Kim (2007). The second columns shows the original solution of Chang and Kim,

the third column shows the corrected solution of Takahashi, and the fourth column shows my solution obtained from the linearized model. Columns 5–7 show three different nonlinear solutions. Column 5 shows he results of a collocation method where the value function at each point in the grid is a quadratic function of the three states aggregate capital, last period's TFP and current TFP shock. In column 7, I have added linear terms in four different moments of the capital distributions, namely the states 3–6 of the balanced reduction. (States 1 and 2 of the balanced reduction are almost perfectly spanned by the three aggregate states listed before.) Column 6 reports my version of the Krusell/Smith solution (cf. Section 4.6), again with a quadratic approximation in the above three aggregate states.[3]

	Data	Chang/Kim	Takahashi	PELin	P2	P2KS	P2B4
σ_H	0.82	0.76	0.57	0.56	0.55	0.55	0.55
σ_{wedge}	0.92	0.76	0.24	0.23	0.21	0.21	0.22
σ_Y	2.06	1.28	1.30	1.24	1.24	1.24	1.24
σ_C	0.45	0.39	0.33	0.33	0.33	0.33	0.33
σ_I	2.41	3.06	3.08	3.12	3.12	3.12	3.12
σ_L	-	0.50	0.41	0.41	0.41	0.41	0.41
$\sigma_{Y/H}$	0.50	0.50	0.49	0.49	0.49	0.49	0.49
$\rho(H, wedge)$	0.85	0.87	0.95	0.95	0.96	0.96	0.96

Table 1: Results Chang/Kim model

It turns out that all solutions, including the linearized one, are very close to the Takahashi solution and therefore distinct from the Chang/Kim solution, as far as second moments are concerned. On the one hand, this is not surprising, because none of these solution methods suffer from the numerical problems pointed out by Takahashi. On the other hand, it is remarkable that linearization, which is based on very different simplifying assumptions than the Krusell/Smith method, gives results that are so close to Takahashi. (The only non-trivial difference is in the absolute size of the GDP fluctuations, which must be due to the fact that I am drawing the TFP shocks differently.)

The fact that linear and nonlinear solution methods yields very similar second moments does not necessarily mean that nonlinearities are not important. A natural check for nonlinearity is to test whether the impulse responses of the model scale linearly with the size of the shock. Figure 2 draws impulse responses for wages, labor input and consumption. The three

[3]There are several differences to the Krusell/Smith algorithm of Takahashi. First, forecast functions are based on three rather than only two variables, namely beginning of period capital, last period's TFP and the current shock to TFP. With full information, it is only current TFP that matters, but in a limited-information environment, past value of TFP provide additional information. Second, I approximate the forecast functions as quadratic rather than as log-linear functions in the aggregate states. Third, in the simulation I solve for temporary equilibrium as defined in Section 4.2. This involves solving for end-of-period capital, not only the current wage.

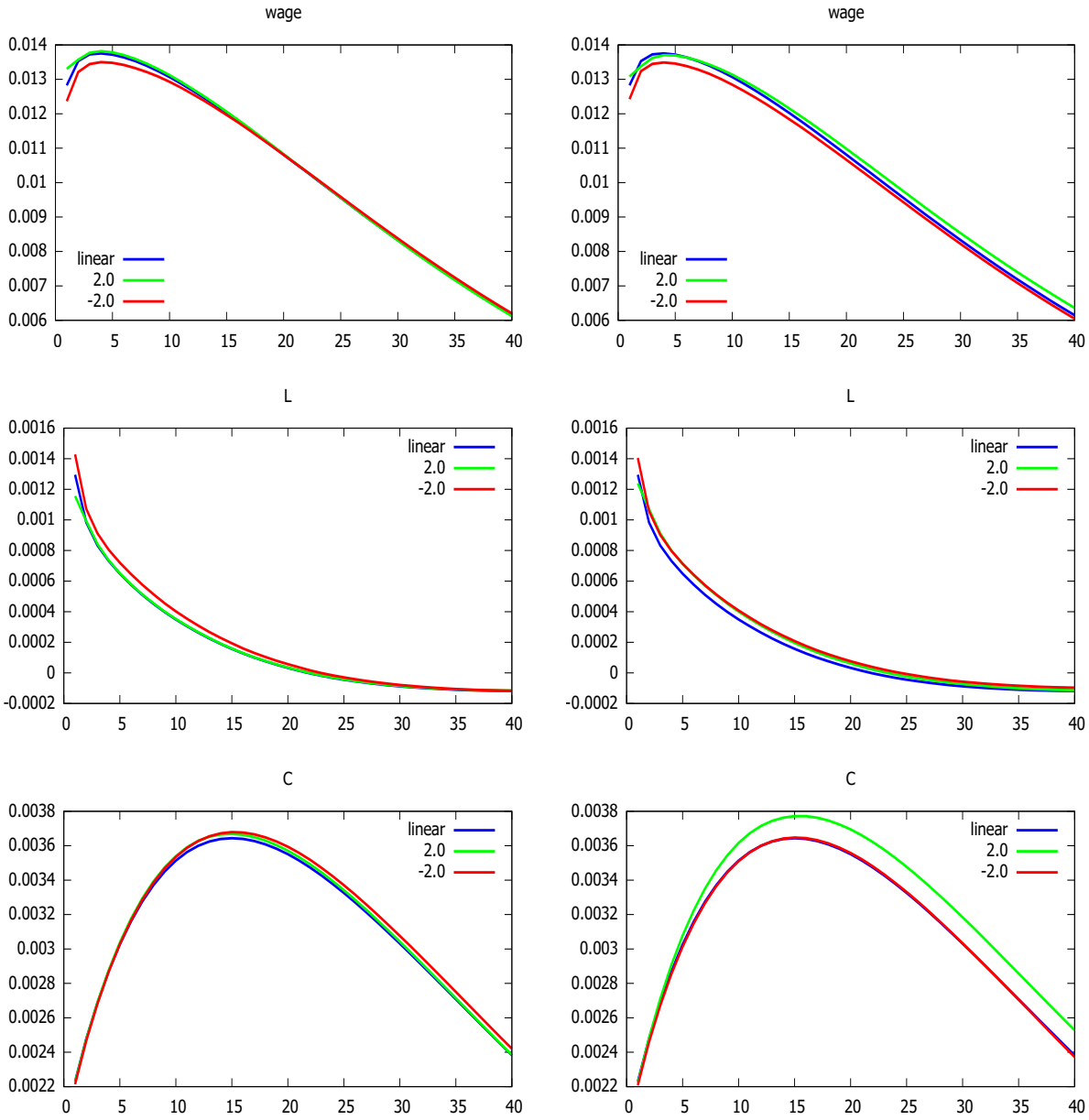
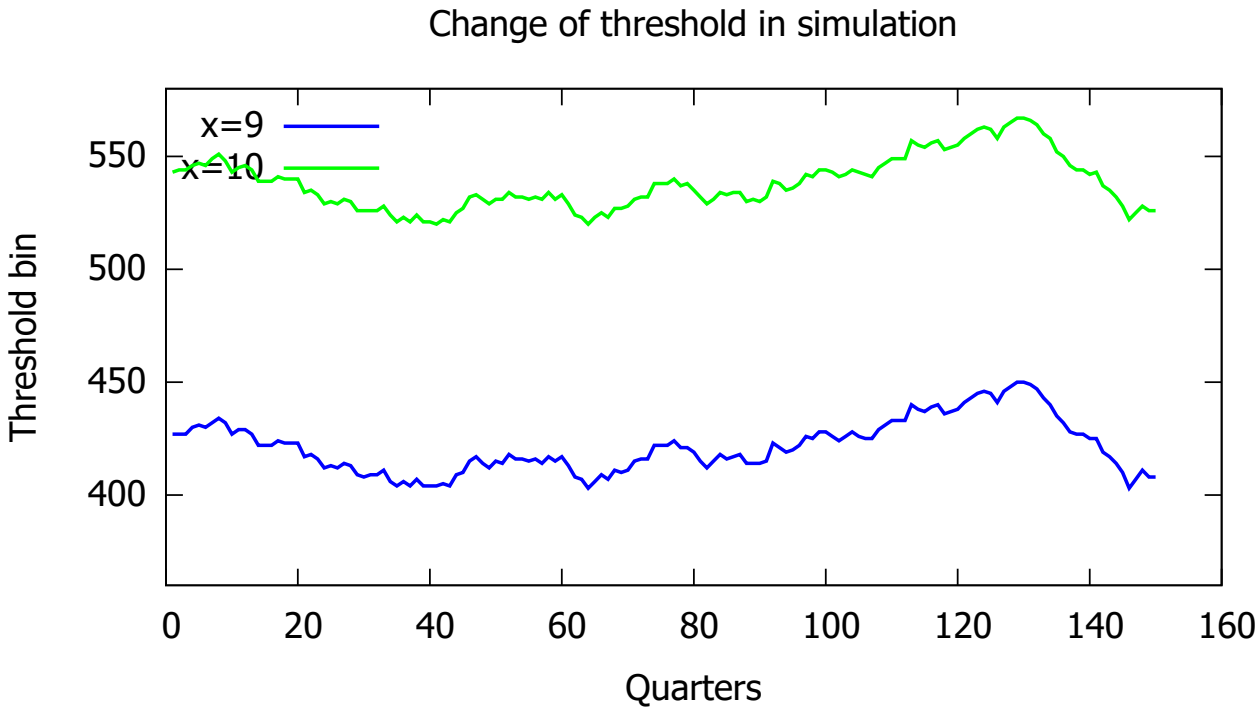


Figure 2: Impulse response to technology shock, PELIN (left) and P2 (right)

left-hand panels were obtained with the method "PELin", which takes the continuation value from the linearized solution. The three right-hand panels draw the same impulse responses, this time calculated with the method "P2", which we already saw above. For each variable, three lines are drawn. The green line gives the response to a positive shock of two standard deviations. The red line gives the response to a negative shock of two standard deviations, multiplied by -1. The blue line gives the linear impulse response as a benchmark, multiplied by 2. If the response was linear, the three lines would exactly coincide. We see a slightly asymmetric response. Let us first focus on the left panels. In response to a positive shock, when the participation threshold moves up, labor supply increases by less and wages increase by more than in the linear benchmark. In response to a negative shock, when the participation threshold moves down, labor supply increases by more and wages increase by less than in the linear benchmark. Qualitatively, the result is the same in the panels on the right side, but it appears that the endogenous price reactions in general equilibrium dampen the asymmetry to some extent.

To understand this pattern, recall that the linearized solution is a *local* solution, where the response to a big shock is just a scaled up version of the response to very small shocks. the labor supply response is therefore determined by the density of households in the bin where the threshold lies. In a nonlinear solution, however, the participation threshold jumps around bins quite a bit in response to realistically sized shocks. This is illustrated in the upper panel of Figure 3. In Figure 1 we have seen that the cross-sectional density is much lower in the bins at the right of the steady state threshold. This means that fewer households are affected by an upward movement of the threshold than be a downward movement. The labor supply response is therefore smaller with a positive shock, and equilibrium wages react more to productivity. The lower panel of Figure 3 illustrates the labor supply response. It shows the response to a one percent increase in TFP. The blue line shows the increase in the threshold point following the technology shock. The green line shows the density of households at the threshold point, and the red line shows the response of hours, which is the product of the two earlier lines. It was divided by the total number of hours in steady state, so the sum over all the points of the red line is the aggregate elasticity of hours to TFP. One can see that the bulk of the labor supply response comes with households at the intermediate productivity levels (level 7-11 out of 17 possible levels). For these households, the lower panel of Figure 1 has shown that the density around threshold points varies sharply.

Despite the sharp variations in the cross-sectional distribution, the deviations from symmetry appear quantitatively small in the Chang/Kim model. This is probably because labor supply reacts very strongly to the current wage, for a *given* continuation value. Since there are no labor market frictions, and most households hold enough assets to be well insured against TFP fluctuations, they react very elastically to transitory changes in the real wage. However, it is hard to say ex ante whether the deviation from linearity is significant. It



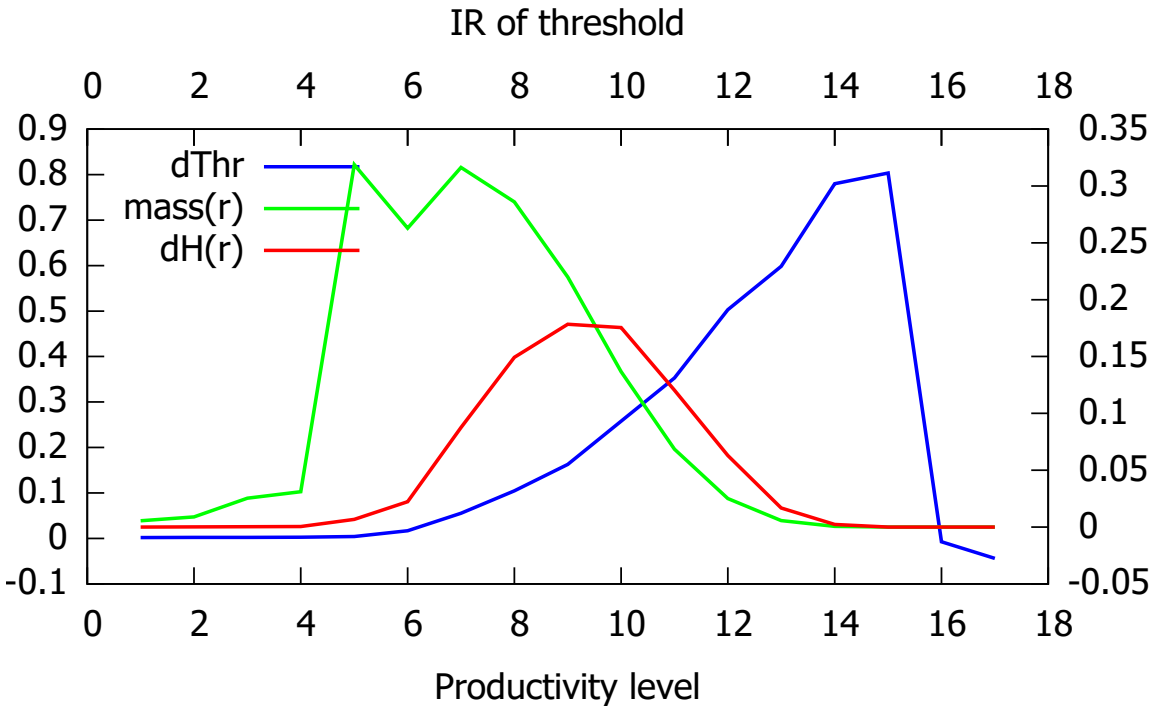


Figure 3: Chang/Kim model, density at the threshold

therefore appears necessary to sole the model nonlinearly in some way.[4]

5.2 Accuracy of solutions in the Chang/Kim-model

	PELin	P1	P2	P2KS	P3	PQ2	PQ4	P2B2	P2B4
σ_H	0.5629	0.5619	0.5549	0.5570	0.5509	0.5555	0.5556	0.5547	0.5560
σ_{wedge}	0.2238	0.2242	0.2131	0.2155	0.2077	0.2142	0.2145	0.2138	0.2165
σ_Y	1.2743	1.2672	1.2609	1.2623	1.2573	1.2615	1.2615	1.2603	1.2608
σ_C	0.3270	0.3279	0.3279	0.3269	0.3289	0.3280	0.3281	0.3282	0.3281
σ_I	3.1326	3.1291	3.1264	3.1302	3.1231	3.1264	3.1263	3.1252	3.1273
σ_L	0.4222	0.4160	0.4100	0.4112	0.4065	0.4107	0.4107	0.4100	0.4110
$\sigma_{Y/H}$	0.4798	0.4817	0.4871	0.4855	0.4906	0.4872	0.4872	0.4883	0.4881
$\rho(H, wedge)$	0.9681	0.9658	0.9642	0.9643	0.9629	0.9634	0.9633	0.9612	0.9588
$\rho(H, Y/H)$	0.8390	0.8355	0.8413	0.8394	0.8432	0.8389	0.8386	0.8378	0.8337
$\max_{t,i} \delta V_t[i]$	5.8768	3.6056	2.7624	2.7553	9.2665	2.7797	2.7648	3.0497	2.4885
$mean_t \max_i \delta V_t[i]$	1.1891	0.7501	0.3387	0.3409	0.3199	0.3423	0.3483	0.3501	0.3131
$mean_{t,i} \delta V_t[i]$	0.0895	0.0687	0.0191	0.0134	0.0207	0.0201	0.0188	0.0259	0.0156
$\max_{t,i} \delta V'_t[i]$	65.8110	34.0251	35.9933	36.0131	66.1480	36.0953	35.6717	31.7983	30.3361
$mean_t \max_i \delta V'_t[i]$	9.0638	5.1157	3.1356	3.1314	2.5569	3.1989	3.2365	3.0417	2.8578
$mean_{t,i} \delta V'_t[i]$	0.0305	0.0209	0.0135	0.0134	0.0122	0.0141	0.0144	0.0146	0.0140

Table 2: Extended results Chang/Kim model

To examine the validity and robustness of these results, the rest of this section reports further accuracy checks. Table 2 presents more detailed results for 9 different solutions, showing the same statistics as in Table 1, and in addition some accuracy statistics. Again, they come from a simulation of 4000 periods, and are not yet free of sampling error. The first column takes the continuation value from the linearized solution, as explained in Section 4.4. As the linearized value is a function of very many states, we solve only for current equilibrium variable K/L in the simulation, not for end-of-period values (what I have called "partial temporary equilibrium" above). Columns 2,3 and 5 give the solution from polynomial approximations of the basic states (lagged aggregate capital, lagged TFP and current shock) up to degree 3. Column 4 gives my implementation of the Krusell/Smith method, using a quadratic approximation both for the value function and the perceived aggregate law of motion. Columns 4–8 give nonlinear solutions, where the first number in parentheses indicates the number of aggregate states used, and the second number indicates the degree of approximation in these states, varying between 1 and 3. If the number of states is higher than 3, the additional states

[4]Boppart, Krusell, and Mitman (2018) provide a test for linearity where they solve nonlinearly for the impulse response to a one-time ("MIT") shock.

are the first states from the balanced reduction of the linearized model. The main results of Table 1 extend to the larger set of solution methods. The second moments are very similar across all solution methods, independent of the details of the approximation (number of states, order of approximation). They are especially close between the quadratic collocation method P2 and the Krusell/Smith method with quadratic approximation. For example, the correlation between GDP in the quadratic collocation and in Krusell/Smith is $1 - 2.6e - 7$, while the correlation between quadratic and cubic collocation has a correlation of $1 - 4.0e - 6$, The table also reports six different accuracy measures as defined in Section 3.7. It turns out that different measures yield a different ranking of the solutions. What might be surprising is that the cubic approximation is not clearly superior to the quadratic approximation, in particular it generates higher maximum errors. The true value function is not very well described by a polynomial, and higher order approximations turn out be less robust than a quadratic approximation. In any case, it is clear that the linear approximation has lower accuracy. This is true both for the partial temporary equilibrium with linear continuation value ("PELIN") as well as for the full temporary equilibrium with linear approximation in the three main states ("P1"). What this also shows is that for this simple model, the linearization error, which is present in "PELIN" is more important than the aggregation error, which is present in all other solutions. The next question is: can accuracy be further improved by including more statistics of the distribution, either quantiles (PQ2 and PQ4) or states from balanced reduction (PB2 and PB4)? The answer is somewhat disappointing. The improvements from more states is not systematic, and in any case not big.

To get a better understanding for the sources of the Bellman error, Figure 4 plots the errors in a random simulation of the model, at some step that was chosen as the period with the highest error of the cubic approximation. The upper left panel shows the error in the level, the upper right panel the error in the slope of the value function. For most asset levels, the error is in fact very small, but has sharp spikes in a narrow range of asset values. This is the range where the household switches from working to not-working. The more complicated solution "P2" has very low errors in the level, but shows some variations in the slope. The lower left panel shows the same picture at a different step in the simulation. The question is why the errors around the participation threshold cannot be effectively reduced by using more approximating variables. The lower right panel of Figure 4 sheds lights on this issue. For five different levels of individual capital, it shows the Bellman as a function of the aggregate state. The aggregate state was changed continuously between some point in the simulation and the steady state distribution. The x-axis shows the aggregate level of capital at these different states. For some level of individual capital (for example $k = 7.29$), the errors has a sharp kink. This happens at the point where a household with this asset level switches to not-working. The discontinuity in behavior causes a kink in the value function. This cannot be approximated by a smooth functions in a higher number of states. To approach this problem systematically, one would probably have to resort to more complicated and

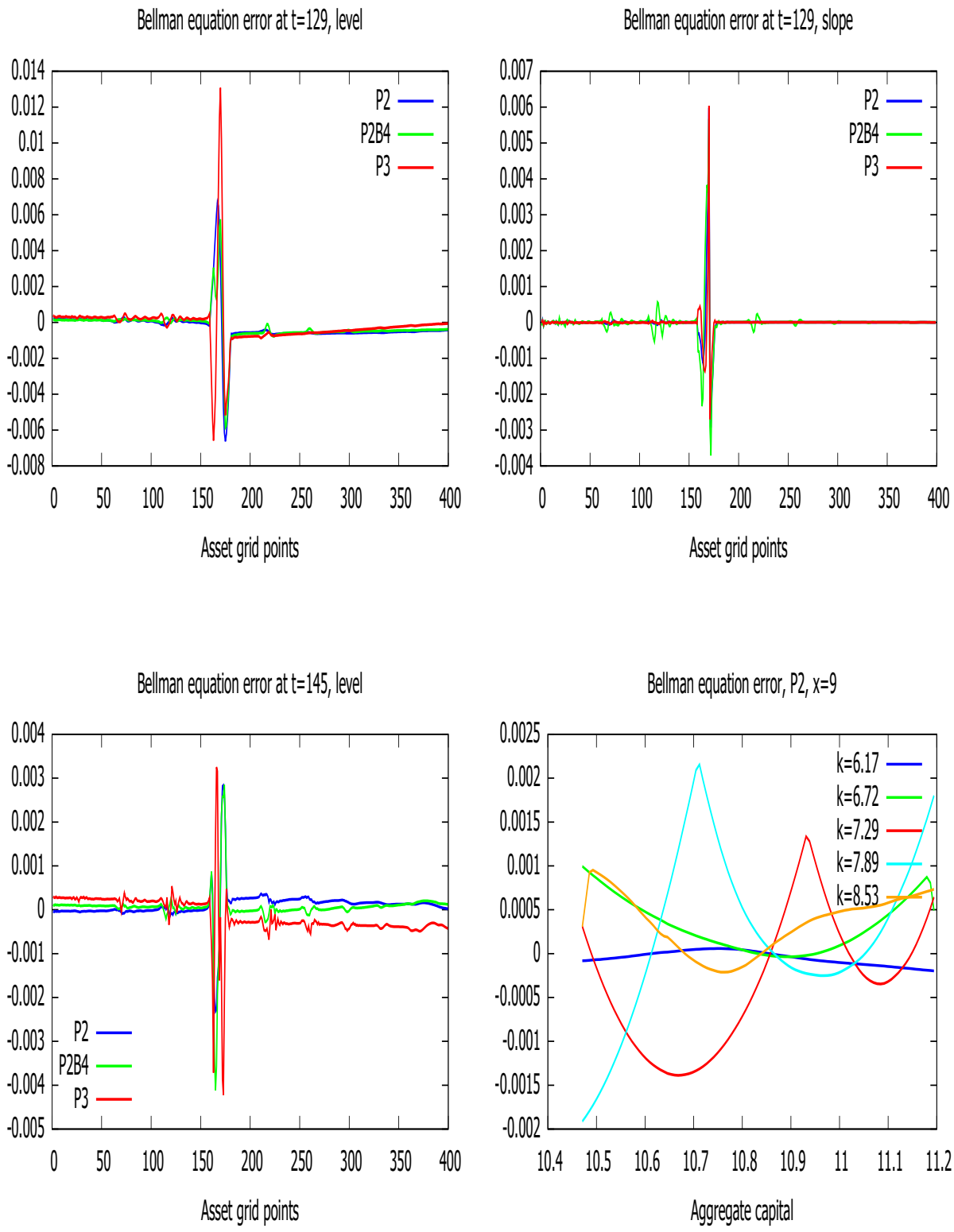


Figure 4: Bellman equation errors in simulation, Chang/Kim model, $x = 9$

more general approximation schemes such as neural networks, but this is beyond the scope of the present paper (for an application of neural networks in a low-dimensional example, cf. Fernández-Villaverde, Hurtado, and Nuno (2018)).

To summarize this discussion, the positive result is that solutions appears to be robust to changes in the details of the value function approximation. The negative result is that the approximation error cannot be systematically reduced by increasing the number of variables or the order of the polynomial approximation, which is due to kinks in the value function at the points in the aggregate state space where the households switch to not-working.

5.3 Variations on the Chang/Kim model

To investigate how different aspects of the model affect accuracy, I consider three variations. First I look at the model with fixed labor supply. This makes the household problem convex and smooth except for the kink at the borrowing constraint, and should allow much higher accuracy. Then I look at the model with divisible labor supply. This makes the household problem again convex but not so smooth, because households with high level of assets hit the boundary of zero labor supply. Finally, I go back to the model with indivisible labor, but introduce an aggregate shock to individual uncertainty, which makes individual volatility time-varying, similar to Bloom, Floetotto, Jaimovich, Eksten, and Terry (2018), This makes it harder to solve the model precisely.

5.3.1 The model with fixed labor supply

I use exactly the same parameters as in the Chang/Kim model, but eliminate the disutility of labor so that all households work.

	Linear	P1	P2	P3	P2KS	PQ4	P2B2	P2B4
σ_H	0.00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
σ_{wedge}	0.88	0.8793	0.8797	0.8797	0.8795	0.8796	0.8796	0.8796
σ_Y	0.82	0.8219	0.8219	0.8219	0.8219	0.8219	0.8219	0.8219
σ_C	0.21	0.2056	0.2051	0.2050	0.2053	0.2052	0.2052	0.2052
σ_I	3.93	3.9298	3.9311	3.9313	3.9306	3.9310	3.9309	3.9309
σ_L	0.00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\sigma_{Y/H}$	1.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\rho(H, wedge)$	0.01	0.0056	0.0060	0.0053	0.0018	0.0036	0.0067	0.0067
$\max_{t,i} \delta V_t[i]$		0.1538	0.0249	0.0251	0.0241	0.0255	0.0221	0.0214
$mean_t \max_i \delta V_t[i]$		0.0182	0.0042	0.0036	0.0042	0.0041	0.0027	0.0026
$mean_{t,i} \delta V_t[i]$		0.0058	0.0009	0.0010	0.0009	0.0008	0.0001	0.0001
$\max_{t,i} \delta V_t'[i]$		0.2650	0.0641	0.0652	0.0641	0.0650	0.0646	0.0655
$mean_t \max_i \delta V_t'[i]$		0.0270	0.0070	0.0063	0.0070	0.0075	0.0070	0.0075
$mean_{t,i} \delta V_t'[i]$		0.0003	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

Table 3: Results fixed labor supply model

We see that Bellman equation errors are two orders of magnitude smaller than in the model with indivisible labor. Especially the solutions that use states from balanced reduction have very small mean Bellman errors. If the solution is smooth, the information from balanced reduction is very helpful in obtaining high-precision solutions.

5.3.2 The model with divisible labor

Here I allow a continuous labor choice, with utility function $u(c, H) = \log(c) + \eta * \log(1 - H)$ and parameter $\eta = 1.5$.

	Linear	P1	P2	P2KS	P3	PQ2	PQ4	P2B2	P2B4
σ_H	0.478	0.4776	0.4779	0.4779	0.4780	0.4779	0.4780	0.4779	0.4778
σ_{wedge}	0.082	0.0784	0.0785	0.0784	0.0786	0.0785	0.0786	0.0784	0.0784
σ_Y	1.186	1.1866	1.1868	1.1869	1.1869	1.1869	1.1869	1.1869	1.1868
σ_C	0.326	0.3270	0.3261	0.3261	0.3261	0.3263	0.3263	0.3262	0.3263
σ_I	3.073	3.0682	3.0689	3.0695	3.0690	3.0691	3.0691	3.0694	3.0691
σ_L	0.409	0.4084	0.4086	0.4087	0.4087	0.4087	0.4087	0.4087	0.4086
$\sigma_{Y/H}$	0.542	0.5423	0.5418	0.5419	0.5417	0.5419	0.5418	0.5419	0.5420
$\rho(H, wedge)$	0.958	0.9902	0.9906	0.9904	0.9906	0.9903	0.9904	0.9903	0.9903
$\rho(H, Y/H)$	0.922	0.9224	0.9234	0.9231	0.9232	0.9228	0.9228	0.9228	0.9228
$\max_{t,i} \delta V_t[i]$		0.7389	0.1415	0.1252	0.2605	0.0795	0.1338	0.0716	0.0770
$mean_t \max_i \delta V_t[i]$		0.1383	0.0320	0.0355	0.0319	0.0240	0.0324	0.0192	0.0199
$mean_{t,i} \delta V_t[i]$		0.0394	0.0083	0.0089	0.0086	0.0059	0.0085	0.0043	0.0045
$\max_{t,i} \delta V_t'[i]$		0.4346	0.1437	0.1440	0.2609	0.1426	0.1328	0.1451	0.1331
$mean_t \max_i \delta V_t'[i]$		0.0641	0.0307	0.0311	0.0159	0.0305	0.0350	0.0310	0.0368
$mean_{t,i} \delta V_t'[i]$		0.0028	0.0006	0.0006	0.0005	0.0005	0.0006	0.0004	0.0005

Table 4: Results divisible labor model

We see that Bellman equation errors are higher than in the model with fixed labor, but still one order of magnitude smaller than in the model with indivisible labor. Again, the solution states from balanced reduction have much smaller mean Bellman errors.

5.3.3 The model with time varying volatility

To model time-varying volatility, I first compute two Markov transition matrices, both with $\rho = 0.929$ as in Chang and Kim, but one with a low variability $\sigma_1 = 0.5 \cdot 0.2272$, the other one with high variance $\sigma_2 = 1.5 \cdot 0.2272$, where $\sigma = 0.2272$ is the standard deviation of the productivity process in Chang and Kim. In any given period, the income process of the household is a convex combination of the two transition matrices, with a weight of the low-variance transition matrix equal to $cdfnormal(-\zeta_t)$, where ζ_t follows an AR(1) process

$$\zeta_t = 0.95\zeta_{t-1} + \epsilon_{\zeta,t} \tag{73}$$

	Linear	P1	P2	P2KS	P3	PQ4	P2B2	P2B4
σ_H	0.76	0.7606	0.7627	0.7678	0.7681	0.7431	0.7410	0.7405
σ_{wedge}	1.15	1.1185	1.1275	1.1227	1.1254	1.0921	1.0886	1.0905
σ_Y	1.28	1.2674	1.2622	1.2689	1.2656	1.2695	1.2692	1.2674
σ_C	0.35	0.3520	0.3528	0.3477	0.3527	0.3545	0.3542	0.3554
σ_I	3.23	3.2053	3.2121	3.2316	3.2219	3.2251	3.2264	3.2200
σ_L	0.52	0.4927	0.4882	0.4918	0.4934	0.5016	0.5012	0.5004
$\sigma_{Y/H}$	0.94	0.9189	0.9289	0.9202	0.9263	0.9069	0.9055	0.9069
$\rho(H, wedge)$	0.82	0.8208	0.8211	0.8201	0.8214	0.8083	0.8070	0.8081
$\rho(H, Y/H)$	-0.32	-0.3026	-0.3137	-0.3088	-0.3148	-0.2780	-0.2749	-0.2761
$\max_{t,i} \delta V_t[i]$		29.9968	13.7011	13.5468	8.9861	14.0621	13.7514	13.5966
$mean_t \max_i \delta V_t[i]$		2.4290	1.0488	1.0418	0.8308	1.0261	0.9907	0.9728
$mean_{t,i} \delta V_t[i]$		0.5538	0.2477	0.2496	0.2433	0.2213	0.1995	0.1964
$\max_{t,i} \delta V_t'[i]$		20.7597	36.8507	36.8598	40.1863	26.5640	25.0701	25.1735
$mean_t \max_i \delta V_t'[i]$		4.8347	4.3551	4.2642	4.6683	4.4078	4.2825	4.2365
$mean_{t,i} \delta V_t'[i]$		0.0477	0.0288	0.0283	0.0336	0.0341	0.0324	0.0324

Table 5: Results indivisible labor model with varying volatility

Table 5 shows that this model is more difficult to solve than the model with constant volatility and the Bellman errors are substantially larger. Nevertheless, the second moments from the linearized solution are pretty close to the nonlinear solutions. This is confirmed by Figure 5 which shows impulse responses to a volatility shock, again comparing the solution with the continuation value from the linearized model (left panels) to the collocation solution with quadratic approximation (right panels). The deviations from linearity and symmetry are visible, but not large.

5.4 The model of Khan and Thomas (2008)

Table 6 compares the results of Table IV in Khan and Thomas (2008) with the results from the linearized solution and from the collocation solution with polynomial approximation up to third degree.

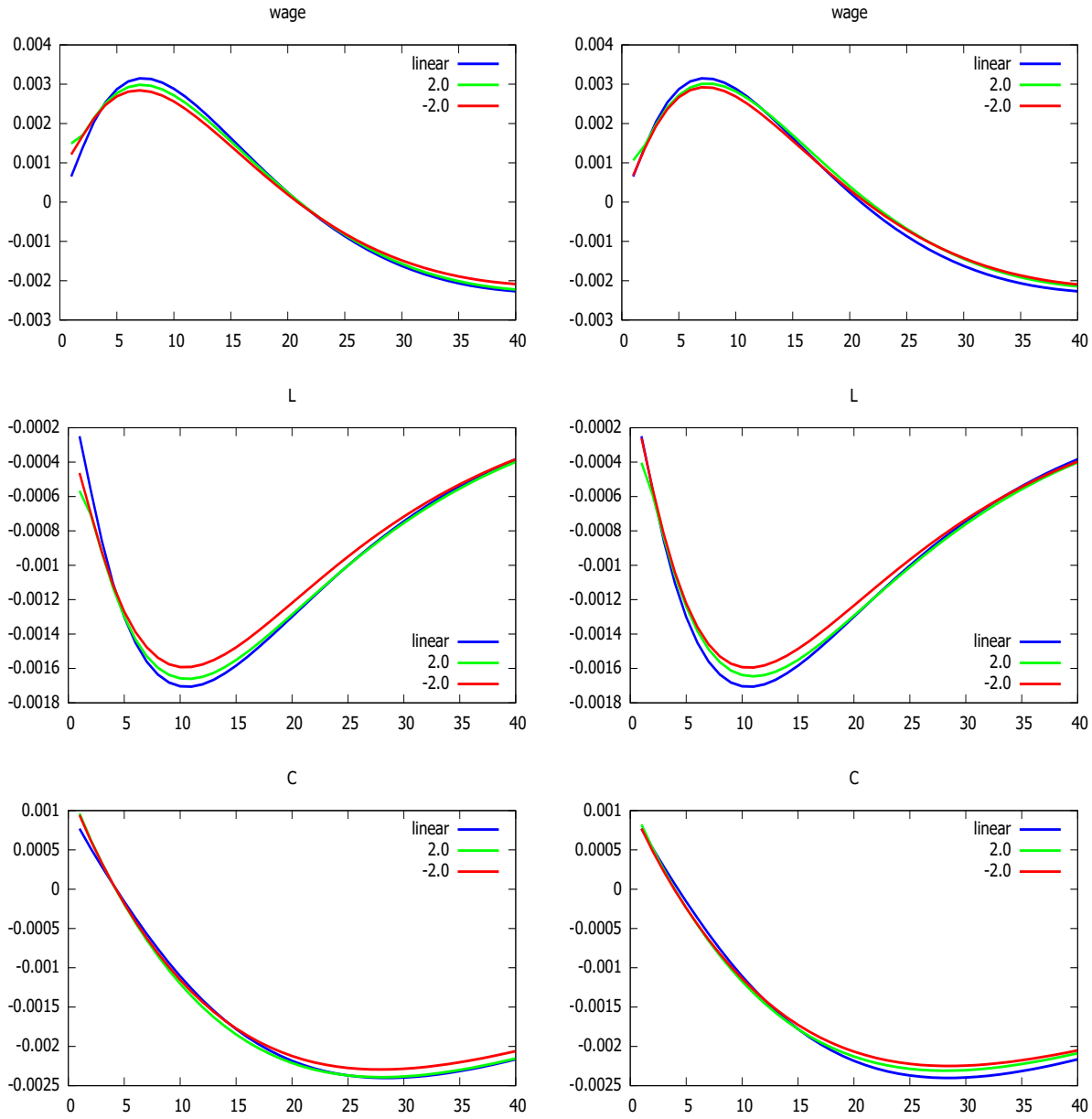


Figure 5: Impulse response to volatility shock, PELIN (left) and P2 (right)

Linear		Khan/Thomas	Linear	P1	P2	P3
σ_Y	2.161	2.264	2.161	2.162	2.173	2.168
σ_{TFP}	0.607	0.605	0.607	0.606	0.603	0.605
σ_H	0.633	0.639	0.633	0.638	0.642	0.640
σ_C	0.436	0.433	0.436	0.437	0.431	0.433
σ_I	3.540	3.539	3.540	3.546	3.550	3.544
σ_K	0.490	0.492	0.490	0.493	0.495	0.494
$cor_{TFP,Y}$	1.000	1.000	1.000	1.000	1.000	1.000
$cor_{H,Y}$	0.955	0.956	0.955	0.953	0.955	0.955
$cor_{C,Y}$	0.903	0.900	0.903	0.897	0.898	0.898
$cor_{I,Y}$	0.975	0.976	0.975	0.973	0.975	0.975
$cor_{K,Y}$	0.038	0.034	0.038	0.040	0.037	0.038
exc.kurtosis(I/K)	0.057	-0.074	0.000	-0.014	-0.074	-0.081

Table 6: Results Khan/Thomas model

Results were obtained from a simulation of 10,000 periods. Excess kurtosis is expressed as difference to the linearized solution, where it is theoretically zero (but was 0.057 due to sampling error). The results of all solution methods are very close (again, the only non-trivial difference is in the absolute size of fluctuations, not the relative standard deviations). The very light excess kurtosis is discovered by the quadratic and the cubic approximation. Near-linearity of the model solution is confirmed by the impulse responses, which are very close to linear in the size of the shock, and therefore not shown. Comparing the Bellman errors (not shown here), a quadratic approximation appears slightly better than a cubic approximation.

6 Conclusions

In this paper I have developed a method of state and value function reduction that allows to solve large heterogeneous agent models by linearization in aggregate states, in particular models with non-convex optimization problems. This reduction process is fully automatized, and independent of the specific structure of the model such as the number of individual states.

Applying these methods to standard models with heterogeneous households or firms, it turns out that linearization provides a very good approximation to global nonlinear solutions in terms of second moments of macroeconomic aggregates. Even with strong deviations from linearity such as non-convexity at the individual level, linearization is still a good approximation at the aggregate level. This will not be true for all models, therefore it is important to check for deviations from linearity, which may arise for example from sharp variations in the cross-sectional density. I have shown how to build on the linearized solution to obtain global nonlinear approximations by collocation methods. They give solutions that are very similar to solutions obtained with a Krusell-Smith method, but are easier to compute.

It is very hard to obtain truly high-precision solutions of heterogeneous agent models with non-convex optimization problems. This is because the value function at a given point of the individual state space shows kinks in aggregate state variables at points where the discrete decision changes. Kinks at unknown points in a higher-dimensional state space cannot be approximated by traditional methods, but can potentially be handled by computationally intensive machine-earning techniques. This is left for future research.

A Interpolating the value function

The problem of the economic agent has one continuous state variable, in which the value function must be interpolated. In a typical household problem, one makes only a small mistake by using piece-wise linear interpolation. This is because the curvature of the period-utility function dominates the curvature of the continuation value function. This is very different in the problem of a firm. Since firm-owners have a diversified portfolio of firms, the objective function is linear in the dividends that the firm pays. The curvature of the problem stems from the curvature of the continuation value function, which arises from decreasing returns to scale, capital adjustment costs etc. If the value function is interpolated linearly, the optimal decision will typically lead to an end-of-period state at one of the knot points of the value function, where the derivatives changes discontinuously. Such a solution cannot be used in combination with the linearization approach, because we have to compute the derivative of the optimal decision with respect to aggregate variables. If the policy is stuck at a knot point, this derivative is artificially zero. A nonlinear interpolation of the firm value function is therefore essential.

The usual nonlinear interpolation techniques, such as polynomial or spline inerpolations, destroy the contraction property of the Bellman operator, and convergence is not guaranteed any longer. This problem is made more severe by the use of acceleration steps, which are very important to speed up the algorithm. I therefore use the following non-standard approach, which guarantees convergence of the dynamic programming algorithm.

1. The first iterations in the DP algorithm are done with piece-wise linear approximation.

2. If the algorithm is close to convergence, I compute a quadratic approximation of the value function between any two knot points x_i, x_{i+1}, where x is the state variable of the agent. I express this quadratic approximation as

$$V(x) = (1-p)V(x_i) + pV(x_{i+1}) + p(p-1)Q_i \tag{74}$$

   where p is defined as

$$p = \frac{x - x_i}{x_{i+1} - x_i} \tag{75}$$

3. In the following iteration steps as well as for the linearization, I keep the quadratic parameters Q_i fixed.

Between knot points, the Q_i keep the curvature fixed at the steady state level. Once set, they do not depend on the value function, so that the interpolation is piecewise linear in the V_{x_i}, which guarantees convergence.

This method was applied for all the examples in the text.

B Finite representation of the distribution

Denote by $\pi_k(i,j)$ the transition probability from histogram bin i to histogram bin j during period t, conditional on individual productivity being x_k at the beginning of t. We now have to approximate the transition probabilities. I assume that the discrete decision is the same at the lower and the upper end of the bin. If there is a threshold point within the bin, I treat this case simply as two separate bins. Denote by $a_0 = a(\bar{a}_i^D, x_k; \mu, \lambda)$ the continuous decision taken at the lower end of the bin, $a_1 = a(\bar{a}_{i+1}^D, x_k; \mu, \lambda)$ the continuous decision taken at the upper end. We make the assumption that the histogram bins are so small that the continuous decision over this range is well approximated by a linear function.

For simplicity of exposition we assume that $a_0 \leq a_1$; the change of formulas for the opposite case is straightforward. Denote by ι_0 and ι_1 the indices of the histogram bin in which a_0 and a_1 lie, respectively. We look for a transition law that preserves the expected value of a', assuming that $a(.)$ is linear on the bin, and that the density is constant on bins. Defining $\hat{a}_j \equiv \frac{\bar{a}_j^D + \bar{a}_{j+1}^D}{2}$, this can be written as

$$\hat{a} \equiv \frac{a_0 + a_1}{2} = \sum_{j=1}^{n_x} \pi_k(i,j)\hat{a}_j \tag{76}$$

We have to distinguish the following three cases.

1. $\iota_1 = \iota_0$. The image of bin i under the mapping $a(., x_k; \mu, \lambda)$ is contained in the bin ι_1. To preserve expected value, we have to allow for a positive probability of going to an adjacent bin. We choose

$$\pi_k(i,\iota_0) = \begin{cases} \frac{\hat{a}-\hat{a}_{\iota_0+1}}{\hat{a}_{\iota_0}-\hat{a}_{\iota_0+1}}, , & \text{if } \hat{a} \geq \hat{a}_{\iota_0} \\ \frac{\hat{a}-\hat{a}_{\iota_0-1}}{\hat{a}_{\iota_0}-\hat{a}_{\iota_0-1}}, & \text{otherwise} \end{cases} \tag{77}$$

   with $\pi_k(i,\iota_0+1) = 1 - \pi_k(i,\iota_0)$ in the first of these case, and $\pi_k(i,\iota_0-1) = 1 - \pi_k(i,\iota_0)$ in the second one.

2. $\iota_1 = \iota_0 + 1$. The image of bin i is contained in two adjacent bins. We preseerve expected values by choosing $\pi_k(i,\iota_0) = \frac{\hat{a}-\hat{a}_{\iota_1}}{\hat{a}_{\iota_0}-\hat{a}_{\iota_1}}$ and $\pi_k(i,\iota_0+1) = 1 - \pi_k(i,\iota_0)$.

3. $\iota_1 = \iota_0 + 2$.k Now we split the probability mass between three intervals. This gives us the flexibility to match both the first and the second moments of the conditional distribution. We first try to do this by solving this as a linear problem. This can fail in the sense that the probability for one of the intervals is negative. This will never be the middle interval, it must be either ι_0 or ι_1. We set this probability to 0 and use the remaining two intervals so as to match the conditional mean. This is the choice that minimizes the cross-sectional variance, and comes as close as possible to the target variance.

4. $\iota_1 > \iota_0 + 2$. If the mean $\hat{a}$ lies within one of the middle intervals, it is certainly possible to match both mean and variance, because one can then match the mean using oinly middle intervals, and this will certainly have a variance that is lower than $Var(a)$. One can then use the outer intervals to achieve a mean preserving spread, until the variance is matched. Since we have more than two free parameters, there is in general a continuum of choices that matches both moments. It would be natural to assign the interior bins probabilities that are proportional to their respective widths; in other words, to treat the interior bins just like one big bin. Since this will not always be possible, I proceed in the following iterative manner:

   (a) I treat the bins $\iota_0 + 1, \ldots, \iota_1 - 1$ as one bin, and assign the probabilities as explain above in the case $\iota_0 + 2 = \iota_1$.

   (b) If one of the outer bins has negative probility, I drop this bin from the list. Then I treat again the inner ones of the remaining bins as one bin and go back to step 4.

   (c) If we are left with only three bins, we proceed with step 3.

   This procedure will always lead to a solution that matches both moments as long as $\hat{a}$ lies within the interior bins. If this is not the case, it is not in general possible to match both moments.

These rules define the transition matrix Π_k, Denote by $\pi_k(i)$ the mass of households with productivity x_k in histogram bin i, and by π_k the vector containing the $\pi_k(i)$'s. The total distribution is then chacterized by stacking all the $\pi_{[}k]$ into the big vector π. The transition matrix for the asset distribution from the beginning to the end of the period is then given by the block-diagonal matrix

$$\Pi = \begin{bmatrix} \Pi_1 & 0 & \ldots & 0 & 0 \\ 0 & \Pi_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \Pi_{n_x-1} & 0 \\ 0 & 0 & \ldots & 0 & \Pi_{n_x} \end{bmatrix} \tag{78}$$

C Approximate Aggregation

C.1 The linearized model

we assume that the model can be written in the form

$$s_t = \mathcal{S}_s s_{t-1} + \mathcal{S}_d d_t + \mathcal{S}_e \epsilon_t \tag{79a}$$

$$d_t = \mathrm{E}_t \left[\mathcal{D}_{s0} s_t + \mathcal{D}_{s1} s_{t+1} + \mathcal{D}_{d1} d_{t+1} \right] \tag{79b}$$

$$y_t = C s_t \tag{79c}$$

with given matrices $\mathcal{S}_s$, $\mathcal{S}_d$, $\mathcal{S}_e$, $\mathcal{D}_{s0}$, $\mathcal{D}_{s1}$, $\mathcal{D}_{d1}$ and C. Here d_t contains both v_t and y_t in the notation of Section 3.2.

I will discuss two approaches to approximate aggregation. First, the well known method of Krusell and Smith (1998), applied to a linearized model. Second, the Proxy Distribution approach of Reiter (2010b). Then I will show that the two are equivalent in a linear setting.

C.1.1 The Linear Krusell/Smith (KS) Algorithm

The idea of Krusell and Smith (1998) is to stipulate an aggregate law of motion for the reduced state vector

$$\hat{x}_t = \hat{A}\hat{x}_{t-1} + \hat{B}\epsilon_t \tag{80}$$

Agents are assumed to solve their individual optimization problem assuming that the aggregate state $\hat{x}_t$ follows (80). The algorithm alternates between solving for the individual decision function given $\hat{A}$ and $\hat{B}$, and updating $\hat{A}$ and $\hat{B}$ from the individual decision functions, until the two are consistent in an OLS sense. That means, if agents see a long realization of the model economy and estimate the model (80)), they obtain the $\hat{A}$ and $\hat{B}$ that they have used in their optimization.

More formally, the KS algorithm solves the following fixed point problem:

1. Guess $\hat{A}$ and $\hat{B}$ where $\hat{A}$ is asymptotically stable (that means, all eigenvalues are smaller than 1 in absolute value).

2. Solve the system of equations (80) and (79b) to get the matrices D and D_E of the decision rule

$$d_t = D\hat{x}_{t-1} + D_E\epsilon_t \tag{81}$$

3. Set A and B as

$$A = \mathcal{S}_s + \mathcal{S}_d D H \tag{82a}$$
$$B = \mathcal{S}_e + \mathcal{S}_d D_E \tag{82b}$$

   and compute Σ_x as the unique symmetric solution of

$$\Sigma_x = A\Sigma_x A' + B\Sigma_\epsilon B' \tag{83}$$

   The numerical computation of Σ_x in cases where A is very large is discussed in Reiter (2010a, Appendix B.3). Notice that the matrix A has to be asymptotically stable for Σ_x to be defined. If it is not, the pair $(\hat{A}, \hat{B})$ is not admissible.

4. Update $\hat{A}$ and $\hat{B}$ by the OLS regression

$$\hat{A} = HA\Sigma_x H' \left(H\Sigma_x H'\right)^{-1} \tag{84a}$$
$$\hat{B} = HB \tag{84b}$$

5. Iterate until the results in (84) are consistent with the guess in Step 1. This can be done by a quasi-Newton algorithm over the elements of $\hat{A}$ and $\hat{B}$.

The household decision rule (81) implies the dynamic equation

$$s_t = As_{t-1} + B\epsilon_t \tag{85}$$

with A and B defined by (82). Σ_x is the unconditional covariance matrix of the state s_t under this dynamics. Notice that, because of the linearity of the setup, the algorithm can use the asymptotic formulas (84) for the OLS estimation (cf. Reiter (2010a, Appendix B.1) for a derivation), and thereby avoid the use of simulation methods. It is therefore not affected by sampling errors, unlike the original (nonlinear) algorithm of Krusell and Smith (1998). In the choice of H, care must be taken that $H\Sigma_x H'$ is regular and not too ill-conditioned. This can be achieved by replacing every row in H by the residual of a weighted least squares regression of this row on all earlier rows, with weighting matrix Σ_x.

Definition 1. *A KS solution of the linearized model (79) consists of matrices* $(\hat{A}, \hat{B}, \Sigma_x, A, B, D, D_E)$*, where* $\hat{A}$ *and* A *are asymptotically stable, such that (81) solves the model (80),(79b) and Equations (82)–(84) are satisfied.*

C.1.2 The Proxy Distribution (PD) Method

The idea of the proxy distribution method is to assume that economic agents perceive a law of motion for the reduced aggregate state $\hat{x}_t$ that is exactly consistent with individual behavior in those states where the cross-sectional distribution of capital is in some sense normal or typical, conditional on the reduced state being equal to $\hat{x}_t$. To make this operational, we need a "proxy distribution function", which assigns to each reduced state a representative cross sectional distribution $\mathbf{\Phi}^{pd}(m_t, z_t)$. In a linear context, we can assume that the proxy distribution function is linear, so that we can write

$$s_t = \begin{bmatrix} \mathbf{\Phi}^{pd}(m_t, z_t) \\ z_t \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ 0 & I \end{bmatrix} \begin{bmatrix} m_t \\ z_t \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ 0 & I \end{bmatrix} \hat{x}_t \equiv S\hat{x}_t \tag{86}$$

for suitable matrices S_{11}, S_{12} and S. Premultiplying (79a) by H, and assuming (86) in period $t-1$, we get the following dynamic equation for $\hat{x}_t$:

$$\hat{x}_t = H\mathcal{S}_s S\hat{x}_{t-1} + H\mathcal{S}_d d_t + H\mathcal{S}_e \epsilon_t \tag{87}$$

For given S, the reduced model consists of (87) and (79b). Notice that (87) will in general not be satisfied for $s_t \neq S\hat{x}_t$.

The PD algorithm iterates over the matrix S, which defines the proxy distribution:

1. Guess S such that $HS = I$. This conditions means that the proxy distribution for (m_t, z_t) really has moments m_t.

2. Solve the system of equations (87) and (79b) to get the matrices D and D_E of the decision rule (81).

3. Define A and B as in (82) and Σ_x as in (83).

4. Update S by
$$S = \Sigma_x H' \left[H \Sigma_x H' \right]^{-1} \tag{88}$$
Return to Step 2 and iterate until convergence.

Again, matrix A in Step 3 is required to be asymptotically stable. The formula for updating S in step 4 is such that $S\hat{x}$ is the expectation of s conditional on $\hat{x}$, under the model (85) with multivariate normal shocks.

Definition 2. *A PD solution of the linearized model (79) consists of matrices $(S, \Sigma_x, A, B, D, D_E)$, where A is asymptotically stable, such that (81) solves the model (87),(79b) and Equations (82), (83) and (88) are satisfied.*

C.1.3 Comparison of KS and PD Algorithms

Proposition 2. *To any set of matrices $(\hat{A}, \hat{B}, \Sigma_x, A, B, D, D_E)$ being a KS solution, there is a matrix S such that $(S, \Sigma_x, A, B, D, D_E)$ is a PD solution.*

Conversely, to any set of matrices $(S, \Sigma_x, A, B, D, D_E)$ being a PD solution, there are matrices $\hat{A}, \hat{B}$ such that $(\hat{A}, \hat{B}, \Sigma_x, A, B, D, D_E)$ is a KS solution.

Proof. For a given KS solution, choose S by (88). For a given PD solution, choose $\hat{A}$ and $\hat{B}$ by (84). Then we have to show that (80) is equivalent to (87). To do this, use (82a) in (84a) to get

$$\begin{aligned} \hat{A} &= H \left(\mathcal{S}_s + \mathcal{S}_d D H \right) \Sigma_x H' \left(H \Sigma_x H' \right)^{-1} \\ &= H \left(\mathcal{S}_s \Sigma_x H' \left(H \Sigma_x H' \right)^{-1} + \mathcal{S}_d D \right) \end{aligned} \tag{89}$$

Then plug (88) and (81) into (87), and then use (84b) and (89). This gives

$$\hat{x}_t = H \left[\mathcal{S}_s \Sigma_x H' \left(H \Sigma_x H' \right)^{-1} + \mathcal{S}_d D \right] \hat{x}_{t-1} + H \mathcal{S}_d D_E \epsilon_t + H \mathcal{S}_e \epsilon_t = \hat{A} \hat{x}_{t-1} + \hat{B} \epsilon_t \tag{90}$$

which is the same as (80).

Then the only thing that remains to be shown is that the stability of A, as required in the definition of a PD solution, also implies the stability of $\hat{A}$, which is required in the definition of a KS solution. This follows from (84a) and (83), as is shown in Reiter (2010a, Appendix B.2). □

The proposition shows that the two approaches are equivalent in the sense of having the same solution. It is nevertheless worthwhile to study both. One reason is that they

lead to different algorithms to find this solution. In this respect, both methods have their advantages. The advantage of the KS algorithm is that the objects over which the algorithm iterates, namely $\hat{A}$ and $\hat{B}$, have only few elements, at least if $\dim(\hat{x})$ is small. This allows to use efficient quasi-Newton algorithms to solve the nonlinear fixed point problem of the algorithm. In contrast, the object S of the PD algorithm has so many elements that the only feasible approach is some kind of fixed point iteration, as outlined in Section C.1.2. The good side of the PD approach is that the fixed point iteration in S converges in a few iterations in many cases. Moreover, if we know the covariance matrix Σ_x from the exact linearized solution, then we might use it in (88) and do not need to iterate on S at all. In my experience, if m_t is big enough so that it contains the information that is essential for household decision making, then the choice of the proxy distribution (the choice of S) has very little impact on the solution. In that case, we solve the model consisting of (87) and (79b) in one step.

However, the main reason why we are interested in the PD approach is that it generalizes more easily to the case of higher order solutions. If household decisions depend nonlinearly on aggregate variables, the law of motion for $\hat{x}_t$ (cf. (80)) should also be nonlinear, in order to be at least approximately consistent with behavior at the micro level. This makes it difficult to apply the Krusell/Smith algorithm if the state vector has medium dimension (10-20, say). Then the number of parameters in the aggregate law of motion is in the range of 10^3 or more. Finding those parameters by an iterative process of guessing the parameters, solving and simulating the model, and reestimating the law of motion, appears very hard to do at the moment.

The proxy distribution method does not suffer from this problem, because it does not stipulate an aggregate law of motion. We rather enforce exact compatibility between individual and aggregate behavior at the proxy distributions. If shocks are small, it is natural to use the proxy distribution function that we have obtained in the linearized solution (cf. (88)). And again, if m_t contains the relevant information for economic agents, the exact choice S should make little difference. Given a proxy distribution function, the model can be solved in one sweep of backward iterations. This has already been shown in Reiter (2010b), although the example there uses a very low-dimensional state vector.

References

Ahn, S., G. Kaplan, B. Moll, T. Winberry, and C. Wolf (2018). When Inequality Matters for Macro and Macro Matters for Inequality. *NBER Macroeconomics Annual 32*(1), 1–75.

Antoulas, A. C. (1999). Approximation of linear dynamical systems. In J. Webster (Ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering, volume 11*. Wiley.

Antoulas, A. C. (2005). *Approximation of Large-Scale Dynamical Systems*. SIAM.

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2018, June). Inequality, Business Cycles, and Monetary-Fiscal Policy. NBER Working Papers 24710, National Bureau of Economic Research, Inc.

Bloom, N., M. Floetotto, N. Jaimovich, I. S. Eksten, and S. J. Terry (2018, May). Really Uncertain Business Cycles. *Econometrica 86*(3), 1031–1065.

Boppart, T., P. Krusell, and K. Mitman (2018). Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative. *Journal of Economic Dynamics and Control 89*(C), 68–92.

Chang, Y. and S.-B. Kim (2007). Heterogeneity and aggregation: Implications for labor-market fluctuations. *American Economic Review 97*(5), 1939–1956.

Childers, D. (2016). On the Solution and Application of Rational Expectations Models with Function-Valued States. 2016 Meeting Papers 807, Society for Economic Dynamics.

Costain, J. and A. Nakov (2011). Distributional dynamics under smoothly state-dependent pricing. *Journal of Monetary Economics 58*(6), 646–665.

Fernández-Villaverde, J., S. Hurtado, and G. Nuno (2018). Financial Frictions and the Wealth Distribution. Technical report. Manuscript.

Judd, K. L., L. Maliar, and S. Maliar (2012). Merging simulation and projection approaches to solve high-dimensional problems. NBER Working Papers 18501, National Bureau of Economic Research, Inc.

Khan, A. and J. K. Thomas (2008). Idiosyncratic shocks and the role of noncenvixities in plant and aggregate investment dynamics. *Econometrica 76*(2), 395–436.

Krusell, P. and A. A. Smith (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy 106*(5), 867–96.

Kubler, F. and S. Scheidegger (2018). Self-justified equilibria: Existence and computation. Manuscript, University of Zurich.

McKay, A. and R. Reis (2016, January). The Role of Automatic Stabilizers in the U.S. Business Cycle. *Econometrica 84*, 141–194.

Mertens, T. M. and K. L. Judd (2017). Solving an incomplete markets model with a large cross-section of agents. ?

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

Reiter, M. (2010a). Approximate and almost-exact aggregation in dynamic stochastic heterogeneous-agent models. IHS Working Paper 258.

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

Reiter, M., T. Sveen, and L. Weinke (2013). Lumpy investment and the monetary transmission mechanism. *Journal of Monetary Economics 60*(7), 821–834.

Takahashi, S. (2014, April). Heterogeneity and Aggregation: Implications for Labor-Market Fluctuations: Comment. *American Economic Review 104*(4), 1446–60.

Winberry, T. (2018). A toolbox for solving and estimating heterogeneous agent macro models. Quantitative Economics, forthcoming.

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.