# Solving OLG Models with Many Cohorts, Asset Choice and Large Shocks *

Michael Reiter, Institute for Advanced Studies, Vienna

December 29, 2015

**Abstract**

The paper presents a computationally efficient method to solve overlapping generations models with asset choice. The method is used to study an OLG economy with many cohorts, up to 3 different assets, stochastic volatility, short-sale constraints, and subject to rather large technology shocks.

On the methodological side, the main findings are that global projection methods with polynomial approximations of degree 3 are sufficient to provide a very precise solution, even in the case of large shocks. Globally linear approximations, in contrast to local linear approximations, are sufficient to capture the most important financial statistics, including not only the average risk premium, but also the variation of the risk premium over the cycle. However, global linear approximations are not sufficient to reliably pin down asset choices.

With a risk aversion parameter of only 4, the model generates a price of risk, measured as the Sharpe ratio, that is almost half of what it is for US stocks. However, the asset price fluctuations and the equity premium are much smaller than in US data.

*JEL classification:* C63, C68, E21, G11.
*Keywords:* OLG models, asset choice, projection methods.

Address of the author:

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

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

*I am grateful to Albert Marcet, Laurence Kotlikoff, Jasmina Hasanhodzic, as well as seminar participants in Barcelona, Oslo, Warsaw and Vienna, for helpful discussions.

# 1 Introduction

Overlapping generations models play a prominent role in economic policy analysis. They are the central tool to evaluate social security systems and address questions of intergenerational redistribution (some recent papers are Krueger and Kubler (2006), Ludwig and Reiter (2010), Hasanhodzic and Kotlikoff (2013)). The quantitative analysis of deterministic models, even those with a large number of cohorts, has been possible since Auerbach and Kotlikoff (1981) and Auerbach and Kotlikoff (1987). But the solution of OLG models with aggregate risk and portfolio choice, like heterogeneous agents models with aggregate risk in general, still poses difficult computational challenges. The number of state variables increases with the number of cohorts, because the wealth of each cohort, and potentially other cohort-specific variables, are state variables. The number of states is a key determinant of the computational complexity of a model. Particularly hard to solve are models which involve aggregate risk and portfolio choice, because portfolio choice is a more subtle decision problem than, say, consumption versus saving, and therefore requires high precision of the approximation.

In principle, methods to solve this kind of problems exist and are well known. First, there are local perturbation methods. They provide insightful results for the case of small shocks (Judd and Guu 2001), but they have problems dealing with large shocks. In particular, it is very difficult to handle inequality constraints, such as short-sale constraints. The main alternative are global projection methods, pioneered in economics by Judd (1992) which are very powerful and flexible. While these methods have been around for a long time, their implementation tends to be complex, and often requires choices that are model-specific.[1] Furthermore, convergence of the algorithm is not guaranteed, and failure is more likely the bigger are the shocks.

The aim of this paper is to provide an efficient implementation of global projection methods, which is general, gives a high-precision solution, and is robust in the sense that it is very likely to converge, even in the case of large shocks. The method is applied to an OLG model with asset choice. I also solve examples including short-sale constraints. Although computationally highly challenging, this model is simple and standard. Much of the discussion in this paper is not specific to OLG models, but applies equally to other

[1] The Matlab software package that accompanies Miranda and Fackler (2002) provides very helpful tools for projecton methods.

heterogeneous agent models with a medium-dimensional state space.

The algorithm I will present allows to solve medium-sized OLG models with several assets in a few minutes on a standard personal computer. It is based on the following key elements:

1. Minimize the number of state variables by an appropriate choice of states.

2. Minimize the number of variables that have to be approximated by parameterized functions.

3. Find the parameters by quasi-Newton methods. To compute Jacobians efficiently, use automatic differentiation and the implicit function theorem.

4. To achieve fast convergence of the quasi-Newton algorithm, use continuation methods, starting from the linearized solution around the deterministic steady state, with small shocks, and transform it continuously into the nonlinear solution with large shocks.

5. Use simulation techniques to obtain a suitable grid on which the solution is approximated.

Let me explain these points in more detail. Since the number of state variables is decisive for the computational complexity of nonlinear algorithms, it is important to reduce it whenever possible. In our specific model, apart from the exogenous states, this means to use only the household net worth of each cohort at the beginning of the period, after the realization of shocks, not the asset portfolio at the end of the last period. This is possible in the absence of asset trading costs. Concerning the number of variables that are parameterized, it turns out that we only need to parameterize the consumption of each cohort and, depending on the asset structure, the prices of some long-lived assets. The portfolio structure (asset choice) is not parameterized, but solved for at each grid point. This has two advantages. First, it reduces the number of parameters needed for function approximation. Second, in the case of short-sale constraints, one avoids approximating functions that have kinks.

Using household net worth as state variables appears very natural. In a model without trading frictions, what matters for the behavior of all economic agents is just the market value of all assets, not the portfolio composition. The idea is not new (see, for example,

Marcet and Singleton (1998)), but the implementation is not as straightforward as it seems. If the net worth of households are the state variables, then we assume that all variables of the solution, including asset prices, are a function of net worth. However, household net worth itself is a function of current asset prices: what is predetermined is households' asset positions at the end of last period, but we need current asset prices to convert those into current net worth. This means that we have to solve a nonlinear fixed point problem in order to compute current states from past states and decisions.

This problem is reinforced by the decision not to parameterize asset choices. It implies that asset choices have to be solved for at each grid point, and in each period of a model simulation, conditional on the parameter vector that determines the consumption function. In combination with the fixed point problem necessary to solve for market wealth of each cohort, this amounts to a formidable equilibrium problem at each grid point. While this is straightforward conceptually, it can be very time consuming. To make this problem tractable, I resort to automatic differentiation and exploit the implicit function theorem.

The biggest challenge lies in finding the parameters that characterize the consumption functions of all cohorts. Quasi-Newton methods are known to converge quickly if good starting values are given. This is where continuation methods come into play. For a model with small shocks, the linearized solution around the deterministic steady state provides an excellent starting point. By increasing the shock variance step by step, the solution in each step provides a starting point for the next step. Still, applying Newton-type methods requires the handling of a very large Jacobian, since we have to solve for thousands of parameters simultaneously. Computing the Jacobian can be done very efficiently using automatic differentiation, explained in more detail in Section 3.4.2. Furthermore, it turns out that we usually have to compute the Jacobian only once. In the course of the iterations, it is then updated using Broyden's low-rank updates, which is inexpensive. The details are explained in Section 3.4.4.

Finally, it is essential to choose a suitable grid on which to compute the residuals of the projection method. With a medium- or high-dimensional state space, it seems unavoidable to use simulation techniques to obtain some information about the region of the state space in which the solution lives, in the spirit of parameterized expectation methods (Marcet and Lorenzoni 1998; Judd, Maliar, and Maliar 2012). I try to minimize the impact of simulation, and use it only to obtain information about the first and second moments of the model solution.

While all these steps together are rather complicated, they can be mostly automatized in a toolkit. The toolkit uses some relatively simple syntax, similar to Dynare, and then generates C++-code which gets compiled and executed. The output is then later post-processed by GNU Octave.[2]

The paper in the literature that is closest to mine is Hasanhodzic and Kotlikoff (2015), who simulate an economy with capital and bonds and 80 cohorts. They also use the vector of household net worth as state variables, and they approximate household consumption functions by globally linear functions. In an earlier version (Hasanhodzic and Kotlikoff 2013), they used past savings as state vector, and solved the model with capital and bonds for 40 cohorts. For the new version, they switched to household net worth as state vector, which turned out to be more robust. Hasanhodzic and Kotlikoff (2015) and my paper were written independently, and provide independent evidence that the vector of household net worth is the most useful choice of state variables. In terms of methodology, the main difference is that Hasanhodzic and Kotlikoff (2015) use a simulation approach in the spirit of Judd, Maliar and Maliar (2011), while I exploit automatic differentiation to provide an efficient implementation of quasi-Newton methods.

In order to keep the discussion focused, I sidestep an issue that is potentially very important, but beyond the scope of this paper, namely state aggregation. In models with many cohorts, it may not be computationally optimal to make the decisions of each cohort a function of the full wealth distribution across cohorts, but rather find statistics of the wealth distribution that comprise the most relevant information in a parsimonius way, in the spirit of Krusell and Smith (1998). Krueger and Kubler (2004) show that state aggregation does not work as easily in OLG models as in the model of Krusell and Smith (1998). They show specifically that one statistic for the distribution is not enough (this has been recently confirmed by Hasanhodzic and Kotlikoff (2013)). However, state aggregation works probably well if enough states are used; for example, in a model with 20 cohorts, one could group 4 adjacent cohorts together and take their joint wealth as state variable (Reiter 2010). Exploring this is left for future work.

Having established a reliable solution algorithm, I am going to address some methodological as well as some economic questions. From a methodological point of view, an interesting question is which polynomial order is necessary to obtain a precise approxima-

[2]Octave is an open-source mathematical language highly compatible with Matlab. I use Octave rather than Matlab because it is very well integrated with the GNU C++-compiler.

tion of the solution. I will measure precision in a variety of ways. First, I compute Euler residuals on many points in the state space. Second, I document how relevant statistics of the solution change with the order of approximation. I focus on aspects of the solution that are related to asset choice. In particular, I look at asset positions, the equity premium and the price of risk (the Sharpe ratio), both in terms of averages and in terms of cyclical properties. From an economic point of view, I will focus on the determinants of the price of risk. I also investigate how closely the solution comes to a complete markets allocation.

The plan of the paper is as follows. Section 2 describes the model, Section 3 describes the algorithm, and Section 4 contains the numerical results. Section 5 concludes.

# 2 An OLG Model With Portfolio Choice

## 2.1 Household structure

Time is discrete. The economy is populated by $J$ overlapping cohorts of equal size. It is subject to aggregate technology shocks, but no idio-syncratic or cohort specific shocks. Each cohort (household) lives for $J$ periods, indexed by $i = 1, \ldots, J$. Households born in $t$ maximize

$$\mathrm{E}_t \sum_{i=1}^{J} \beta^{i-1} \frac{c_{i,t+i-1}^{1-\gamma}}{1-\gamma} \tag{1}$$

Labor supply is assumed to be fixed. In the first two thirds of their life (periods 1 to $2J/3$, where $J$ is a multiple of 3) workers supply their labor endowment, which is denoted by $\zeta_j$ for the cohort of age $j$. In the remaining periods, they are retired and do not work. The decision problem of the household is described in more detail in Section 2.4.

## 2.2 Final good technology

Final output is used for investment and consumption, so that the aggregate resource constraint is

$$y_t = I_t + C_t \tag{2}$$

Final output is the numeraire, so (2) implies that the price of consumption and of investment is 1. Output is produced by competitive firms using capital and labor, with the

constant returns to production function

$$y_t = z_t F(K_{t-1}, L_t) \tag{3}$$

Total factor productivity $z_t$ is a nonstationary process whose growth rate is denoted by $g$:

$$z_t = (1 + g_t) z_{t-1} \tag{4}$$

This growth rate has mean $\bar{g}$, and is subject to time-varying volatility:

$$g_t = \bar{g} + \rho_g (g_{t-1} - \bar{g}) + (1 + \sigma_{t-1}) \epsilon_{g,t} \tag{5}$$

Volatility $\sigma$ itself follows an AR(1) process:

$$\sigma_t = \rho_s \sigma_{t-1} + \epsilon_{s,t} \tag{6}$$

Notice from (5) that a shock to $\sigma_t$ affects growth only in period $t + 1$.

Factor markets are competitive. The wage $w_t$ and the rental rate of capital $r_t^K$ equal marginal productivities in equilibrium:

$$r_t^K = F_k(K_{t-1}, L_t)) \tag{7}$$

$$w_t = F_L(K_{t-1}, L_t) \tag{8}$$

## 2.3 Capital Adjustment Costs

We assume that capital is produced in a perfectly competitive capital sector, which transforms old capital $K_{t-1}$ and investment $I_t$ into new capital $K_t$ according to

$$K_t = \Psi(K_{t-1}, I_t) \tag{9}$$

where $\Psi(K, I)$ has constant returns to scale, and satisfies $\Psi(K, \delta K) = K$, $\Psi_I(K, \delta K) = 1$ and $\Psi_{II}(K, \delta K) < 0$. In period $t$, a capital producing firm maximizes current profits

$$Q_t \Psi(K_{t-1}, I_t) - \tilde{Q}_t K_{t-1} - I_t \tag{10}$$

where $Q_t$ denotes the price of a unit of new capital in period $t$, $\tilde{Q}_t$ denotes the price of a unit of capital that was produced in period $t-1$ and has already been used for production in $t$. Profit maximization gives the first order conditions

$$Q_t \Psi_K(K_{t-1}, I_t) = \tilde{Q}_t \tag{11}$$

$$Q_t \Psi_I(K_{t-1}, I_t) = 1 \tag{12}$$

We assume that capital produced in period $t-1$ is used in period $t$ first for the production of the final good according to the production function (3), and afterwards for the production of new capital according to (9). Therefore, buying a unit of capital in period $t$ costs $Q_t$ and will yield $r^K_{t+1} + \tilde{Q}_{t+1}$ in the next period. Notice that a final good firm rents capital at the price $r^K_t$, while a capital producing firm buys old capital at the price $\tilde{Q}_t$.

## 2.4 Households decisions

Since labor supply is exogenous, household decisions can be described by its asset demand functions, which then determine consumption through the household budget constraint. There is a maximum of three assets in the economy that can be traded:

1. Physical capital. As explained above, it has price $Q_t$ and yields $r^K_{t+1} + \tilde{Q}_{t+1}$ in the next period.

2. The safe asset. It costs $Q^S_t$ and yields one unit of the output good in the next period.

3. A second financial asset. Its cost is denoted by $Q^C_t$, and its yield in the next period equals $(1 + (g_{t+1} - \bar{g})^2)$ units of the output good. This asset is useful for insuring against large (in absolute terms) technology shocks. Since I will allow for a large negative shock (crisis), but not a large positive shock, it can serve primarily as an insurance against crises. So the "C" in the price of the shock is a memo for "crisis".

Denote by $k_{i,t}$ the holdings of the physical asset at the end of period $t$ by cohort $i$, where $i = 1, \ldots, J$. Similarly, $A^S_{i,t}$ and $A^C_{i,t}$ denote the holdings of the first (safe) and the second financial asset. Let $W_{i,t}$ denote market wealth of cohort $i$ at the beginning of perid $t$. End of period capital holdings then satisfy

$$k_{i,t} = (W_{i,t} + w_t\zeta_i - c_{i,t} - Q^S_t A^S_{i,t} - Q^C_t A^C_{i,t})/Q_t \tag{13}$$

Market wealth follows

$$W_{i,t} = (r^K_t + \tilde{Q}_t)k_{i-1,t-1} + A^S_{i-1,t-1} + \left(1 + (g_t - \bar{g})^2\right) A^C_{i-1,t-1} \tag{14}$$

for $i = 2, \ldots, J$, and $W_{1,t} = 0$, because a household is born without assets. Households have no bequest motive and therefore do not accumulate assets in the last period of their life, $k_{J,t} = A^S_{J,t} = A^C_{J,t} = 0$. The budget constraint in the last period is then given by

$$c_{J,t} = W_{J,t} + w_t\zeta_J = W_{J,t} \tag{15}$$

because their labor endowment in the last period of life is zero.

With asset returns defined as above, the household Euler equations are

$$U_c(c_{i,t})Q_t = \beta \, \mathrm{E}_t \left[ (r^K_{t+1} + \tilde{Q}_{t+1}) U_c(c_{i+1,t+1}) \right] \tag{16}$$

$$U_c(c_{i,t})Q^S_t = \beta \, \mathrm{E}_t \left[ U_c(c_{i+1,t+1}) \right] \tag{17}$$

$$U_c(c_{i,t})Q^C_t = \beta \, \mathrm{E}_t \left[ \left(1 + (g_{t+1} - \bar{g})^2\right) U_c(c_{i+1,t+1}) \right] \tag{18}$$

Next to the model with three assets, I also consider the model with only one or two assets. With two assets, the Euler equation (18) is replaced by $A^C_{i,t} = 0$. In the case of only one asset, the Euler equation (17) is replaced by $A^S_{i,t} = 0$.

## 2.5 Aggregation

Aggregate variables are defined as the means of cohort specific variables:

$$L_t = \frac{1}{J} \sum_{i=1}^{J} \zeta_i$$

$$K_t = \frac{1}{J} \sum_{i=1}^{J} k_{i,t}$$

$$Waggr_t = \frac{1}{J} \sum_{i=1}^{J} W_{i,t}$$

$$C_t = \frac{1}{J} \sum_{i=1}^{J} c_{i,t}$$

The financial assets are in zero net supply:

$$\frac{1}{J} \sum_{i=1}^{J} A^S_{i,t} = 0 \tag{20}$$

$$\frac{1}{J} \sum_{i=1}^{J} A^C_{i,t} = 0 \tag{21}$$

## 2.6 Short-scale constraints and government debt

In the version of the model with two assets, I also consider a short-sale constraint on the riskless asset.

$$A^S_i \geq 0 \tag{22}$$

However, households can only save in this asset if there is an aggent that is allowed to hold a short position. I therefore introduce riskless government debt. The amount of outstanding debt is constant over time. The interest on the debt is financed by lump sum taxes on all generations. Denote by $D^G$ the amount of debt per cohort. Then the household budget constraints (13) and (15) are replaced by

$$k_{i,t} = (W_{i,t} + w_t\zeta_i + D^G(Q_t^S - 1) - c_{i,t} - Q_t^S A_{i,t}^S - Q_t^C A_{i,t}^C)/Q_t \tag{23}$$

and

$$c_{J,t} = W_{J,t} + D^G(Q_t^S - 1) \tag{24}$$

The asset market clearing condition (20) is replaced by

$$\frac{1}{J}\sum_{i=1}^{J} A_{i,t}^S = D^G \tag{25}$$

# 3 Algorithm

I describe the algorithm in several steps. In Section 3.1, I describe the basic aspects of the method, such as the choice of approximating functions and the choice of a discrete grid. Section 3.2 provides the outer loop, where the model is solved several times, starting with small shock variances, and increasing the variance to its required level. Section 3.3 gives the exact definition of the residual that has to be set to zero in the solution. Section 3.4 provides some more details. For convenience of exposition, I first describe the algorithm for the case without short-sale constraints. Section 3.5 explains the modifications that are necessary to handle those constraints.

## 3.1 Setting up the algorithm

1. Choose a set of variables as state variables. We will use the exogenous states $\sigma$ and $g$,[3] as well as cohort wealth at the beginning of period, $W_i$ for $i = 2, \ldots, J$. Collect the state variables into the vector $X$.

2. Choose variables to be approximated. We wil use consumption $c_i$ for cohorts $i = 1, \ldots, J-1$ as well as the price of the safe asset, $Q^S$.

[3]Notice that $g$ could be dropped from the list of state variables in the uncorrelated case $\rho_g = 0$, because of scale invariance of the economy. Nevertheless, in the computations below I keep $g$ as state in all cases.

3. Choose a set of $n_b$ basis functions $B_l(X)$ of the vector of states $X$, $l = 1, \ldots, n_b$. We approximate the consumption functions and the asset price as linear combinations of the basis functions, with parameter vector $\theta$:

$$c_j = \mathcal{P}(X; \theta_j) \equiv \sum_{l=1}^{n_b} \theta_{j,l} B_l(X), \qquad j = 1, \ldots, J-1 \tag{26}$$

and

$$Q^S(X) = \mathcal{P}(X; \theta_J) \equiv \sum_{l=1}^{n_b} \theta_{J,l} B_l(X) \tag{27}$$

Stack the $\theta_j$ into the big parameter vector $\Theta$. The number of elements of $\Theta$ is $n_b \cdot J$.

4. Choose a set of grid points $X_i^\circ$ for $i = 1, \ldots, n_g$ with $n_g \geq n_b$. These points should cover the $n_g$-dimensional unit ball more or less uniformly. The details on how to choose this grid are discussed in Section 3.4.1).

5. Choose a long sequence of uniformly distributed random numbers $\epsilon_t$, for $t = 1, \ldots, T$. Choose a $T_{skip} < T$; this will be the number of initial observations in the model simulations that will be skipped for the computation of averages.

6. For each variable, a residual equation is needed that is supposed to be (close to) zero in equilibrium. We choose household Euler equations and an asset market clearing conditions. For details, cf. Section 3.3. We denote the j-th residual at point $X$, given parameter vector $\Theta$, by $Res(X, j; \Theta)$.

7. Solving the model means to find a parameter vector $\Theta$ by setting weighted sums of residuals equal to zero.[4] We denote the $n_g \times n_b$-matrix of weights by $\Omega$. If we write the residuals for parameter vector $\theta$ as the $n_g \times n_b$-matrix $Res(., .; \Theta)$, the objective is to find $\Theta$ such that $\Omega' Res(., .; \Theta) = 0$. We choose $\Omega$ as matrix of basis functions the the grid points, that means, $\Omega_{i,l} = B_l(\bar{X}_i)$. Since $\Omega' Res(., .; \Theta) = 0$ is equivalent to $(\Omega'\Omega)^{-1} \Omega' Res(., .; \Theta) = 0$, we effectively set the projection of the residuals into the space spanned by our basis functions equal to zero.

[4] If $n_g = n_b$, one can set the residuals at all grid points equal to zero: $Res(\bar{X}_i, j; \Theta) = 0$ for all $i = 1, \ldots, n_g$ and all $j$. I will always use $n_g > n_b$.

## 3.2 The outer loop of the algorithm

The grand scheme of the algorithm is the following. First, we solve for the deterministic steady state of the model. Then we solve the stochastic model several times. We start with a very small variance of the shock, such that the linearized solution around the deterministic steady state provides an excellent starting point. Then we increase the shock variance slowly until we reach the desired variance. A problem with this procedure is that portfolio choice is a numericall ill-conditioned problem when shocks are small (in the limit of no shocks, the portfolio problem is undetermined). We overcome this problem by introducing asset holding costs into the model, parameterized by a parameter $\kappa$, which will be gradually reduced towards zero over the course of the solution. This will be explained in detail in the next section.

The whole procedure is an application of the continuation method (Judd 1998, Section 5.7). We denote the number of steps on the continuation path by $n_H$ and the iteration count by $k$, where $k = 1, \ldots, n_H$. The outer loop then goes as follows.

1. Solve the model with only one asset (physical capital) by linearization around the deterministic steady state. Notice that this model has the same states as the model with several assets.

   The linearized solution can be written in the form

   $$\begin{aligned} X_t &= \mu_X^* + A(X_{t-1} - \mu_X^*) + B\epsilon_t \\ Y_t &= \mu_Y^* + C(X_t - \mu_X^*) \end{aligned} \tag{28}$$

   where $X$ is the state vector and $Y$ is the vector of decision variables, namely consumption levels of cohorts $j = 1, \ldots, J-1$.

   The linear solution implies a covariance matrix for the state vector, $\Sigma_X^{lin}$, which satisfies

   $$\Sigma_X^{lin} = A\Sigma_X^{lin}A' + B\Sigma_\epsilon B' \tag{29}$$

   where $\Sigma_\epsilon$ denotes the covariance matrix of the exogenous shock vector $\epsilon$.

2. Set the iteration count to $k = 1$.

3. At any iteration $k$, choose a variance scaling parameter $\lambda_k$. The sequence of variance scaling parameters should satisfy $\lambda_k > 0$, $\lambda_k \leq \lambda_{k+1}$ and $\lambda_{n_H} = 1$.

4. Choose an asset holding cost parameter $\kappa_k$. The sequence of asset holding cost parameters should satisfy $\kappa_k \geq \kappa_{k+1} \geq 0$. The role of these parameters is explained in more detail in Section 3.3. $\kappa_{n_H}$ should equal zero or be very close to zero, cf. Step 9 below.

5. Estimate the mean $\mu_{X,k}$ and the covariance matrix $\Sigma_{X,k}$ of the states.

   - In the first iteration ($k = 1$), set

   $$\begin{aligned} \mu_{X,k} &= \mu_X^* \\ \Sigma_{X,k} &= \lambda_1^2 \Sigma_X^{lin} \end{aligned} \tag{30}$$

   - In later iterations ($k > 1$), use the parameter vector $\theta_{k-1}$ and the sequence of exogenous shocks $\epsilon_t$, $t = 1, \ldots, T$, to simulate a series of state variables, denoted by $X_t^{k,+\epsilon}$ for $t = 1, \ldots, T$. Simulate a second series of state variables, denoted by $X_t^{k,-\epsilon}$ with $t = 1, \ldots, T$, by using the negative of these shocks, $-\epsilon_t$ with $t = 1, \ldots, T$. Then set

   $$\begin{aligned} \mu_{X,k} &= \frac{1}{2(T - T_{skip})} \sum_{t=T_{skip}+1}^{T} \left( X_t^{k-1,+\epsilon} + X_t^{k-1,-\epsilon} \right) \\ \Sigma_{X,k} &= \left( \frac{\lambda_k}{\lambda_{k-1}} \right)^2 \frac{1}{2(T - T_{skip}) - 1} \sum_{t=T_{skip}+1}^{T} \Big[ (X_t^{k-1,+\epsilon} - \mu_{X,k})(X_t^{k-1,+\epsilon} - \mu_{X,k})' \\ &\quad + (X_t^{k-1,-\epsilon} - \mu_{X,k})(X_t^{k-1,-\epsilon} - \mu_{X,k})' \Big] \end{aligned} \tag{31}$$

6. Set the grid of state vectors $\bar{X}_i$ to

   $$\bar{X}_i = \mu_{X,k} + \psi \Gamma_k X_i^{\circ}, \qquad i = 1, 2, \ldots, n_g \tag{32}$$

   where $\Gamma_k$ is such that $diag(\Gamma_k \Gamma_k') = diag(\Sigma_{X,k})$, and $\psi > 0$ is a factor that scales the size of the state space. These grid points should cover, at least, the part of the state space in which the economy lives most of the time. Choosing high enough $\psi$ (I use $\psi = 6$) allows the grid to occupy a rather large part of the state space, which turned out to be useful for accuracy. For a detailed description of this step, see Section 3.4.1.

7. Choose an initial guess of the parameter vector $\Theta$, denoted by $\Theta_k^{init}$.

- In the first iteration ($k = 1$), set $\Theta_k^{init}$ as the parameter vector that conforms to the linear approximation of the $c_j$ with respect to $X$ which comes out of the linearized model solution. The linear approximation should of course be nested in the family of approximation functions (26).
- For $k > 1$, set $\Theta_k^{init} = \Theta_{k-1}$, i.e., the vector obtained as the solution of the last iteration step $(k-1)$, as described in Step 8.

8. Starting from the initial guess $\Theta_k^{init}$, we find a parameter vector $\Theta$ so as to set the $n_b$ sums of weighted residuals, defined in Step 7 of Section 3.1, equal to zero. Written more extensively, we have

$$\sum_{i=1}^{n_g} \Omega_{i,l} Res(\bar{X}_i, j; \Theta) = 0, \qquad j = 1 \dots J-1, \quad l = 1, \dots, n_b \tag{33}$$

   To find $\Theta$, I use a quasi-Newton method; for details, see Section 3.4.4.

9. If $k < n_H$, increment $k$ by one and go to Step 5.

At the final parameter vector $\theta = \theta_{n_H}$, all of the following should be satisfied: First, $\lambda_{n_H} = 1$. Second, $|\Theta_{n_H} - \Theta_{n_H-1}|_\infty$ should be smaller than a suitable convergence criterion, which I have chosen to be $10^{-6}$. Third, $\kappa_{n_H}$ should be zero or very close to zero, such that reducing $\kappa$ does not significantly change the portfolio allocation. I have used $\kappa_{n_H} \leq 10^{-7}$. For further discussion of $\kappa$, see Section 3.3.

## 3.3 Defining the residual

For a given vector of parameters $\Theta$, at any given state $X = (W_2, \dots, W_J, z, \sigma)$, we have to define the vector of residuals $Res(X, j; \Theta)$, which we have introduced in Section 3.1. For each variable to be approximated by polynomials, we need a residual equation that should be set to zero by appropriate choice of $\Theta$. For consumption, we take the Euler residual of capital holdings:

$$Res(X, j; \Theta) = \frac{\beta \sum_{\xi=1}^{n_\epsilon} \pi_\xi \left[ (r^{K,\xi} + \tilde{Q}^\xi) U_c \left( \mathcal{P} \left( X^\xi; \theta_{j+1} \right) \right) \right]}{U_c \left( \mathcal{P} \left( X; \theta_j \right) \right)} - Q, \qquad j = 1, \dots, J-1 \tag{34}$$

For the price of the safe asset, we choose the corresponding market clearing condition:

$$Res(X, J; \Theta) = \sum_{j=1}^{J-1} A_j^S(X) \tag{35}$$

As explained in the introduction, we have to solve an equilibrium problem at each point of the state space, in order to compute these residuals. In particular, we have to find the $2J+n_\epsilon$ variables $Q^C$, $R_K$, $R_K^\xi$ for $\xi = 1, \ldots, n_\epsilon$, and $A_j^C$, $A_j^S$ for $j = 1, \ldots, J-1$, such that

$$\frac{1}{J}\sum_{j=1}^{J-1} A_{j,t}^C = 0 \tag{36}$$

$$R_K = \tilde{Q}\Psi_K(K_{-1}, I) + r^K \tag{37}$$

$$R_K^\xi = \tilde{Q}^\xi \Psi_K(K, I^\xi) + r^{K,\xi}, \qquad \xi = 1, \ldots, n_\epsilon \tag{38}$$

$$\frac{\beta \sum_{\xi=1}^{n_\epsilon} \pi_\xi \left[U_c\left(\mathcal{P}\left(X^\xi; \theta_{j+1}\right)\right)\right]}{U_c\left(\mathcal{P}\left(X; \theta_j\right)\right)} = \left(1 - \kappa A_{j,t}^S\right) Q_t^S, \qquad j = 1, \ldots, J-1 \tag{39}$$

$$\frac{\beta \sum_{\xi=1}^{n_\epsilon} \pi_\xi \left[(1 + (g_\xi - \bar{g})^2)\, U_c\left(\mathcal{P}\left(X^\xi; \theta_{j+1}\right)\right)\right]}{U_c\left(\mathcal{P}\left(X; \theta_j\right)\right)} = \left(1 - \kappa A_{j,t}^S\right) Q_t^C, \qquad j = 1, \ldots, J-1 \tag{40}$$

where we define

$$K_{-1} \equiv \frac{1}{R_K}\sum_{j=2}^{J} W_j\left(X\right)$$

$$w \equiv z F_L(K_{-1}, L)$$
$$r^K \equiv z F_K(K_{-1}, L)$$

$$I \equiv F(K_{-1}, L) - \sum_{j=1}^{J-1} \mathcal{P}\left(X; \theta_j\right) - \left(W_J + w\zeta_J\right)$$

$$Q \equiv \frac{1}{\Psi_I(K_{-1}, I)}$$
$$\tilde{Q} \equiv \frac{\Psi_K(K_{-1}, I)}{\Psi_I(K_{-1}, I)}$$

$$k_j \equiv \left(W_j + w\zeta_j - \mathcal{P}\left(X; \theta_j\right) - \mathcal{P}\left(X; \theta_J\right) A_j^S - Q^C A_j^C\right)/Q, \qquad j = 1, \ldots, J-1$$

$$W_j^\xi \equiv R_K^\xi k_{j-1} + A_{j-1}^S + \left(1 + (g - \bar{g})^2\right) A_{j-1}^C, \qquad j = 1, \ldots, J-1, \quad \xi = 1, \ldots, n_\epsilon$$

$$K \equiv \sum_{j=1}^{J-1} k_j$$

and the following equations all hold for $\xi = 1, \ldots, n_\epsilon$:

$$w^\xi \equiv z^\xi F_L(K,L)$$
$$r^{K,\xi} \equiv z^\xi F_K(K,L)$$
$$I^\xi \equiv F(K,L) - \sum_{j=1}^{J-1} \mathcal{P}\left(X^\xi;\theta_j\right) - (W_J^\xi + w^\xi \zeta_J)$$
$$Q^\xi \equiv \frac{1}{\Psi_I(K,I^\xi)}$$
$$\tilde{Q}^\xi \equiv \frac{\Psi_K(K,I^\xi)}{\Psi_I(K,I^\xi)}$$

This completes the definition of the residual. To solve this equilibrium problem, I use Powell's method, as described in Nocedal and Wright (2006, Chapter 11). The Jacobian is computed using automatic differentiation, cf. Section 3.4.2.

Notice that (39) and (40) each contain a term in the parameter $\kappa$, which can be interpreted as an asset holding cost parameter, creating a wedge in the Euler equation, such that the asset holding choice is determined even in the limit of no aggregate shocks. This makes the problems numerically well-behaved in intermediate steps of the solution; in the end, we will have $\kappa \leq 10^{-7}$, so that there is at most a tiny distortion of asset holdings. To illustrate the effect of asset holding costs, write (39) as

$$U_c(c_{i,t})(1 - \kappa A_{i,t}^S)Q_t^S = \beta \, \mathrm{E}_t \left[U_c(c_{i+1,t+1})\right] \tag{41}$$

Dividing (16) by $U_c(c_{i,t})Q_t$, (41) by $U_c(c_{i,t})Q_t^S$, and subtracting we get

$$\kappa A_{i,t}^S Q_t^S = \beta \, \mathrm{E}_t \left[ \frac{U_c(c_{i+1,t+1})}{U_c(c_{i,t})} \left( \frac{r_{t+1}^K + \tilde{Q}_{t+1}}{Q_t} - \frac{1}{Q_t^S} \right) \right] \tag{42}$$

We see that $\kappa A_{i,t}^S Q_t^S$ is like a wedge on the risk premium for capital, the rhs of (42). By choosing $\kappa \leq 10^{-7}$, we make sure that this wedge is several orders of magnitude smaller than the risk premium in the model solution.

## 3.4 Details

### 3.4.1 The choice of a discrete grid

Global methods require the researcher to fix a region of the state space as the domain of the approximating functions (cf. Step 4 in Section 3.1). On the one hand, this region

should be as small as possible, since the approximating functions have a limited number of degrees of freedom. On the other hand, the approximation only makes sense if the dynamic system spends most of the time in this region. Since it is not known a priori in which part of the state space the economy lives, one has to resort to simulation techniques, to find this out iteratively.

Two methods to choose grid points have been prominent in the recent literature. The first ones are sparse grid methods (Malin, Krueger, and Kubler 2007; Judd, Maliar, Maliar, and Valero 2014). The algorithm presented below is compatible with sparse grid techniques. Whether these are useful in the present context is left for future research. The second method uses $\epsilon$-distinguishable sets obtained from simulations (Judd, Maliar, and Maliar 2012). I have used those as well, but I found the method that I present below to be slightly superior in my numerical experiments. This might differ from one application to the next. I do not follow this up here, because the choice of grid is not the main focus of this paper.

The method that I use has the following steps:

1. Choose a set of points $\hat{X}_i$, $i = 1, \ldots, n_g$ which fill the $d$-dimensional unit cube approximately equally. For the results reported below, I simply use a set of uniformly distributed random numbers, with 3 times as many grid points as there are basis functions. Of those, half are uniformly distributed on $[-1, 1]^{n_x}$, and half are uniformly distributed on $[-0.5, 0.5]^{n_x}$. There are more sophisticated alternatives to this simple approach.[5] I have tried several alternatives and found no significant improvement over the simple approach.

2. Transform the grid points $\hat{X}_i$ into grid points $X_i^\circ$, $i = 1, \ldots, n_g$, which fill the $d$-dimensional unit ball rather than the unit cube. This is done in the following way.

   (a) Find a sequence $\bar{X}_i$, $i = 1, \ldots, n_g$ that fills the $d$-dimensional hypercube approximately equally (for details, see Section 3.4.1).

   (b) Denote by $F_n^{-1}(x) : [0, 1] \to [-\infty, \infty]$ the inverse of the cumulative normal distribution function. Then define a new set of points $Y_i$ by transforming each

[5] For example, one can use low-discrepancy sequences, such as Sobol or Faure points. Although these sequences are designed to fill the hypercube approximately equally, practical experience shows that this is still done very imperfectly, in the sense that there are still many points that are close to each other. One can improve on this by eliminating points that are too close, using the idea of $\epsilon$-distinguishable sets as in Judd et al. (2012).

element of each $\hat{X}_i$:

$$Y_{i,j} = F_n^{-1}(\hat{X}_{i,j}), \qquad i = 1, \ldots, n_g, j = 1, \ldots, d \tag{43}$$

(c) Set $X_i^\circ = \frac{1}{\|Y_i\|_2} \left[F_{\chi^2;d}\left(\|Y_i\|_2\right)\right]^{1/d}$ where $F_{\chi^2;d}(.)$ denotes the $\chi^2$ distribution function with $d$ degrees of freedom, and $\|Y_i\|_2$ denotes the sum of squares of the vector $Y_i$.

3. Finally, we choose an augmentation factor $\psi$ and a matrix $\Gamma_k$ to transform the $X_i^\circ$ into the grid points $\bar{X}_i$ (Step 6 in Section 3.2). $\Gamma_k$ should be such that $diag(\Gamma_k \Gamma_k') = diag(\Sigma_{X,k})$. I have chosen $\Gamma_k$ as the diagonal matrix with the square root of $diag(\Sigma_{X,k})$ as entries. This means, I stretch the domain to account for the standard deviation of states, but do not rotate it.

   I have also used more general $\Gamma_k$ which rotate the domain so as to account for the correlations between states. Since correlations are very high, this yields a narrow stretch, and turned out to be unstable. I have also tried intermediate cases; they sometimes gave higher accuracy, but not consistently so, and therefore I decided to stick with the simpler choice of a diagonal matrix $\Gamma_k$.

   I have chosen $\psi = 6$. This means that the domain covers about $\pm 6$ estimated standard deviations of the state variables. This may appear rather large, but in my applications this has given higher accuracy (measured by the Euler residuals at the simulated points) than an approximation on a narrower grid.

### 3.4.2 The use of automatic differentiation

"Automatic differentiation" means the numerical computation of derivatives for a given value of the independent variables, by using the exact differentiation rules rather than finite-difference approximations. A simple form of automatic differentiation (called the forward mode) can be easily implemented in object-oriented programming languages such as C++. The forward mode, combined with the technique described in Appendix A.1, is sufficient for our purpose. For a general description of automatic differentiation, see Griewank and Walther (2008).

To see how automatic differentiation is useful in our setup, write the problem in general terms as searching for a parameter vector $\Theta$ and a set of endogenous variables $w$ so as to

solve the following nonlinear system of equations:

$$R(y(\Theta), w) = 0 \tag{44}$$

$$H(y(\Theta), w) = 0 \tag{45}$$

In the outer loop (Section 3.2), we use an iterative method to find the $\Theta$ that satisfies (44). For any guess $\Theta$ in the iterative process, we solve an inner loop problem (Section 3.3) at each grid point, where we choose $w$ so as to satisfy (45).

Automatic differentiation is useful in two ways. First, to solve the inner-loop problem efficiently by quasi-Newton methods based on the exact Jacobian. The second way is more interesting. If we solve for $\Theta$ in the outer loop by a quasi-Newton method, we need to compute the derivative of $R$ with respect to $\Theta$. Applying the implicit function theorem on (45), we get

$$\begin{aligned} \frac{\partial R}{\partial \Theta} &= R_y \frac{\partial y}{\partial \Theta} + R_w \frac{\partial w}{\partial y} \frac{\partial y}{\partial \Theta} \\ &= \left( R_y - R_w H_w^{-1} H_y \right) \frac{\partial y}{\partial \Theta} \end{aligned} \tag{46}$$

The important lesson from (46) is that, once we have found an $w$ that satisfies (45), no further search over $w$ is necessary to compute the Jacobian $\frac{\partial R}{\partial \Theta}$. If solving for $w$ is a non-trivial task, as it is in our model, computing the whole Jacobian of $R$ is not much more costly, and may even be cheaper than evaluating $R$. This gives a big advantage to Newton-type methods, which exploit the Jacobian, over alternative methods such as fixed point iteration.

### 3.4.3 Time iteration vs. quasi-Newton algorithms

There are several ways to solve the fixed point problem for the parameter vector $\Theta$ in the outer loop of the algorithm. I have stressed the use of automatic differentiation, to make quasi-Newton algorithms competitive. However, in a model with few cohorts, the discount factor $\beta$ of each cohort is so low that time-iteration methods converge rather quickly. The trade-off between time iteration vs. quasi-Newton algorithms depends mainly on two parameters:

1. The number of cohorts: more cohorts means higher frequency, higher $\beta$, and more time steps for time iteration to converge.

2. The approximation order: higher order implies higher dimension of the parameter vector $\Theta$. Quasi-Newton methods involve the solution of a dense linear equation system of the same dimension as $\Theta$. The computational complexity of this problem is cubic in the dimension.

I have found time iteration to be only competitive in the model with 6 cohorts. With 12 cohorts or more, quasi-Newton methods turned out to be clearly superior.

In terms of reliability, I found no difference between the two methods. If one of them converges, so does the other one. Non-convergence of either algorithm seems to indicate a problem with the equation system: there is no guarantee that the nonlinear fixed point problem has a solution. And if a solution exists, but is too far away from the starting point, then it may be impossible to find it.

### 3.4.4 Implementation of the quasi-Newton algorithm

In the outer loop of the algorithm (Step 8 in Section 3.2), we use a quasi-Newton method to find the parameter vector $\Theta$. For the results below, I have used Broyden's algorithm. For a general description, see Press et al. (1986, Section 9.7). The key elements of the algorithm are the following.

1. At the initial guess $x_0$, the equation system is linearized, either by finite differencing, or in our case, by automatic differentiation. Denote the Jacobian by $J_0$.

2. After each iteration, having found a new point $x_k$, the Jacobian is changed by a rank-one update: $J_k = J_{k-1} + u_k v_k'$ where $u_k$ and $v_k$ are column vectors.

3. In iteration $k$, the search direction is determined by the Newton step

$$dx_k = -J_k^{-1} F(x_k) \tag{47}$$

4. If the search direction computed by $J_k$ does not lead to an improvement, the Jacobian is computed again by automatic differentiation.

Since we deal with a very big equation system (close to 7000 parameters in one case), solving the linear system in (47) is computationally expensive, and it becomes important to use an efficient method. The following approach has turned out to work very well. First, we compute an LU-decomposition of $J_0$ (or any $J_k$ where the Jacobian has been

newly computed, not just updated). In the steps where the Jacobian gets updated, we exploit that $J_k = J_m + UV'$ where $U = [u_{m+1}, u_{m+2}, \ldots, u_k]$ and $V = [v_{m+1}, v_{m+2}, \ldots, v_k]$, if $J_k$ was the last computation of the Jacobian $J$. Then we can use the Sherman-Morrison-Woodbury formula (Higham 2002, p. 258)

$$(A + UCV')^{-1} = A^{-1} - A^{-1}U\left(C^{-1} + V'A^{-1}U\right)^{-1}V'A^{-1} \tag{48}$$

and the LU-decomposition of $J_m$ to solve (47) very fast. Usually it turns out that the LU decomposition has to be computed only once for the solution of the model with a given degree of approximation.

### 3.5 Short-sale constraints

Introducing short-sale constraints as described in Section 2.6 requires first some changes in the formulas. The market clearing condition (35) is replaced by

$$Res(X, J; \Theta) = \sum_{j=1}^{J-1} A_j^S - D^G \tag{49}$$

and the household budget constraint (13) is replaced by

$$k_j = (W_j + w\zeta_j + D^G(Q^S - 1) - \mathcal{P}(X; \theta_j) - \mathcal{P}(X; \theta_J) A_j^S - Q^C A_j^C)/Q \tag{50}$$

More importantly, the Euler equations (39) and (40) must be replaced by inequalities. Occasionally binding inequality constraints mean that standard nonlinear root finders cannot be applied. Therefore, in solving for the residuals as described in Section 3.3, I adapt Powell's method such that the search directions are chosen by Lemke's algorithm (Murty 1988). Simply speaking, we solve the problem iteratively by piecewise linearization, and treat each linearized problem by a standard method (Lemke's algorithm) for linear complementarity problems. It turns out that this procedure slows down the computation compared to the case without inequality constraints (cf. Section 4.6), but is very reliable.

## 4 Numerical Results

### 4.1 Parameter values

Since the focus of this paper is on the computational method, I do not aim for the most realistic calibration of the model, in particular not for a realistic size of the shocks. Shock

variances are chosen to be rather high, to demonstrate the ability of the method to handle large shocks. All shocks are approximated by finite distributions. The i.i.d. shock to technology growth, $\epsilon_g$, takes the values $(-\Phi, -0.05, 0, 0.05 + \Phi/3)$ with probabilities $(0.1, 0.3, 0.3, 0.3)$, respectively. The volatility shock $\epsilon_s$ takes the values $(-\Psi, 0, \Psi)$ with probabilities $(0.3, 0.4, 0.3)$, respectively. I consider the following parameter constellations. In most cases, $\Psi$ is set to 0.5. Then I let $\Phi$ vary between 0.05 (no crisis state) and 0.12. In the latter case, the maximum negative shock to growth, when volatility is high, equals $-1.5 \times 0.12 = -0.18$. This means an 18 percent permanent reduction in the level of technology.

The other parameters are mostly set to standard values from the literature. The length of the working plus the retirement life is thought to be 60 years. The length of one model period is then given by $n_y = 60/J$ years, where $J$ denotes the number of cohorts. I choose as time discount factor $\beta = 0.98^{n_y}$. Household labor endowment increases linearly from 0.5 in the first period to 1 in the last period before retirement. The risk aversion parameter in (1) is varied between $\gamma = 2$ and $\gamma = 6$.

On the production side, I use a Cobb-Douglas production function

$$F(k, l) = k^{\alpha} l^{(1-\alpha)} \tag{51}$$

with the output share of capital set to $\alpha = 0.4$. The depreciation rate for capital is set to $\delta = 1 - 0.9^{n_y}$. For the capital accumulation function, I use

$$\Psi(K, I) = (1 - \delta)K + I - \frac{1}{\eta_I} \frac{(I - \delta K)^2}{K} \tag{52}$$

with $\eta_I = 3$.

## 4.2 Accuracy

Table 1 reports results for the model with 6 cohorts, 2 assets and large shocks ($\Phi = 0.12$). The first part of the table shows accuracy results, for different degrees of approximation order. Notice that 6 functions need to be approximated: the consumption functions of cohort 1–5 (the last cohort simply consumes all its wealth), as well as the price of the safe asset. I approximate them as complete polynomials of the 7 state variables (the exogenous states $g_t, \sigma_t$ and market wealth of cohorts 1–5). In the table, degree 1 means linear approximation, and results are shown up to degree 4. The parameters are chosen so as to satisfy 6

equations: the Euler equations for cohorts 1–5, as well as the market clearing condition for the safe asset. As explained in Section 3, all the other variables, in particular asset choices, are not approximated as functions of the state, but their equilibrium values are computed at each point in the aggregate grid. Their first order condition is therefore satisfied almost exactly, and not reported here. The results indicate that a linear approximation does not yet give a very precise approximation: both mean and max absolute errors are in the range of $10^{-3}$, rather uniformly for all the residuals. Notice that Euler residuals in Euq. (34) are written as fractions of marginal utilities, so they should be interpreted as relative errors. Increasing the degree of approximation to 4 reduces the mean absolute errors to about $10^{-7}$ and the max absolute errors to about $10^{-6}$. Do these differences in accuracy matter for the economic results? The second part of the table reports mean values, standard deviations and correlation with outpout for 7 different variables: the holdings of the safe asset for cohorts 1–5, the excess return of capital relative to the safe asset, and the risk premium, defined as the expected value of the excess return. Both risk premium and excess return are in annualized terms. I will give a precise definition of these terms, and discuss in more detail the quantitative significance, in Section 4.3. For the moment, the focus is on how precisely these statistics are computed by approximations of varying degree.

We see that asset holdings are difficult to pin down accurately: the results for the linear approximation are quite different from the more precise results that are obtained with degrees 3 or 4. This is true for mean values, standard deviations and also the correlation with output. Perhaps surprisingly, both the mean and the standard deviation of the risk premium, as well as its correlation with output, are captured well by the linear approximation. This shows that a *global* linear approximation is very different from the *local* linear approximations that arise from linearization about the deterministic steady state. Local linearizations give certainty equivalence results, and local second order perturbations give a local approximation to the mean, but not the variance of the risk premium. To capture the second moments of the risk premium, one needs at least a third-order perturbation around the steady state.[6] It is worth noting that degree 3 and 4 lead to practically identical results. This is important, because degree 4 cannot be easily achieved for the case of a larger number of cohorts.

Table 2 provides the same information for the model with 3 assets. Asset holdings are now reported for the third asset, which is thought to protect against the crisis state.

[6]For a perturbation analysis of a portfolio choice model, see Judd and Guu (2001).

| Degree | **Mean absolute error** | | | | | | |
|---|---|---|---|---|---|---|---|
| | $Euler_1$ | $Euler_2$ | $Euler_3$ | $Euler_4$ | $Euler_5$ | $\sum_i A_i^S$ | |
| 1 | 9.77e-4 | 1.05e-3 | 1.07e-3 | 1.31e-3 | 1.03e-3 | 1.77e-3 | |
| 2 | 3.48e-5 | 3.66e-5 | 4.42e-5 | 2.67e-5 | 3.03e-5 | 1.26e-4 | |
| 3 | 7.46e-6 | 7.90e-6 | 7.96e-6 | 7.61e-6 | 3.27e-6 | 3.78e-5 | |
| 4 | 2.12e-7 | 2.25e-7 | 2.22e-7 | 2.12e-7 | 1.05e-7 | 1.23e-6 | |
| | **Max absolute error** | | | | | | |
| 1 | 2.84e-3 | 3.34e-3 | 3.23e-3 | 3.63e-3 | 2.83e-3 | 5.98e-3 | |
| 2 | 3.26e-4 | 3.80e-4 | 4.24e-4 | 3.08e-4 | 2.86e-4 | 8.59e-4 | |
| 3 | 4.84e-5 | 5.94e-5 | 6.89e-5 | 6.78e-5 | 2.74e-5 | 2.88e-4 | |
| 4 | 2.62e-6 | 3.80e-6 | 4.88e-6 | 7.43e-6 | 3.96e-6 | 2.50e-5 | |
| | $A_1^S$ | $A_2^S$ | $A_3^S$ | $A_4^S$ | $A_5^S$ | $RiskPr$ | $ExcRet$ |
| | **100× Means** | | | | | | |
| 1 | -0.456 | 3.509 | 4.680 | -2.469 | -5.265 | 0.098 | 0.084 |
| 2 | -0.633 | 3.640 | 4.816 | -2.416 | -5.407 | 0.099 | 0.100 |
| 3 | -0.659 | 3.614 | 4.816 | -2.381 | -5.389 | 0.099 | 0.100 |
| 4 | -0.659 | 3.612 | 4.814 | -2.380 | -5.387 | 0.099 | 0.100 |
| | **100× Standard deviations** | | | | | | |
| 1 | 0.040 | 0.091 | 0.120 | 0.089 | 0.144 | 0.068 | 0.540 |
| 2 | 0.112 | 0.244 | 0.165 | 0.256 | 0.263 | 0.068 | 0.547 |
| 3 | 0.121 | 0.248 | 0.162 | 0.250 | 0.278 | 0.068 | 0.547 |
| 4 | 0.122 | 0.251 | 0.163 | 0.252 | 0.280 | 0.068 | 0.547 |
| | **Correlation with output** | | | | | | |
| 1 | -0.952 | -0.473 | -0.230 | 0.492 | 0.447 | -0.006 | -0.601 |
| 2 | 0.945 | 0.946 | 0.903 | -0.951 | -0.923 | -0.012 | -0.601 |
| 3 | 0.946 | 0.954 | 0.907 | -0.954 | -0.935 | -0.012 | -0.601 |
| 4 | 0.948 | 0.954 | 0.907 | -0.954 | -0.935 | -0.012 | -0.601 |

Table 1: Model with 6 cohorts and 2 assets, $\Phi = 0.12$

The risk premium and excess return still refer to the excess return of capital. The table confirms the earlier results. For given order of approximation, the solution has about the same accuracy as in the model with 2 assets. Again, the linear approximation does not accurately pin down asset holdings. Perhaps surprisingly, introduction of the third asset has no significant effect on the risk premium. This reflects the fact that the safe asset already provides most of the insurance that can be achieved by asset trade, so that the third asset does not change the allocation much. We will see this in more detail in Section 4.5.

Table 3 provides summary results for the model with two assets and large shocks, with a widely varying number of cohorts. The information provided are the mean and max (over all the residuals reported in the earlier tables) of the errors, the mean and standard deviation of the risk premium, as well as the correlation of risk premium and excess return with output. In addition, a measure of the price of risk, namely the Sharpe ratio, is reported (cf. Section 4.3 for a definition). I have computed the model for up to 60 cohorts, such that the model period can be interpreted as one year. The larger the number of cohorts, the smaller is the degree of approximation that is feasible on a personal computer. For 12 cohorts, I have computed the solution up to degree 3. This involves solving simultaneously for 6720 parameters, the fourth-degree approximation would require 28560 parameters, which needs more memory than what a personal computer currently has. For the model with 20 cohorts I compute a quadratic approximation. The model with 60 cohorts I can only compute with a linear approximation. We see that, for a given degree of approximation, the accuracy goes down slightly in the number of cohorts, but for $J = 60$ errors are still of the same order of magnitude as for $J = 6$. Furthermore, the equity premium and the price of risk increase in the number of cohorts. This is because, with more cohorts, the same shock in growth happens with higher frequency. What is perhaps most important, we can confirm the result that the global linear approximation gives good approximations to aggregate statistics, including financial statistics.

Table 4 provides some more checks on these results. The first part of the table refers to the model with constant volatility. This is interesting, because in the benchmark calibration, the variation in the risk premium is largely driven by changes in volatility, while growth is driven by the technology shock. Therefore, the risk premium had very little correlation with output growth. With constant volatility, the risk premium is almost perfectly negatively correlated with output growth. It is still true that the linear approximation

| Degree | **Mean absolute error** | | | | | | |
|---|---|---|---|---|---|---|---|
| | $Euler_1$ | $Euler_2$ | $Euler_3$ | $Euler_4$ | $Euler_5$ | $\sum_i A_i^C$ | |
| 1 | 9.77e-4 | 1.05e-3 | 1.07e-3 | 1.31e-3 | 1.03e-3 | 1.77e-3 | |
| 2 | 1.22e-4 | 1.31e-4 | 1.27e-4 | 1.74e-4 | 9.71e-5 | 1.40e-4 | |
| 3 | 7.46e-6 | 7.90e-6 | 7.96e-6 | 7.61e-6 | 3.27e-6 | 3.78e-5 | |
| 4 | 2.12e-7 | 2.26e-7 | 2.22e-7 | 2.12e-7 | 1.05e-7 | 1.23e-6 | |
| | **Max absolute error** | | | | | | |
| 1 | 2.84e-3 | 3.34e-3 | 3.23e-3 | 3.63e-3 | 2.83e-3 | 5.99e-3 | |
| 2 | 1.48e-3 | 1.71e-3 | 1.71e-3 | 2.46e-3 | 1.25e-3 | 6.76e-4 | |
| 3 | 4.83e-5 | 5.92e-5 | 6.88e-5 | 6.78e-5 | 2.73e-5 | 2.87e-4 | |
| 4 | 2.61e-6 | 3.83e-6 | 4.88e-6 | 7.45e-6 | 3.95e-6 | 2.52e-5 | |
| | $A_1^C$ | $A_2^C$ | $A_3^C$ | $A_4^C$ | $A_5^C$ | $RiskPr$ | $ExcRet$ |
| | **100× Means** | | | | | | |
| 1 | -0.580 | 4.205 | 5.693 | -2.952 | -6.367 | 0.098 | 0.084 |
| 2 | -1.080 | 3.493 | 4.812 | -1.977 | -5.247 | 0.099 | 0.099 |
| 3 | -1.077 | 3.647 | 5.331 | -2.178 | -5.723 | 0.099 | 0.100 |
| 4 | -1.080 | 3.640 | 5.326 | -2.172 | -5.715 | 0.099 | 0.100 |
| | **100× Standard deviations** | | | | | | |
| 1 | 0.049 | 0.231 | 0.330 | 0.220 | 0.371 | 0.068 | 0.540 |
| 2 | 0.138 | 0.333 | 0.354 | 0.341 | 0.478 | 0.068 | 0.547 |
| 3 | 0.097 | 0.296 | 0.262 | 0.278 | 0.369 | 0.068 | 0.547 |
| 4 | 0.105 | 0.304 | 0.266 | 0.288 | 0.379 | 0.068 | 0.547 |
| | **Correlation with output** | | | | | | |
| 1 | -0.859 | -0.239 | -0.108 | 0.234 | 0.220 | -0.006 | -0.600 |
| 2 | 0.948 | 0.841 | 0.806 | -0.873 | -0.833 | -0.012 | -0.600 |
| 3 | 0.916 | 0.836 | 0.691 | -0.843 | -0.766 | -0.012 | -0.601 |
| 4 | 0.925 | 0.845 | 0.704 | -0.854 | -0.779 | -0.012 | -0.601 |

Table 2: Model with 6 cohorts and 3 assets, $\Phi = 0.12$

| Degree | Error | | RiskPr | | Sharpe | Corr. with Y | |
|---|---|---|---|---|---|---|---|
| | Mean | Max | Mean | StdDev | | RiskPr | ExcRet |
| 2 assets, 6 cohorts, $\gamma = 4$, $\Phi = 12$ | | | | | | | |
| 1 | 1.20e-3 | 5.98e-3 | 0.098 | 0.0677 | 0.058 | -0.0061 | -0.6005 |
| 2 | 4.97e-5 | 8.59e-4 | 0.099 | 0.0679 | 0.057 | -0.0119 | -0.6012 |
| 3 | 1.20e-5 | 2.88e-4 | 0.099 | 0.0679 | 0.057 | -0.0122 | -0.6013 |
| 4 | 3.68e-7 | 2.50e-5 | 0.099 | 0.0679 | 0.057 | -0.0123 | -0.6013 |
| 2 assets, 12 cohorts, $\gamma = 4$, $\Phi = 12$ | | | | | | | |
| 1 | 9.51e-4 | 1.38e-2 | 0.129 | 0.0930 | 0.085 | -0.0271 | -0.5375 |
| 2 | 7.24e-5 | 2.63e-3 | 0.129 | 0.0932 | 0.086 | -0.0244 | -0.5354 |
| 3 | 1.23e-5 | 7.05e-4 | 0.129 | 0.0932 | 0.086 | -0.0251 | -0.5369 |
| 2 assets, 20 cohorts, $\gamma = 4$, $\Phi = 12$ | | | | | | | |
| 1 | 1.17e-3 | 2.42e-2 | 0.146 | 0.1150 | 0.114 | -0.0449 | -0.4715 |
| 2 | 1.28e-4 | 6.27e-3 | 0.146 | 0.1150 | 0.114 | -0.0349 | -0.4612 |
| 2 assets, 60 cohorts, $\gamma = 4$, $\Phi = 12$ | | | | | | | |
| 1 | 2.08e-3 | 3.24e-2 | 0.179 | 0.1913 | 0.197 | -0.1044 | -0.3273 |

Table 3: Different cohort sizes

| Degree | Error | | RiskPr | | Sharpe | Corr. with Y | |
|---|---|---|---|---|---|---|---|
| | Mean | Max | Mean | StdDev | | RiskPr | ExcRet |
| 6 cohorts, constant shock variance | | | | | | | |
| 1 | 3.75e-4 | 3.22e-3 | 0.095 | 0.0031 | 0.052 | -0.9988 | -0.5913 |
| 2 | 7.91e-6 | 2.37e-4 | 0.095 | 0.0032 | 0.052 | -0.9982 | -0.5906 |
| 3 | 3.43e-7 | 1.70e-5 | 0.095 | 0.0032 | 0.052 | -0.9982 | -0.5907 |
| 4 | 2.60e-8 | 1.79e-6 | 0.095 | 0.0032 | 0.052 | -0.9982 | -0.5907 |
| 12 cohorts, $\rho_g = 0.2$ | | | | | | | |
| 1 | 1.52e-3 | 2.22e-2 | 0.128 | 0.0924 | 0.095 | -0.0143 | -0.4213 |
| 2 | 1.22e-4 | 5.87e-3 | 0.128 | 0.0926 | 0.097 | -0.0252 | -0.4070 |
| 3 | 1.39e-5 | 1.48e-3 | 0.128 | 0.0927 | 0.096 | -0.0249 | -0.4077 |
| 12 cohorts, $\rho_g = -0.5$ | | | | | | | |
| 1 | 7.63e-4 | 1.57e-2 | 0.127 | 0.0922 | 0.068 | -0.0568 | -0.7681 |
| 2 | 9.00e-5 | 3.44e-3 | 0.127 | 0.0923 | 0.068 | -0.0515 | -0.7696 |
| 3 | 9.08e-6 | 4.26e-4 | 0.127 | 0.0922 | 0.068 | -0.0524 | -0.7706 |

Table 4: Various results, $\gamma = 4$, $\Phi = 12$

gives a very good approximation for these statistics. Positive or negative correlation in technology growth, for which results are reported in the second and third part of the table, substantially affect the correlation of excess returns with output growth. Still, the linear approximation gets it approximately right in all cases.

## 4.3 The price of risk

Table 5 provides summary information for various degrees of risk aversion and shock size. Our focus here is on the price of risk. We measure the risk premium as the expected value of the excess return of capital versus the riskless bond, expressed in annualized percentage points. The excess return is defined as

$$ExcRet_{t+1} \equiv \frac{100}{n_y} \cdot \left( \mathrm{E}_t \frac{r^K_{t+1} + \tilde{Q}_{t+1}}{Q_t} - \frac{1}{Q^S_t} \right) \tag{53}$$

The risk premium is the expected excess return:

$$RiskPr_t \equiv \mathrm{E}_t \, ExcRet_{t+1} \tag{54}$$

| Degree | Error | | RiskPr | | Sharpe | Corr. with Y | |
|---|---|---|---|---|---|---|---|
| | Mean | Max | Mean | StdDev | | RiskPr | ExcRet |
| | **2 assets, 12 cohorts,** $\gamma = 2$**,** $\Phi = 12$**:** | | | | | | |
| 3 | 3.34e-6 | 2.19e-4 | 0.066 | 0.0475 | 0.039 | -0.0242 | -0.6427 |
| | **2 assets, 12 cohorts,** $\gamma = 4$**,** $\Phi = 12$**:** | | | | | | |
| 3 | 1.23e-5 | 7.05e-4 | 0.129 | 0.0932 | 0.086 | -0.0251 | -0.5369 |
| | **2 assets, 12 cohorts,** $\gamma = 6$**,** $\Phi = 12$**:** | | | | | | |
| 3 | 2.25e-5 | 8.41e-4 | 0.182 | 0.1315 | 0.134 | -0.0224 | -0.4767 |
| | **2 assets, 12 cohorts,** $\gamma = 4$**,** $\Phi = 8$**:** | | | | | | |
| 3 | 1.34e-5 | 1.34e-4 | 0.085 | 0.0605 | 0.068 | -0.0175 | -0.5388 |
| | **3 assets, 12 cohorts,** $\gamma = 4$**,** $\Phi = 12$**:** | | | | | | |
| 3 | 1.22e-5 | 6.85e-4 | 0.129 | 0.0932 | 0.086 | -0.0251 | -0.5370 |

Table 5: Different risk aversion

For the Sharpe ratio, I report the ratio of the unconditional mean and unconditional standard deviation of the excess return:

$$Sharpe \equiv \frac{1}{\sqrt{n_y}} \cdot \frac{\mathrm{E}\left[ExcRet_t\right]}{\sqrt{\mathrm{Var}\left[ExcRet_t\right]}} \tag{55}$$

It is annualized by dividing by the square root of the time period, since the risk premium increases linearly in the time period, and the standard deviation only with the square root of time. As expected, we see that the risk premium increases in the degree of risk aversion. With $\gamma = 6$, it is 0.182 annualized percentage points. While not trivial this is still an order of magnitude below what is observed in the data for stocks. This is mainly due to the fact that the price of capital varies much less than the price of stocks in the data. Looking at the price of risk, as measured by the annualized Sharpe ratio, the value reaches 0.134. This is about half the observed value for the US stock market, which is roughly 0.25. This can be considered a success, given that a risk aversion parameter of $\gamma = 6$ is rather moderate by the standards of the finance literature on the equity premium (see for example Bansal and Yaron (2004)). The Sharpe ratio increases somewhat more than linearly in the risk aversion parameter $\gamma$.

| Degree | Error | | RiskPr | | Sharpe | Corr. with Y | |
|---|---|---|---|---|---|---|---|
| | Mean | Max | Mean | StdDev | | RiskPr | ExcRet |
| 12 cohorts, short-sale constraints | | | | | | | |
| 1 | 9.33e-4 | 1.26e-2 | 0.146 | 0.1055 | 0.092 | -0.0250 | -0.5606 |
| 2 | 7.35e-5 | 1.86e-3 | 0.146 | 0.1054 | 0.092 | -0.0139 | -0.5587 |
| 3 | 1.17e-5 | 6.39e-4 | 0.146 | 0.1053 | 0.092 | -0.0159 | -0.5599 |

Table 6: Model with short-sale constraints

## 4.4 Results with short-sale constraints

Table 6 provides summary information on the model with 12 cohorts, two assets, and a lower bound on asset holdings. The level of government debt was chosen to be 0.1, which means that the outstanding sum of safe assets is only about 3 percent of total wealth. Comparing Table 6 with the corresponding part of Table 3, we see that the degree of accuracy is very similar. The risk premium and the Sharpe ratio have increased, but not dramatically. Figure 1 displays the minimum and the maximum (over a long simulation) of safe asset holdings, both for the linear approximation (dashed line) and the cubic approximation (solid line). We see that the constraint is binding always for the young and the old cohorts, never binding for cohorts 4 and 5, and occasionally binding for cohorts 3 and 6. Even with short-selling constraints, the linear approximation does not give accurate results for asset holdings.

## 4.5 Intergenerational Insurance

Recently, Hasanhodzic and Kotlikoff (2013) have argued that ingerentional insurance is not an issue of primary importance. I briefly look at this issue in the context of my model. My results do not contradict Hasanhodzic and Kotlikoff (2013), but show a somewhat more detailed picture, since we now have an additional type of shock, namely to volatility.

Figure 2 displays impulse responses to two different shocks, for the model with one, two and with 3 assets. The first shock is a technology shock of size $\Phi = -0.12$, the crisis shock. The second shock is a volatility shock, an increase of 50% in next period's volatility. Results are shown for the model with 6 cohorts, for better readability of the graphs. With 12 cohorts, the same qualitative conclusions hold.

In the graphs, each line follows a specific cohort. The single point refers to the cohort

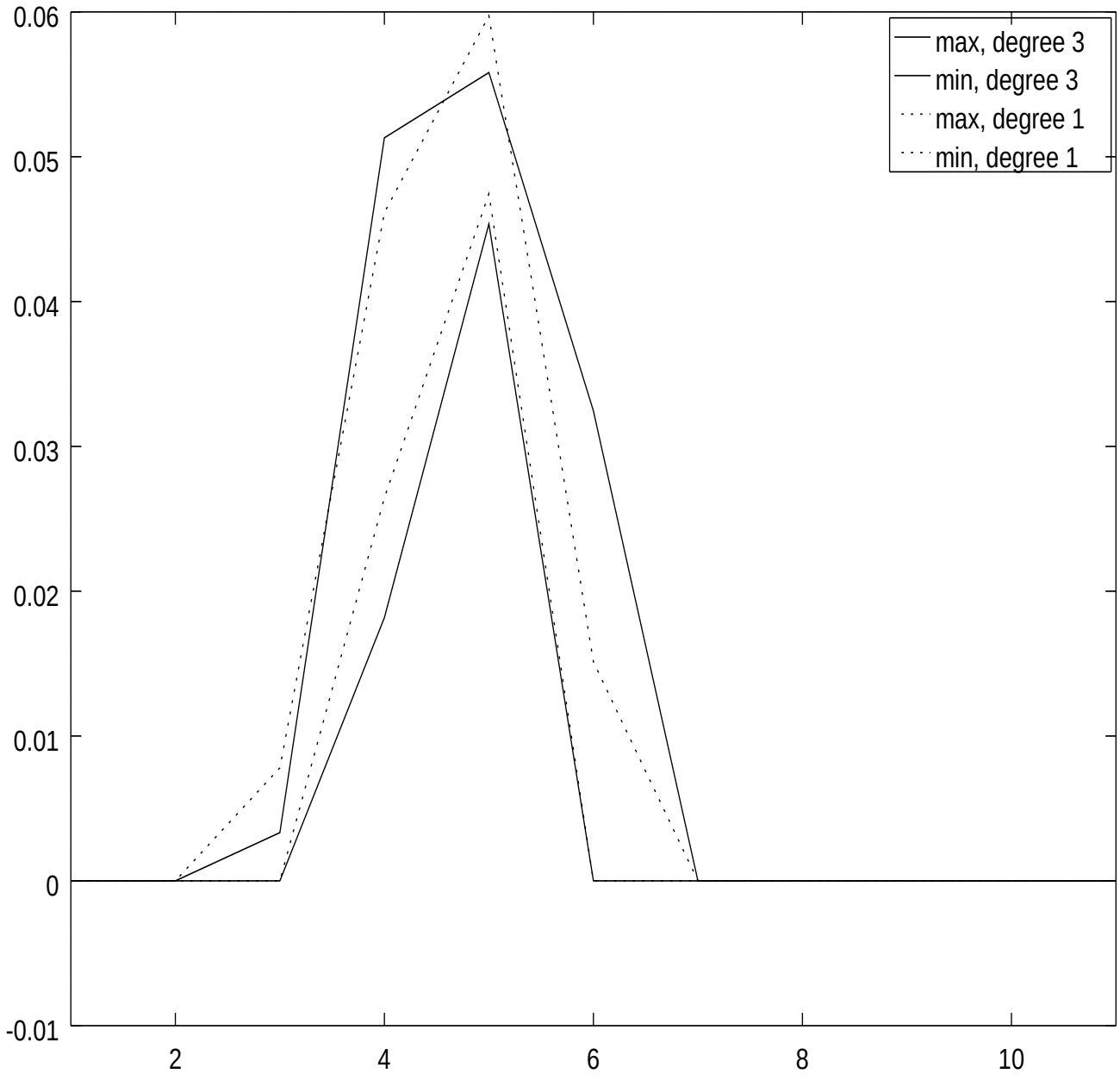


Figure 1: Holdings of safe assets in model with short-sale constraints

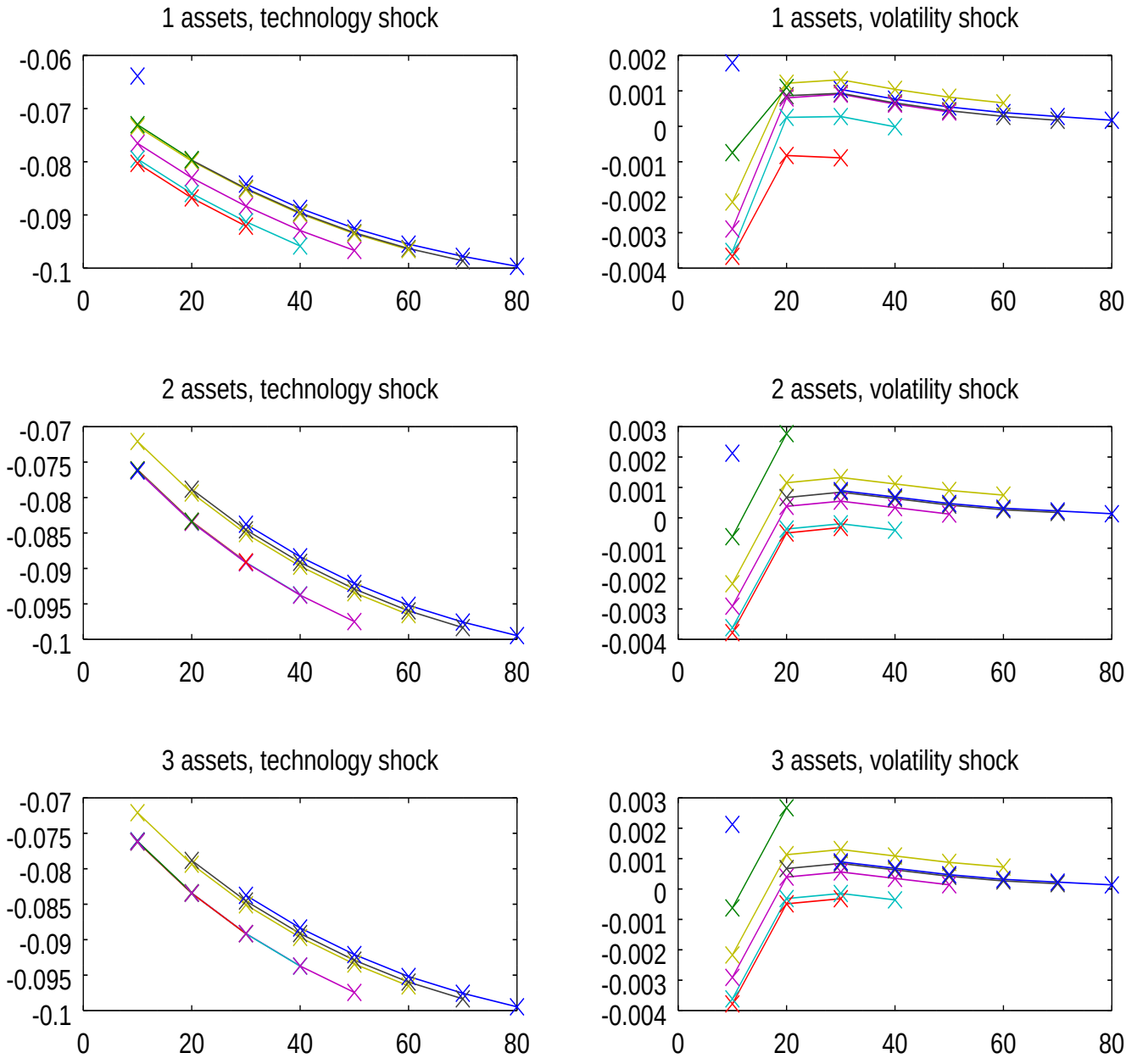


Figure 2: Impulse response consumption, 6 cohorts

that is in the last period of life when the shock hits. The line with two points refers to the cohort that has two periods to live when the shock hits, etc. In a model of sequentially complete markets, cohorts that could trade assets with each other in the period before the shock (cohorts 2 to $J$) use this to insure each other against aggregate shocks. With iso-elastic utility, all of these cohorts then face the same percentage reduction in consumption. In the graph, these lines would be on top of each other. However, for new cohorts, i.e., those born at the time of the shock or later, the change in consumption can be very different.

The picture shows that markets are not complete. In response to a technology shock, the departure from perfect insurance is clearly visibile in the case of one asset. In particular, the oldest cohort ($j = 6$) is less affected by the negative shock. With two assets, the lines for cohorts 2 to 6 are almost on top of each other. With three assets, this is true to higher precision. The generations born at the time of the shock or later also face a reduction in consumption, which is not much affected by the number of assets.

Against the volatility shock, however, the three assets provide little insurance, as can be seen from the graphs in the right part of the figure. This is because none of the three assets is particularly suitable as an insurance against a volatility shock. Furthermore, the welfare effect of this shock is rather small compared to the technology shock, as can be seen from the small change in consumption. Thus there is little incentive to utilize any of the few available assets to hedge against this shock.

The results show that it is not easy to come close to market completeness when the economy is subject to several shocks. Further results (not reported here) show that the allocation is very close to the complete markets allocation when the economy is only subject to a stationary neutral technology shock, even if physical capital is the only asset. This result is similar to Rios-Rull (1994), and more recently Hasanhodzic and Kotlikoff (2013). But things are different when shocks are non-neutral, if they affect the productivity of some cohorts more than others. For questions of inter-generational insurance, it is therefore essential to know the exact nature of the shocks.

## 4.6 Computing times

Below, I report computation times for an Intel i5-4690 CPU at 3.50GHz. I do not deal with parallelization here, and report computing times using only one core. Since most calculations are done separately for each point on a grid, parallelization is rather straight-

forward, using tools such as MPI. Some computations, for example LU decomposition on a big Jacobian matrix, could benefit from a linear algebra library using a graphics processor. I think it is more transparent if I report computing times on a single processor.

In the model with 12 cohorts and two assets, computation time was 69 seconds for the global linear approximation (168 parameters in the outer loop), 16 seconds for the quadratic (1272 parameters), and 208 seconds for the cubic approximation (6720 parameters). This includes the time for the accuracy checks. Notice that the linear approximation needs more time than the quadratic approximation, because the model was solved 40 times in the linear case (the continuation method described in Section 3.2). The quadratic approximation starts from the results in the linear case; parameters and the size of the state space are updated only a few times.

In the same model with three assets, which has the same number of parameters in the outer loop, but requires to solve for more parameters in the equilibrium step, computation time was 100 seconds for the global linear approximation, 24 seconds for the quadratic, and 234 seconds for the cubic approximation. In the model with 12 cohorts, two assets and short-sale constraints, computation time was 275 seconds for the global linear approximation, 54 seconds for the quadratic, and 301 seconds for the cubic approximation.

For the global linear approximation with 60 cohorts (3720 parameters in the outer loop), total computing time was 86 minutes. This model requires to solve for a large number of asset values in the equilibrium step. Again, computing time is high because the model is solved 40 times.

The importance of reducing the number of states and approximated functions can be illustrated with the example of 12 cohorts and three assets. In the reduced state space, we approximate 12 functions (cubic approximation of 11 consumption functions and the price of the safe asset) in 13 states ($z$, $g$ and 11 wealth levels). A cubic function in 13 states has $1 + 13 + 13 \cdot 14/2 + 13 \cdot 14 \cdot 15/6 = 560$ basis functions. This makes a total $560 \cdot 12 = 6720$ parameters. If we took the predetermined portfolios as states, we would have $2 + 33$ state variables. A cubic function in 35 states has $1 + 35 + 35 \cdot 36/2 + 35 \cdot 36 \cdot 37/6 = 8436$ basis functions. If we approximate portfolio choices, we would approximately have 34 functions to approximate, which requires to solve nonlinearly for almost 286824 parameters.

# 5 Conclusions

The paper has shown that it is possible to compute OLG models with a medium number of cohorts, several assets and sort-selling constraints with high precision. Computing times range from several minutes (6 or 12 cohorts) to more than an hour (60 cohorts) on a regular personal computer. Increasing the degree of the approximating polynomials by one brings an increase in accuracy, measured by Euler residuals, of almost one order of magnitude. Global linear approximations are not precise enough to pin down asset choices. However, their business cycle predictions about aggregate statistics, including financial statistics such as the Sharpe ratio, turn out to be very close to the results from higher-order approximations.

# A Details of automatic differentiation

The purpose of this Appendix is to show how automatic differentiation can be efficiently implemented in the context of our model. I do this in two steps. Section A.1 describes how to differentiate the outer-loop part of the algorithm, by breaking it into suitable subproblems. Section A.2 deals with the implicit differentiation of the equilibrium problem in the inner loop.

## A.1 Breaking the problem into parts

The theory of automatic differentiation has established that the computation of the gradient of a function requires not more than five times the number of operations needed for function evaluation, irrespective of the number of independent variables (Griewank and Walther 2008, p.85). However, this can be achieved only in the socalled "reverse mode", which is often difficult to implement. In our application, similar efficiency can be achieved in a much simpler way. The general idea is the following. We are given a function $R(\Theta)$ and want to compute the gradient $\frac{\partial R}{\partial \Theta}$, where $\Theta$ is a high-dimensional vector of parameters. Assume that $R(\Theta)$ can be written as $R(x(\Theta))$, where $x$ is a vector of much lower dimension than $\Theta$. While $R$ can be a complicated function of $x$, we require that $x$ is an easy-to-compute function of $\Theta$. Then the gradient $\frac{\partial R}{\partial \Theta}$ can be efficiently computed as $\frac{\partial R}{\partial x} \cdot \frac{\partial x}{\partial \Theta}$. This is exactly the case in our model. The residuals are complicated functions of the states and of the approximated functions, but the number of states and approximated functions is small, compared to the number of parameters. The approximated functions depend on the parameters in a very simple way, namely $a_j = \sum_l B_l(x)\Theta_{lj}$. Since the basis functions $B_l(x)$ are already computed for function evaluation, they do not impose any additional burden for the gradient evaluation.

In the following, denote by $x$ the vector of states, with dimension $n_x$, by $a$ the vector of approximated variables, with dimension $n_a$, and by $e$ a vector of endogenous variables, with dimension $n_e$, which is related to $x$ and $a$ by

$$e_j = \phi_j(x, a), \qquad j = 1, \ldots, n_e \tag{56}$$

where $\phi_j$ is a known function. The role of $e$ will become clear in the next subsection. We further assume that the exogenous shock $\epsilon$ can only take a finite number of realizations,

$\epsilon^\xi$, where $\xi = 1, \ldots, n_\epsilon$. Next period's states $\hat{x}$ are assumed to follow

$$\hat{x}_j^\xi = \chi_j(x, a, \epsilon^\xi), \qquad j = 1, \ldots, n_x \tag{57}$$

where again the $\chi_j$ are are known functions. For any variable $y$, we adopt the notation $y^\xi$ for the value of $y$ in the next period, conditional on the shock being $\epsilon^\xi$.

At a given point in the state space, $x$, written as a column vector, any equation of the model can be written as

$$\sum_{\xi=1}^{n_\epsilon} \pi_\xi R\left(x, a, e, \hat{x}^\xi, \hat{a}^\xi\right) = 0 \tag{58}$$

where $a_j = \sum_{l=1}^{n_b} B_l(x)\Theta_{lj}$ and $n_b$ is the number of basis functions. Next period's decision function are then given by $\hat{a}_j^\xi = \sum_{l=1}^{n_b} B_l(\hat{x}^\xi)\Theta_{lj}$.

Understanding that $R$ is always evaluated at $\left(x, a, e, \hat{x}^\xi, \hat{a}^\xi\right)$, and $\chi$ is always evaluated at $(x, a, \hat{\epsilon})$, we drop arguments in the following formulas. Differentiating the lhs of (58) w.r.t $\Theta_{i,k}$ then gives

$$\sum_{\xi=1}^{n_\epsilon} \pi_\xi \left[ \left( \frac{\partial R}{\partial a_k} + \sum_{j=1}^{n_e} \frac{\partial R}{\partial \phi_j}\frac{\partial \phi_j}{\partial a_k} + \sum_{j=1}^{n_x} \frac{\partial R}{\partial \hat{x}_j^\xi}\frac{\partial \chi_j}{\partial a_k} \right) B_i(x) + \sum_{j=1}^{n_a} \frac{\partial R}{\partial \hat{a}_j^\xi}\frac{d\hat{a}_j^\xi}{d\Theta_{ik}} \right] \tag{59}$$

and

$$\frac{d\hat{a}_j^\xi}{d\Theta_{ik}} = \sum_{l=1}^{n_b} \frac{d\left[B_l(\hat{x}^\xi)\Theta_{lj}\right]}{d\Theta_{ik}} = B_i(\hat{x}^\xi)\delta_{j,k} + \sum_{l=1}^{n_b}\sum_{m=1}^{n_x} \Theta_{lj}\frac{\partial B_l(\hat{x}^\xi)}{\partial x_m}\frac{\partial \chi_m}{\partial a_k}B_i(x) \tag{60}$$

The gradient w.r.t. the column vector $\Theta_{:,k}$ therefore has two components:

1. The scalar

$$\sum_{\xi=1}^{n_\epsilon} \pi_\xi \left[ \frac{\partial R}{\partial a_k} + \sum_{j=1}^{n_e} \frac{\partial R}{\partial \phi_j}\frac{\partial \phi_j}{\partial a_k} + \sum_{j=1}^{n_x} \frac{\partial R}{\partial \hat{x}_j^\xi}\frac{\partial \chi_j}{\partial a_k} + \sum_{j=1}^{n_a}\sum_{l=1}^{n_b}\sum_{m=1}^{n_x} \frac{\partial R}{\partial \hat{a}_j^\xi}\Theta_{lj}\frac{\partial B_l(\hat{x}^\xi)}{\partial x_m}\frac{\partial \chi_m}{\partial a_k} \right] \tag{61}$$

   multiplied with the vector of basis functions at $x$, denoted by $\mathcal{B}(x)$.

2. The vector

$$\sum_{\xi=1}^{n_\epsilon} \pi_\xi \frac{\partial R}{\partial \hat{a}_k^\xi}\mathcal{B}(\hat{x}^\xi) \tag{62}$$

Given $x, a, \hat{x}^\xi, \hat{a}^\xi$, the vectors $\frac{\partial R}{\partial a}$, $\frac{\partial R}{\partial \hat{x}^\xi}$, $\frac{\partial R}{\partial \hat{a}}$ and $\frac{\partial \chi}{\partial a}$ are small and relatively cheap to compute (in particular, independent of the dimension of $\Theta$). The triple sum in (61) can be written as $(\Delta R\Theta^T)(\Delta B \Delta\chi_k)$, where the matrix $\Delta R$ is $1 \times n_a$, $\Theta$ is $n_b \times n_a$, $\Delta B$ is $n_b \times n_x$,

and $\Delta\chi_k$ is $n_x \times 1$. The triple sum can be done with $n_b(n_a + n_x + 1)$ multiplications. It is linear in $n_b$ with a small multiple.

The basis function vectors $\mathcal{B}(x)$ and $\mathcal{B}(\hat{x}^\xi)$ are big, but they are already needed for the evaluation of $R$. The extra cost for the evaluation of the gradient therefore comes mainly from the term $\frac{\partial B}{\partial x}(\hat{x}^\xi)$ and from the multiplications involved in the last term on the rhs of (60). Computing $\frac{\partial B}{\partial x}(\hat{x}^\xi)$ costs a small multiple of the cost of $\mathcal{B}(\hat{x}^\xi)$.

## A.2 Implicit differentiation of the inner loop problem

We now consider the case that the state transition function $\chi$ and the function $\phi$ are not known, but are given implicitly by systems of nonlinear equations. In particular, we assume that next periods state satisfies

$$T(x, a, e, \hat{x}^\xi, \hat{a}^\xi, \hat{\epsilon}^\xi) = 0, \qquad \xi = 1, \ldots, n_\epsilon \tag{63}$$

and the endogenous variables $e$ satisfy

$$\sum_{\xi=1}^{n_\epsilon} \pi_\xi H(x, a, e, \hat{x}^\xi, \hat{a}^\xi) = 0 \tag{64}$$

To apply the formulas of Section A.1, we have to compute the derivatives $\frac{\partial \phi_j}{\partial a_k}$ and $\frac{\partial \chi_j}{\partial a_k}$. Given $x$ and $a = \mathcal{B}(x)\Theta$, and using $\hat{a}^\xi = \mathcal{B}(\hat{x}^\xi)\Theta$, the endogenous variables $e$ and next period's states can be solved for simultaneously using (63) and (64). This is a big nonlinear system of equations, solving simultaneously for $n_b + n_\epsilon n_x$ variables. The problem could become intractable (in particular if there are many possible realizations of shocks) if it didn't have a sparsity structure: each $\hat{x}^\xi$ depends on $a$, $e$ and $\epsilon^\xi$, but not on the $\hat{\epsilon}^\zeta, \hat{x}^\zeta$ with $\xi \neq \zeta$.

To make this explicit, write(63) and (64) as a stacked vector of equations:

$$\begin{bmatrix} \sum_{\xi=1}^{n_\epsilon} \pi_\xi H(x, a, e, \hat{x}^\xi, \mathcal{B}(\hat{x}^\xi)\Theta) \\ T(x, a, e, \hat{x}^1, \mathcal{B}(\hat{x}^1)\Theta, \epsilon^1) \\ \vdots \\ T(x, a, e, \hat{x}^{n_\epsilon}, \mathcal{B}(\hat{x}^{n_\epsilon})\Theta, \epsilon^{n_\epsilon}) \end{bmatrix} = 0 \tag{65}$$

The Jacobian of this system, w.r.t to the stacked vector of variables $[e, \hat{x}^1, \ldots, \hat{x}^{n_\epsilon}]$, can be written in partitioned form as

$$J = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{66}$$

where $A$, $B$ and $C$ have dimensions $n_a \times n_a$, $n_a \times n_\epsilon n_x$ and $n_\epsilon n_x \times n_a$, respectively. The $n_\epsilon n_x \times n_\epsilon n_x$ matrix $D$ is block-diagonal, having $n_\epsilon$ blocks of size $n_x \times n_x$ along the diagional. The matrix $J$ is used for two purposes:

1. to solve for the $[e, \hat{x}^1, \ldots, \hat{x}^{n_\epsilon}]$ by quasi-Newton methods;

2. to compute the derivative of the equilibrium $[e, \hat{x}^1, \ldots, \hat{x}^{n_\epsilon}]$ by the implicit function theorem:
$$\frac{\partial [e, \hat{x}^1, \ldots, \hat{x}^{n_\epsilon}]}{\partial a} = -J^{-1} J_a \tag{67}$$
where $J_a$ denotes the Jacobian of (65) w.r.t $a$, and both $J^{-1}$ and $J_a$ are evaluated at the solution point $[e, \hat{x}^1, \ldots, \hat{x}^{n_\epsilon}]$.

To solve a linear system $Jx = b$ efficiently, we first use the partitioned-inversion formula (Press, Flannery, Teukolsky, and Vetterling 1986, p. 77)

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}, \tag{68}$$

Notice that $A$ is of the small dimension $n_a \times n_a$, but $D - CA^{-1}B$ is of the possibly large dimension $n_\epsilon n_x \times n_\epsilon n_x$. To invert the latter matrix, notice that $D$ is block-diagonal, therefore easy to invert, and $CA^{-1}B$ is of small rank. Then we can use again the Sherman-Morrison-Woodbury formula (48) to obtain[7]

$$\left(D - CA^{-1}B\right)^{-1} = D^{-1} + D^{-1}C\left(A - BD^{-1}C\right)^{-1} BD^{-1} \tag{69}$$

Equation system (64) is formulated in terms of the full state vectors $\hat{x}^\xi$. Notice that it will typically be possible to reduce the dimension of the numerical root finding problem by expressing the future states $\hat{x}^\xi$ as a function of past states, decisions and and a small number of current prices. For an example, see Section 3.3 where $\hat{x}^\xi$ is computed using only $R_K^\xi$. This can be exploited for the efficient solution of the equilibrium problem.

[7]Notice that $D^{-1}$ is sparse, while $D^{-1}C\left(A - BD^{-1}C\right)^{-1} BD^{-1}$ is in general not, therefore one should not compute this matrix, but compute $D^{-1}C\left(A - BD^{-1}C\right)^{-1} BD^{-1}X$, for any $X$, from right to left. I describe the computations in terms of matrix inversion for notational simplicity. The computationally efficient way to solve the system is to use LU-decompositions rather than inverses.

# References


Auerbach, A. and L. Kotlikoff (1987). *Dynamic Fiscal Policy*. Cambridge University Press.

Auerbach, A. J. and L. J. Kotlikoff (1981). National Savings, Economic Welfare, and the Structure of Taxation. NBER Working Papers 0729, National Bureau of Economic Research, Inc.

Bansal, R. and A. Yaron (2004, 08). Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. *Journal of Finance 59*(4), 1481–1509.

Griewank, A. and A. Walther (2008). *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation, Second Edition*. SIAM e-books. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104).

Hasanhodzic, J. and L. J. Kotlikoff (2013). Generational Risk  Is It a Big Deal?: Simulating an 80-Period OLG Model with Aggregate Shocks. NBER Working Papers 19179, National Bureau of Economic Research, Inc.

Hasanhodzic, J. and L. J. Kotlikoff (2015). Generational Risk  Is It a Big Deal?: Simulating an 80-Period OLG Model with Aggregate Shocks. Manuscript.

Higham, N. (2002). *Accuracy and Stability of Numerical Algorithms* (2nd ed.). SIAM.

Judd, K. L. (1992). Projection methods for solving aggregate growth models. *Journal of Economic Theory 58*(2), 410–52.

Judd, K. L. (1998). *Numerical Methods in Economics*. Cambridge and London: MIT Press.

Judd, K. L. and S.-M. Guu (2001). Asymptotic methods for asset market equilibrium analysis. *Economic Theory 18*(127-57), 127–157.

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

Judd, K. L., L. Maliar, S. Maliar, and R. Valero (2014). Smolyak method for solving dynamic economic models: Lagrange interpolation, anisotropic grid and adaptive domain. *Journal of Economic Dynamics and Control 44*, 92–123.

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

Krueger, D. and F. Kubler (2006). Pareto improving social security reform when financial markets are incomplete!? *American Economic Review, forthcoming 96*(3), 737–755.

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

Ludwig, A. and M. Reiter (2010). Sharing demographic risk – who is afraid of the baby bust? *American Economic Journal: Economic Policy 2*, 83–118.

Malin, B., D. Krueger, and F. Kubler (2007). Computing stochastic dynamic economic models with a large number of state variables: A description and application of a smolyak-collocation method. Manuscript.

Marcet, A. and G. Lorenzoni (1998). Parametrized expectations approach; some practical issues. In R. Marimon and A. Scott (Eds.), *Computational Methods for the Study of Dynamic Economies*. Oxford University Press.

Marcet, A. and K. J. Singleton (1998). Equilibrium asset prices and savings of heterogeneous agents in the presence of incomplete markets and portfolio constraints. UPF Working Paper 319.

Miranda, M. J. and P. L. Fackler (2002). *Applied Computational Economics and Finance*. MIT Press.

Murty, K. G. (1988). *Linear Complementarity, Linear and Nonlinear Programming*. Helderman-Verlag.

Nocedal, J. and S. J. Wright (2006). *Numerical Optimization, second edition*. World Scientific.

Press, W., B. Flannery, S. Teukolsky, and W. Vetterling (1986). *Numerical Recipes*. Cambridge University Press.

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

Rios-Rull, J.-V. (1994). On the quantitative importance of market completeness. *Journal of Monetary Economics 34*(3), 463–496.