Solving OLG Models with Asset Choice

Michael Reiter, Institute for Advanced Studies, Vienna

February 15, 2015

PRELIMINARY VERSION!

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 about two thirds of that of US stocks. Being subject to three types of shocks, the equilibiurm allocation, even with 3 assets, differs substantially from an allocation under sequentially complete markets. In particular, the oldest cohorts are more more heavily exposed to negative shocks.

Keywords: OLG models, asset choice, projection methods.

Address of the author:

Michael Reiter
Institute for Advanced Studies
Stumpergasse 56
A-1060 Vienna
Austria

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

1 Introduction

Overlapping generation models play a prominent role in economic policy analysis. They are the central tool for evaluating social security systems, and questions of inter-generational redistribution (some more recent papers are Krueger and Kubler (2006),Hasanhodzic and Kotlikoff (2013),Ludwig and Reiter (2010)).

The quantitative analysis of deterministic models, even those with a large number of cohorts, has been possible at least since 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 can be easily applied using the appropriate tools, such as Dynare. However, they have problems accounting for large shocks, and in particular it is very difficult to deal with inequality constraints, such as short-sale constraints. The main alternative are global projection methods, pioneered in economics by Judd (1992) (see also (Miranda and Fackler 2002)), 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. Furthermore, convergence of the algorithms is not guaranteed, in particular they may fail if one allows for big shocks.

The aim of this paper is to outline an efficient implementation of a solution strategy that is very general, gives a high-precision solution, and is robust in the sense that it is very likely to converge, even in the case of big shocks.

The method is applied to an OLG model with asset choice, large shocks and short-sale constraints. This is a simple and natural, but computationally highly challenging model. Much of the discussion in this paper is not specific to OLG models, but applies equally to other heterogeneous agent models.

A fast and reliable algorithm can be obtained adopting the following ideas:

1. Minimize the number of state variables.

2. Minimize the number of variables that are parameterized, that means, that are approximated by smooth functions for which we have to find the optimal parameters.

3. Use automatic differentiation and the implicit function theorem.

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

5. Use continuation methods, starting from the linearized solution around the deterministic steady state, with small shocks, and transform it continuously into the fully nonlinear solution with large shocks.

6. Fast evaluation of approximating functions.

Let me explain some of these issues in more detail. Since the number of state variables is decisive for the computational complexity of nonlinear algorithms, it is obvious that one should minimize the number of state variables. Concretely, 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 computed optimally at each grid point.

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. However, the implementation is not as straightforward as it seems. If the net wealth of households are the state variables, then we assume that all variables of the solution, including asset prices, are a function of household net worth. However, household net worth is a function of current asset prices: what is predetermined is households' asset posisions at the end of last periods, but we need current asset prices to convert those into given last period's asset holdings In practice, this implies that we have to solve a nonlinear fixed point problem if we weant to compute next period's states (household net worth) from current period states and decisions (asset choice).

This is reinforced by the decision not to parameterize asset choices. This implies that asset choices have to be solved for at each grid point, and for each value of the parameters that determine 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 enormously time consuming. To make this problem tractable, I resort to automatic differentiation and exploit the implicit function theorem. This is explained in more detail in Section 3.4.

With a medium- or high-dimensional state space, the choice of a grid on which one computes the residuals of the projecton method is essential. For this, one has to use simulation techniques to obtain information about the state space in which the solution lives. This information can be used in several ways: one can either use information on means and (co-)variances of the state variables to construct an ellipsoid in which the solution lives most of the time. Or one directly use a set of points that arises from simulations of the model, in the spirit of parameterized expectation methods (Marcet and Lorenzoni 1998; Judd, Maliar, and Maliar 2012). I explore both approaches.

In order to highlight the points that are the focus of this paper, I neglect two issues that are potentially important, and might be used in the future for further improvements to the algorithm. First, I do not deal with state aggregation. In models with many cohorts, it may not be computationally optimal to make the decisions of each cohort a function of th 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). Although Krueger and Kubler (2004) have shown that this does not work as easily in OLG models as in the model of Krusell and Smith (1998), it is likely that there are refinements of this method that are helpful to OLG models. This is beyond the scope of the present paper.

The second issue that I neglect are sparse grids (Malin, Krueger, and Kubler 2007; Judd, Maliar, Maliar, and Valero 2013). I rather use complete polynomials for function approximation, and either simple random grids, or ergodic grids obtained from simulations (Judd, Maliar, and Maliar 2012). Whether sparse grid techniques are useful in the present context is left for future research.

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 approximation 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 investigate how the possibility of large negative shocks affects the price of risk.

2 An OLG Model With Portfolio Choice

Time is discrete. Each cohort (household) lives for J periods, denoted by $j = 0, \ldots, J-1$. Household born in t maximize

$$\mathrm{E}_t \sum_{j=0}^{J-1} \beta^j \frac{c_{j,t+j}^{1-\gamma}}{1-\gamma} \tag{1}$$

Labor supply is assumed to be fixed. In the first two thirds of their life, workers supply their labor endowment, which is denoted by ζ_j for the cohort of age j. In the last third of their life, they are retired.

2.1 Technology

Output is used for investment and consumption, so that aggregate resource constraint is

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

The production function is given by

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

where TFP_t is a nonstationary process whose growth rate is denoted by G:

$$TFP_t = g_t TFP_{t-1} \tag{4}$$

This growth rate has an i.i.d. component ϵ and an autoregressive component z:

$$g_t = 1 + z_t + \epsilon_t \tag{5}$$

The autoregressive component is subject to time-varying volatility (Bansal and Yaron 2004; Caldara, Fernandez-Villaverde, Rubio-Ramirez, and Yao 2012):

$$z_t = \bar{g} + \rho(z_{t-1} - \bar{g}) + (1 + \sigma_{t-1})\epsilon_{z,t} \tag{6}$$

Volatility σ itself follows an AR(1) process:

$$\sigma_t = \rho_s \sigma_{t-1} + \epsilon_{s,t} \tag{7}$$

The wage w_t and rental rate of capital rK_t equal marginal productivities in equilibrium:

$$rK_t = F_k(K_{t-1}, L_t)) \tag{8}$$

$$w_t = F_L(K_{t-1}, L_t) \tag{9}$$

2.2 Households decisions

Denote by $k_{i,t}$ the holdings of the physical asset at the end of period t by cohort i, where $i = 1, \ldots, J-1$. Similarly, $A_{i,t}$ and $A2_{i,t}$ denote the holdings of the first (safe) and the second financial asset.

The return on Physical capital costs Q^K_{t-1} in period $t-1$ and yields in t

$$R^k_t \equiv qK_t gK_t + rK_t \tag{10}$$

The first financial asset, which is safe, costs Q^S_{t-1} and gets a guaranteed yield which is normalized to 1. The second financial asset is rather artificially constructed. Since the return on physical capital is approximately linear in technology, we choose the second asset as quadratic in the deviation of z from its mean:

$$R^q_t = 1 + (z_t - \bar{g})^2 \tag{11}$$

$W_{i,t}$ denotes market wealth of the same cohort at the beginning of the next perid. It evolves from assets as

$$W_{i,t} = k_{i,t-1}R^k_t + A_{i,t-1}1 + A2_{i,t-1}R^q t \tag{12}$$

End of period capital satisfies

$$k_{i,t} = (W_{i-1,t} + w_t\zeta_i - c_{i,t} - qS_tA_{i,t} - qS2_tA2_{i,t})/qK_t \tag{13}$$

Since a household is born without assets, capital at the end of the first period can also be written as

$$k_{1,t} = (w_t\zeta_1 - c_{1,t} - qS_tA_{1,t} - qS2_tA2_{1,t})/qK_t \tag{14}$$

Since a household leaves no assets, consumption in the last period is given by

$$c_{J,t} = W_{J-1,t} + w_t\zeta_J \tag{15}$$

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

$$U_c(c_{i,t})qK_t = \beta\,\mathrm{E}_t((R^k_{t+1})U_c(c_{i+1,t+1})) \tag{16}$$

for capital,

$$U_c(c_{i,t})qS_t+ = \beta\,\mathrm{E}_t(1U_c(c_{i+1,t+1})) \tag{17}$$

for the safe assets, and

$$U_c(c_{i,t})qS2_t = \beta\,\mathrm{E}_t(R^q t + 1)U_c(c_{i+1,t+1})) \tag{18}$$

for the second financial asset.

2.3 Aggregation

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

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

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

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

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

The financial assets are in zero net supply:

$$\frac{1}{J}\sum_{i=1}^{J-1}A_{i,t} = 0 \tag{19}$$

$$\frac{1}{J}\sum_{i=1}^{J-1}A2_{i,t} = 0 \tag{20}$$

2.4 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

$$K_t = G(K_{t-1}, I_t, \delta_t) \tag{21}$$

where $G(K, I)$ is constant returns to scale, and satisfies $G(K, \delta K) = K$, $G_I(K, \delta K) = 1$ and $G_{II}(K, \delta K) < 0$.

The sector maximizes

$$\mathrm{E}\sum_{t=0}^{\infty}\Lambda_{0,t}\left[K_{t-1}R_t^k - I_t + \nu_t\left(-K_t + G(K_{t-1}, I_t)\right)\right] \tag{22}$$

which gives the FOCs

$$1 = \nu_t G_I(K_{t-1}, I_t)$$

$$\Lambda_{0,t}\nu_t = \mathrm{E}_t\,\Lambda_{0,t+1}\left(R_{t+1}^k + \nu_{t+1}G_K(K_t, I_{t+1})\right) \tag{23}$$

$$G_I^{-1}(K_{t-1}, I_t) = \mathrm{E}_t\,\Lambda_{t,t+1}\left(R_{t+1}^k + G_I^{-1}(K_t, I_{t+1})G_K(K_t, I_{t+1})\right) \tag{24}$$

3 Algorithm

3.1 Outline of the algorithm

For simplicity of exposition, I describe the algorithm for the case without short-sale constraints. Section 3.3 outlines the change that is necessary to handle short-sale constraints.

I assume that the dividend of security a at time t, denoted by D^a, only depends on the realization of the exogenous shock z_t, therefore can be written as $D^a(z_t)$. The total return of as asset is given by $R^a(z_t, q_t) = D^a(z_t) + q_t$.

Denote by Σ_ϵ the covariance matrix of the exogenous shock vector ϵ.

1. Solve the model with only one asset (physical capital) by linearization around the deterministic steady state.

   The linearized solution can be written in the form

   $$\begin{aligned} X_t &= \mu_X^* + A(X_{t-1} - \mu_X^*) + B\epsilon_t \\ Y_t &= \mu_Y^* + C(X_t - \mu_X^*) \end{aligned} \tag{25}$$

   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, Σ_X^{lin}, which satisfies

   $$\Sigma_X^{lin} = A\Sigma_X^{lin}A' + B\Sigma_\epsilon B' \tag{26}$$

2. Choose a set of n_p basis functions $\phi_l(X)$ of the vector of states x. The basis functions serve to approxe the consumption functions for cohorts 1 to $J-1$:

   $$c_j = \mathcal{P}(X; \theta_j) = \sum_{j=1}^{n_p} \theta_{j,l}\phi_l(X), \qquad l = 1, \ldots, J-1 \tag{27}$$

   Stack the θ_j into the big vector Θ. The number of elements of Θ is $n_p(J-1)$.

3. Choose a set of points X_i° for $i = 1, \ldots, n_g$ with $n_g \geq n_p$ such that they fill the d-dimensional unit-sphere approximately equally. 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. This can be done by low-discrepancy sequences, such as Sobol or Faure points.[1]

[1] 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, just as is described in ... for simulated sequences.

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

$$Y_{i,j} = F_n^{-1}(\hat{X}_{i,j}), \qquad i = 1, \ldots, n_g, j = 1, \ldots, d \tag{28}$$

(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 χ^2 distribution with d degrees of freedom, and $\|Y_i\|_2$ denotes the sum of squares of the vector Y_i.

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

5. Choose a sequence of variance scaling parameters λ_1, λ_2, …, λ_{n_ϵ} with $\lambda_k \leq \lambda_{k+1}$ and $\lambda_{n_\epsilon} = 1$.

6. Choose a sequence of asset holding cost parameters κ_1, κ_2, …, κ_{n_ϵ} with $\kappa_k \geq \kappa_{k+1}$ and $\kappa_{n_\epsilon} = 0$.

7. Set the iteration count (determining the variance scaling parameter) to $k = 1$.

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

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

$$\begin{aligned} \mu_{X,k} &= \mu_X^* \\ \Sigma_{X,k} &= \lambda_1^2 \Sigma_X^{lin} \end{aligned} \tag{29}$$

   - In later iterations ($k > 1$), set

$$\begin{aligned} \mu_{X,k} &= \frac{1}{2(T - T_{sk})} \sum_{t=T_{sk}+1}^{T} \left(X_t^{k-1,+\epsilon} + X_t^{k-1,-\epsilon}\right) \\ \Sigma_{X,k} &= \left(\frac{\lambda_k}{\lambda_{k-1}}\right)^2 \frac{1}{2(T - T_{sk} - 1)} \sum_{t=T_{sk}+1}^{T} \Big[(X_t^{k-1,+\epsilon} - \mu_{X,k})(X_t^{k-1,+\epsilon} - \mu_{X,k})' \\ &\quad + (X_t^{k-1,-\epsilon} - \mu_{X,k})(X_t^{k-1,-\epsilon} - \mu_{X,k})'\Big] \end{aligned} \tag{30}$$

9. Choose a grid of state vectors $\bar{X}_i$ as

$$\bar{X}_i = \mu_{X,k} + \psi \Gamma_k X_i^\circ, \qquad i = 1, 2, \ldots, n_g \tag{31}$$

where Γ_k is such that $\Gamma_k \Gamma_k' = \Sigma_{X,k}$, and $\psi \geq 1$ is an augmentation factor, which allows the grid to fill a larger state space then what is suggested by the estimated covariance matrix $\Sigma_{X,k}$.

10. If $n_g = n_p$, we will find a parameter vector θ by setting the residuals at all grid points equal to zero. If $n_g > n_p$, we will set weighted sums of residuals equal to zero. Denote these weights by $\phi_{l,i}$. We choose them in the following way. Define B as the $n_g \times n_p$ matrix of basis functions, that means, $B_{i,l} = \phi_l(\bar{X}_i)$. Then define Φ as the projection matrix into the space of basis functions, $\Phi = (B'B)^{-1} B'$. Choose $\phi_{l,i} = \Phi_{i,l}$.

    Other choices of weights are possible.

11. Choose an initial guess of the parameter vector Θ, denoted by θ_k^{init}.

    - In the first iteration ($k = 1$), set θ_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 (we assume here that the linear approximation is nested in the family of approximation functions (27).
    - In later iterations ($k > 1$), 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 13 below.[2]

12. Given a parameter vector θ_k, compute residuals $R(i, j; \theta, k)$ as follows. For each state $\bar{X}_i$ do the following:

    - Set beginning-of-period capital $\widehat{K}$ as

    $$\widehat{K} = \frac{1}{\mathcal{R}(z\left(\bar{X}_i\right), \widehat{K}, L)} \sum_{j=2}^{J} W_j\left(\bar{X}_i\right) \tag{32}$$

    - Set $\widehat{w} = \mathcal{W}(z\left(\bar{X}_i\right), \widehat{K}, L)$. and $\widehat{R^k} = \mathcal{R}(z\left(\bar{X}_i\right), \widehat{K}, L)$.
    - Set

    $$\widehat{S_1} = \widehat{w}\zeta_1 L_1 - \mathcal{P}\left(\bar{X}_i; \theta_1\right) \tag{33}$$

    $$\widehat{S_j} = W_{j-1}\left(\bar{X}_i\right) + \widehat{w}\zeta_{j+1} L_{j+1} - \mathcal{P}\left(\bar{X}_i; \theta_{j+1}\right) \tag{34}$$

    - Define $z_\xi \equiv \mathcal{T}\left(z\left(\bar{X}_i\right), \lambda_k \bar{\epsilon}_\xi\right)$.

      Find asset prices $\widehat{q^a}$, $a = 1, \ldots, n_a$, and portfolio choices $\widehat{\omega_j^a}$ for $a = 0, \ldots, n_a$, $j = 1, \ldots, J-1$, such that

    $$\sum_{j=1}^{J-1} \widehat{S_j}\widehat{\omega_j^a} = 0, \qquad a = 1, \ldots, n_a \tag{35}$$

[2]Notice that this implies that the initial value of Θ is used to extrapolate the approximated functions beyond the state space on which Θ was determined, because a new iteration uses higher Σ, and therefore a larger state space. This works well with polynomials if the change in the state space is not too big.

$$\sum_{a=0}^{n_a} \widehat{\omega_j^a} = 1, \qquad j = 1, \ldots, J-1 \tag{36}$$

and

$$U_c\left(\mathcal{P}\left(\bar{X}_i; \theta_j\right), L_j\right) \widehat{q^a} = \beta \sum_{\xi=1}^{n_\epsilon} U_c\left(\mathcal{P}\left(\hat{X}_\xi; \theta_{j+1}\right), L_{j+1}\right) R^a\left(z_\xi, \widehat{q}\right), \quad a = 1, \ldots, n_a \tag{37}$$

where $\hat{X}_\xi$ is defined as next period's state that emerges if next period's shock equals $\lambda_k \bar{\epsilon}_\xi$. More precisely,

$$\begin{aligned} z\left(\hat{X}_\xi\right) &= z_\xi \\ W_{j+1}\left(\hat{X}_\xi\right) &= \widehat{S_j}\left[\widehat{\omega_j^0}\widehat{R^k} + \sum_{a=1}^{n_a} \widehat{\omega_j^a} R^a\left(z_\xi, \widehat{q}\right)\right], \quad j = 1, \ldots, J-1 \end{aligned} \tag{38}$$

- The residual of cohorts $j = 1, \ldots, J-1$ at grid point i are then defined as the Euler residual of capital holdings:

$$R(i, j; \theta, k) = -U_c\left(\mathcal{P}\left(\bar{X}_i; \theta_j\right), L_j\right) + \beta \sum_{\xi=1}^{n_\epsilon} U_c\left(\mathcal{P}\left(\hat{X}_\xi; \theta_{j+1}\right), L_{j+1}\right) \mathcal{R}(z\left(\hat{X}_\xi\right), \widehat{K}', L) \tag{39}$$

Denote the parameter vector found by θ_k.

13. Starting from the initial guess θ_k^{init}, use a quasi-Newton method to find a parameter vector θ such that for each cohort j the n_p sums of weighted residuals are set equal to zero:

$$\sum_{i=1}^{n_g} \phi_{l,i} R(i, j; \theta, k) = 0, \qquad j = 1 \ldots J-1, \quad l = 1, \ldots, n_p \tag{40}$$

Specifically, I use Powell's method, as described in Nocedal and Wright (2006, Chapter 11).

14. Using the parameter vector θ_k and the sequence of exogenous shocks ϵ_t, $t = 1, \ldots, T$, 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 $t = 1, \ldots, T$.

15. If $k = n_\epsilon$, stop. Take θ_{n_ϵ} as the final solution vector.

    Otherwise, increment k by one and go to Step 9.

3.2 Time iteration vs. quasi-Newton algorithms

There are several ways to solve the fixed point problem for the parameter vector Θ 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 β 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 variables:

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

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

In the model with only 6 cohorts, time iteration is competitive. With 12 cohorts, it is faster for the cubic approximation. In the model with 60 cohorts and global linear approximation, quasi-Newton is clearly superior.

In terms of reliability, I found no difference between the methods. If one of them converges, so does the other ones. 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; if a solution exists, but is too far away from the starting point (linearized solution around the steady state), then it may be impossible to find it.

3.3 Short sale constraints

It is not the point of this paper to investigate what are reasonable constraints on households portfolio decisions, but rather to explore what are the technical challenges and how they can be overcome. I therefore impose simple bound constraints on assets:

$$A_{i,t} \geq \underline{A}_i \tag{41}$$

for some exogenous $\underline{A}_i$. Occasionally binding inequality constraints mean that standard non-linear root finders cannot be applied. Therefore, in Step 13 of the algorithm above, 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 for linear complementarity problems. It turns out that this procedure slows down the computation by a factor of about five, compared to the case without inequality constraints, but is very reliable.

3.4 The use of automatic differentiation

Automatic differentiation means the numerical computation of derivatives for a given value of the independent variables, but not by finite-difference approximations, but by using the exact differentiation rules. Simple forms of automatic differentiation (forward mode) can be easily implemented in object-oriented programming languages such as C++, which I have used for the current project. For a general description of automatic differentiation, see Griewank and Walther (2008).

To see how automatic differentiation is useful in our setup, consider the equilibrium problem in Step 12 of the algorithm described in Section 3.1. In general terms, we can describe this problem as searching for a parameter vector Θ and a set of endogenous variables w so as to solve the following nonlinear system of equations:

$$F(y(\Theta), w) = 0 \tag{42}$$

$$G(y(\Theta), w) = 0 \tag{43}$$

In the overall problem, we use an iterative method to find the Θ that satisfies (42). For any guess $\hat{\Theta}$ in the iterative process, we have an inner loop problem, where we choose the $w = \hat{w}$ so as to satisfy (43).

Automatic differentiation is the useful in two ways. First, to solve the inner-loop problem efficiently by quasi-Newton methods. The second way is more interesting. If we solve for Θ in the outer loop by a quasi-Newton method, we need to compute the derivative of F with respect to Θ. Applying the implicit function theorem on (43), we get

$$\begin{aligned} \frac{\partial F}{\partial \Theta} &= F_y \frac{\partial y}{\partial \Theta} + F_w \frac{\partial w}{\partial y} \frac{\partial y}{\partial \Theta} \\ &= \left(F_y \frac{\partial y}{\partial \Theta} - F_w G_w^{-1} G_y \right) \frac{\partial y}{\partial \Theta} \end{aligned} \tag{44}$$

It should be understand that all derivatives of F and G are taken at the point $y(\hat{\Theta}), \hat{w}$. The important lesson from (44) is that, once we have found a $\hat{w}$ that satisfies (43), no further search over w is necessary to compute the whole Jacobian $\frac{\partial F}{\partial \Theta}$. If solving for w is a non-trivial task, as it is in our model, computing the Jacobian is not much more costly than evaluating F. This means that Newton-type methods are much more competative with respect to fixed point iteration methods.

To fully exploit the power of automatic differentiation in the outer loop, one has to apply some techniques of elimination of common subexpressoins. This is explained in more detail in Appendix [to be completed].

4 Parameter values

Since the focus of this paper is on the computational methodology, 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 growth, ϵ, takes the values $(-\Phi, -0.02, 0, 0.02 + \Phi/3,)$ with probabilities $(0.1, 0.3, 0.3, 0.3)$, respectively. The shock to the AR process for z, ϵ_z, takes the values $(-0.05, 0, 0.05)$ with probabilities $(0.3, 0.4, 0.3)$ respectively. The shock to variability, ϵ_s, takes the values $(-0.5, 0, 0.5,)$ with probabilities $(0.3, 0.4, 0.3)$ respectively. The parameter Φ is varied between 0.06 and 0.2, to capture the idea of large negative shocks (disaster or crisis). A value of 0.2, occuring with probability 0.1 in a model with 12 cohorts, means that every 50 years there is a shock that lowers TFP permanently by 20 percent.

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 nc 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.

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

$$F(k, l) = k^{\alpha} l^{(1-\alpha)} \tag{45}$$

with the output share of capital set to $\alpha = 0.4$.

The depreciation rate for capital is set to $\delta = 1 - 0.9^{n_y}$. The risk aversion parameter is varied between $\gamma = 2$ and $\gamma = 6$.

For the capital accumulation function, I follow Christiano and Fisher (2003) and use the functional form

$$g(k, i) = \left[a_1 k^{\psi} + a_2 i^{\psi}\right]^{1/\psi} \tag{46}$$

with parameter values $a_2 = \delta^{1-\psi}$ and $a_1 = (1 - \delta)$. This gives $g(k, \delta k) = k$ and $g_i(k, \delta k) = 1$, which implies that the adjustment cost structure does not affect the steady state.

$$g_i(k, i) = \left[a_1 k^{\psi} + a_2 i^{\psi}\right]^{\frac{1-\psi}{\psi}} a_2 i^{\psi-1} \tag{47}$$

this requires $\left[a_1 + a_2 \delta^{\psi}\right] = 1$ and $\left[a_1 + a_2 \delta^{\psi}\right]^{\frac{1-\psi}{\psi}} a_2 \delta^{\psi-1} = 1$ which gives $a_2 = \delta^{1-\psi}$ and $a_1 = (1 - \delta)$.

For the parameter ψ, I use their estimate $\psi = 0.23$.

Degree	**Mean absolute error**					
	$Euler_1$	$Euler_2$	$Euler_3$	$Euler_4$	$Euler_5$	$\sum_i A_i$
1	8.38e-3	8.04e-3	7.42e-3	9.47e-3	4.33e-3	2.14e-2
2	7.83e-4	7.29e-4	6.69e-4	7.84e-4	3.14e-4	1.35e-3
3	7.97e-6	7.50e-6	7.31e-6	8.74e-6	3.90e-6	1.58e-5
4	2.26e-6	2.19e-6	2.13e-6	2.13e-6	2.63e-6	4.30e-6
	Max absolute error					
1	1.77e-2	1.67e-2	1.50e-2	1.86e-2	9.94e-3	4.57e-2
2	3.74e-3	3.59e-3	3.25e-3	3.88e-3	1.32e-3	5.09e-3
3	8.34e-5	7.86e-5	7.69e-5	8.90e-5	3.63e-5	1.14e-4
4	2.96e-5	2.89e-5	2.66e-5	3.05e-5	3.16e-5	5.32e-5
	100× Mean asset and risk premium					
	A_1	A_2	A_3	A_4	A_5	$RiskPr$
1	0.607	-1.103	-3.478	0.686	3.289	4.138
2	0.869	-0.480	-2.566	-0.083	2.260	4.301
3	0.940	-0.384	-2.408	-0.209	2.061	4.310
4	0.940	-0.385	-2.408	-0.209	2.062	4.309
	100× Standard dev. asset and risk premium					
1	0.063	0.274	0.432	0.173	0.475	2.642
2	0.091	0.222	0.453	0.145	0.459	2.746
3	0.093	0.216	0.454	0.138	0.458	2.750
4	0.093	0.216	0.454	0.138	0.458	2.750
	Correlation with output, asset and risk premium					
1	0.375	0.036	-0.016	-0.085	-0.026	-0.100
2	0.303	-0.090	-0.146	0.024	0.120	-0.092
3	0.304	-0.067	-0.128	-0.047	0.111	-0.091
4	0.304	-0.066	-0.128	-0.047	0.111	-0.091

Table 1: Accuracy in the model with 6 cohorts, 2 assets, $\Phi = 0.06$

5 Results

Table 1 reports results for the model with only 6 cohorts, 2 assets and moderately sized shocks ($\Phi = 0.06$). The first two parts of the table show 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 8 state variables (the exogenous states g_t, z_t, σ_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.

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^-2, rather uniformly for all the residuals. (One should notice that Euler residuals are expressed here as absolute errors, not relative; since martinal utilities are larger than 1, relative errors are smaller by about one order of magnitude.) Increasing the degree of approximation to 4 reduces the mean absolute errors to about 10^-6 and the max absolute errors to about 10^-5. Do these differences in accuracy matter for the economic results? The last 3 parts of the table report mean values, standard deviations and correlation with outpout for 6 different variables: the holdings of the safe asset for cohorts 1–5, as well as the risk premium, defined as the expected returns to capital relative to the return of the safe asset. We see that asset holdings are a senstive variable: The results for linear approximation are quite different than the precise results that are obtained with degrees 3 or 4. This is true for mean values, standard deviations and also the correlation with output. Interestingly, the risk premium is already captured well by the linear approximation. This might come as a surprise, and one should notice that a *global* linear approximation is very different from the *local* linear approximations that arise from linearization about the deterministic steady state. The risk premium in this model will be analyzed in more detail below.

A noteworthy result result here is 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 1 provides the same information for the model with 12 cohorts, and with a larger shock ($\Phi = 12$). For reasons of space, results are reported for odd-aged cohorts. The increase in the number of cohorts brings a reduction in accuracy, by a factor of about 10, for the same

Degree	**Mean absolute error**						
	$Euler_1$	$Euler_3$	$Euler_5$	$Euler_7$	$Euler_9$	$Euler_{11}$	$\sum_i A_i$
1	2.39e-3	2.42e-3	2.41e-3	2.25e-3	1.94e-3	1.07e-3	7.08e-3
2	3.39e-4	3.11e-4	2.66e-4	2.18e-4	7.53e-5	1.04e-4	2.10e-4
3	8.39e-5	8.29e-5	7.96e-5	7.36e-5	7.82e-5	1.08e-4	1.29e-4
	Max absolute error						
1	7.44e-3	7.37e-3	7.69e-3	5.55e-3	8.01e-3	2.43e-3	3.01e-2
2	5.16e-3	5.59e-3	5.76e-3	3.48e-3	1.15e-3	9.81e-4	1.34e-3
3	7.79e-4	8.14e-4	7.81e-4	6.59e-4	6.24e-4	9.40e-4	1.10e-3
	100× Mean asset and risk premium						
	A_1	A_3	A_5	A_7	A_9	A_{11}	$RiskPr$
1	7.832	16.643	18.138	-0.718	-25.403	-13.557	1.555
2	8.585	16.483	17.481	-1.537	-24.619	-13.156	1.580
3	8.465	16.358	17.554	-1.210	-24.658	-13.221	1.581
	100× Standard dev. asset and risk premium						
1	1.631	2.065	4.510	2.163	4.283	3.852	0.873
2	2.447	2.998	4.652	2.114	4.880	3.852	0.888
3	2.142	2.585	4.537	2.096	4.521	3.710	0.888
	Correlation with output, asset and risk premium						
1	0.011	0.103	0.069	0.042	-0.089	-0.065	-0.054
2	0.545	0.520	0.220	-0.347	-0.396	-0.222	-0.052
3	0.459	0.491	0.207	-0.256	-0.360	-0.191	-0.053

Table 2: Accuracy in the model with 12 cohorts, 2 assets, $\Phi = 0.12$

Shock size	Risk prem.	Stdev. ExcRet	Annual Sharpe
0.08	0.0143	0.0578	0.1108
0.12	0.0158	0.0589	0.1200
0.15	0.0174	0.0599	0.1297
0.20	0.0210	0.0619	0.1516

Table 3: Risk premium as a function of disaster size

approximation order. Looking at more results (not reported here), the loss in accuracy is more due to the increased number of cohorts than the increased size of the shock. The reason is probably that the larger number of cohorts makes it harder to find a reasonable grid, but this conjecture remains to be confirmed.

I have computed the solution only up to degree 3. This already involves solving simultaneously for 8160 parameters, the fourth-degree approximation would require 36720 parameters, which needs more RAM than what a normal personal computer currently has.

5.1 The price of risk

Table 3 displays the risk premium as a function of the disaster size Φ. Notice that the risk premium is counted for one model period, which in the case of 12 cohorts should be interpreted as 5 years. Converted to annual levels, the premium is around 0.4 percent. While not trivial this is still an order of magnitude below what is observed in the data for stocks. However, this is mainly due to the fact that the price of capital varies much less than the price of stocks in the data. Measuring the price of risk as the annualized Sharpe ratio (defined as the risk premium divided by the standard deviation of the excess return, and divided by the square root of the model period, which is 5), the value ranges between 0.11 and 0.15. This is not too far away from realistic values for the stock market, which are around 0.25. This is somewhat remarkable, given that we use a risk aversion parameter $\gamma = 4$, which is a rather moderate degree of risk aversion, compared to what is common in the financial literature on the equity premium (Bansal and Yaron 2004).

5.2 Short-sale constraints

Table 4 provides information on the model with 12 cohorts, two assets, and a lower bound on asset holdings. I impose a very tight constraint: $A_i \geq -0.01$. Unconstrained debt is much higher than what is allowed here (compare Tables 4 and Tables 2). The constraint is always binding for the older cohorts, never binding for the young cohorts, and occasionally binding for intermediate cohorts.

Degree	**Mean absolute error**						
	$Euler_1$	$Euler_3$	$Euler_5$	$Euler_7$	$Euler_9$	$Euler_{11}$	$\sum_i A_i$
1	2.39e-3	2.44e-3	2.45e-3	2.26e-3	1.92e-3	1.06e-3	7.36e-3
2	3.60e-4	3.29e-4	2.85e-4	2.37e-4	6.06e-5	1.00e-4	2.38e-4
3	1.26e-4	1.24e-4	1.22e-4	1.15e-4	1.20e-4	1.53e-4	1.84e-4
	Max absolute error						
1	6.40e-3	7.11e-3	7.39e-3	5.79e-3	6.87e-3	3.06e-3	3.03e-2
2	5.37e-3	5.48e-3	5.08e-3	3.27e-3	6.17e-4	5.17e-4	1.25e-3
3	1.46e-3	1.44e-3	1.40e-3	1.33e-3	1.33e-3	2.11e-3	2.04e-3
	100× Mean asset and risk premium						
	A_1	A_3	A_5	A_7	A_9	A_{11}	$RiskPr$
1	-0.135	2.647	-0.184	-1.000	-1.000	-1.000	1.690
2	-0.027	2.492	-0.005	-1.000	-1.000	-1.000	1.713
3	-0.002	2.477	0.016	-1.000	-1.000	-1.000	1.715
	100× Standard dev. asset and risk premium						
1	0.822	0.337	1.089	0.000	0.000	0.000	0.936
2	0.878	0.408	1.271	0.000	0.000	0.000	0.950
3	0.885	0.393	1.283	0.000	0.000	0.000	0.951
	Correlation with output, asset and risk premium						
1	-0.036	-0.028	0.037	0.000	0.000	0.000	-0.051
2	0.141	0.241	-0.208	-0.000	-0.000	-0.000	-0.045
3	0.077	0.218	-0.145	0.000	0.000	0.000	-0.047

Table 4: Accuracy in the model with 12 cohorts, 2 assets, short-sale constraints, $\Phi = 0.12$

The inequality constraint makes it harder to obtain high precision. The third degree aproximation is not consistently better than second-degree approximation, and precision is somewhat lower than in the unconstrained case, but only by a factor of about two.

The risk premium goes up when households are constrained in asset holdings. For example, with $\Phi = 0.12$, it increases from 0.158 to 0.171. A careful modeling of the constraints is therefore important for calibrating the price of risk.[3]

5.3 Intergenerational Insurance

Figure 1 displays impulse responses to all three shocks, both for the model with 1 and for the model with 3 assets. Results are shown for the model with 6 cohorts, for better readibility of the graphs. With 12 cohorts, the results are qualitatively very similar.

In the graphs, each line follows a specific cohort. The single point refers to the cohort 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 last period (cohorts 2 to J) insure each other as good as they can; as a consequence, given the iso-elastic utility, all of these cohorts face the same percentage reduction in consumption. In the graph, these lines would be on top of each other. The new cohorts (born at the time of the shock or later) can potentially face a very different change in consumption.

The picture shows that markets are not complete. In particular, the oldest cohorts behave very differently than the younger cohorts. Going from one asset to three, this difference is partically closed, in particular in reponse to ϵ_z. This is not surprising, because the return on the third asset is conditional on this shock.

The rsults show that it is not easy to come close to market completeness when the economy is subject to several shocks. Further results (not reported in detail here), 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 reminiscent of Rios-Rull (1994), and more recently Hasanhodzic and Kotlikoff (2013). But things are very different when shocks are non-neutral, such as affecting the productivity of some cohorts more than others. For questions of inter-generational insurance, it is therefore essential to pin down the exact nature of the shocks.

[3]In results not reported here, the risk premium goes up substantially if young cohorts are excluded from investing in risky capital, thereby imposing all the risk on the older cohorts.

5.4 Many cohorts

It is possible to compute the model with 60 cohorts, which can be interpreted as annual frequency, but only with a linear approximation. Models with higher frequency can be more easily brought to the data, and global linear approximation turned out to give meaningful approximations to aggregate statistics, including financial statistics, in the model with fewer cohorts, although the results for household asset choice were not high. [MORE DETAILS TO BE INCLUDED.]

6 Conclusions

The paper has shown that it is possible to compute OLG with a medium number of cohorts, several assets and sort-selling constraints with high precision. Computing times range from several minutes to half an hour on a regular personal computer.

If subject to infrequent but large shocks, the price of risk in these models can be similar to what is observed in the data, with a relatively low degree of risk aversion.

References

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

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.

Caldara, D., J. Fernandez-Villaverde, J. Rubio-Ramirez, and W. Yao (2012, April). Computing dsge models with recursive preferences and stochastic volatility. *Review of Economic Dynamics 15*(2), 188–206.

Christiano, L. J. and J. D. M. Fisher (2003, October). Stock Market and Investment Goods Prices: Implications for Macroeconomics. NBER Working Papers 10031, National Bureau of Economic Research, Inc.

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, June). 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.

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

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 (2013). Smolyak method for solving dynamic economic models: Lagrange interpolation, anisotropic grid and adaptive domain. NBER Working Papers 19326, National Bureau of Economic Research, Inc.

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 96*, 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.

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.

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

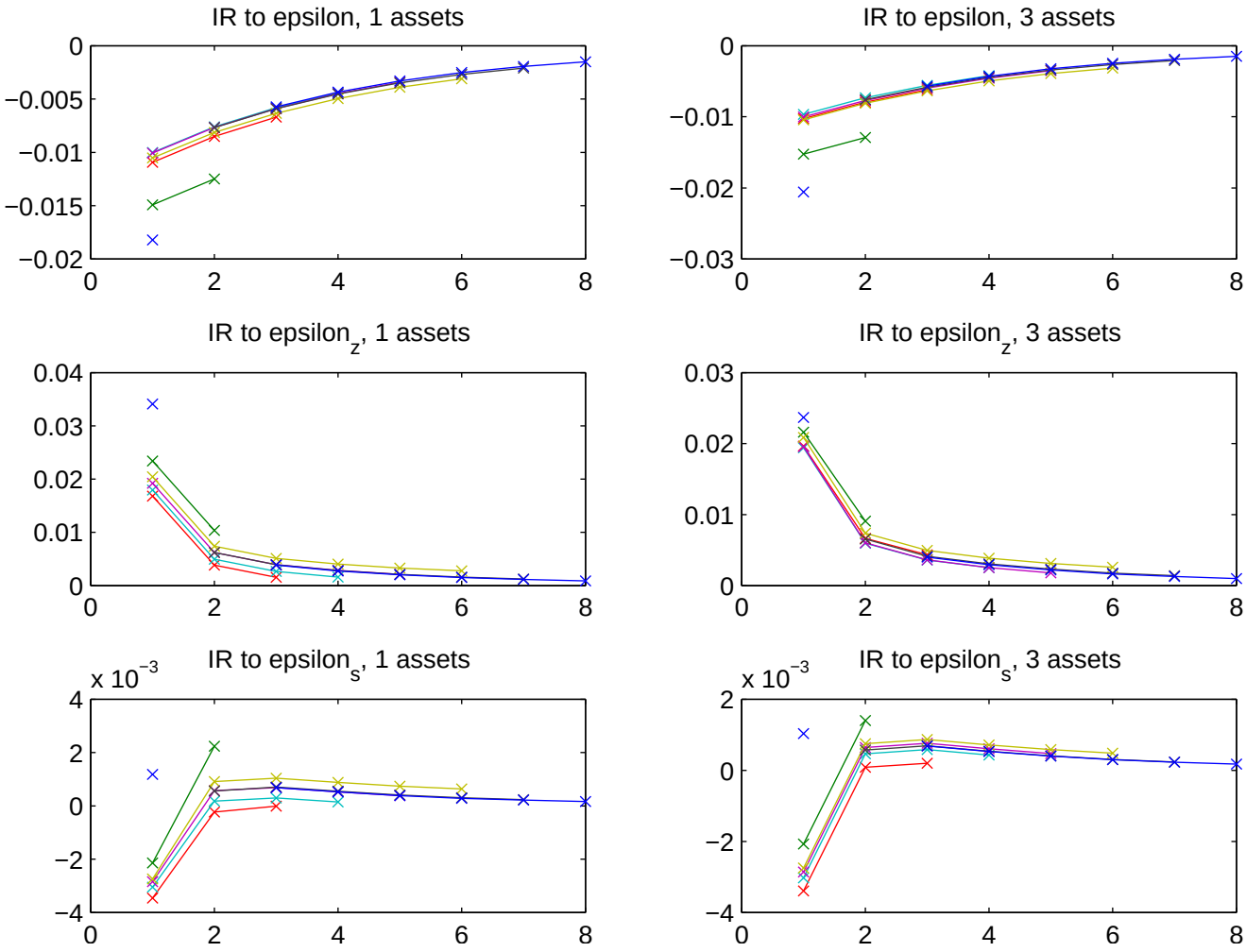

Figure 1: Impulse responses