Nonlinear Solution of Heterogeneous Agent Models by Approximate Aggregation*

Michael Reiter, Institute for Advanced Studies, Vienna

February 2010

Abstract

The paper deals with the computation of DSGE models with a large number (or continuum) of heterogeneous agents and incomplete markets. Solving this model requires approximate aggregation, representing the cross-sectional distribution by a finite number of state variables. The existing literature computes nonlinear solutions with a very low-dimensional state vector, or high-dimensional approximations where the solution is linear in aggregate states. This paper shows how to compute precise higher-order approximations with a medium-dimensional state vector. This is made possible by using a backward induction algorithm that exploits the information obtained from the high-dimensional linear solution.

The method is illustrated with a version of the Krusell/Smith model. A quadratic approximation with up to 15 state variables can be computed in a few minutes on a PC, running Matlab.

JEL classification: C63, C68, E21
Keywords: heterogeneous agents; projection methods; perturbation methods; approximate aggregation

Address of the author:

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

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

*I have benefited from discussions with seminar participants in Vienna and Zurich. Support from the Austrian National Bank under grant Jubiläumsfonds 13038 is gratefully acknowledged.

1 Introduction

Stochastic general equilibrium models with incomplete markets and a continuum of heterogeneous agents play a more and more important role in macroeconomics. Early contributions to the numerical solution of these models are Krusell and Smith (1998) and Den Haan (1996). The former approach is still the most widely used method to handle those models. The last few years have seen an increasing effort to develop methods that can handle more complicated models or yield an even more accurate solution (cf. Preston and Roca (2007) and Den Haan and Rendahl (2010) and the contributions summarized in Den Haan (2010)). Common to all those approaches so far is that they use approximate aggregation with a very small state vector. The relevant cross-sectional distribution in the model is summarized by a small number of statistics.

The present paper is a further step in a research program that aims at obtaining accurate solutions with a higher-dimensional state vector. The first step was done in Reiter (2009b), developing a method that uses a high-dimensional, non-parametric approximation to the cross-sectional distribution. The method allows for a nonlinear relationship between individual decisions and individual states, so that it can handle, for example, a consumption function with borrowing constraints. However, the relationship between individual decisions and aggregate states is linearized, which allows to compute a solution with many state variables, in the range of 1000-2000 on a PC. This method can be seen as an analogue to the linearization methods that have been widely used for business cycle models, generalized to models of heterogeneous agents and incomplete markets.

The method in Reiter (2009b) avoids aggregation problems by using a very high-dimensional state vector. One could say that this is overkill: studying the aggregation problem, Reiter (2009a) shows that it is possible to reduce the number of state variables substantially, and still get a solution that is precise basically up to machine precision. Reducing accuracy a little bit, the paper further shows how to obtain very precise solutions of the linearized model with a medium-dimensional state vector.

The topic of the present paper is the computation of higher-order approximations with a medium-dimensional state vector. In a concrete example, I compute a solution where the decision of economic agents is a quadratic function in the aggregate state variables (extensions to higher-order polynomials are straightforward). The state vector includes up to 15 statistics of the cross-sectional distribution. The method is based on the backward-iteration approach of Reiter (2010), which was developed as an alternative to the Krusell/Smith algorithm, but so far has only been applied with a low-dimensional state (one or two moments of the distribution; cf. (Den Haan 2010) for a comparison of the performance of the different algorithms). Here I am going to show how to use this approach for higher-order solutions with a medium-dimensional state vector. To make this possible, it is crucial to exploit the information obtained in the linearized solution. In this paper, I restrict attention to the case of small shocks, studying the solution in a small neighborhood of the steady state. This can be seen as the analogue to the second-order approximations that are now widely used in the analysis of DSGE models (Judd and Guu 1993; Collard and Juillard 2001; Schmitt-Grohé and Uribe 2004). In the conclusions I will briefly discuss how

to extend this approach to the case of big shocks.

The plan of the paper is as follows. Section 2 presents the example economy to which we apply the method, a variant of the Krusell/Smith model. Section 3 discusses approximate aggregation in a linearized setup. Section 4 presents the method for higher-order approximations. Section 5 presents numerical results, and Section 6 concludes. The Matlab programs are available at `http://elaine.ihs.ac.at/~mreiter/nlaa.tar.gz`.

2 The Stochastic Growth Model With Heterogeneous Agents and Incomplete Markets

To keep the exposition as transparent as possible, I apply the method to a simple version of the Krusell/Smith model. The model is very similar to Reiter (2009b) and Reiter (2009a). The following exposition of the model closely follows the latter paper.

There is a continuum of infinitely lived households of unit mass. Households are ex ante identical, and differ ex post through the realization of their individual labor productivity. They supply their labor inelastically. Production takes place in competitive firms with constant-returns-to-scale technology. A government is introduced into this model to the sole purpose of creating some random redistribution of wealth. This helps to identify the effect of the wealth distribution on the dynamics of aggregate capital.

2.1 Production

Output is produced by perfectly competitive firms, using the Cobb-Douglas gross production function

$$Y_t = \mathcal{Y}(K_{t-1}, L_t, \theta_t) = A\theta_t K_{t-1}^{\alpha} L_t^{1-\alpha}, \qquad 0 < \alpha < 1 \tag{1}$$

where A is a constant. Production at the beginning of period t uses K_{t-1}, the aggregate capital stock determined at the end of period $t-1$. Since labor supply is exogenous, and individual productivity shocks cancel due to the law of large numbers, aggregate labor input is constant and normalized to $L_t = 1$, cf. Section 2.3. Aggregate capital is obtained from summing over all households, cf. (19).

The aggregate resource constraint of the economy is

$$K_t = (1-\delta)K_{t-1} + Y_t - C_t \tag{2}$$

where δ is the depreciation rate and C_t is aggregate consumption. The aggregate productivity parameter θ_t follows the AR(1) process

$$\log \theta_{t+1} = \rho_\theta \log \theta_t + \epsilon_{\theta,t+1} \tag{3}$$

where ϵ_θ is an i.i.d. shock with expectation 0 and standard deviation σ_z. The before tax gross interest rate $\bar{R}_t$ and wage rate $\bar{W}_t$ are determined competitively:

$$\bar{R}_t = 1 + \mathcal{Y}_K(K_{t-1}, L_t, \theta_t) - \delta \tag{4}$$

$$\bar{W}_t = \mathcal{Y}_L(K_{t-1}, L_t, \theta_t) \tag{5}$$

2.2 The Government

The only purpose of the government is to create some random redistribution between capital and labor. In period t, the government taxes the capital stock accumulated at the end of period $t-1$ at rate τ_t^k, and labor at rate τ_t^l, so that after tax gross interest rate R_t and wage W_t are related to before tax prices by

$$R_t = \bar{R}_t - \tau_t^k \tag{6}$$
$$W_t = \bar{W}_t(1-\tau_t^l) \tag{7}$$

The tax rate on capital follows an AR(1) process around its steady state value τ^{k*}:

$$\tau_{t+1}^k - \tau^{k*} = \rho_\tau(\tau_t^k - \tau^{k*}) + \epsilon_{\tau,t+1} \tag{8}$$

where ϵ_τ is an i.i.d. shock with expectation 0 and standard deviation σ_τ. The labor tax is determined by a balanced-budget requirement

$$\tau_t^k K_{t-1} + \tau_t^l \bar{W}_t L_t = 0 \tag{9}$$

2.3 The Household

There is a continuum of households, indexed by h. Each household supplies inelastically one unit of labor. Households differ ex post by their labor productivity $\xi_{t,h}$, which is assumed to be i.i.d. both over households and over time. It is normalized to have unit mean:

$$\mathrm{E}\,\xi_{t,h} = \mathrm{E}_{t-1}\,\xi_{t,h} = 1 \tag{10}$$

Net labor earnings are then given by

$$y_{t,h} = W_t(1-\tau_t^l)\xi_{t,h} \tag{11}$$

Household h enters period t with asset holdings $k_{t-1,h}$ left at the end of the last period. It receives the after tax gross interest rate R_t on its assets, such that the available resources after income of period t ("cash on hand") are given by

$$\chi_{t,h} = R_t k_{t-1,h} + y_{t,h} \tag{12}$$
$$\tag{13}$$

Cash on hand is split between consumption and asset holdings:

$$k_{t,h} = \chi_{t,h} - c_{t,h} \tag{14}$$

We impose the borrowing constraint

$$k_{t,h} \geq \underline{k} = 0 \tag{15}$$

2.4 Finite Approximation of the Model Equations

2.4.1 Household Productivity, Consumption Function and the Euler Equation

Individual productivity $\xi_{t,h}$ has a discrete distribution, taking on the values $\bar{\xi}_j$ with probabilities $\omega_\xi(j)$ for $j = 1, \ldots, n_y$.[1]

In each period t, I represent the savings function $k_t(\chi)$ by n_s+1 numbers: a critical level χ_t^c where the borrowing constraint starts binding, and the values $k_{t,i}$, $i = 1, \ldots, n_s$ at n_s knot points $\bar{\chi}_{t,i}$ with $\bar{\chi}_{t,i} > \chi_t^c$. The knot points are chosen as $\bar{\chi}_{t,0} = \chi_t^c$ and $\bar{\chi}_{t,i} = \chi_t^c + \bar{\bar{\chi}}_i$, $i = 1, \ldots, n_s$ with some fixed set of grid points $\bar{\bar{\chi}}_i$. These $n_s + 1$ numbers are stacked into the decision vector d_t.

Off the knot points $\bar{\chi}_{t,i}$, the savings function is approximated by

$$\hat{k}(\chi; d_t) = \begin{cases} \underline{k} & \text{for } \chi \leq \chi_t^c \\ CSI(\chi; d_t) & \text{for } \chi_t^c < \chi \leq \bar{\chi}_{t,n_s} \\ k_{t,n_s} + CSI'(\bar{\chi}_{t,n_s}; d_t)(\chi - \bar{\chi}_{t,n_s}) & \text{for } \chi > \bar{\chi}_{t,n_s} \end{cases} \tag{16}$$

where $CSI(\chi; d_t)$ stands for "cubic spline interpolation", the natural cubic spline that interpolates the points $(\chi_t^c, \underline{k})$, $(\bar{\chi}_{t,i}, k_{t,i})$, $i = 1, \ldots, n_s$. Beyond the last knot point $\bar{\chi}_{t,n_s}$, we approximate the savings function by a straight line, with the slope given by the derivative of the spline at k_{t,n_s}.

The consumption function is then given by $\hat{\mathcal{C}}(\chi; d_t) \equiv \chi - \hat{k}(\chi; d_t)$. The approximation of the saving function has $n_s + 1$ degrees of freedom. We therefore apply a collocation method and require the household Euler equation to hold at the knot points $\bar{\chi}_{t,i}$:

$$U'\left(\hat{\mathcal{C}}(\bar{\chi}_{t-1,i}; d_{t-1})\right) = \beta \sum_{j=1}^{n_y} \omega_\xi(j) \left[(\bar{R}_t - \tau_t^k) U'\left(\hat{\mathcal{C}}(\bar{\chi}^{ij}; d_t)\right)\right] + \eta_{it}^c, \qquad i = 0, \ldots, n_s \tag{17a}$$

where

$$\bar{\chi}^{ij} \equiv (\bar{R}_t - \tau_t^k)\left(\bar{\chi}_{t-1,i} - \hat{\mathcal{C}}(\bar{\chi}_{t-1,i}; d_{t-1})\right) + \bar{W}_t(1 - \tau_t^l)\bar{\xi}_j \tag{17b}$$

For any $\chi < \chi_t^c$, the Euler equation does not hold with equality, but we know that $\hat{\mathcal{C}}(\chi; d_t) = \chi$. Notice that Equ. (17a) uses the notation of Sims (2001): the η_{it}^c are the expectation errors that result from the aggregate shocks (idiosyncratic shocks are handled by summing over the quadrature points). They are determined endogenously in the solution of the system.

2.4.2 Wealth Distribution

In the model with a continuum of agents, the ergodic cross-sectional distribution of wealth has an infinite number of discrete mass points, because the distribution of idiosyncratic productivity is discrete, so that households at the borrowing constraint $\underline{k} = 0$ return to

[1] The main difference between the present model and Reiter (2009b) is that in the latter, productivity is assumed to have a continuous distribution. The current setup is considerably simpler.

the region of positive k in packages of positive mass. I approximate this complicated distribution by a finite number of mass points at a predefined grid $\underline{k} = \bar{k}_1, \bar{k}_2, \ldots, \bar{k}_{n_k} = k^{max}$. The maximum level k^{max} must be chosen such that in equilibrium very few households are close to it.

The key element of the approximation is the following. If the mass ϕ of households in period t saves the amount $\tilde{k}$ with $\bar{k}_i \leq \tilde{k} \leq \bar{k}_{i+1}$, I approximate this by assuming that $\phi \cdot \psi(i, \tilde{k})$ households end up at grid point $\bar{k}_i$, while $\phi \cdot \psi(i+1, \tilde{k}) = \phi \cdot (1 - \psi(i, \tilde{k}))$ households end up at grid point $\bar{k}_{i+1}$. This random perturbation of capital is done such that aggregate capital is not affected, so we require that $\psi(i, \tilde{k})\bar{k}_i + \psi(i+1, \tilde{k})\bar{k}_{i+1} = \tilde{k}$. This is achieved by defining

$$\psi(i,k) \equiv \begin{cases} 1 - \frac{k - \bar{k}_i}{\bar{k}_{i+1} - \bar{k}_i} & \text{if } \bar{k}_i \leq k \leq \bar{k}_{i+1} \\ \frac{k - \bar{k}_{i-1}}{\bar{k}_i - \bar{k}_{i-1}} & \text{if } \bar{k}_{i-1} \leq k \leq \bar{k}_i \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

The function $\psi(i,k)$ gives the fraction of households with savings k which end up at grid point $\bar{k}_i$. For any k, $\psi(i,k)$ is non-negative and $\psi(i,k) > 0$ for at most two values of i.

Define $\phi_t(i)$ as the fraction of households at time t that have capital level $\bar{k}_i$. Then we can write aggregate capital as

$$K_t = \sum_{i=1}^{n_k} \bar{k}_i \phi_t(i) \tag{19}$$

Stack the $\phi_t(i)$ into the column vector

$$\mathbf{\Phi}_t \equiv \begin{bmatrix} \phi_t(0) \\ \vdots \\ \phi_t(n_k) \end{bmatrix} \tag{20}$$

which describes the cross-sectional distribution of capital at time t. We can now describe the dynamics of the capital distribution for a given savings function. The transition from the end-of-period distribution $\mathbf{\Phi}_{t-1}$ to $\mathbf{\Phi}_t$ is given by the linear dynamic equation

$$\mathbf{\Phi}_t = \Pi\left(d_t; w_t, R_t\right) \mathbf{\Phi}_{t-1} \tag{21}$$

where the elements of the transition matrix Π_t are given by

$$\Pi_{i',i}\left(d_t; w_t, R_t\right) = \sum_{j=1}^{n_y} \omega_\xi(j) \psi(i', R_t \bar{k}_i + W_t \xi_{t,j} - \hat{\mathcal{C}}(R_t \bar{k}_i + W_t \xi_{t,j}; d_t)) \tag{22}$$

From the properties of $\psi(i,k)$, each column of Π has at most $2n_y$ non-zero elements. In should be understood in (22) that factor prices R_t and W_t are functions of $(\mathbf{\Phi}_{t-1}, z_t)$.

2.4.3 The Discrete Model

In the discrete model, aggregate capital K is given by (19). Aggregate consumption can be written in the same way. Because of inelastic labor supply and the assumptions (10) about labor productivity, the law of large numbers[2] implies that aggregate effective labor is given by $L_t = 1$.

With the approximations of Sections 2.4.2 and 2.4.1 the model is reduced to a finite set of equations in each period t. To write it with the minimum number of variables, we can say that the discrete model consists of the equations (3), (8), (17) and (21). These equations define, for each period t, a system of $n_s + n_k + 3$ equations in just as many variables: d_t, $\mathbf{\Phi}_t$, θ_t and τ_t^k, understanding that the variables τ_t^l, $\bar{R}_t$, $\bar{W}_t$ and K are defined through (4), (5), (9) and (19). Notice that one of those variables is redundant, since $\sum_i \phi_t(i) = 1$. Correspondingly, one equation in (21) is linearly dependent from the others, since all rows in Π_t add up to one.

It should be stressed that the accuracy checks of Section 5 always refer to this discrete model. The question whether the discrete model is a good approximation to the theoretical model with a continuum of agents is beyond the scope of this paper.

2.5 Parameter Values and Functional Forms

The frequency of the model is quarterly. Standard values are used for most of the the model parameters: $\beta = 0.99$, $\alpha = 1/3$, $\delta = 0.025$. I use log utility, $U(c) = \log(c)$. For the technology shock I choose $\rho_\theta = 0.95$ and $\sigma_z = 0.007$, which again are standard values. I choose the tax shock as uncorrelated, $\rho_\tau = 0$, to create unpredictable short-run redistributions. Taxes fluctuate around zero, so $\tau^{k*} = 0$. The variability of the tax shock is set, rather arbitrarily, to $\sigma_\tau = 0.01$.

Individual productivity $\xi_{t,h}$ is modeled to have only two realizations of equal probability. The two realizations were chosen such that $Var(\xi_{t,h}) = 0.061/4$, corresponding to the size of the transitory shock in the RIP income specification of Guvenen (2009).

I use $n_k = 1000$ grid points for the distribution of capital and a spline of order $n_s = 100$ for the consumption function, so that the discrete model has around 1100 variables. The linearized version of the model can then be solved exactly, for example by the method of Sims (2001).

3 Approximate Aggregation in the Linearized Model

Before computing a higher-order solution of the model, which will be done in Section 4, it is first necessary to compute the solution of the linearized model. This involves the following steps:

1. Finding the steady state of the discretized model without aggregate shocks.

[2] A law of large numbers for economies with a continuum of agents, using standard analysis and measure theory, is given in Podczeck (2009).

2. Linearizing the model equations around the steady state.

3. Solving the complete linearized model, and/or solving the linearized model with a reduced state vector (approximate state aggregation).

This has been explained in detail in Reiter (2009b), except for approximate aggregation, which is the topic of the rest of this section.

3.1 The Linearized Model

In the discrete model of Section 2.4.3, the natural state variables are $\mathbf{\Phi}_t$, describing the cross-sectional distribution of capital at the end of t, and the exogenous driving forces, θ_t and τ_t^k. To make the notation more compact, we define $z_t \equiv (\theta_t, \tau_t^k)'$ and $\rho_z = \begin{pmatrix} \rho_\theta & 0 \\ 0 & \rho_\tau \end{pmatrix}$, so that $z_t = \rho_z z_{t-1} + \epsilon_t$ Furthermore, we stack all state variables into the vector[3] $x_t \equiv \begin{bmatrix} \mathbf{\Phi}_t \\ z_t \end{bmatrix}$. We see from the household Euler equation (17) that private decisions depend on the aggregate state only through the after tax factor prices. Those in turn only depend on aggregate capital K and the exogenous driving forces θ and τ^k. We therefore introduce a vector of statistics of the capital distribution, m_t, which contains, as a minimum, information about average capital K, but may contain additional variables. The aggregate state then enters the Euler equation only through m_t and z_t, which we stack into the reduced state $\hat{x}_t$. We assume that m_t only contains linear functions of the distribution $\mathbf{\Phi}_t$:

$$\hat{x}_t \equiv \begin{bmatrix} m_t \\ z_t \end{bmatrix} = H x_t = \begin{bmatrix} H_{11} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_t \\ z_t \end{bmatrix} \tag{23}$$

for some given matrix H_{11}. In the following, I will often refer to m_t as "moments", although it may include statistics of the distribution that are not moments, as long as they are of the linear form (23).

After linearization around the steady state, the discrete model can be written as

$$x_t = T x_{t-1} + D d_t + F \epsilon_t \tag{24a}$$

$$0 = \mathrm{E}_{t-1}\left[\mathcal{E}_{m1}\hat{x}_{t-1} + \mathcal{E}_{d1} d_{t-1} + \mathcal{E}_{m0}\hat{x}_t + \mathcal{E}_{d0} d_t\right] \tag{24b}$$

The matrices T, D, F, $\mathcal{E}_{m1}$, $\mathcal{E}_{m0}$, $\mathcal{E}_{d1}$ and $\mathcal{E}_{d0}$ are the result from the linearization. In (24), all variables are defined as deviations from steady state.

The dynamic equation (24a) still contains the high-dimensional state vector x_t (in the theoretical model with a continuum of agents, the state vector is of infinite dimension). The idea behind approximate aggregation is that economic agents base their decisions at

[3] A remark about the timing of variables. At the time when households make the decisions of period t, the natural state variables are $\mathbf{\Phi}_{t-1}$, the cross-sectional distribution of capital at the end of $t-1$, and the exogenous vector z_t. In the nonlinear part of the algorithm in Section 4, we therefore group $\mathbf{\Phi}_{t-1}$ and z_t together. In the linearized model, it is more convenient to use $\mathbf{\Phi}_t$ and z_t together.

time t not on the full vector $\mathbf{\Phi}_t$, but only on the reduced state vector $\hat{x}_t$. Then we can replace (24) by some equation defining the dynamics of $\hat{x}_t$, which has to be consistent with (24) in some approximate sense. This will be discussed in Section 3.3. Before doing this, we discuss how to choose the matrix H, and thereby the reduced state vector $\hat{x}$.

3.2 Choosing the Reduced State Vector

The literature following Krusell and Smith (1998) has used the first cross-sectional moments (typically one or two) of the distribution of capital $\mathbf{\Phi}_t$ for the statistics m_t. I will follow this approach with one modification: for higher moments, I use the moments of $\log(k)$ rather than k. Then we take the relevant part of H as

$$H_{11} = \begin{pmatrix} \bar{k}_1 & \bar{k}_2 & \dots & \bar{k}_{n_k} \\ \log(\bar{k}_1)^2 & \log(\bar{k}_2)^2 & \dots & \log(\bar{k}_{n_k})^2 \\ \vdots & \vdots & \ddots & \vdots \\ \log(\bar{k}_1)^m & \log(\bar{k}_2)^m & \dots & \log(\bar{k}_{n_k})^m \end{pmatrix} \tag{25}$$

To avoid multi-collinearity problems, I orthogonalize the rows of H.

The use of cross-sectional moments as state variables is not optimal. The problem of the household is to predict future values of aggregate capital, because this is what determines factor prices in this model. Reiter (2009a) uses techniques from systems and control theory to identify state vectors that are nearly optimal for that purpose. However, it is also shown that in the Krusell/Smith model, cross-sectional moments perform reasonably well in this respect. Therefore, I stick here to cross-sectional moments as state variables, in order to simplify the presentation. The topic of the present paper is not the choice of the reduced state vector, but rather the computation of higher-order approximations with a given, medium sized state vector. For those models where it is not clear what a reasonable set of statistics is, one should resort to the techniques developed in Reiter (2009a).

3.3 Approximate Aggregation

I will discuss two approaches to approximate aggregation. First, the well known method of Krusell and Smith (1998). Second, the Proxy Distribution approach of Reiter (2010). Then I will show that the two are equivalent in a linearized setting, but that the latter approach generalizes more easily to higher-order approximations.

3.3.1 The Linear Krusell/Smith (KS) Algorithm

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

$$\hat{x}_t = \hat{A}\hat{x}_{t-1} + \hat{B}\epsilon_t \tag{26}$$

Agents are assumed to solve their individual optimization problem assuming that the aggregate state $\hat{x}_t$ follows (26). The algorithm alternates between solving for the individual

decision function given $\hat{A}$ and $\hat{B}$, and updating $\hat{A}$ and $\hat{B}$ from the individual decision functions, until the two are consistent in an OLS sense. That means, if agents see a long realization of the model economy and estimate the model (26)), they obtain the $\hat{A}$ and $\hat{B}$ that they have used in their optimization.

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

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

2. Solve the system of equations (26) and (24b) to get the matrices $D_{\hat{X}}$ and D_E of the decision rule
$$d_t = D_{\hat{X}} \hat{x}_{t-1} + D_E \epsilon_t \tag{27}$$

3. Set A and B as
$$A = T + D D_{\hat{X}} H \tag{28a}$$
$$B = F + D D_E \tag{28b}$$
and compute Σ_x as the unique symmetric solution of
$$\Sigma_x = A \Sigma_x A' + B \Sigma_\epsilon B' \tag{29}$$
The numerical computation of Σ_x in cases where A is very large is discussed in Reiter (2009a, Appendix B.3). Notice that the matrix A has to be asymptotically stable for Σ_x to be defined. If it is not, the pair $(\hat{A}, \hat{B})$ is not admissible.

4. Update $\hat{A}$ and $\hat{B}$ by the OLS regression
$$\hat{A} = H A \Sigma_x H' \left(H \Sigma_x H' \right)^{-1} \tag{30a}$$
$$\hat{B} = H B \tag{30b}$$

5. Iterate until the results in (30) are consistent with the guess in Step 1. This can be done by a quasi-Newton algorithm over the elements of $\hat{A}$ and $\hat{B}$.

The household decision rule (27) implies the dynamic equation
$$x_t = A x_{t-1} + B \epsilon_t \tag{31}$$
with A and B defined by (28). Σ_x is the unconditional covariance matrix of the state x_t under this dynamics. Notice that, because of the linearity of the setup, the algorithm can use the asymptotic formulas (30) for the OLS estimation (cf. Reiter (2009a, Appendix B.1) for a derivation), and thereby avoid the use of simulation methods. It is therefore not affected by sampling errors, unlike the original (nonlinear) algorithm of Krusell and Smith (1998). In the choice of H, care must be taken that $H \Sigma_x H'$ is regular and not too ill-conditioned. This can be achieved by replacing every row in H by the residual of a weighted least squares regression of this row on all earlier rows, with weighting matrix Σ_x.

Definition 1. *A KS solution of the linearized model (24) consists of matrices* $(\hat{A}, \hat{B}, \Sigma_x, A, B, D_{\hat{X}}, D_E)$*, where* $\hat{A}$ *and* A *are asymptotically stable, such that (27) solves the model (26),(24b) and Equations (28)–(30) are satisfied.*

3.3.2 The Proxy Distribution (PD) Method

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

$$x_t = \begin{bmatrix} \mathbf{\Phi}^{pd}(m_t, z_t) \\ z_t \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ 0 & I \end{bmatrix} \begin{bmatrix} m_t \\ z_t \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ 0 & I \end{bmatrix} \hat{x}_t \equiv S\hat{x}_t \tag{32}$$

for suitable matrices S_{11}, S_{12} and S. Premultiplying (24a) by H, and assuming (32) in period $t-1$, we get the following dynamic equation for $\hat{x}_t$:

$$\hat{x}_t = HTS\hat{x}_{t-1} + HDd_t + HF\epsilon_t \tag{33}$$

For given S, the reduced model consists of (33) and (24b). Notice that (33) will in general not be satisfied for $x_t \neq S\hat{x}_t$.

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

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

2. Solve the system of equations (33) and (24b) to get the matrices $D_{\hat{X}}$ and D_E of the decision rule (27).

3. Define A and B as in (28) and Σ_x as in (29).

4. Update S by
$$S = \Sigma_x H' \left[H\Sigma_x H'\right]^{-1} \tag{34}$$
Return to Step 2 and iterate until convergence.

Again, matrix A in Step 3 is required to be asymptotically stable. The formula for updating S in step 4 is such that $S\hat{x}$ is the expectation of x conditional on $\hat{x}$, under the model (31) with multivariate normal shocks.

Definition 2. *A PD solution of the linearized model (24) consists of matrices* $(S, \Sigma_x, A, B, D_{\hat{X}}, D_E)$*, where A is asymptotically stable, such that (27) solves the model (33),(24b) and Equations (28), (29) and (34) are satisfied.*

3.3.3 Comparison of KS and PD Algorithms

Proposition 1. *To any set of matrices* $(\hat{A}, \hat{B}, \Sigma_x, A, B, D_{\hat{X}}, D_E)$ *being a KS solution, there is a matrix* S *such that* $(S, \Sigma_x, A, B, D_{\hat{X}}, D_E)$ *is a PD solution.*

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

Proof. For a given KS solution, choose S by (34). For a given PD solution, choose $\hat{A}$ and $\hat{B}$ by (30). Then we have to show that (26) is equivalent to (33). To do this, use (28a) in (30a) to get

$$\begin{aligned}\hat{A} &= H\left(T + DD_{\hat{X}}H\right)\Sigma_x H'\left(H\Sigma_x H'\right)^{-1} \\ &= H\left(T\Sigma_x H'\left(H\Sigma_x H'\right)^{-1} + DD_{\hat{X}}\right)\end{aligned} \tag{35}$$

Then plug (34) and (27) into (33), and then use (30b) and (35). This gives

$$\hat{x}_t = H\left[T\Sigma_x H'\left(H\Sigma_x H'\right)^{-1} + DD_{\hat{X}}\right]\hat{x}_{t-1} + HDD_E\epsilon_t + HF\epsilon_t = \hat{A}\hat{x}_{t-1} + \hat{B}\epsilon_t \tag{36}$$

which is the same as (26).

Then the only thing that remains to be shown is that the stability of A, as required in the definition of a PD solution, also implies the stability of $\hat{A}$, which is required in the definition of a KS solution. This follows from (30a) and (29), as is shown in Reiter (2009a, Appendix B.2).

□

The proposition shows that the two approaches are equivalent in the sense of having the same solution. It is nevertheless worthwhile to study both. One reason is that they lead to different algorithms to find this solution. In this respect, both methods have their advantages. The advantage of the KS algorithm is that the objects over which the algorithm iterates, namely $\hat{A}$ and $\hat{B}$, have only few elements, at least if $\dim(\hat{x})$ is small. This allows to use efficient quasi-Newton algorithms to solve the nonlinear fixed point problem of the algorithm. In contrast, the object S of the PD algorithm has so many elements that the only feasible approach is some kind of fixed point iteration, as outlined in Section 3.3.2. The good side of the PD approach is that the fixed point iteration in S converges in a few iterations in many cases. Moreover, if we know the covariance matrix Σ_x from the exact linearized solution, then we might use it in (34) and do not need to iterate on S at all. In my experience, if m_t is big enough so that it contains the information that is essential for household decision making, then the choice of the proxy distribution (the choice of S) has very little impact on the solution. In that case, we solve the model consisting of (33) and (24b) in one step.

However, the main reason why we are interested in the PD approach is that it generalizes more easily to the case of higher order solutions. If household decisions depend nonlinearly on aggregate variables, the law of motion for $\hat{x}_t$ (cf. (26)) should also be nonlinear, in order to be at least approximately consistent with behavior at the micro level. This makes it

difficult to apply the Krusell/Smith algorithm if the state vector has medium dimension (10-20, say). Then the number of parameters in the aggregate law of motion is in the range of 10^3 or more. Finding those parameters by an iterative process of guessing the parameters, solving and simulating the model, and reestimating the law of motion, appears very hard to do at the moment.

The proxy distribution method does not suffer from this problem, because it does not stipulate an aggregate law of motion. We rather enforce exact compatibility between individual and aggregate behavior at the proxy distributions. If shocks are small, it is natural to use the proxy distribution function that we have obtained in the linearized solution (cf. (34)). And again, if m_t contains the relevant information for economic agents, the exact choice S should make little difference. Given a proxy distribution function, the model can be solved in one sweep of backward iterations. This has already been shown in Reiter (2010), although the example there uses a very low-dimensional state vector. An improved algorithm for this problem will be presented in Section 4.

4 Higher-Order Approximations

Higher-order approximation means that the decision function of economic agents, and therefore the aggregate dynamics, depend nonlinearly on aggregate variables, in particular on the vector that describes the cross-sectional distribution of capital. As already mentioned above, using the proxy distribution method, we do not specify an aggregate law of motion. The only object that we approximate explicitly is the individual decision function (here, the household saving function; in other applications, it may be the value function of the agents). We will approximate the dependence of consumption on aggregate variables by polynomials.

"Solving" the model of Section 2 then means finding a consumption function that satisfies the aggregate consistency requirements at the proxy distributions. The nonlinear part of the algorithm has the following steps:

1. Setting up the framework:
   (a) Choose an approximation for the consumption function (Section 4.1).
   (b) Choose a grid of end-of-period individual capital, for the solution of the household problem. (Section 4.2.1).
   (c) Approximate the vector of aggregate shocks ϵ by a distribution with finite support, taking on values $\bar{\epsilon}_a$ with probability $\omega_a(a)$, for $a = 1, \ldots, n_a$. For the results in Section 5, I choose independent 2-point distributions for both $\epsilon_{\theta,t}$ and $\epsilon_{\tau,t}$, giving $n_a = 4$.
   (d) Choose a discrete grid in aggregate states $\hat{x}_t$ (Section 4.2.2).
   (e) Choose a proxy distribution function (Equ. (43)).

2. Initialization: choose an initial estimate for the household savings function, $\mathcal{D}_T$ (Section 4.2.3).

3. Backward iteration: Given $\mathcal{D}_t$, compute $\mathcal{D}_{t-1}$, and iterate until convergence (Section 4.3).

4.1 Approximation of the Savings Function

For given aggregate state, consumption is a nonlinear function of k, which is characterized by the (n_s+1)-vector d_t, cf. Section 2.4.1. From the linearized solution, we get each of the (n_s+1) components of d_t as a linear function of the reduced aggregate state (m_{t-1}, z_t).[4] "Higher-order approximation" means that we approximate the components of d_t as complete polynomials in the aggregate states. In the examples below, I will consider quadratic approximations. The generalization to higher orders is straightforward.

In the quadratic approximation, each polynomial has $1 + n + n(n+1)/2$ coefficients, where $n = \dim m + \dim z$. The complete savings function is then given by the $(n_s+1)(1+n+n(n+1)/2)$ coefficients of the (n_s+1) quadratic functions. We stack those coefficients into the vector $\mathcal{D}_t$. We denote by $d(m, z; \mathcal{D})$ the decision vector d obtained from $\mathcal{D}$ by evaluating the quadratic functions at the point (m, z).

To summarize, the interpolation of $k(\chi; m, z)$ can be seen as a two-stage process. First, we compute all components of d by evaluating the polynomial approximations at (m, z). Given d, we compute $k(\chi; d)$, the cubic spline interpolation defined in (16). Since $c = \chi - k$, the savings function $k(\chi; m, z)$ also defines a consumption function $c(\chi; m, z)$.

4.2 Setting up the Computations

4.2.1 The Grid in Individual Capital

To speed up the computations, we solve the individual problem using the endogenous grid point method of Carroll (2006). For this, we choose a grid of end-of-period asset levels $\bar{k}_l$, $l = 1, \ldots, n_w$ with $\bar{k}_1 = \underline{k}$. This should not be confused with the knot points of the savings spline, which are in cash-on-hand χ, not in end-of-period capital k.

4.2.2 The Aggregate Grid

Although we consider small fluctuations around the steady state, technically I am not using perturbation methods, but least squares projection methods (Judd 1998, p. 382). I compute the equilibrium on a on a finite grid of points in the reduced aggregate state space. Then the household savings function is fitted by least squares. The projection equations are not solved simultaneously by a quasi-Newton method, because the household savings function has too may parameters to do that. We rather solve it by backward iteration, cf. Section 4.3.

[4] See Footnote 3 for the timing of the state variables.

With more than one or two moments in the state vector, it is very difficult to find a reasonable grid. For example, using a Cartesian grid in moments, one would have many grid points that are very unusual, in the sense of having very low probability in the ergodic set of the solution process. In earlier experiments, I found it very hard to compute numerically precise results with a state space of more than 3 or 4 moments. The decisive improvement comes from the use of the solution of the linearized model. The matrix Σ_x in (29) gives the covariance matrix of the fluctuations of the full state vector x_t around the steady state. Then $\Sigma_{\hat{x}} \equiv H\Sigma_x H'$ gives the covariance matrix of the reduced state vector $\hat{x}_t$. Choose S such that $SS' = \Sigma_{\hat{x}}$. Then construct a grid G_{MZ} of (m, z)-values as follows:

1. Construct a grid on the unit ball in $(\dim(m) + \dim(z))$-dimensional Euclidean space. Denote the points on this grid by ω_i.

2. Transform this grid into a grid in (m, z)-space by $(m_i, z_i) = S\omega_i$.

The first step can be done in many ways; in the examples below, I was using all points on the unit circle of the form

$$(0, \ldots, 0, \pm 1, 0, \ldots, 0) \qquad \text{or} \qquad (0, \ldots, 0, \pm\sqrt{1/2}, 0, \ldots, 0, \pm\sqrt{1/2}, 0, \ldots, 0)$$

This is not an optimal way to choose an aggregate grid. Most likely one could obtain more precise results, for example by using Smolyak grids and polynomials (Malin, Krueger, and Kubler 2007). I choose this very simple setup to make clear that the working of the method does not depend on using sophisticated projection methods.

4.2.3 Initializing the Savings Function

From the linearized solution, we can get each of the $(n_s + 1)$ components of d_t as a linear function of the state vector m_{t-1}, z_t. This linear approximation can be used to initialize the quadratic approximation, setting the coefficients of the quadratic components to zero. All the coefficients are stacked into the vector $\mathcal{D}_T$. The subscript T here stands for a fictitious last period where we start our iterations. Since the iteration step is interpreted as a step backward in time, the next iteration will then be called $\mathcal{D}_{T-1}$, etc.

4.3 Backward Iteration Step

The basic idea behind iterating backward in time is the following. Assume we know the decision function of the agents at time $t + 1$, which is given here by the household consumption function $\hat{\mathcal{C}}(\chi; d_{t+1})$. Then we can, in principle, solve for the equilibrium at time t separately for each point of the state space, $(\mathbf{\Phi}_{t-1}, z_t)$, and thereby obtain the decision function of period t, $\hat{\mathcal{C}}(\chi; d_t)$. With approximate aggregation, however, we work in the lower-dimensional state space (m_{t-1}, z_t), which does not contain the complete information about $\mathbf{\Phi}_{t-1}$. We solve this problem by choosing a proxy distribution $\mathbf{\Phi}_{t-1} = \mathbf{\Phi}^{pd}(m_{t-1}, z_t)$, cf. Step 2 below, which should represent a "typical distribution" for the reduced aggregate

state (m_{t-1}, z_t). Then we can solve for the equilibrium separately for each point in the grid of reduced states.

So assume that after t iterations we are given the vector $\mathcal{D}_t$, representing an approximation of the household savings function. The aim is to obtain a new approximation $\mathcal{D}_{t-1}$ by iteration backward in time. Separately for each aggregate grid point $(\bar{m}, \bar{z}) \in G_{MZ}$, do the following:

1. For each point $\bar{k}_l$, $l = 1, \ldots, n_w$, on the grid of end-of-period assets, compute

$$\mathcal{E}_l \equiv \beta \sum_{j,a} \omega_\xi(j)\omega_a(a)(1 + r(\bar{m}, \bar{z}(a)))u_c\left(\hat{\mathcal{C}}(\chi^{lja}; d(\bar{m}, \bar{z}(a); \mathcal{D}_t))\right) \tag{37}$$

where χ^{lja} is next period's cash-on-hand,

$$\chi^{lja} \equiv \bar{k}_l\left(1 + r(\bar{m}, \bar{z}(a))\right) + W(\bar{m}, \bar{z}(a))(1 - \tau^l(\bar{m}, \bar{z}(a))\bar{\xi}_j \tag{38}$$

and

$$\bar{z}(a) \equiv \rho_z \bar{z} + \bar{\epsilon}_a \tag{39}$$

$\mathcal{E}_l$ is the expectation of $u_c(c_{t+1})$, conditional on having assets $\bar{k}_l$ at the end of period t, and the aggregate state at the end of the period being $(\bar{m}, \bar{z})$. From the household Euler equation, we then see that consumption at time t and cash-on-hand at time t of this household are given by

$$\bar{c}_l = u_c^{-1}\left[\mathcal{E}_l\right] \tag{40}$$

$$\bar{\chi}_l^{endog} = \bar{k}_l + \bar{c}_l \tag{41}$$

This gives us optimal consumption $\bar{c}_l$ at the endogenous grid points of cash-on-hand, $\bar{\chi}_l^{endog}$, $l = 1, \ldots, n_w$. By construction, the critical level χ^c of the household equals $\bar{\chi}_0^{endog}$. Consumption at the exogenous knot points $\bar{\chi}_{t,i} = \chi^c + \bar{\bar{\chi}}_i$ can be obtained by cubic spline interpolation from the $\left\{\bar{\chi}_l^{endog}, \bar{c}_l\right\}_{l=1,\ldots,n_w}$. This then defines the consumption function $\bar{d}$ for all the households, conditional on the aggregate state at the end of the period being $(\bar{m}, \bar{z})$.

2. Find the vector of past moments $\bar{m}_{-1}$ such that

$$H_{11}\Pi\left(\bar{d}; w(\bar{m}_{-1}, \bar{z}), R(\bar{m}_{-1}, \bar{z})\right)\mathbf{\Phi}^{pd}\left(\bar{m}_{-1}, \bar{z}\right) = \bar{m} \tag{42}$$

Here $\mathbf{\Phi}^{pd}\left(\bar{m}_{-1}, \bar{z}\right)$ is defined as

$$\mathbf{\Phi}^{pd}\left(\bar{m}_{-1}, \bar{z}\right) = \mathbf{\Phi}^* + \Sigma_{x_{-1}}H'\left[H\Sigma_{x_{-1}}H'\right]^{-1}\begin{bmatrix}\bar{m}_{-1} - m^* \\ \bar{z} - z^*\end{bmatrix} \tag{43}$$

In (43), starred variables refer to steady state values. $\Sigma_{x_{-1}}$ is the covariance matrix of m_{t-1}, z_t, which is obtained from Σ_x by post-multiplying the part relating to the covariance between m_{t-1} and z_t by ρ_z. The interpretation of $\mathbf{\Phi}^{pd}\left(\bar{m}_{-1}, \bar{z}\right)$ is the expected value of $\mathbf{\Phi}_{t-1}$ conditional on $\bar{m}_{-1}$ and $\bar{z}$, under the dynamics generated from the linearized solution of the model.

From the above, we have obtained for each $(\bar{m}, \bar{z}) \in G_{MZ}$, a $\bar{d}$ and a $\bar{m}_{-1}$ such that $\bar{d}$ characterizes the consumption function at $(\bar{m}_{-1}, \bar{z})$. Then we regress each component of the $(n_s + 1)$ components of $\bar{d}$ on a quadratic function in the $(\bar{m}_{-1}, \bar{z})$. The coefficients of these regressions are stacked into the updated vector $\mathcal{D}_{t-1}$.

4.4 Discussion of the Algorithm

4.4.1 The Endogenous Grid Point Method

For computational speed, the critical step is backward iteration. The key element that makes the algorithm of Section 4.3 fast is the use of endogenous grids.

First, we apply the endogenous grid point method of Carroll (2006) in Step 1. It allows to compute optimal consumption in the current period, given next period's household consumption function, without resorting to numerical optimization.

Second, a similar kind of trick is used to solve for equilibrium in Step 2 of Section 4.3. Rather than fixing an $\bar{m}_{-1}$ and search for the $\bar{m}$ that we go to, we fix an end-of-period $\bar{m}$ and search for the $\bar{m}_{-1}$ from which we reach $\bar{m}$. Knowing the *end-of-period* aggregate state $(\bar{m}, \bar{z})$, we compute the household consumption function of the same period. The equilibrium problem (42) is then easy to solve, because it depends on $\bar{m}_{-1}$ only through $\mathbf{\Phi}^{pd}(\bar{m}_{-1}, \bar{z})$, which is a linear function, and through the effect of the interest rate on $\Pi\left(\bar{d}; w(\bar{m}_{-1}, \bar{z}), R(\bar{m}_{-1}, \bar{z})\right)$. We do not have to recompute the household savings function for each $\bar{m}_{-1}$.

4.4.2 The Arguments of the Consumption Function

If the state vector m_{t-1} contains all the relevant information about the distribution of capital at the end of last period, $\mathbf{\Phi}_{t-1}$, then last period's exogenous state z_{t-1} is irrelevant for equilibrium in t, once we know z_t. With approximate aggregation, m_{t-1} provides only incomplete information about $\mathbf{\Phi}_{t-1}$, and z_{t-1} can still affect the equilibrium in t through the additional information it may provide $\mathbf{\Phi}_{t-1}$. One might therefore consider several values z_{t-1} that are possible predecessors of z_t, and compute an equilibrium for each of those z_{t-1}. Then it is natural to write consumption at time t as a function of $(\mathbf{\Phi}_{t-1}, z_{t-1}, z_t)$ rather than $(\mathbf{\Phi}_{t-1}, z_t)$.

However, if the state m_{t-1} contains many variables and is well chosen, the effect of z_{t-1} should be minimal. Then it appears better to average over possible z_{t-1} by using as proxy distribution the expectation of $\mathbf{\Phi}_{t-1}$ conditional on m_{t-1} and z_t. We denote this by $\mathbf{\Phi}^{pd}(m_{t-1}, z_t)$. Consumption is then approximated as a function of $(\mathbf{\Phi}_{t-1}, z_t)$. This is the route that I have followed here.

4.4.3 The Role of the Linearized Solution

The solution to the linearized model is an essential ingredient in obtaining a higher-order solution with medium-dimensional state vector. In the algorithm described above, the linearized solution serves the following purposes:

1. It gives a proxy distribution function, cf. (43).

2. The variance-covariance matrix of the state variables in the linearized model can be used to construct a good aggregate grid, which means a grid that covers the part of the state space in which the solution really lives (Section 4.2.2).

3. The solution from the linearized model will be a good starting point for higher-order approximations, reducing the number of iteration steps in the nonlinear procedure.

4. To obtain higher order approximations of the solution, it is necessary to substantially limit the number of state variables. Currently, one can handle up to about 20 state variables for a second-order approximation on a PC, with the above method. For many models, it is not clear what these state variables should be. Reiter (2009a) shows how to choose a reduced state vector in an almost optimal way, based on methods from linear control theory.

4.4.4 The Assumption of Small Shocks

Although the algorithm is based on projection methods, I see the approach as being in the spirit of perturbation methods, solving the model in a small neighborhood of the steady state without aggregate shocks. The assumption of very small shocks plays several roles:

1. It makes sure that the proxy distribution computed by (43) has positive entries.

2. The grid of points in aggregate state space is relevant, in the sense that they come from a part of the state space that is frequently visited in the ergodic distribution.

3. The algorithm can be expected to converge fast, starting from household decision function obtained in the linearized solution.

This list mirrors the list in Section 4.4.3 of the different roles of the linearized solution. Small shocks guarantee that the information obtained from linearization is in fact useful.

4.5 Simulating the Model

Since we cannot compare the approximate solutions to the exact nonlinear solution, it is necessary to find alternative accuracy checks. In the literature it is common to simulate the model solution in two different ways and then check for consistency. This is often called "simulation using aggregate law of motion" versus "simulation without using aggregate law of motion" (Den Haan 2010). I will do something similar:

- Simulation following the cross-sectional distribution (abbreviated "SimD")

  In each period, we start from the aggregate state $(\mathbf{\Phi}_{t-1}, z_t)$, which we update as follows.

1. Set $m_{t-1} = H_{11}\mathbf{\Phi}_{t-1}$ as in (23), and compute d_t by interpolation, as described in Section 4.1.
2. Compute $\Pi(d_t; w_t, R_t)$ by (22) and set $\mathbf{\Phi}_t$ by (21).
3. Draw random numbers $\epsilon_{\tau,t+1}$ and $\epsilon_{\theta,t+1}$ and update z_{t+1} using (3) and (8).

- Simulation using the proxy distribution ("Prox")

  Here, we start from the reduced aggregate state (m_{t-1}, z_t), which we update as follows.

  1. Set $\mathbf{\Phi}_{t-1} = \mathbf{\Phi}^{pd}(m_{t-1}, z_t)$, as defined by (43).
  2. Compute d_t by interpolation, as described in Section 4.1.
  3. Compute $\Pi(d_t; w_t, R_t)$ by (22) and set $\mathbf{\Phi}_t$ by (21). Set $m_t = H_{11}\mathbf{\Phi}_t$.
  4. Draw random numbers $\epsilon_{\tau,t+1}$ and $\epsilon_{\theta,t+1}$ and update z_{t+1} using (3) and (8).

Simulation Prox is the same as SimD, except that in each period we replace $\mathbf{\Phi}_{t-1}$ by the corresponding proxy distribution, $\mathbf{\Phi}^{pd}(m_{t-1}, z_t)$. This corresponds to what is usually called "simulation using aggregate law of motion", because the economic agents implicitly suppose that the distribution is given by the proxy distribution when they compute the aggregate behavior of the economy. If the two types of simulation give almost the same results, then the agents do not make a big mistake in using the proxy distribution when computing their optimal behavior.

5 Numerical Results

The purpose of this section is to measure the precision of the solution that is obtained by the second-order approximation described in Section 4. We measure precision in two ways. First, we compare the outcome from the two types of simulation of the model, as described in Section 4.5. Second, we compare the simulation outcomes using a low number of moments to the outcome using 15 moments, which for this purpose can be considered the "precise" solution.

We look at the precision of the impulse responses generated by the model, both to technology and tax shocks. Each impulse is normalized to 10^{-4}. This is a very small shock, in the spirit of a perturbation solution. We will discuss below how things change with a larger, more realistic, size of the shocks. Figure 1, Panel a), plots the impulse response to a technology shock of the model with only one moment. Both the SimD- and the Prox simulation are shown. Panel b) plots the error, first as difference between SimD and Prox solutions ("DiffSim"), and then as the difference between the SimD- and the solution with 15 moments ("Diff15Mom"). We see that the error is 2 orders of magnitude smaller than the impulse response. This confirms the finding of Krusell-Smith, that even the approximate model with only one moment gives a reasonably accurate solution.

Panels c) and d) present the same statistics for the sum of the impulse responses to a positive and a negative shock, both of absolute size 10^{-4}. This represents the nonlinear part of the impulse response: in an approximation that is linear in aggregate variables (such as Reiter (2009b)), the sum would be identically zero. We see that here the error is at least one order of magnitude smaller than the estimated response, so that we can say that the nonlinear part is also modeled with reasonable accuracy.

Panels e)–h) present the same information for the second moment (which is not a state variable in the one-moment solution). The accuracy is comparable to the one obtained for the first moment.

Summary statistics on the accuracy of this solution are also presented in the first two lines of Table 1, for predictions of the first four moments. Accuracy is measured as the maximum absolute error relative to the maximum value of the impulse response. For each solution, the first line refers to the difference between SimD and Prox solutions, the second line to the difference to the 15-moments solution. The following lines of the table present results for solutions with up to 15 moments. We see that accuracy can be increased by three to four orders of magnitude, both for the positive impulse response and for the sum of the positive and the negative impulse.

Figure 2 looks at the impulse response to a tax shock. It confirms what has already been found in Reiter (2009b) and in Reiter (2009a), namely that the one-moment solution is totally inadequate to capture the effect of this redistributional shock. Figures 3 and 4, showing the 4-moments and 8-moments solutions, respectively, illustrate the gradual increase in accuracy from using more moments. Table 1 shows a gain of three to five orders of magnitude in the accuracy if we go to a 15-moments solution. Most of the time, results for the prediction of the second to fourth moments are comparable in precision to the prediction of mean capital.

The accuracy statistics of Table 1 essentially measure the aggregation error, the error that results from using a small number of moments in the state vector, rather than a high-dimensional representation of the cross-sectional distribution. We can compare this to another interesting type of error, the linearization error, that arises if we use a linear approximation, as in Reiter (2009b) and in Reiter (2009a). This error is effectively given by the sum of the impulse responses (Panels c) and f) in each figure), which measures the deviation from linearity in the impulse responses. Comparing the size of the impulse response in Panel c) to the size of the error in the positive response, Panel b), we get an idea about which type of error is more important. Obviously, this depends on the size of the shock. If the shock gets scaled up by a factor of 100, so that it approximately equals the empirical standard error of this shock, the positive impulse response, and its error, will get scaled up by 100 as well. The sum of the responses, being a quadratic effect, will be scaled up by 10^4. Taking this into account, we see from Figure 1) that the quadratic effect in the response to a technology shocks has the same order of magnitude as the aggregation error, already in the solution with one moment. For the tax shock, the aggregation error is much more important. The quadratic effect has about the same size as the aggregation error if we use a solution with 4 moments. (Panels b) and c) of Figure 3).

Figures 5) and 6) present similar statistics for the prediction of the third and fourth

moment. It is interesting to see that the tax shock generates very persistent deviations in second, third and fourth moments (notice that here the x-axis stretches to 1000 periods). For the prediction of the positive shock, the 4-moments solution appears good enough. To predict correctly the nonlinear part of the impulse response (sum of shocks), we need 8 moments.

Similarly, we could estimate the error from using a quadratic approximation, by computing a cubic approximation. I leave this for future work that will consider large shocks.

6 Conclusions

The paper has shown a method to compute higher-order solutions of heterogeneous agent models around the steady state. The method uses approximate aggregation, and enforces consistency between individual policy rules and aggregate behavior on a selected grid of aggregate states. The advantage of this method over existing approaches is that it can handle a higher-dimensional state vector; in the numerical examples, up to 15 moments were used to characterize the dynamics of the cross-sectional distribution. To make this possible, it is crucial to exploit the information obtained from the linearization of the model around the steady state. To speed up the computations, the backward induction steps use endogenous grid point methods similar to Carroll (2006). As a result, a solution using 15 moments in the state vector can be obtained within a few minutes on a PC, using Matlab.

The method is in the spirit of perturbation methods, assuming small shocks. The same techniques may be applied to handle the case of big shocks, which means solving the model on a larger part of the state space. This needs a method to select proxy distributions in a way that guarantees the non-negativity of the cross-sectional probability (or density) function. A simple way to do this is described in Reiter (2010). It may be worthwhile to apply simulation methods more extensively, to learn about the ergodic set of the model, and to obtain a better proxy distribution function for the nonlinear solution. This procedure will probably be very application-specific, and should be done in the context of more interesting examples than the stochastic growth model, which is rather easy to solve anyway. This is left for future work.

		1st mom.		2nd mom.		3rd mom.		4th mom.	
Shock	#Moms	pos.	sum	pos.	sum	pos.	sum	pos.	sum
Tech	1	6.45e-3	7.14e-2	2.22e-2	1.59e-1	7.75e-2	4.31e-1	1.18e-1	5.65e-1
		2.32e-3	3.18e-2	2.07e-3	2.83e-2	1.38e-3	1.75e-2	2.42e-3	1.20e-2
	2	2.56e-3	2.93e-2	7.67e-3	4.29e-2	1.63e-2	8.11e-2	3.74e-2	1.19e-1
		2.50e-3	4.80e-2	2.52e-3	4.25e-2	2.22e-3	2.60e-2	1.91e-3	1.77e-2
	4	4.65e-4	5.66e-3	2.60e-3	2.04e-2	1.19e-2	7.70e-2	2.88e-2	1.47e-1
		3.21e-3	7.22e-2	3.19e-3	6.38e-2	2.71e-3	3.84e-2	2.27e-3	2.58e-2
	6	1.00e-4	4.54e-4	8.65e-5	6.50e-4	2.54e-4	1.51e-3	6.53e-4	2.33e-3
		4.32e-5	2.08e-2	4.45e-5	1.86e-2	4.18e-5	1.15e-2	3.75e-5	7.72e-3
	8	7.53e-6	2.57e-4	1.01e-5	1.58e-4	1.72e-5	2.02e-4	4.22e-5	2.32e-4
		1.05e-5	3.34e-2	1.30e-5	2.94e-2	1.58e-5	1.74e-2	1.61e-5	1.14e-2
	12	6.84e-7	9.51e-5	9.80e-7	9.81e-5	8.95e-7	1.20e-4	1.24e-6	2.09e-4
		4.30e-6	2.24e-2	5.13e-6	1.98e-2	5.93e-6	1.18e-2	5.90e-6	7.72e-3
	15	4.57e-7	7.50e-5	1.22e-6	1.10e-4	3.73e-7	9.19e-5	7.73e-7	1.67e-4
Tax	1	9.85e-1	9.17e-1	9.88e-1	9.26e-1	9.94e-1	9.38e-1	9.96e-1	9.58e-1
		6.32e-1	3.82e-1	1.81e-1	1.94e-1	3.93e-2	9.98e-2	4.56e-2	1.45e-1
	2	7.34e-1	8.67e-1	4.15e-1	5.53e-1	1.52e-1	1.20e+0	2.62e-1	1.38e+0
		3.54e-1	3.47e-1	6.31e-2	1.75e-1	1.29e-2	8.07e-2	8.31e-3	6.06e-2
	4	2.00e-1	4.89e-1	3.45e-2	7.88e-1	3.00e-2	2.03e+0	5.76e-2	3.44e+0
		7.11e-2	1.81e-1	8.29e-3	7.73e-2	1.69e-3	3.56e-2	1.25e-3	2.30e-2
	6	3.07e-2	1.67e-1	1.94e-3	1.76e-1	2.95e-3	6.93e-2	5.82e-3	1.47e-1
		7.97e-3	5.38e-2	9.34e-4	1.91e-2	1.89e-4	8.41e-3	1.69e-4	4.98e-3
	8	4.88e-3	6.22e-2	2.46e-4	1.73e-1	1.44e-4	9.69e-3	2.91e-4	1.99e-2
		5.31e-4	2.61e-2	8.80e-5	9.42e-3	4.68e-5	4.29e-3	3.84e-5	4.51e-3
	12	2.79e-4	7.71e-2	1.41e-4	1.74e-1	3.42e-6	1.10e-2	1.82e-6	1.39e-2
		1.49e-4	2.80e-3	2.49e-5	8.30e-4	1.41e-5	4.48e-4	1.20e-5	5.61e-4
	15	2.00e-4	7.41e-2	1.40e-4	1.70e-1	2.92e-6	6.92e-3	1.28e-6	1.09e-2

Table 1: Relative impulse response errors

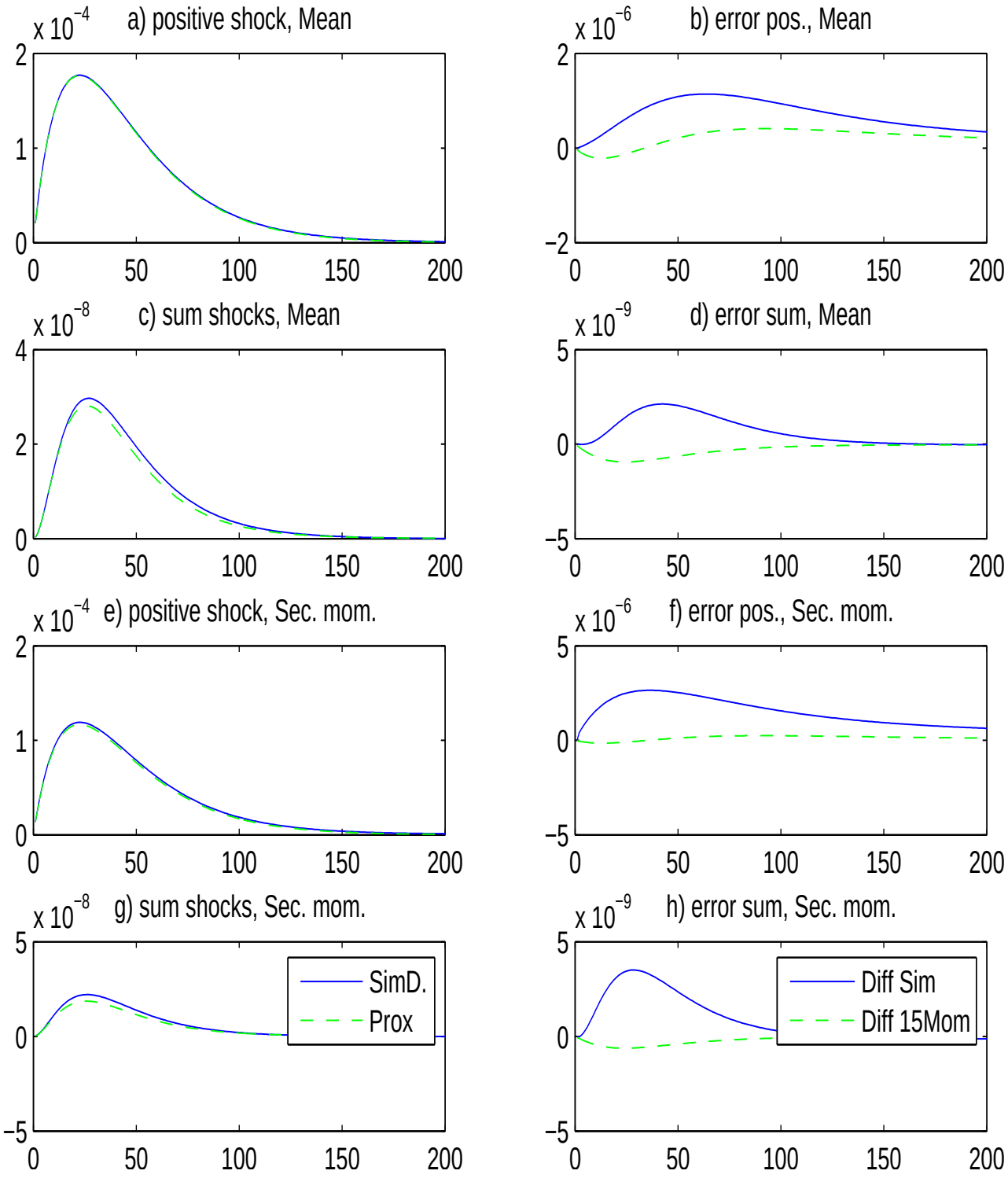

Figure 1: Response of distribution of capital to technology shock, 1 moment

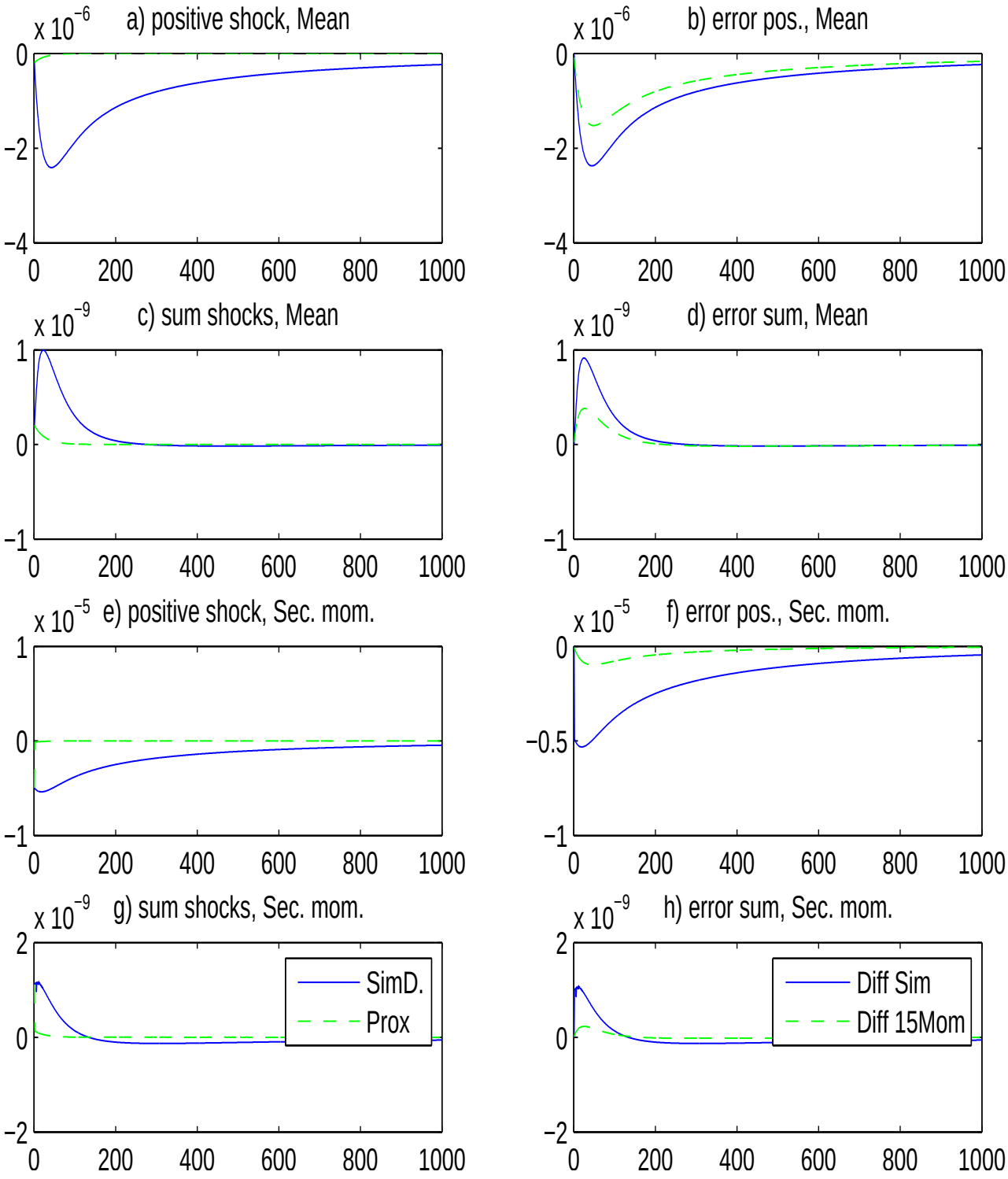

Figure 2: Response of distribution of capital to tax shock, 1 moment

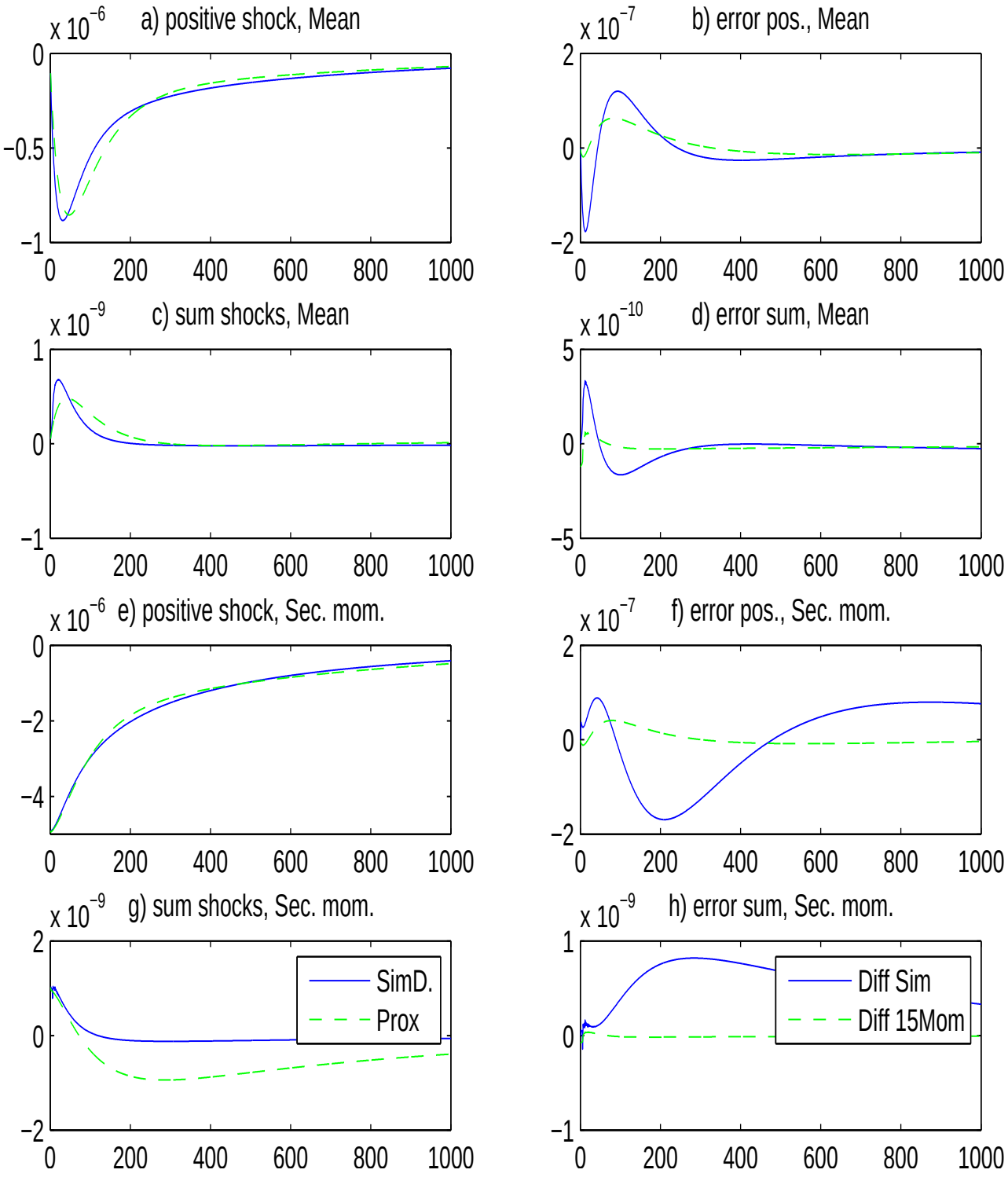

Figure 3: Response of distribution of capital to tax shock, 4 moments

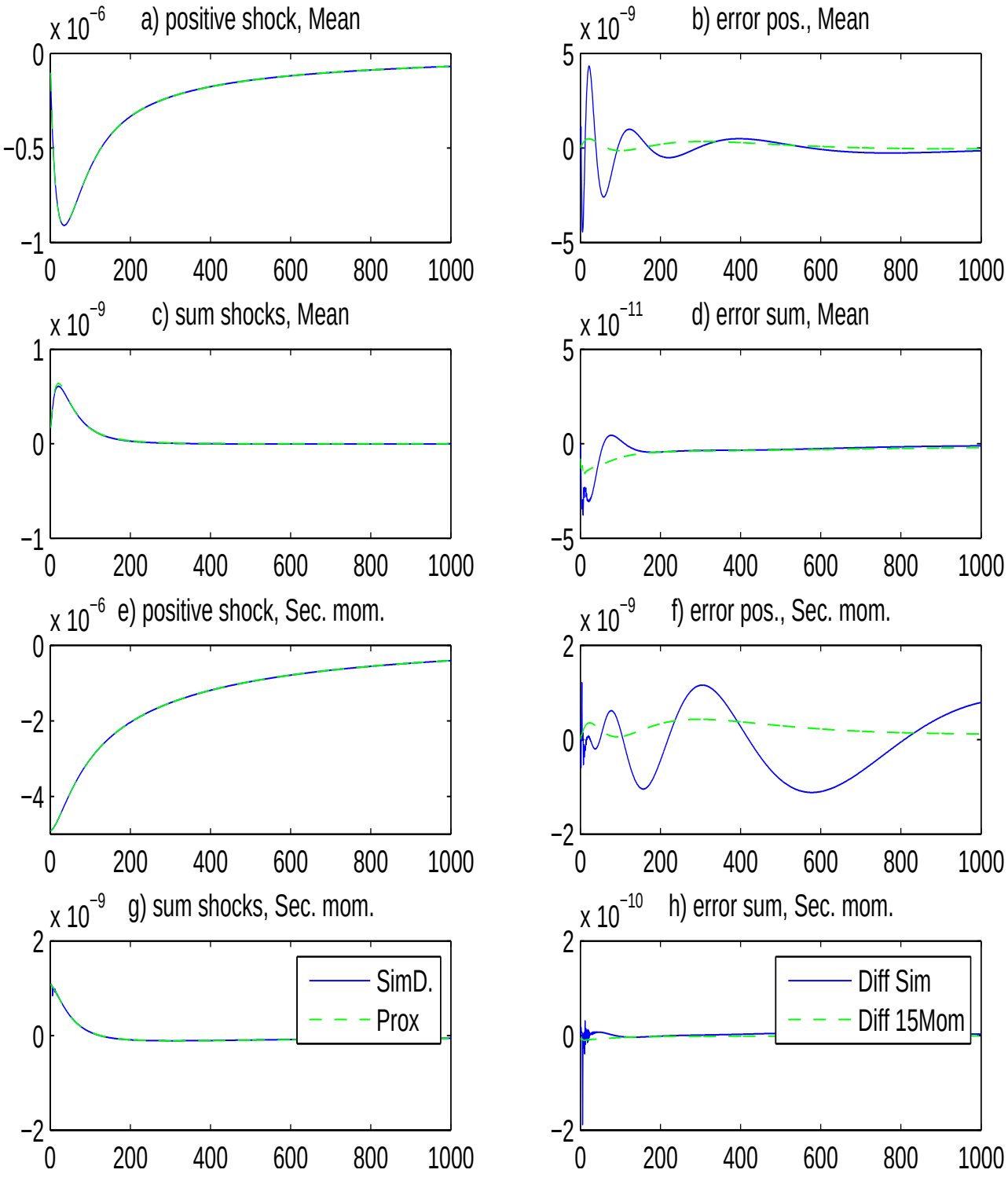

Figure 4: Response of distribution of capital to tax shock, 8 moments

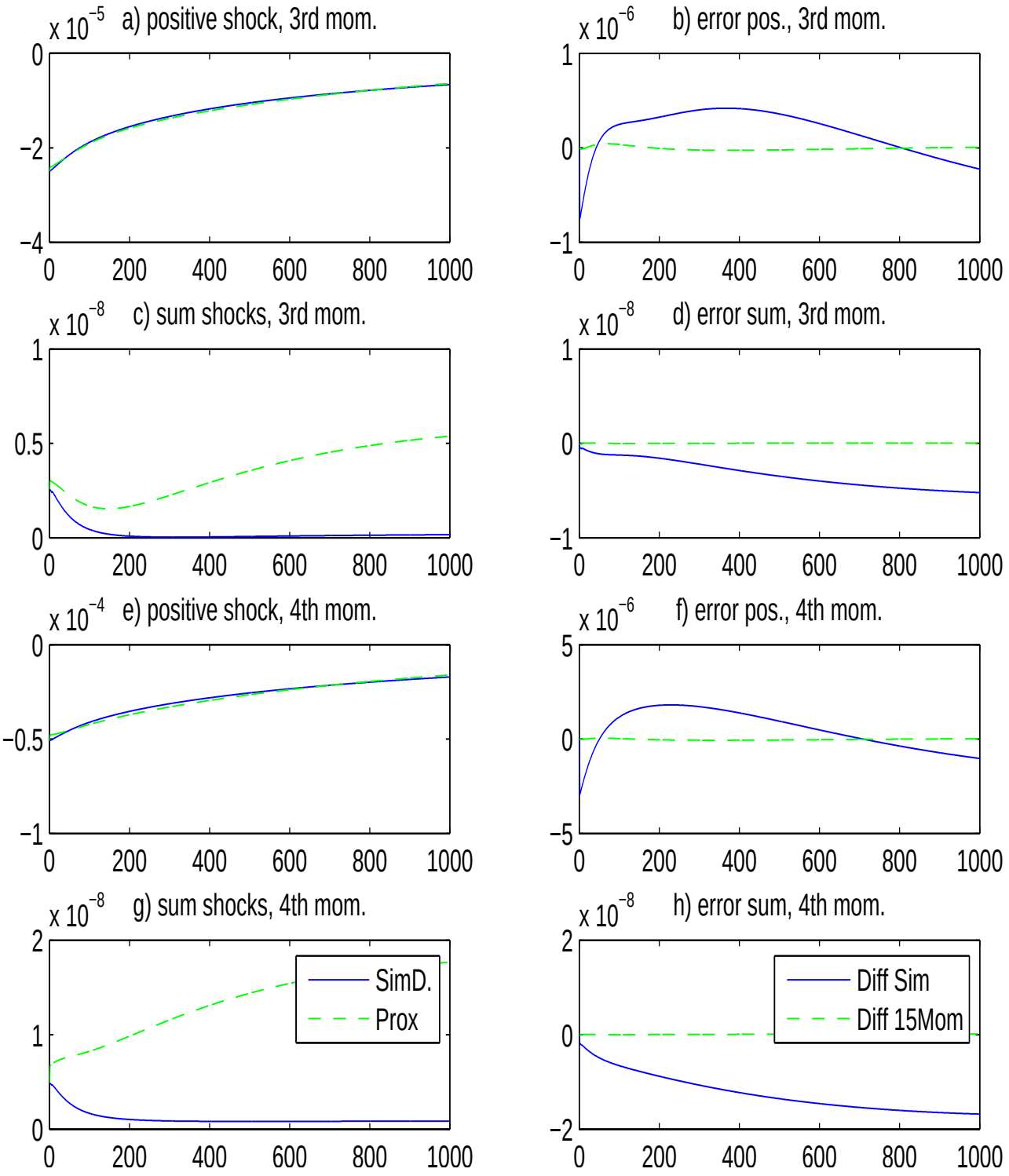

Figure 5: Response of distribution of capital to tax shock, 4 moments

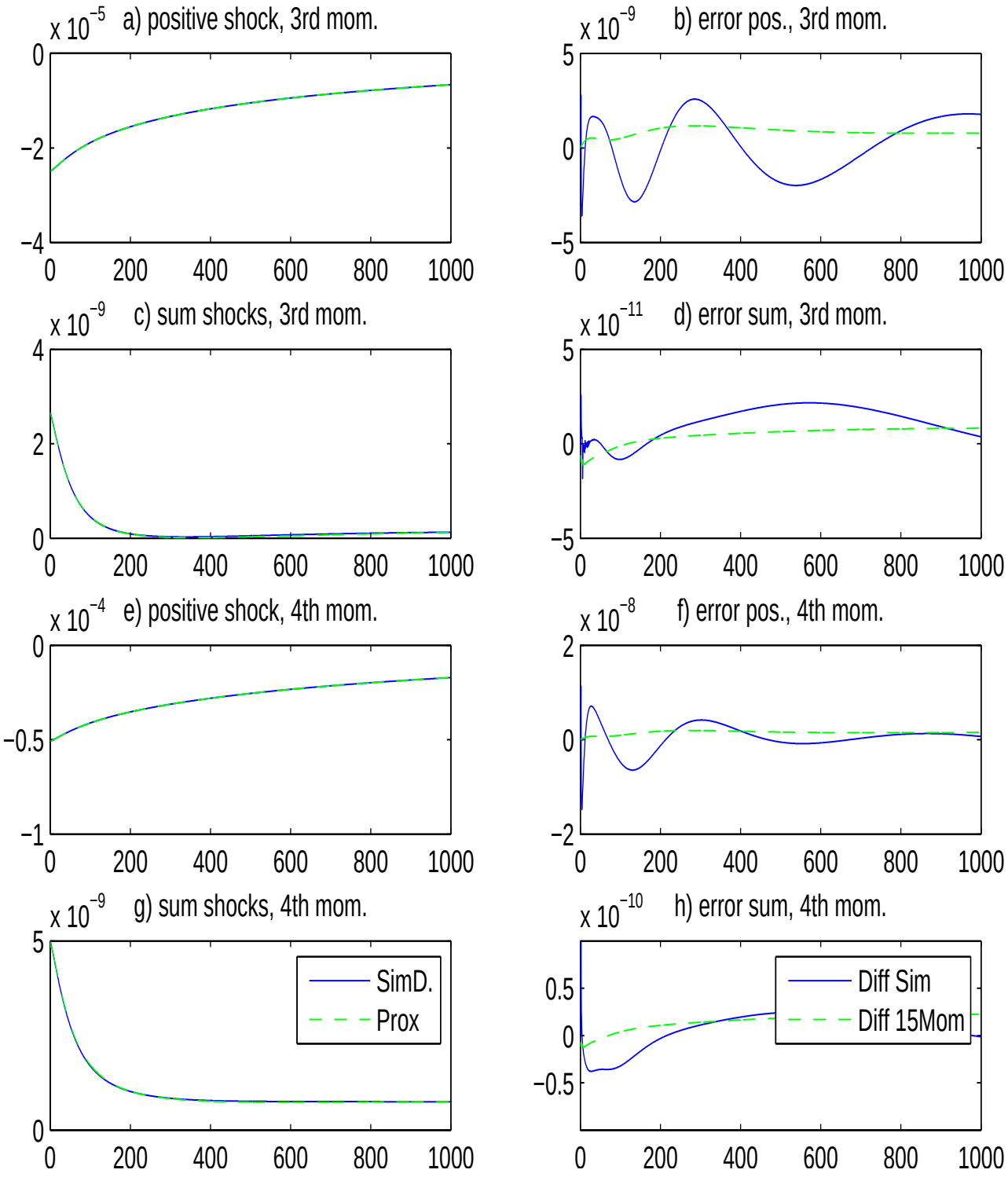

Figure 6: Response of distribution of capital to tax shock, 8 moments

References

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

Collard, F. and M. Juillard (2001). A higher-order taylor expansion approach to simulation of stochastic forward-looking models with an application to a nonlinear phillips curve model. *Computational Economics 17*, 125–139.

Den Haan, W. J. (1996). Heterogeneity, aggregate uncertainty, and the short-term interest rate. *Journal of Business and Economic Statistics 14*(4), 399–411.

Den Haan, W. J. (2010). Comparison of solutions to the incomplete markets model with aggregate uncertainty. *Journal of Economic Dynamics and Control 34*(1), 4–27.

Den Haan, W. J. and P. Rendahl (2010). Solving the incomplete markets model with aggregate uncertainty using explicit aggregation. *Journal of Economic Dynamics and Control 34*(1), 69–78.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics 12*(1), 58–79.

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

Judd, K. L. and S.-M. Guu (1993). Perturbation solution methods for economic growth models. In H. R. Varian (Ed.), *Economic and Financial Modeling with Mathematica.* New York: Springer.

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

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.

Podczeck, K. (2009). On existence of rich Fubini extensions. forthcoming in Economic Theory.

Preston, B. and M. Roca (2007). Incomplete markets, heterogeneity and macroeconomic dynamics. NBER Working Paper 13260.

Reiter, M. (2009a). Approximate and almost-exact aggregation in dynamic stochastic heterogeneous-agent models. Manuscript.

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

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

Schmitt-Grohé, S. and M. Uribe (2004). Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control 28*, 755–75.

Sims, C. A. (2001). Solving linear rational expectations models. *Computational Economics 20*(1-2), 1–20.