Approximate Aggregation in Heterogeneous-Agent Models*

Michael Reiter, Institute for Advanced Studies, Vienna

VERY PRELIMINARY AND INCOMPLETE VERSION!
February 14, 2009

Abstract

The paper shows how to get precise approximations to the solution of DSGE models with a great number of heterogeneous agents. The examples studied are a stochastic growth model with incomplete markets similar to Krusell and Smith (1998), and a model of heterogenous firms with state-dependent pricing. The aggregation problem is studied in a linearized version of those models. It is shown that an approximation with only one or two state variables, which is often used in the macroeconomic literature, is generally not sufficient to get a high-precision approximation. Then methods of model reduction from systems and control theory are used to identify optimal sets of state variables. Using medium-dimensional approximations (20 to 50 states) are sufficient to get very good approximations. The paper develops a collocation approach that allows to compute this kind of approximations. The results can also be used to obtain higher-order approximations of the model solutions with a reduced set of state variables.

JEL classification: C63, C68, E21
Keywords: heterogeneous agents; projection methods; perturbation methods; invariant distribution

Address of the author:

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

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

*Support from the Austrian National Bank under grant Jubiläumsfonds 13038 is gratefully acknowledged.

1 Introduction

Stochastic general equilibrium models with incomplete markets and a large number (or even a continuum) of heterogeneous agents have recently become more widely used in economics. In most cases, only a numerical solution to these models can be computed. The main computational challenge in solving these models is that, theoretically, the state vector includes the whole cross-sectional distribution of individual state variables of the agents in the economy (which in the case of a continuum of agents is an infinite-dimensional object).

Krusell and Smith (1998) and Den Haan (1997) have shown that in some versions of the stochastic growth model with heterogenous agents, a very good approximate solution can be obtained by some form of approximate aggregation: agents are supposed to base their decision not on the whole state vector of the economy, but only on a very small set of aggregate statistics (such as moments of the cross-sectional distribution of capital). This is a deviation from the assumption of strictly rational expectations, but all the available accuracy checks indicate that the resulting approximation is very good (cf. Den Haan (2008)). Krusell and Smith (2006) explains and extends this approximate aggregation result in detail.

These results raise a number of questions. Does the approximate aggregation result carry over to other classes of heterogenous agent models? Can even better approximations be obtained by using medium-dimensional (10 to 100 dimensions, say) approximations? How can we solve models with medium-dimensional state space? Which state variables (statistics of the cross-sectional distribution) should be used in that case? In the present paper I address those issues, taking as examples two widely used models, a version of the Krusell/Smith model, and a model of heterogeneous firms with state-dependent pricing and idiosyncratic firm-productivity shocks.

The starting point of the analysis is the method developed in Reiter (2008a). It uses a high-dimensional, non-parametric approximation to the cross-sectional distribution, and computes a solution that is nonlinear in individual decision variables, but linear in aggregate variables. The nonlinearity in individual variables allows to handle inequality constraints such as borrowing constraints, while the linearity in aggregate variables allows to handle a large number of aggregate state variables (on a PC, around 1000 to 2000) which characterize the cross-sectional distribution. This method can be seen as an analogue of the well-known linearization methods for business cycle models, generalized to models of heterogenous agents and incomplete markets. While linearization has important shortcomings (for example, one cannot study portfolio choice problems in such a setting), a linearized model is an ideal laboratory to study approximate aggregation. Even high-dimensional linear models can be accurately analyzed, and the statistical analysis can be done analytically, so that one never has to resort to simulation techniques as they were used, for example, by Krusell and Smith (1998). Furthermore, there is a large literature in systems- and control theory that deals with model reduction in linear settings, this means, the approximation of a given high-dimensional dynamical system by a lower-dimensional one. One can exploit this literature, even if the approximation problem in a general equilibrium

heterogeneous-agent economy is conceptually more difficult than finding an approximate model for an ex-ante given dynamical system, which is what the control literature usually considers.

The paper makes five contributions. First it investigates the possibility of approximate aggregation in a simple version of the Krusell/Smith model. Similar to what the literature has found, a one-moment approximation is a very good approximation if the only shock to the economy is a technology shock (this is not true for the state-dependent pricing model). However, if there are shocks that affect the cross-sectional distribution of wealth rather directly (such as a shock to redistributive taxation), then a one-moment approximation is not precise any longer. The good news is that commonly used accuracy statistics would indicate the failure of the approximation in these cases.

Second, it discusses and compares two approaches to low-dimensional approximation: the well known method of Krusell and Smith (1998), and a collocation approach that builds on Reiter (2002) and Reiter (2008b). It shows that the two approaches are equivalent in a linearized setting, but that the latter one has important computational advantages, which make it feasible to handle medium-dimensional approximations efficiently.

The question then arises whether approximate aggregation can give a high-precision solution if a larger set of statistics is used. The third contribution of the paper is to discuss various ways to select a set of statistics of the distribution such that these statistics help to predict future relevant aggregate variables (such as factor prices). For this, I draw on the engineering literature on model reduction, to choose those statistics in an optimal way. It turns out, both for the Krusell/Smith model and the sticky price model, that between 15 and 50 statistics are needed to get a high-precision approximation. This is probably not due to the particular aspects of these economic models, but rather mirrors the findings in the engineering literature, saying that high-dimensional dynamic systems with only a few inputs (here, shocks) and few outputs (here, relevant macroeconomic variables) can usually be well approximated by a much lower-dimensional approximate model.

Fourth, it develops a method to compute medium-dimensional approximations of the linearized model even in those cases where the number of state variables is so big (up to 100000, say) that not even an exact linearized solution can be computed. This part of the paper draws heavily on advanced sparse-matrix methods.

Finally, the paper outlines how the results of the linearized model can be used to get precise (relative to the existing literature) higher-order approximations to the nonlinear models, using a medium-dimensional set of statistics. The details of the non-linear method, however, are left to a companion paper.

2 The Stochastic Growth Model

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 $L_t = 1$. Capital accumulates as

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

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 the wage W_t are determined competitively:

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

$$W_t = \mathcal{Y}_L(K_{t-1}, L_t, \Theta_t) \tag{5}$$

2.2 The Government

The only purpose of the government is to create some random redistribution between capital and labor. The government levies a tax τ_t^k at the beginning of t on the capital stock accumulated at the end of period $t-1$. The after tax gross interest rate R_t is then related to the before tax rate by

$$R_t = \bar{R}_t - \tau_t^k \tag{6}$$

Furthermore, the government taxes The revenues are redistributed lump sum, providing a transfer T_t to every household. The government has to balance its budget every period:

$$T_t = \tau_t^k K_{t-1} \tag{7}$$

The tax rate 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.

2.3 The Household

There is a continuum of households, indexed by i. Households differ ex post by their labor productivity $\xi_{t,i}$. Labor productivity of household i is the product of a permanent component $\theta_{t,i}$, and an i.i.d. component $\xi_{t,i}$. Both are normalized to have unit mean:

$$\mathrm{E}\,\theta_{t,i} = 1 \tag{9}$$

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

Permanent productivity $\theta_{t,i}$ follows an n_P-state Markov chain and takes on the values $\bar{\theta}_1, \ldots, \bar{\theta}_{n_P}$. Transitory productivity $\xi_{t,i}$ has a discrete distribution, taking on the values $\bar{\xi}_i$ with probabilities ω_i^{ξ}, $i = 1, \ldots, n_y$.[1]

The household supplies inelastically one unit of labor. Its labor earnings are therefore given by

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

Household i enters period t with asset holdings $k_{t-1,i}$ 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

$$x_{t,i} = R_t k_{t-1,i} + y_{t,i} \tag{12}$$

$$\tag{13}$$

Cash on hand is split between consumption and asset holdings:

$$k_{t,i} = x_{t,i} - c_{t,i} \tag{14}$$

We impose the liquidity constraint

$$k_{t,i} \geq \underline{k} \tag{15}$$

The solution of the household problem is given by a consumption function $\mathcal{S}_t(x, p)$, or equivalently a savings function

$$k_t(x, p) \equiv x - \mathcal{S}_t(x, p) \tag{16}$$

where the subscript t denotes the dependence on a vector of aggregate state variables, which will be specified later. The first order condition of the household problem is the Euler equation

$$\begin{aligned} U'\left(\mathcal{S}_t(x_t, \theta_t)\right) \geq \beta\,\mathrm{E}_t\left[R_{t+1}U'\left(\mathcal{S}_{t+1}(R_{t+1}\left(x_t - \mathcal{S}_t(x_t, \theta_t)\right) + y_{t+1}, \theta_{t+1})\right)\right] \\ \text{and } \mathcal{S}_t(x_t, \theta_t) = x_t - \underline{k} \end{aligned} \tag{17a}$$

or

$$\begin{aligned} U'\left(\mathcal{S}_t(x_t, \theta_t)\right) = \beta\,\mathrm{E}_t\left[R_{t+1}U'\left(\mathcal{S}_{t+1}(R_{t+1}\left(x_t - \mathcal{S}_t(x_t, \theta_t)\right) + y_{t+1}, \theta_{t+1})\right)\right] \\ \text{and } \mathcal{S}_t(x_t, \theta_t) < x_t - \underline{k} \end{aligned} \tag{17b}$$

[1] In Reiter (2008a), productivity is assumed to have a continuous distribution. This leads to more complex transition processes and requires a more sophisticated discretization procedure. All this just distracts from the focus of the present paper, so I use a simpler model setup.

where $R_{t+1} \equiv 1 + r(\Omega_{t+1})$. The expectation in (17) is over the distribution of θ_{t+1}, ξ_{t+1} and the aggregate state Ω_{t+1}, which together determine next period's net income y_{t+1}. Since the household problem is concave, Equ. (17) together with the constraint (15) are both necessary and sufficient for a solution of the household problem. Notice that the solution depends on the aggregate state Ω_t only through the expected future dynamics of the factor prices.

2.4 Finite Approximation of the Model Equations

2.4.1 Savings, Consumption and the Euler Equation

At each point in time, and for any value of productivity $\bar{\theta}_p$, the savings behavior is characterized by a critical level χ_{pt} where the borrowing constraint starts binding, and a smooth function for $x > \chi_{pt}$. I approximate the savings function by a piecewise linear approximation.[2] The savings function $k_t(x, p)$ can then be represented by $(n_s + 1)n_P$ numbers, representing the critical level and the level of saving at n_s grid points of x, for each $p = 1, \ldots, n_P$. These numbers are collected into the vector $\mathbf{s}_t$. The details of the approximation are given in Appendix **??**.

Since we approximate the saving function by a piecewise linear function with $n_s + 1$ degrees of freedom, we apply a collocation method and require the Euler equation (17) to hold with equality at the knot points $\bar{x}_{it}$:

$$U'\left(\hat{\mathcal{S}}_p(\bar{x}_{it-1}; \mathbf{s}_{t-1})\right) = \beta \sum_{q=1}^{n_P} \sum_{j=1}^{n_y} \omega_{p,q}^{\theta} \omega_j^{\xi} \left[\hat{R}(\mathbf{P}_{t-1}, \Theta_t) U'\left(\hat{\mathcal{S}}_q(\bar{x}^{iqjq}; \mathbf{s}_t)\right)\right] + \eta_{ipt}^c, \qquad i = 0, \ldots, n_s \quad (18a)$$

where $\hat{\mathcal{S}}_p(x; \mathbf{s}_t)$ is short for $x - \hat{k}_p(x; \mathbf{s}_t)$ and

$$\bar{x}^{ipjq} \equiv \hat{R}(\mathbf{P}_{t-1}, \Theta_t)\left(\bar{x}_{it-1} - \hat{\mathcal{S}}_p(\bar{x}_{it-1}; \mathbf{s}_{t-1})\right) + \hat{W}(\mathbf{P}_{t-1}, \Theta_t)\bar{\theta}_q \bar{\xi}_j + T_t \quad (18b)$$

The η_{ipt}^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.

We use the fact that gross interest rate $\hat{R}(\mathbf{P}_{t-1}, \Theta_t)$ and wage rate $\hat{W}(\mathbf{P}_{t-1}, \Theta_t)$ are a function of the end of last period's capital and today's productivity. Equ. (18a) uses the notation of Sims (2001): the $\eta_{i,t}^c$ 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 (cf. Section **??**).

[2] In Reiter (2008a, Section 3.1.1), I use cubic splines with 100 knot points, here I use piecewise linear approximation with a higher number of knot points; later the number of variables will be reduced by a specific projection technique, cf.Section 8.

2.4.2 Wealth Distribution

In the theoretical model, the ergodic cross-sectional distribution of wealth has an infinite number of discrete mass points, because households at the borrowing constraint $\underline{k} = 0$ return to the region of positive k in packages of positive mass, since the distribution of idiosyncratic productivity is discrete. I approximate this complicated distribution by a finite number of mass points at a predefined grid $\underline{k} = \bar{k}_0, \bar{k}_1, \ldots, \bar{k}_{n_d} = k^{max}$. 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 p of households in period t saves the amount $\tilde{k}$ with $\bar{k}_i \leq \tilde{k} \leq \bar{k}_{i+1}$, we approximate this by assuming that $p \cdot \phi(i, \tilde{k})$ households end up at grid point $\bar{k}_i$, while $p \cdot (1 - \phi(i, \tilde{k}))$ households end up at grid point $\bar{k}_{i+1}$, such that $\phi(i, \tilde{k})\bar{k}_i + (1 - \phi(i, \tilde{k}))\bar{k}_{i+1} = \tilde{k}$. Somewhat more formally, define

$$\phi(i,k) \equiv \begin{cases} 1 - \frac{k - \bar{k}_i}{\bar{k}_{i+1} - \bar{k}_i} & \text{if } \bar{k}_i \leq k \leq \bar{k}_{i+1} \\ \frac{k - \bar{k}_{i-1}}{\bar{k}_i - \bar{k}_{i-1}} & \text{if } \bar{k}_{i-1} \leq k \leq \bar{k}_i \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

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

Define $P_t(p,j)$ as the fraction of households at time t that have productivity level $\bar{\theta}_p$ and capital level $\bar{k}_j$. Then define $P_t(p)$ as the column vector $(P_t(p,0), \ldots, P_t(p,n_y))^T$, and stack all the $P_t(p)$'s into the column vector

$$\mathbf{P}_t \equiv \begin{bmatrix} P_t(1) \\ \vdots \\ P_t(n_P) \end{bmatrix} \tag{20}$$

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. For a level of permanent productivity $\bar{\theta}_p$, the transition from the distribution $\tilde{P}_t(j)$ at the beginning of period t (after the shock to productivity has realized) and the end-of-period distribution P_t(,j) is given by

$$P_t(j) = \Pi_j^k \left(\mathbf{s}_{t,j}, w_t, R_t, T_t)\right) \tilde{P}_t(j) \tag{21}$$

where the elements of the transition matrix Π_j^k are given by

$$\Pi_p^k(l,j) = \sum_{j=1}^{n_y} \phi(l, \hat{\mathcal{S}}_p(R_t k_{t-1,j} + W_t(1 - \tau_t^l)\theta_{t,j}\xi_{t,j}; \mathbf{s}_{t-1})) \tag{22}$$

From the properties of $\phi(i,k)$, each column of Π_j^k has at most $2n_y$ non-zero elements.

Using the Markov transition matrix of permanent productivity Π^p, we can write the transition from end-of-period distributions $t-1$ and t as the linear dynamic equation $\mathbf{P}_t$:

$$\mathbf{P}_t = \Pi\left(\mathbf{s}_t, w_t, R_t, T_t)\right) \mathbf{P}_{t-1} = \Pi(\hat{\Omega}_t)\mathbf{P}_{t-1} \tag{23}$$

where

$$\Pi(\hat{\Omega}_t) = (I_{n_d} \otimes \Pi^p) \begin{bmatrix} \Pi_1^k & 0 & \dots & 0 & 0 \\ 0 & \Pi_2^k & \dots & 0 & 0 \\ 0 & 0 & \dots & \Pi_{n_P-1}^k & 0 \\ 0 & 0 & \dots & 0 & \Pi_{n_P}^k \end{bmatrix} \tag{24}$$

The transition matrix Π is a function of $\mathbf{s}_t$, w_t, R_t and T_t, which in turn are functions of the current aggregate state vector $\hat{\Omega}_t \equiv (\mathbf{P}_{t-1}, \Theta_t, \tau_t^k)$.

The details are given in Appendix **??**.

2.4.3 The Discrete Model

With the finite approximations defined above, we can now reduce the model to a finite set of equations in each period t. The **Discrete Model** consists of the equations (3), (8), (18), (23) and the government budget constraint

$$T_t = \tau_t^k \hat{K}(\mathbf{P}_t) = \tau_t^k \sum_{j=0}^{n_d} \left(\sum_{p=1}^{n_P} P_t(p,j) \right) \bar{k}_j \tag{25}$$

where $\hat{K}(\mathbf{P}_t)$ is aggregate capital, the first moment of the cross-sectional distribution of capital.

These equations define, for each period t, a system of $n_s + n_d + 4$ equations in just as many variables: $\mathbf{s}_t$, $\mathbf{P}_t$ (note that one of the $P_t(j,i)$ is redundant since probabilities sum to unity), T_t, Θ_t and τ_t^k. Of those, $\mathbf{P}_t$ are endogenous state variables, Θ_t and τ_t^k are exogenous state variables, while $\mathbf{s}_t$ and T_t are control or "jump" variables.

2.5 Parameter Values and Functional Forms

I will show numerical results for two different calibrations of the model. The first version follows Reiter (2008a). Variations in individual productivity are i.i.d., which means that $n_P = 1$. To make those fluctuations more important, the frequency of the model is annual. This calibration serves two purposes. First, it is similar to the original model (Krusell and Smith 1998) and most of the following literature, such that our results are well comparable to earlier findings. Second, by using $n_P = 1$ and $n_d = 1000$, we have around 1000 state variables, such that the exact linearized solution can be computed. This allows a rigorous measurement of the approximation errors induced by approximate aggregation. Therefore I will document the failure of low-dimensional approximations in this framework.

The second calibration allows for persistent variations in individual productivity, using $n_P = 11$. The frequency of the model is quarterly, which leads to a sparser transition probability matrix than in an annual model. The purpose of this calibration is to show that the CM allows to achieve a high-precision solution even in a case where the number of state variables (here around 11000) is too high to permit an exact linearized solution. From an economic point of view, this calibration is interesting because it leads to a realistic distribution of income, consumption and wealth, with a calibration of income that is similar to what microeconometricians find.

The Distributions of Earnings

	Gini	Quintiles					Top Groups (%)		
Economy		1st	2nd	3rd	4th	5th	90–95	95–99	99–100
U.S.	0.63	-0.40	3.19	12.49	23.33	61.39	12.38	16.37	14.76
CDR	0.63	0.00	3.74	14.59	15.99	65.68	15.15	17.65	14.93

The Distributions of Wealth (%)

	Gini	Quintiles					Top Groups (%)		
Economy		1st	2nd	3rd	4th	5th	90–95	95–99	99–100
U.S.	0.78	-0.39	1.74	5.72	13.43	79.49	12.62	23.95	29.55
CDR	0.79	0.21	1.21	1.93	14.68	81.97	16.97	18.21	29.85

Table 1: Earnings and wealth distribution
Source: Castaneda, Diaz-Gimenez, and Rios-Rull (2003).

2.5.1 The small calibration

Here I follow Reiter (2008a). The frequency of the model is annual. Standard values are used for most of the the model parameters: $\beta = 0.95$, $\alpha = 1/3$, $\delta = 0.1$. For the utility function I use CRRA

$$U(c) = \frac{c^{1-\gamma} - 1}{1 - \gamma} \tag{26}$$

with risk aversion parameter $\gamma = 1$ (the reader can use the computer programs to explore other degrees of risk aversion). The parameter A in the production function is set to $A = (\beta^{-1} - 1 + \delta)/\alpha$, such that the steady state capital stock of the representative agent model equals 1.

For the technology shock I choose $\rho_\Theta = 0.8$ and $\sigma_z = 0.014$, which are standard values, in annual terms. 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 to $\sigma_\tau = 0.02$.

The labor income distribution is assumed to be log-normal. Following Carroll (2001), I choose standard deviation and mean of $\log(\xi_{t,i})$ as $\sigma = 0.2$ and $\mu = -0.5\sigma^2$, which gives an expected value of productivity of 1, as required by (10). Reiter (2008a) also considers a Pareto distribution, to obtain a more realistic cross-sectional variation of capital, but here I will achieve this by persistent productivity shocks, cf. Section 2.5.2.

2.5.2 The big calibration

With a persistent component in individual productivity, the model should be able to explain broadly the features of the distribution of earnings and wealth, that are detailed in Table 1. [DETAILS TO BE FILLED IN.]

3 The State-Dependent Pricing Model

The model of this section is similar to Reiter, Sveen, and Weinke (2009) insofar as price setting and monetary policy are concerned. There the model also features lumpy investment decision of the firms. In the present paper, I abstract from capital, and rather consider firm-specific stochastic productivity.

3.1 Households

The representative household maximizes expected discounted utility

$$E_t \sum_{k=0}^{\infty} \beta^k U\left(C_{t+k}, L_{t+k}\right),$$

where β is the subjective discount factor, L_t is the number of hours worked and C_t denotes a Dixit-Stiglitz consumption aggregate

$$C_t \equiv \left(\int_0^1 C_t\left(i\right)^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}}, \tag{27}$$

with corresponding price index

$$P_t \equiv \left(\int_0^1 P_t\left(i\right)^{1-\epsilon} di \right)^{\frac{1}{1-\epsilon}} \tag{28}$$

where ϵ is the elasticity of substitution between different varieties of goods $C_t\left(i\right)$. With this price index, and with households allocating their spending optimally on the available goods, consumption expenditure can be written as $P_t C_t$. I use the period utility function

$$U\left(C_t, L_t\right) = \ln C_t + \eta \ln\left(1 - L_t\right)$$

Even though financial assets are not traded in equilibrium, it is convenient to assume complete financial markets in the model. The household budget equation then reads

$$P_t C_t + E_t\left\{Q_{t,t+1} D_{t+1}\right\} \leq D_t + P_t W_t L_t + T_t, \tag{29}$$

where $Q_{t,t+1}$ denotes the stochastic discount factor for random nominal payments and D_{t+1} gives the nominal payoff associated with the portfolio held at the end of period t. We have also used the notation W_t for the real wage and . T_t is nominal dividend income resulting from ownership of firms. Note that the stochastic discount factor is unambiguous in equilibrium since households are identical.

Denoting the real stochastic discount factor by $Q^R_{t,t+1} \equiv Q_{t,t+1}\left(\frac{P_{t+1}}{P_t}\right)$, the household Euler equation is given by

$$Q^R_{t,t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-1}, \tag{30}$$

The first order condition for labor supply is

$$\phi C_t \left(1 - N_t\right)^{-\phi} = W_t, \tag{31}$$

3.2 Firms

There is a continuum of firms and each of them is the monopolistically competitive producer of a differentiated good. Each firm $i \in [0,1]$ is assumed to maximize its market value subject to constraints implied by the demand for its good and the production technology it has access to. Moreover each firm faces random fixed costs of price adjustment. This implies generalized (S,s) rules for price-setting. Productivity shocks and monetary policy shocks represent the sources of aggregate uncertainty. In each period the time line is as follows.

1. The cost of price adjusting, c_p, and the idiosyncratic productivity shock α realize.
2. The firm changes its price (or not).
3. Production takes place.
4. The cost of adjusting the capital stock, c_k, realizes.
5. The firm invests (or not).

Each firm produces with the Cobb-Douglas production function

$$Y_t(i) = Z_t \alpha_t(i) L_t(i) \tag{32}$$

where α_L and α_K denote the shares of labor and capital in production. The aggregate level of technology, Z_t, is assumed to be given by the following process

$$\ln Z_t \equiv z_t = \rho_z z_{t-1} + e_{z,t}, \tag{33}$$

where $e_{z,t}$ is i.i.d. with zero mean.

In order to change its price the firm must pay a fixed cost, denoted by $C_{p,t}(i)$, which is measured in units of the aggregate good and given by

$$C_{p,t}\left(P_t(i), P_{t+1}(i), c_p\right) = \begin{cases} 0 \text{ if } P_{t+1}(i) = P_t(i), \\ c_p \text{ otherwise,} \end{cases} . \tag{34}$$

Each period, c_p is drawn independently, from a distribution with a tent-shaped density.

Cost-minimization on the part of households and firms implies that demand for good i is given by

$$Y_t^d(i) = \left(\frac{P_t(i)}{P_t}\right)^{-\epsilon} Y_t^d, \tag{35}$$

where aggregate demand is $Y_t^d = C_t + C_{p,t}$, the sum of consumption and aggregate price-setting costs, $C_{p,t} = \int_0^1 C_{p,t}(i)\, di$.

Each firm maximizes its market value

$$E_t \sum_{k=0}^{\infty} Q_{t,t+k}^R \left\{\Phi_{t+k}(i) - C_{p,t+k}(i)\right\}, \tag{36}$$

where $\Phi_t(i) \equiv P_t(i) Y_t(i) - W_t L(i) - \zeta$ is the gross operating surplus. The maximization is done subject to the constraints in equations (32), (**??**), (34) and (35).

3.3 Market Clearing and Monetary Policy

The goods and labor market clearing conditions are

$$Y_t(i) = Y_t^d(i) \text{ for all } i. \tag{37}$$

$$\int_0^1 L_t(i)\, di = L_t. \tag{38}$$

Monetary policy takes the form of a simple interest rate rule

$$R_t = R_{t-1}^{\phi_r} \left(\beta^{-1} \left(\frac{P_t}{P_{t-1}} \right)^{\phi_\pi} \right)^{1-\phi_r} e^{e_{r,t}}, \tag{39}$$

where parameters ϕ_π and ϕ_r measure the responsiveness of the nominal interest rate to changes in current inflation and past nominal interest rates, respectively, and $e_{r,t}$ is i.i.d. with zero mean.

3.4 Parameter values

I use standard parameter values. [DETAILS TO BE FILLED IN.]

4 Models of approximate aggregation

In this paper I do not ask whether the discrete model is a good approximation to the theoretical model. This is an important question, which I leave for future work. I rather focus on the question whether the solution to the discrete model, which has a high-dimensional state vector, can be approximated well by a lower-dimensional model, in the spirit of Krusell and Smith (1998) and a large subsequent literature. In the context of linearized solutions, I will therefore consider the solution to the discrete model as the "exact solution", and the solutions to the low-dimensional models as approximate solutions. We will find that an approximate model with a state vector of dimension about 20 will give a very precise solution, much more precise than the one-moment solution advocated by Krusell and Smith (1998).

Investigating the accuracy of low- or medium-dimensional approximations is useful in at least two respects. First, it helps in getting high-accuracy solutions in cases where the discrete model has an even larger state vector, such that the exact solution can no longer be computed (cf. Section 8). And furthermore, the low-dimensional model can be used to obtain higher-order approximations (cf. Section 9).

4.1 State-space reduction by bounded rationality

To reduce the dimension of the state space, we make the bounded-rationality assumption that the household bases its saving decision on a small vector of state variables, which

contains only partial information about the true high-dimensional state vector. Assume for now that household savings of time t depend only on the lagged exogenous state Z_{t-1}, current shocks ϵ_t, and a vector M_{t-1} which contains statistics on the beginning-of-period distribution $\mathbf{P}_{t-1}$.

$$\mathbf{s}_t = \mathbf{s}(M_{t-1}, Z_{t-1}, \epsilon_t) \tag{40}$$

In the literature, $\mathbf{P}_{t-1}$ usually contains moments such as the cross-sectional mean and variance of capital. Here we allow all linear functions of the cross-sectional distribution:

$$M_{it} = H_i \mathbf{P}_t \tag{41}$$

with H_i a known vector. Since the aggregate capital stock plays a crucial role in the economy, determining factor prices, we always use it as the first statistic, such that

$$K_t = H_1' \mathbf{P}_t \tag{42}$$

the choice of other statistics will be discussed in Section 5.

Notice that in (44) we allow that M_t to depend on Z_{t-1} and ϵ_t separately, not just through Z_t. Since M_t contains icomplete information about the true state $\mathbf{P}_t$, past variables such as Z_{t-1} carry additional information on M_t that goes beyond what is contained in Z_t.

Given any household savings function (40), one can then derive what this implies for the dynamics of the aggregate distribution of wealth. A linear approximation to aggregate dynamics is of the form

$$\begin{bmatrix} \mathbf{P}_t - \mathbf{P}^* \\ Z_t \end{bmatrix} = A(\mathcal{S}) \begin{bmatrix} \mathbf{P}_{t-1} - \mathbf{P}^* \\ Z_{t-1} \end{bmatrix} + B(\mathcal{S})\epsilon_t \tag{43}$$

The specification (40) is a deviation from the assumption of household optimization under perfectly rational expectations. The question is then: how should household expectations be formulated so as to achieve approximate consistency between individual behavior and expectations, and the aggregate dynamics in equilibrium. Section 4.2 and 4.3 describe two different approaches, and Section 4.4 will be shown that they are very closely related.

4.2 A linear Krusell/Smith algorithm

Key to the Krusell/Smith algorithm is the assumption that economic agents suppose the following aggregate law of motion for the aggregate variables that affect their decisions:

$$\begin{bmatrix} M_t - M^* \\ z_t \end{bmatrix} = \begin{bmatrix} A_p & C_p \\ 0 & \rho_z \end{bmatrix} \begin{bmatrix} M_{t-1} - M^* \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} B_p \\ I \end{bmatrix} \epsilon_t \tag{44}$$

This assumption is part of the bounded-rationality argument, because (44) is not exactly true in our economy. The high-dimensional dynamic system (23) does not exactly aggregate to the low-dimensional system (44). The idea of Krusell and Smith (1998) is to find the transition matrices A_p and B_p such that (44) is approximately true in equilibrium. Notice

that a nonlinear version of the KS algorithm could (and should) replace (44) by a nonlinear law of motion, but in our linearized context, the VAR(1) specification (44) is the most natural formulation.

The linear Krusell/Smith algorithm is then:

1. For given matrices A_p and B_p, solve the model consisting of the equations (3),(8), (18), (44) and (25) by linearization about the steady state. Effectively, this solve the household problem for the given aggregate law of motion (ALM) in Equ. (44).

2. From the solution, compute the linear dynamic system (43). This describes the distribution dynamics, given the household behavior computed in step 1.

3. Estimate the VAR model (44), using as data generating process the aggregate distribution dynamics computed in step 2. This is not done by simulating the model, but by computing the asymptotic regression coefficients, as explained in Appendix **??**. Call the estimation results $\hat{A}_p$ and $\hat{B}_p$.

4. Iterate over A_p and B_p (using either fixed point iteration or a quasi-Newton method) until consistency is achieved, that means $\hat{A}_p = A_p$ and $\hat{B}_p = B_p$.

Unlike the original Krusell/Smith algorithm, this linearized algorithm is not affected by any sampling error.

Given the matrix A_p, we can replace Equ. (23) by Equ. (44), and again have a system of $n_s + n_d + 4$ equations, namely (3),(8), (18), (44) and (25), in just as many variables. I call this system the modified Krusell/Smith model. Again we collect all those variables in the vector x_t, and solve the model by linear approximation. This gives us a solution $\mathbf{s}_t \left(M_{t-1}, \log \Theta_t, \tau_t^k\right)$.

4.3 Collocation method approach

Under "bounded rationality", household behavior cannot be consistent with rational expectations for all realizations of the cross-sectional distribution. The idea of the Collocation method is to assume that behavior is consistent with rational expectations in situations where the cross-sectional distribution is "typical" in some sense. To be more precise, for any vector of statistics M, we consider a "typical realization" $\mathbf{P}^{pr}(M)$ of the cross-sectional distribution. Of course, we require that it really has those characteristics:

$$H\mathbf{P}^{pr}(M) = M \tag{45}$$

What does "typical" mean? We will interpret it here as the expected value of $\mathbf{P}$ conditional on (45), where the expectation is taken over the ergodic distribution of $\mathbf{P}$, according to some approximate model solution. This will be made precise below. In the context of linearized solutions, $\mathbf{P}^{pr}(M)$ can be written as

$$\mathbf{P}^{pr}(M) = \mathbf{P}^* + D^{pr}(M - M^*) \tag{46}$$

Then we require that for any M_{t-1} next period's statistic M_t are computed by applying the aggregate transition matrix $\Pi(M_{t-1}, z_{t-1}, \epsilon_t)$ to the proxy distribution $\mathbf{P}^{pr}(M_{t-1})$:

$$M_t = H\Pi(M_{t-1}, z_{t-1}, \epsilon_t)D^{pr}M_{t-1} \tag{47}$$

For a given D^{pr}, the model then consists of the equations (3),(8), (18), (25) and (47). This can again be solved by linearization about the steady state.

An iterative approach is needed in order to find a suitable matrix D^{pr} (unless we assume we have already solved the exact model; this is relevant if we want to go on to a nonlinear solution, cf. Section 9):

1. Set $n_m = 1$ and set the according H. For given m, choose the proxy distribution that is closest to the the steady state among all distributions satisfying the moment condition (cf. Section **??**).

2. Solve the Collocation method model by linearization. This gives us a savings function $\mathcal{S}$.

3. Compute the $A(\mathcal{S})$ and $B(\mathcal{S})$ that describe the dynamics of $x_t \equiv \begin{bmatrix}\mathbf{P}_t \\ z_t\end{bmatrix}$, which was explained in Section 4.1.

4. Compute the unconditional covariance matrix of x as implied by the model solution. For that, let us define $\mathcal{L}(A, B, \Sigma)$ as the unique symmetric solution to the discrete Lyapunov equation

$$\mathcal{L}(A, B, \Sigma) = A\mathcal{L}(A, B, \Sigma)A^T + B\Sigma B^T \tag{48}$$

   where we assume that the matrix A is stable in the sense that all eigenvalues are smaller than unity in absolute value. Then the covariance matrix of x is given by $\mathcal{L}(A(\mathcal{S}), B(\mathcal{S}), \Sigma_\epsilon)$.

5. Choose a new n_m and corresponding H_m.

6. Compute a new D^{pr} in the following way. Define $H_z \equiv \begin{bmatrix}0 & I\end{bmatrix}$ as the matrix that selects Z from xx. Then define $H \equiv \begin{bmatrix}H_m \\ H_z\end{bmatrix}$ and set

$$D^{pr} = \mathcal{L}(A(\mathcal{S}), B(\mathcal{S}), \Sigma_\epsilon)H'\left[H\mathcal{L}(A(\mathcal{S}), B(\mathcal{S}), \Sigma_\epsilon)H'\right]^{-1} \tag{49}$$

   The interpretation of (49) is that $D^{pr}\begin{bmatrix}M \\ Z\end{bmatrix}$ gives the expected value of x conditional on $Hx = \begin{bmatrix}M \\ Z\end{bmatrix}$, assuming a joint normal distribution for x. Notice that $HD^{pr} = I$, which guarantees that the proxy distribution for the statistics M really has these statistics. Obviously, the last n_z rows of D^{pr} are equal to H_z, because the Z are among the conditioning variables. [DEFINE nz!!??]

7. Repeat steps 2–6 until convergence is reached, in some suitable sense.

4.4 Equivalence between KS and Collocation method

Define $x_t \equiv (M_{t-1}, Z_{t-1}, \epsilon_t)$ as the vector of all state variables that affect the household saving function. Notice that Z_{t-1} and ϵ enter the savings function separately, not just through Z_t, because Z_{t-1} affects the proxy distribution while ϵ does not. Define $\hat{M}_t \equiv \begin{bmatrix} M_t \\ Z_t \end{bmatrix}$. Equ. (18) defines a set of $n_s + 1$ Euler equations, at the $n_s + 1$ knot points in the approximation of the savings function. These equations depend on the variables $\mathbf{s}_{t-1}$, x_t and $\mathbf{s}_t$. After linearization, they can be written as

$$\mathcal{E}_1 \mathbf{s}_{t-1} + \mathrm{E}_{t-1}\left[\mathcal{E}_2 \hat{M}_{t-1} + \mathcal{E}_3 \mathbf{s}_t + \mathcal{E}_4 \epsilon_t\right] = 0 \tag{50}$$

where the $\mathcal{E}_i$ are constant matrices (they depend on the choice of knot points $\bar{x}_{it}$, the Gaussian quadrature points $\bar{\xi}_j$ etc., but not on any variables of the model). Note that future shocks do not appear in (51) because they have zero condition expectation and (51) is linear. The solution of the linear model including (40) delivers the decision rule $\mathbf{s}_t = S_x \hat{M}_{t-1} + S_\epsilon \epsilon_t$, so that (50) becomes

$$\mathcal{E}_1 \left(S_x \hat{M}_{t-2} + S_\epsilon \epsilon_{t-1}\right) + \mathrm{E}_{t-1}\left[(\mathcal{E}_2 + \mathcal{E}_3 S_x)\hat{M}_{t-1} + (\mathcal{E}_4 + \mathcal{E}_3 S_\epsilon)\epsilon_t\right] = 0 \tag{51}$$

A solution to the KS algorithm consists in matrices S_x, A_p, B_p, such that

1. S_x satisfies the linearized Euler equation (51).

$$\mathcal{E}_1 S_x \hat{M} + S_\epsilon \epsilon + (\mathcal{E}_2 + \mathcal{E}_3 S_x)(A_p \hat{M} + B_p \epsilon) = 0, \qquad \forall \hat{M}, \epsilon \tag{52}$$

2. $A(\mathcal{S})$ and A_p are stable [3]

3. The matrices A_p and B_p satisfy the normal equations

$$H A(\mathcal{S}) \mathcal{L}(\mathcal{S}) H' = A_p H \mathcal{L}(\mathcal{S}) H' \tag{53}$$

$$H B(\mathcal{S}) = B_p \tag{54}$$

where $A(\mathcal{S})$ and $B(\mathcal{S})$ are defined as in (**??**), and $\mathcal{L}(\mathcal{S})$ is defined as the solution of

$$\mathcal{L}(\mathcal{S}) = A(\mathcal{S}) \mathcal{L}(\mathcal{S}) A(\mathcal{S})^T + B(\mathcal{S}) \Sigma_\epsilon B(\mathcal{S})^T \tag{55}$$

Note that B_p plays no essential role in this algorithm because of certainty-equivalence.

For a given D^{pr}, a solution to the Collocation method algorithm consists in a matrix S_x such that

1. $A(\mathcal{S})$ is stable

[3] The stability of $A(\mathcal{S})$ implies the stability of A_p, cf. Appendix **??**.

2. S_x satisfies the linearized Euler equation (51).

$$\mathcal{E}_1 S_x \hat{M} + S_\epsilon \epsilon + (\mathcal{E}_2 + \mathcal{E}_3 S_x) H (A(\mathcal{S}) D^{pr} \hat{M} + B(\mathcal{S}) \epsilon) = 0, \qquad \forall \hat{M}, \epsilon \tag{56}$$

Proposition 1. *If (S_x, A_p, B_p) are a solution of the KS algorithm , then S_x is a solution to the Collocation method algorithm if for the proxy distribution we use*

$$D^{pr} = \mathcal{L}(\mathcal{S}) H' \left[H \mathcal{L}(\mathcal{S}) H' \right]^{-1} \tag{57}$$

Notice that D^{pr} in (57) is not known ex ante, because it depends on S_x, so that a iterative algorithm is necessary. To interpret (57), notice that D^{pr} delivers as proxy distribution the expected value of $\mathbf{P}$ condition on M, under the assumption that the joint distribution of the $\mathbf{P}$ is given by (55).

4.5 Iterative solutions: KS versus Collocation method

The Krusell/Smith approach and the Collocation method approach are theoretically equivalent in the sense that if they both converge, they converge to the same solution. Nevertheless, in order to find the fixed point, they give rise to different iterative schemes with very different convergence properties. My experience from solving the model of Section 2 shows that the KS approach is fine as long as we use one or two moments. Using more than two moments, it becomes very difficult to implement, for the following reasons:

1. Fixed point iteration does not converge, because some eigenvalues of the Jacobian at the solution are strongly positive
2. quasi-Newton methods are
   - slow, many function calls for Jacobian
   - difficult to make converge.

In contrast, the Collocation method approach works fine. Fixed point iteration leads us to a very precise solution after only two or three steps.

To understand the differences in the iterative procedures, notice that the linearized Euler equation in the fixed point boils down to

$$\mathcal{E}_1 S_x + (\mathcal{E}_2 + \mathcal{E}_3 S_x) H (A(\mathcal{S}) \mathcal{L}(\mathcal{S}) H' \left[H \mathcal{L}(\mathcal{S}) H' \right]^{-1} = 0 \tag{58a}$$

$$S_\epsilon + H B(\mathcal{S}) = 0 \tag{58b}$$

With the Collocation method, the fixed point iteration becomes

$$\mathcal{E}_1 S_x^k + (\mathcal{E}_2 + \mathcal{E}_3 S_x^k) H (A(\mathcal{S}^k) \mathcal{L}\left(\mathcal{S}^{k-1}\right) H' \left[H \mathcal{L}\left(\mathcal{S}^{k-1}\right) H' \right]^{-1} = 0 \tag{59a}$$

$$S_\epsilon^k + H B(\mathcal{S}^k) = 0 \tag{59b}$$

With Krusell/Smith, the fixed point iteration becomes

$$\mathcal{E}_1 S_x^k + (\mathcal{E}_2 + \mathcal{E}_3 S_x^k) H (A(\mathcal{S}^{k-1}) \mathcal{L}\left(\mathcal{S}^{k-1}\right) H' \left[H\mathcal{L}\left(\mathcal{S}^{k-1}\right) H'\right]^{-1} = 0 \tag{60a}$$

$$S_\epsilon^k + HB(\mathcal{S}^{k-1}) = 0 \tag{60b}$$

In step k of the KS fixed point iteration (60), we take aggregate behavior $A(\mathcal{S}^{k-1})$ and $B(\mathcal{S}^{k-1})$ as given from the last iteration, while in the Collocation method fixed point iteration, we solve for $A(\mathcal{S}^k)$ and $B(\mathcal{S}^k)$ simultaneously with $\mathcal{S}^k$, and use the ALM of the last period only through the term $D^{pr} = \mathcal{L}\left(\mathcal{S}^{k-1}\right) H' \left[H\mathcal{L}\left(\mathcal{S}^{k-1}\right) H'\right]^{-1}$. Why is this better? The idea is that D^{pr} should have only a small influence on individual behavior. This can be seen most clearly in two limiting cases:

- "Exact aggregation": $\mathcal{S}^k$ allows exact aggregation if there exists a matrix $\hat{A}$ such that
$$\hat{A}H = HA(\mathcal{S}^k) \tag{61}$$
Then (59a) reduces to $\mathcal{E}_1 S_x^k + (\mathcal{E}_2 + \mathcal{E}_3 S_x^k)\hat{A} = 0$, such that the effect of $\mathcal{S}^{k-1}$ disappears. This is trivially the case if H has full rank, which means that we use as many statistics as there are state variables in the discrete model.

- "Ergodic aggregation": If $\mathcal{L}(\mathcal{S})$ does not have full rank, it means that the economy effectively lives in a lower-dimensional subspace of the n-dimensional space of distributions. Assume that H spans this subspace, so that $H\mathcal{L}(\mathcal{S})H'$ has the same rank as $\mathcal{L}(\mathcal{S})$. This implies that there exists a matrix L such that $\begin{bmatrix} H \\ L \end{bmatrix}$ has full rank and $L\mathcal{L}(\mathcal{S})H' = 0$. Then

$$\mathcal{L}(\mathcal{S})H'\left[H\mathcal{L}(\mathcal{S})H'\right]^{-1} = \begin{bmatrix} H \\ L \end{bmatrix}^{-1} \begin{bmatrix} H \\ L \end{bmatrix} \mathcal{L}(\mathcal{S})H'\left[H\mathcal{L}(\mathcal{S})H'\right]^{-1} = \begin{bmatrix} H \\ L \end{bmatrix}^{-1} \begin{bmatrix} I \\ 0 \end{bmatrix} \tag{62}$$

(62) shows that the exact choice of $\mathcal{L}(\mathcal{S})$ does not matter as long as H spans the ergodic set of (technically: H spans the set of eigenvectors of $\mathcal{L}(\mathcal{S})$ with strictly positive eigenvalues).

Notice that ergodic aggregation is weaker than exact aggregation. The latter condition assumes exact aggregability for the whole state space, while the former assumes aggregability only in the ergodic distribution. Exact aggregation then need not hold starting from an arbitrary initial distribution.

5 Model reduction: finding a set of statistics

There is a large literature in the systems and control theory that deals with model reduction in linear settings, this means, the approximation of a given high-dimensional dynamical system by a lower-dimensional one. To find a general equilibrium of the heterogeneous-agent economy, the approximation problem is conceptually more complicated. We assume

that economic agents use a simplified description of the high-dimensional economic reality for their decision making. The approximate model that agents use then feeds back into their decisions, and therefore into the high-dimensional dynamics again. This leads to a fixed point problem: consistency is achieved if the approximate model that agents use for their decision making is a good approximation to the true high-dimensional system that results from agents' decisions. The results from system theory help in finding optimal low- or medium dimensional approximations to the high-dimensional system. In order to solve for general equilibrium, ideas from the literature on heterogeneous agents model will have to be brought into the picture.

Here we define the approximation problem as a problem of approximating the dynamics of an exogenously given stochastic process. Our approximation problem can be described as follows. Assume that the true dynamics is given by

$$x_t = Ax_{t-1} + B\epsilon_t \tag{63}$$

We search for a matrix H such that the vector of statistics

$$m = Hx \tag{64}$$

gives a good approximation to the dynamic system (63) in the sense that there exist matrices $\hat{A}$ and $\hat{B}$ such that the dynamics of

$$m_t = \hat{A}m_{t-1} + \hat{B}\epsilon_t \tag{65}$$

is similar to the dynamics of m implied by (63) together with (64).

There are a number of approaches to constructing a set of statistics m:

1. **Higher moments**, either of k, or $\log k$ or some other useful transformation of capital. Care must be taken to avoid multicollinearity, for example by using Chebyshev polynomials rather than monomials. This is the approach usually taken in the macroeconomics literature.

2. **Principal component analysis (PCA)**. Given (63), we can compute the unconditional covariance matrix $\mathcal{L}(A, B, \Sigma_\epsilon)$ of the state vector x. PCA gives us the components of $\mathcal{L}(A, B, \Sigma_\epsilon)$ that are most important, in the sense that they are the components of the state vector that are really moving over time. It turns out that the numerical rank of $\mathcal{L}(A, B, \Sigma_\epsilon)$ is much lower than the dimension of the system, so that we can obtain an almost perfect solution if the number of statistics we use equals the numerical rank of the covariance matrix.

3. As we said above, what is really important in the stochastic growth model is to predict future capital stocks, since they determine prices, which is all that the economic agents have to know. (In the sticky price model, the situation is a bit more complicated, since more variables have to be taken into account at any point in time.) To fix ideas, let us concentrate for now on the growth model. From the viewpoint of a

household that has to make his consumption and savings decision, the approximation problem can be considered as perfectly solved if

$$\mathrm{E}_{t|\hat{A},\hat{B}}[m_{1,t+i}] = \mathrm{E}_{t|A,B}[h_1 x_{t+i}], \qquad i = 1, 2, \ldots \tag{66}$$

Notice that in the context of linearized solutions, we do not worry about predictions of higher moments of m_1, because the solution is of the certainty-equivalence type anyway. Since $\mathrm{E}_{t|A,B}[h_1 x_{t+i}] = h_1 A^i x_t$, it is natural to consider the **Krylov subspace**

$$\hat{m}_i \equiv \hat{H}_i x, \qquad \hat{H}_i \equiv h_1 A^i \tag{67}$$

where i goes from 1 up to some chosen forecasting horizon T. With this choice, we obtain

$$\hat{m}_{t,1} = \mathrm{E}_{t|A,B}[h_1 x_{t+1}] \tag{68}$$
$$\hat{m}_{t,2} = \mathrm{E}_{t|A,B}[\hat{m}_{t+1,1}] = \mathrm{E}_{t|A,B}[h_1 x_{t+2}] \tag{69}$$
$$\ldots \tag{70}$$
$$\hat{m}_{t,T} = \mathrm{E}_{t|A,B}[\hat{m}_{t+1,T-1}] = \mathrm{E}_{t|A,B}[h_1 x_{t+T}] \tag{71}$$

This means that the $\hat{m}_i$ predict the future capital stock as well as some future intermediate $\hat{m}_j$. A problem remains. To obtain high accuracy, a long forecasting horizon T is needed, so that it seems that we need a large number of statistics. Fortunately, and similar to what we have said in the discussion of PCA, the numerical rank of this Krylov subspace is relatively small. We don't need more states than what is the numerical rank of the Krylov subspace. The statistics are then obtained as an orthonormal basis of the Krylov subspace.

4. The concept of **Balanced Reduction** combines the ideas of PCA and the Krylov subspace method. One can show that there is a variable transformation, applied to the state vector x, such that in the transformed state space the states are ordered such that earlier states are more important both in the sense that they have higher ergodic variance and that they are more useful in predicting future capital stocks. [MORE DETAILED EXPLANATIONS AND FORMULA TO BE FILLED IN.]

6 Results I: Approximating the exact solution by lower-dimensional systems

First we look at cases where we can compute the exact solution of the linearized discrete model, and ask whether it is possible to approximate the dynamics of the exact solution by a lower-dimensional model. In Section 7 we will then see whether a similar accuracy can be obtained by a bounded-rationality solution, where we assume that the agents in the economy, not just the outside observer, use a simplified model.

6.1 Measuring the Accuracy of Approximate Models

When measuring accuracy, two different questions can be asked. First, how well does a model predict the relevant aggregate variables (for example capital) on average?, where the average is taken over the ergodic distribution of the model solution. Second, how well does the approximate model predict the transition path from an arbitrary starting distribution x_0. For this we need two accuracy measures.

Accuracy measure 1: unconditional mean squared error

The first measure, (77), addresses the main concern in Den Haan (2007), who suggests to use long-term forecasts to assess accuracy. In fact, the statistic (77) is the unconditional variance of an infinite horizon forecast.

While the state variables of the approximate model are not the same as of the exact model, both models are driven by the same shocks ϵ.

We combine (63) and (65) into

$$x_t^* = A^* x_{t-1}^* + B^* \epsilon_t \tag{72}$$

$$x^* \equiv \begin{bmatrix} x_t \\ \hat{x}_t \end{bmatrix}, \qquad A^* \equiv \begin{bmatrix} A & 0 \\ 0 & \hat{A} \end{bmatrix}, \qquad B^* \equiv \begin{bmatrix} B \\ \hat{B} \end{bmatrix} \tag{73}$$

The unconditional variance-covariance matrix Σ_{x^*} of x^* is obtained by solving the discrete Lyapunov equation

$$\Sigma_{x^*} = A^* \Sigma_{x^*} A^{*T} + B^* \Sigma_\epsilon B^{*T} \tag{74}$$

The i-th non-central moment $m_{i,t}$ is a linear function of the state variables, both in the exact and the approximate models. So we can write the moments of the exact and any approximate model as $m_{i,t} = H_i x_t$ and $\hat{m}_{i,t} = \hat{H}_i \hat{x}_t$, respectively, with known row vectors H_i and $\hat{H}_i$. From (**??**) and (**??**) it is clear that the unconditional expected error is zero:

$$\mathrm{E}(m_{i,t} - \hat{m}_{i,t}) = 0 \tag{75}$$

The unconditional mean squared error is then given by

$$\mathrm{E}[(m_{i,t} - \hat{m}_{i,t})^2] = \begin{bmatrix} H_i & -\hat{H}_i \end{bmatrix} \Sigma_{x^*} \begin{bmatrix} H_i^T \\ -\hat{H}_i^T \end{bmatrix} \tag{76}$$

Below we will report the root mean squared error, normalized by the standard deviation of $m_{i,t}$:

$$\sqrt{\frac{\begin{bmatrix} H_i & -\hat{H}_i \end{bmatrix} \Sigma_{x^*} \begin{bmatrix} H_i^T \\ -\hat{H}_i^T \end{bmatrix}}{H_i \Sigma_x H_i^T}} \tag{77}$$

This statistic can be interpreted as the average of an infinite-horizon forecast error under a typical realization of the driving shocks.

Accuracy measure 2: precision of transition path

For a n-step prediction of the statistics $m = Hx$, the difference between the exact and the approximate model is given by $HA^n x - \hat{A}^n Hx$. The maximum in the prediction error over all possible x is then given by the matrix norm

$$\|HA^n - \hat{A}^n H\|_2 \tag{78}$$

The results below report the maximum of (78) over $n = 1, \ldots, 200$.

6.2 Results for the stochastic growth model

Figure 1 considers the small calibration (idiosyncratic productivity is i.i.d.) of the model of Section 2. The exact model characterizes the cross-sectional distribution by 1000 variables, so with the exogenous processes it has in total 1002 state variables. Figure 1 reports approximation errors using 4 different sets of statistics, in each case varying the number of statistics used from 1 to 80. All statistics refer to the approximation of mean capital. This is the relevant statistic in this model, because it determines all the prices in the economy. Numbers are decimal logs. The upper left panel measures the approximation error as the infinite horizon forecast error (accuracy measure 1 in Section 6.1). With 1 moment only (the mean capital stock, as in Krusell and Smith (1998)), the forecast error is around 0.1 percent (of steady state capital), similar to what people have found in the literature following Krusell/Smith. Using more statistics, this error can be reduced to 10^{-8}. The polynomial approximations become numerically unstable for more than 30 statistics. This is because the higher moments are more and more stochastically dependent among each other. The other model reduction schemes suffer much less from this problem. The Krylov subspace approximation and the balanced reduction perform about equally well. The same results hold also if we look at the approximation error in the two impulse responses. An interesting insight emerges from the lower right panel, which gives the maximum error from a transition path (that is, the maximum error that could result starting from an arbitrary initial distribution). Using more and more moments of the distribution, or using PCA for the approximation, does not reduce this error. Even if we use dozens of statistics to characterize the distributions, there We can avoid this problem by using the Krylov subspace or the balanced reduction method. Those methods were designed so as to contain the necessary information for predicting the future capital stock.

Figure 2 provides similar information for the sticky price model. The same conclusions arise, except that an approximation with only 1 or 2 moments is more than one order of magnitude worse than in the case of the stochastic growth model.

7 Results II: Bounded-rationality solution of the models

In the last section we have seen that approximate models with moderate dimension (around 50) can yield very precise approximations to our high-dimensional model. In this exercise, the model solution is obtained under strictly rational expectations, and only the outside observer tries to approximate the given model by a lower-order model. In a bounded-rationality solution, as explained in Section 4, it is the economic agents who use an approximation to form a decision. The approximation therefore directly affects the solution of the model. The question is whether this solution comes close to the true rational-expectations solution.

We start by looking at the stochastic growth model using only one statistic, the aggregate capital stock. This is what most of the literature following Krusell and Smith (1998)) has done. Figure 3 shows impulse responses to a technology shock of four types of solutions: the exact (rational expectations) solution, the best approximation to the exact solution (called "VAR"; this is the solution considered already Section 6); the approximate model solution ("boundedly-rational", either with the Krusell-Smith method or the Collocation method method; the two solutions are identical, as should be clear from Proposition 1); the true dynamics resulting from the model if the household saving function of the approximate model is used. The figure shows that the four solutions are visually indistinguishable. This is the famous result that "one moment is enough". Figure 4 shows the analogous result for the tax shock. It looks quite different. All the approximations are far off the exact solution. What happens here? Let's first understand the impulse response to a tax shock:

1. On impulse, the tax shock is a redistribution from the rich to the poor.

2. First round effect: for *given expected interest rates*, aggregate consumption goes up, saving goes down, because the poor (in terms of assets) have a larger propensity to consume out of their wealth.

3. Second-round effect: the reduction in aggregate savings means that future capital goes down, and therefore expected future interest rates go up. This has the tendency to increase savings, and therefore partly offsets the first-round effect.

Where goes the one-moment approximation wrong? By looking at the mean only, households forget about the effect of the redistribution as soon as they forget the shock itself. Then they underestimate the future reduction in saving and the corresponding increase in interest rates, which results in a smaller second-round effect. This means that the first round effect (reduction in saving) is even stronger.

The good news from Figure 4 is that a careful accuracy check reveals the failure of the approximation. While the comparison to the exact solution is usually not available (because the exact solution cannot be computed), it is always possible to compare the model outcome to the simulation outcome that results from simulating the distribution of capital using the decision functions of the approximate model (compare lines "CM ALM"

and "CM sim." in the figure). They differ by the same order of magnitude as the difference between exact and approximate solution. A careful application of approximate-aggregation methods should always do this check.

Figure 5, which shows both the one-moment and the 8-moments approximate solution (Collocation method method) next to the exact solution, shows that 8 moments are enough to get a good approximation.

To get a very precise approximation, even more moments are needed. Using Krylov subspace approximation and 50 statistics, the infinite-horizon standard deviation (distance of approximate solution to exact solution) is $6 \cdot 10^{-6}$, which is not quite as good as the approximations in Figure 1, but one has to consider that the numerical errors involved in the bounded-rationality solution are much Since the square root of the machine precision is 10^{-8}, which is the maximum one could expect in such a complicated computation, an error of $6 \cdot 10^{-6}$ comes reasonably close.

Results for the sticky price model are similar [DETAILS TO BE FILLED IN.]

The following conclusions emerge:

- The bounded-rationality solution with 1 moment fails on model with substantial redistribution.
- This failure is visible from usual accuracy checks.
- Problem can be solved by using more moments (8 are fine)

8 Handling very big systems

So far we have looked at models where the exact solution of the linearized discrete model can be computed. In more complicated models this will not be true, for example in the model of Section 2, if we approximate household productivity by a high-dimensional Markov chain, and still want to maintain a fine representation of the cross-sectional distribution, the number of states may be between 10000 and 100000. To handle such a system, we have to do the following:

- Make sure that the transition matrix is sparse. From each point in the state space, there is only a limited set of points that an economic agent can reach in the next period with positive probability. This requires, for example, that idiosyncratic shocks have small support. If the transition matrix is very sparse, it is still possible to compute the invariant distribution of the model without aggregate shocks.
- Reduce the state space (number of state variables) by the methods described in Section 5.
- Reduce the number of decision variables; this could be done by general projection methods. However, one can also use approximation ideas that are parallel to the methods used in state-space reduction.

[DETAILS OF THE METHOD AND RESULTS TO BE FILLED IN. SO FAR I HAVE SOLVED MODELS WITH UP TO 30000 STATE VARIABLES IN THIS WAY.]

9 Nonlinear solution

The linearized model can be seen as an analogue of the well-known linearization methods for business cycle models, generalized to models of heterogenous agents and incomplete markets. It also shares the shortcoming of those methods. If we want to model the nonlinear effects of big aggregate shocks, or if we want to model asset choice, we may need a solution that is nonlinear not just in individual but also in aggregate variables. Nevertheless, even in cases where a linearized solution is not sufficient, it turns out that the solution to the linearized model is an essential ingredient in obtaining a precise higher solution. The linearized solution serves the following purposes:

1. To obtain higher order approximations of the solution, it is necessary to substantially limit the number of state variables (I can currently handle up to 20 state variables for a second-order approximation, with a method that combines the ideas from Reiter (2008a) and from Reiter (2008b)). This can be done by the methods of Section 5.

2. Projection approaches to higher order approximations need grids for the aggregate state variables. It is usually very difficult to construct reasonable grids in a medium-sized state space. Knowing the variance-covariance matrix of the state variables in the linearized model allows to construct a reasonable grid (that means, a grid that describes the part of the state space in which the solution really lies).

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

A detailed description of the nonlinear algorithm is left to a companion paper.

10 Conclusions

Using two prominent examples of heterogeneous agent models, the paper has shown that a medium-dimensional (20 to 50 state variables) approximation of the model is sufficient to provide a solution that comes very close to the exact solution. A collocation method has been provided that allows the computation of solutions with this size of the state space.

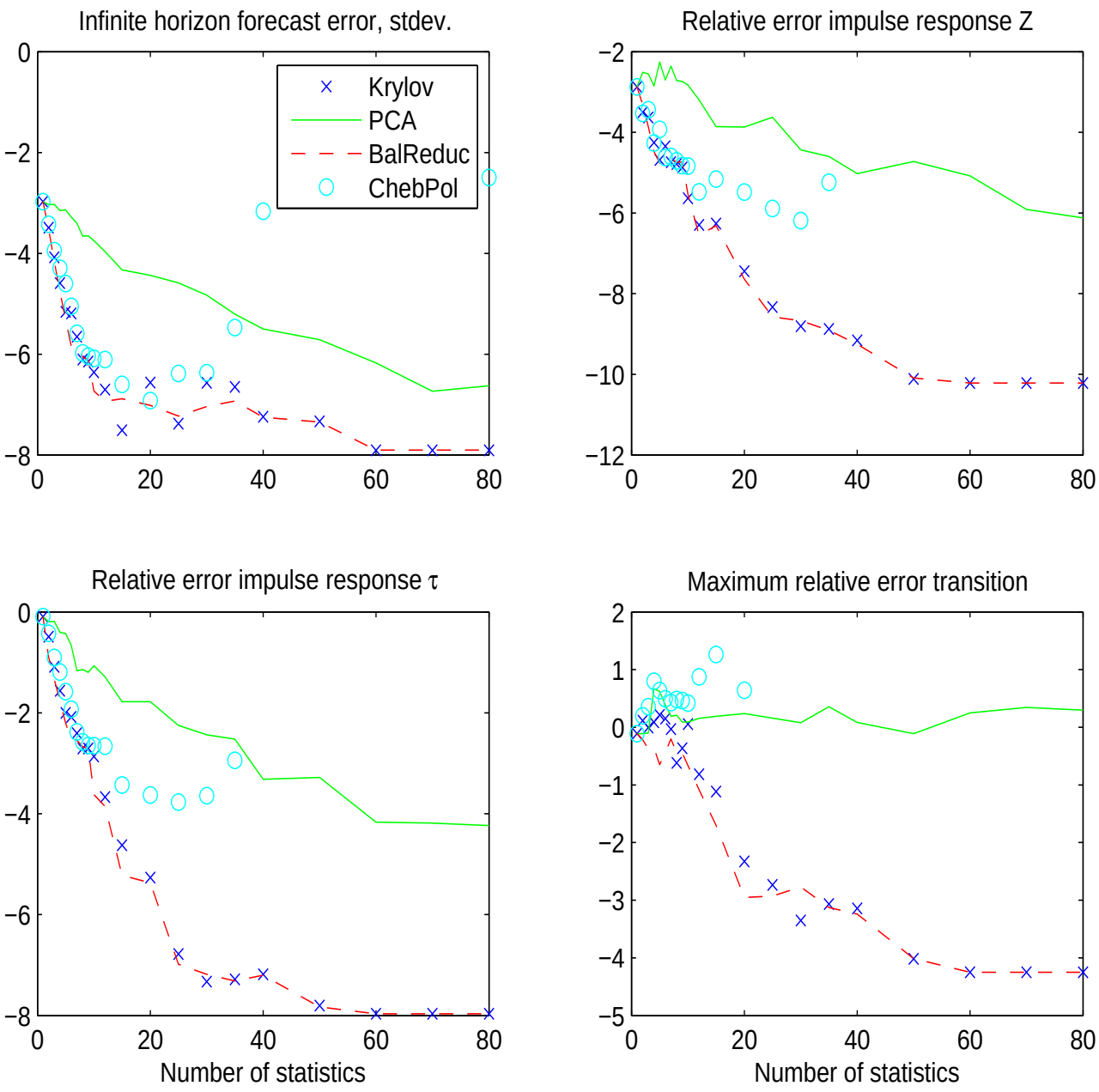

Figure 1: log10 of approximation errors, stochastic growth model

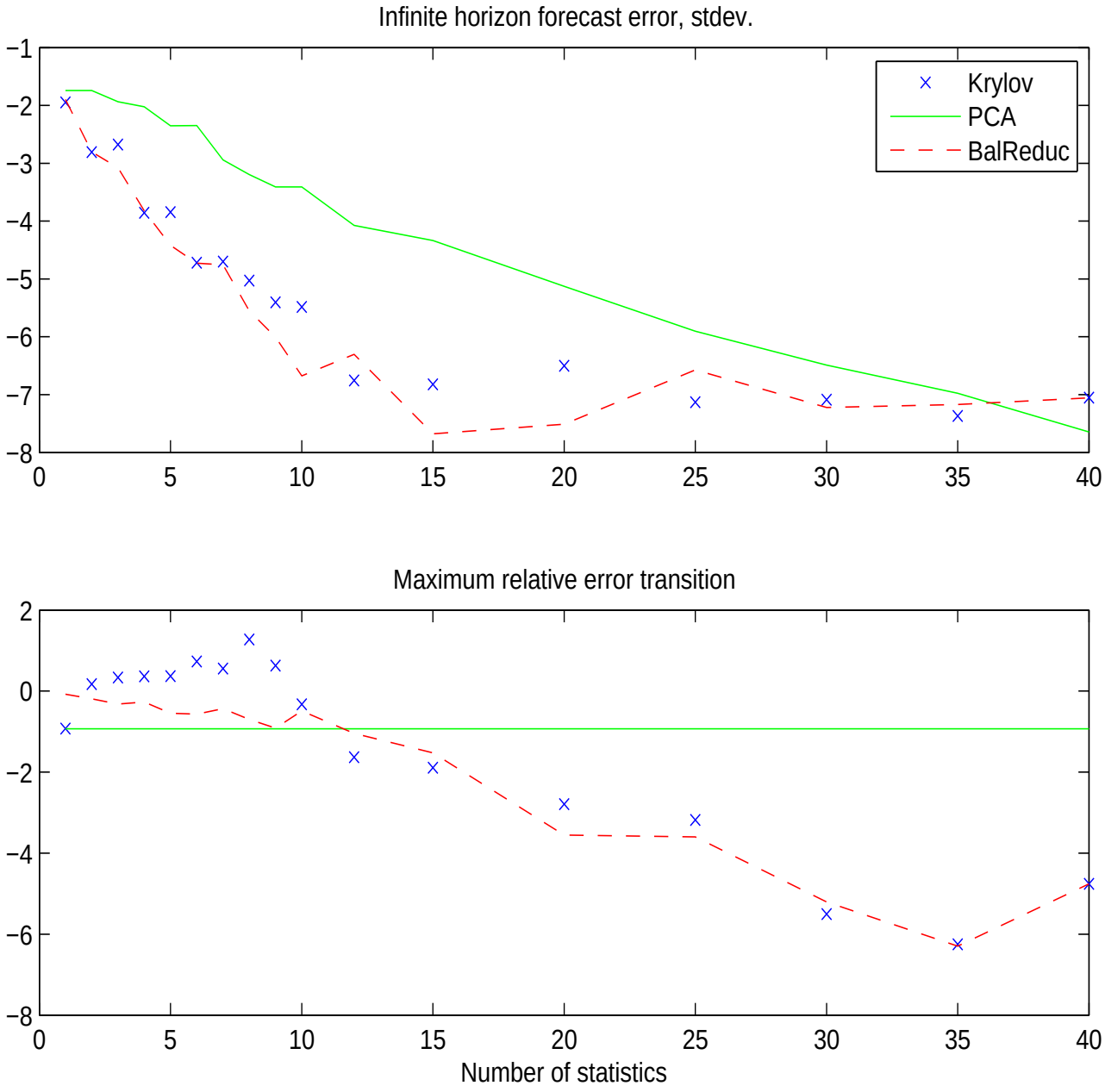

Figure 2: log10 of approximation errors, state-dependent pricing model

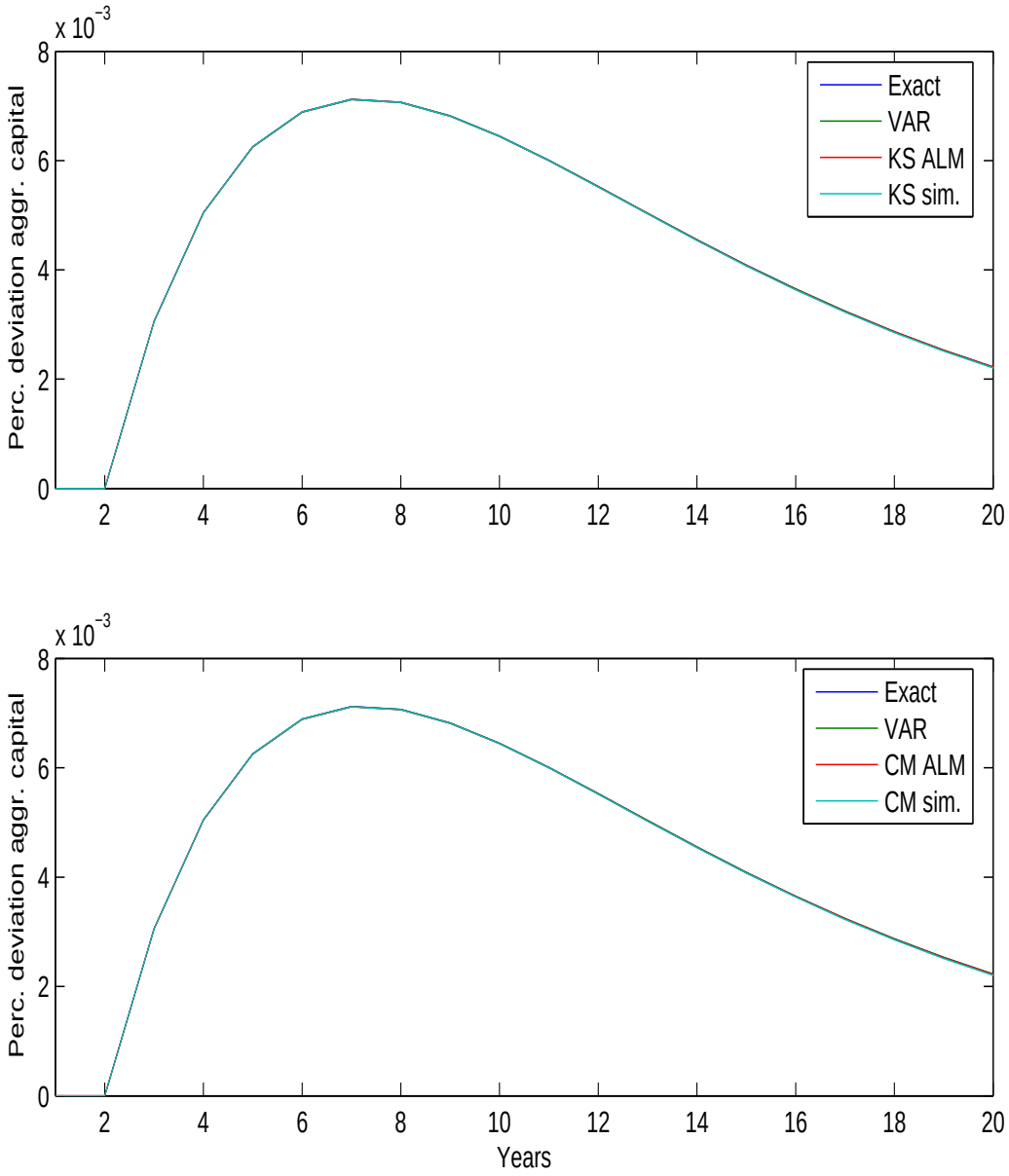

Figure 3: Response to technology shock, 1 moment

3

Figure 3:

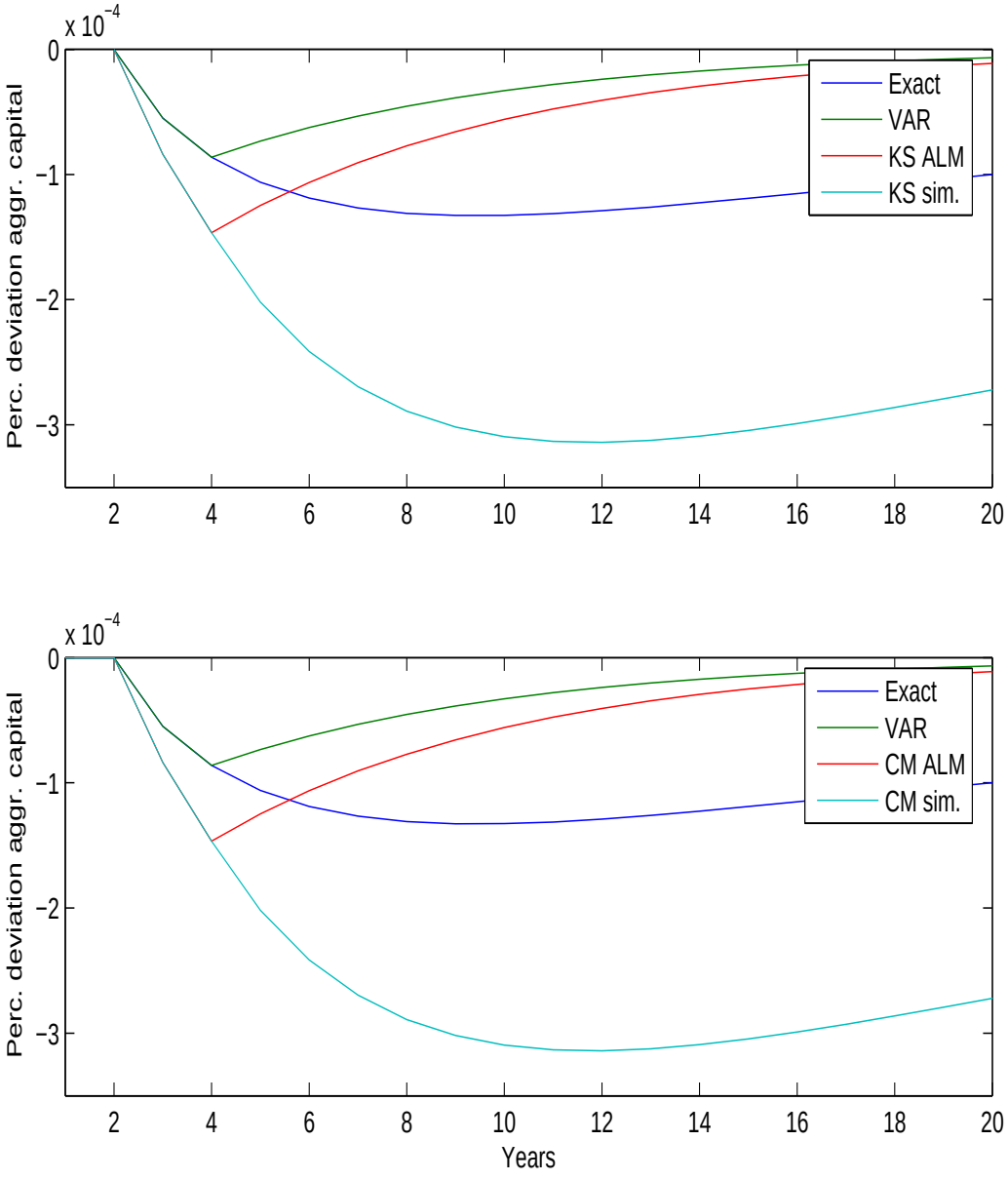

Figure 4: Response to tax shock, 1 moment

4

Figure 4:

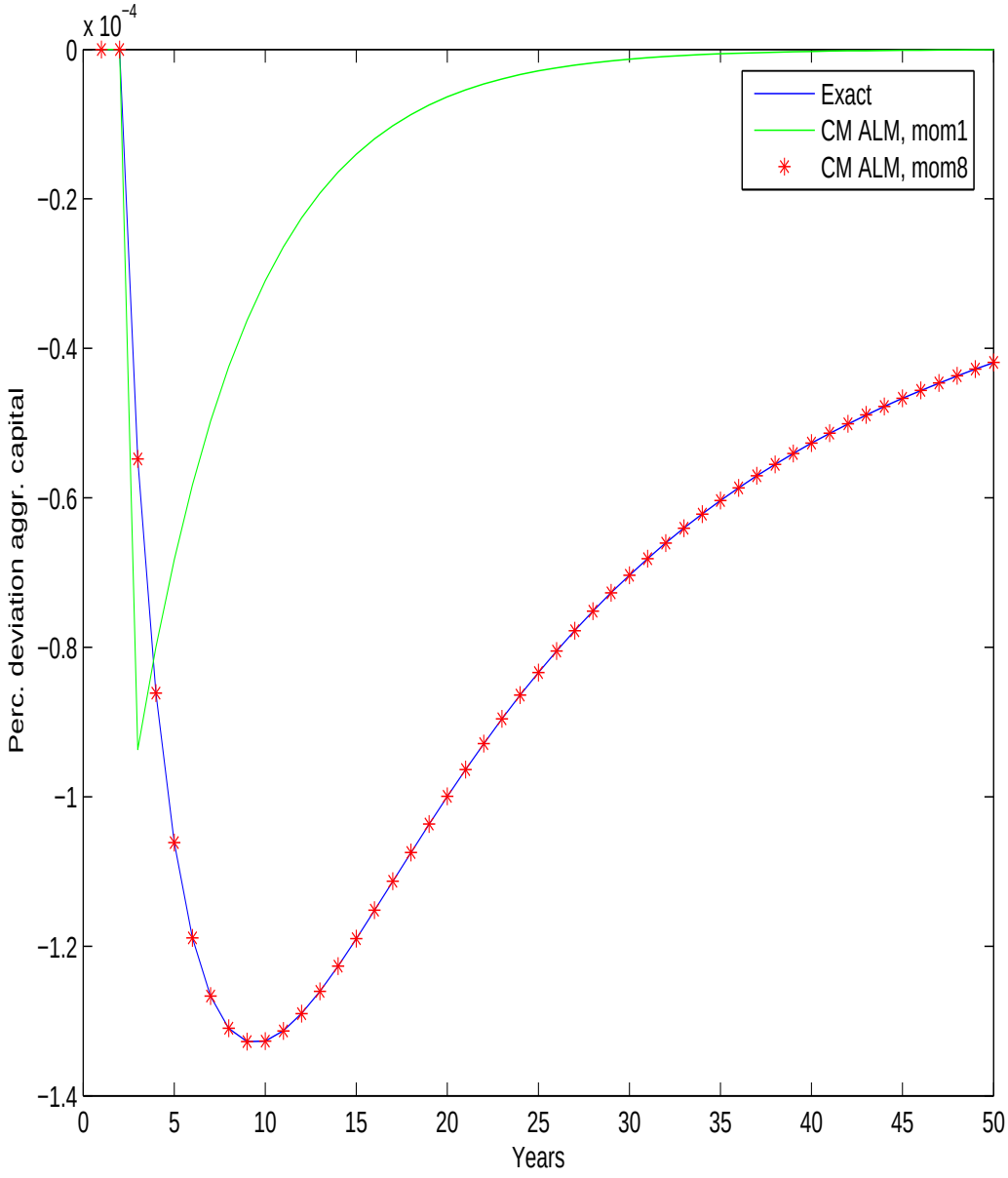

Figure 7: Response to tax shock

Figure 5:

References

Carroll, C. D. (2001). A theory of the consumption function, with and without liquidity constraints. *Journal of Economic Perspectives 15*(3), 23–45.

Castaneda, A., J. Diaz-Gimenez, and J.-V. Rios-Rull (2003). Accounting for earnings and wealth inequality. *Journal of Political Economy 111*(4), 818–857.

Den Haan, W. J. (1997). Solving dynamic models with aggregate shocks and heterogenous agents. *Macroeconomic Dynamics 1*, 355–86.

Den Haan, W. J. (2007). Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents. Manuscript.

Den Haan, W. J. (2008). Comparison of solutions to the incomplete markets model with aggregate uncertainty. forthcoming.

Krusell, P. and A. Smith (2006). Quantitative macroeconomic models with heterogeneous agents. In R. Blundell, W. K. Newey, and T. Persson (Eds.), *Advances in economics and econometrics: theory and applications, Ninth World Congress*. New York, NY: Cambridge University Press.

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

Reiter, M. (2002). Recursive computation of heterogeneous agent models. Manuscript.

Reiter, M. (2008a). Solving heterogenous agent models by projection and perturbation. forthcoming "Journal of Economic Dynamics and Control".

Reiter, M. (2008b). Solving the incomplete markets model with aggregate uncertainty by backward induction. prepared for special issue of JEDC.

Reiter, M., T. Sveen, and L. Weinke (2009). Lumpy investment and state-dependent pricing in general equilibrium. Manuscript.

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