

Approximate and Almost-Exact Aggregation in Dynamic Stochastic Heterogeneous-Agent Models*

Michael Reiter, Institute for Advanced Studies, Vienna

December 2009

Abstract

The paper presents a new method to solve DSGE models with a great number of heterogeneous agents. Using tools from systems and control theory, it is shown how to reduce the dimension of the state and the policy vector so that the reduced model approximates the original model with high precision. The method is illustrated with a stochastic growth model with incomplete markets similar to Krusell and Smith (1998), and with a model of heterogeneous firms with state-dependent pricing. For versions of those models that are nonlinear in individual variables, but linearized in aggregate variables, approximations with 50 to 200 state variables deliver solutions that are precise up to machine precision. The paper also shows how to reduce the state vector even further, with a very small reduction in precision.

JEL classification: C63, C68, E21

Keywords: heterogeneous agents; aggregation; model reduction

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 Istanbul, Barcelona, 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 large number (or even a continuum) of heterogeneous agents are now widely used in economics. Except for some special 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 is a high-dimensional or even infinite-dimensional object.

Krusell and Smith (1998) and Den Haan (1996) have shown, in some versions of the stochastic growth model with heterogeneous agents and incomplete markets, that a very good approximate solution can be obtained through 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 the available accuracy checks indicate that the resulting approximation is very good (a recent comparison of several solution approaches is cf. Den Haan (2008)). Krusell and Smith (2006) explain and extend this approximate aggregation result.

However, we cannot expect such a favorable result to go through in all relevant heterogeneous agent models. As an example, we will see below that even in the Krusell/Smith model we need approximate solutions with more state variables if we want to study not just technology shocks, but also tax shocks that directly affect the wealth distribution. Reiter (2009b) has developed a method to handle those cases, using a high-dimensional, non-parametric approximation to the cross-sectional distribution.¹ 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 normal PC.

While Krusell and Smith (1998) and most of the subsequent literature use an extreme form of aggregation (only one state variable) to obtain a reasonably precise solution,² Reiter (2009b) goes the opposite way and avoids aggregation by using a high-dimensional state vector. The latter approach has two important limitations. First, a state space of around 1000 state variables appears sufficient to give a precise discrete approximation of a univariate cross-sectional distribution, which is the case of one individual state variable. However, the constraint on the number of states becomes restrictive if each individual has two or more state variables. Second, linearization in aggregate variables is sufficient for

¹Midrigan (2009) shows in a model of state dependent pricing that higher-order moments matter for aggregate dynamics. He solves the model using only one state variable, because those higher-order moments do not vary much. Nevertheless, the precision of the forecasting rule is considerably smaller than in the Krusell/Smith model. To avoid those problems, Costain and Nákov (2008) apply the method of Reiter (2009b) to a similar model.

²Recent innovative approaches such as Preston and Roca (2007) and Den Haan and Rendahl (2009) can handle a somewhat larger number of variables in the state vector.

many interesting applications, but it is inadequate for others, for example models with asset choice. The question then is whether it is possible to construct aggregate models which give a similarly precise solution, but use a much lower number of state variables, so that higher-order approximations in the aggregate variables can be obtained. Furthermore, for a precise analysis of the properties of the model it is very useful to have a lower-dimensional representation of the solution.

In the present paper I address those issues, by taking a closer look at the aggregation problem in typical dynamic macroeconomic heterogeneous agent models. It turns out that these models do not allow exact aggregation in the theoretical sense, but a substantial reduction in the number of state variables is possible in a way that the approximate solution is still precise basically up to machine precision. Building on results from the systems and control literature on model reduction, I provide a general algorithm to construct such an aggregate model and choose the appropriate state space. The algorithm is independent of the underlying structure of the original state space (whether we have a univariate or a bivariate distribution, etc.). Moreover, the algorithm does not depend on whether the cross-sectional distribution is smooth, unlike methods that use smooth parameterizations of the distribution (for example Den Haan (1996)). The distribution of the model in Appendix A, for example, is quite irregular.

To give an idea of the magnitudes involved, we are going to study models with between 700 and more than 30000 state variables, while the reduced models have in the range of 50 and 200 state variables and are extremely accurate. Giving up some precision, the same method can be used to select an even lower number of state variables. Such an approximation can serve as the basis for higher-order solutions, an issue that is explored in Reiter (2009a). Numerical results are provided for several models. The main application is a stochastic growth model similar to Krusell and Smith (1998), but allowing for persistent individual productivity shocks.³ A second application is a standard OLG model with many cohorts. This gives an interesting comparison, because aggregation turns out to be much harder in OLG models. A third application, provided in the appendix, is a model of state-dependent pricing (SDP) with firm-specific productivity shocks. Beyond the examples provided here, the same method has been applied successfully in Reiter, Sveen, and Weinke (2009) and Haefke and Reiter (2009).

It should be stressed that the goal of the proposed method is to compute the best possible approximation to the rational-expectations equilibrium with full information. While the Krusell/Smith approximation has the flavor of a “bounded-rationality” solution of the model, the outcome of the method presented here cannot be interpreted in this way.

The plan of the paper is as follows. Section 2 describes the heterogeneous-agents stochastic growth model that will be the main example to test the algorithm. Subsection 2.6 gives the OLG version of the model. Section 3 presents the theory of aggregation and the numerical algorithms. Section 4 provides numerical results for the examples economies. Section 5 concludes. Appendix A describes a model of heterogeneous firms with state-

³The Matlab programs to solve this model are available at <http://elaine.ihs.ac.at/~mreiter/appraggr.tar.gz>.

dependent pricing as another test case. Appendices B and C present technical details.

2 Example 1: The Stochastic Growth Model With Heterogeneous Agents and Incomplete Markets

The first example is a variation on the well-known model of Krusell and Smith (1998). The model of this section is very similar to Reiter (2009b); the main difference is that here we allow for permanent changes in household productivity. Furthermore, the discretization procedure (Section 2.4) is simpler than in Reiter (2009b).

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

$$Y_t = \mathcal{Y}(K_{t-1}, L_t, Z_t) = AZ_t K_{t-1}^\alpha L_t^{1-\alpha}, \quad 0 < \alpha < 1 \quad (1)$$

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

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

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

$$\log Z_{t+1} = \rho_Z \log Z_t + \epsilon_{Z,t+1} \quad (3)$$

where ϵ_Z 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:

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

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

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

$$R_t = \bar{R}_t - \tau_t^k \quad (6)$$

$$W_t = \bar{W}_t(1 - \tau_t^l) \quad (7)$$

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

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

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

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

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 $\xi_{t,h}$. Labor productivity of household h is the product of a permanent component $\theta_{t,h}$, and an i.i.d. component $\xi_{t,h}$. Both are normalized to have unit mean:

$$E \theta_{t,h} = 1 \quad (10a)$$

$$E \xi_{t,h} = E_{t-1} \xi_{t,h} = 1 \quad (10b)$$

Net labor earnings are therefore given by

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

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

$$x_{t,h} = R_t k_{t-1,h} + y_{t,h} \quad (12)$$

$$(13)$$

Cash on hand is split between consumption and asset holdings:

$$k_{t,h} = x_{t,h} - c_{t,h} \quad (14)$$

We impose the borrowing constraint

$$k_{t,h} \geq \underline{k} = 0 \quad (15)$$

The consumption decision of the household depends on aggregate variables, which are reflected in a time subscript and will be specified in detail later, and the individual state variables x and θ . The solution of the household problem is then given by a consumption function $\mathcal{C}_t(x, \theta)$. The first order condition of the household problem is the Euler equation

$$U'(\mathcal{C}_t(x_t, \theta_t)) \geq \beta E_t [R_{t+1} U'(\mathcal{C}_{t+1}(R_{t+1}(x_t - \mathcal{C}_t(x_t, \theta_t)) + y_{t+1}, \theta_{t+1}))] \\ \text{and } \mathcal{C}_t(x_t, \theta_t) = x_t - \underline{k} \quad (16a)$$

or

$$U'(\mathcal{C}_t(x_t, \theta_t)) = \beta E_t [R_{t+1} U'(\mathcal{C}_{t+1}(R_{t+1}(x_t - \mathcal{C}_t(x_t, \theta_t)) + y_{t+1}, \theta_{t+1}))] \\ \text{and } \mathcal{C}_t(x_t, \theta_t) < x_t - \underline{k} \quad (16b)$$

The expectation in (16) is over the distribution of θ_{t+1} , ξ_{t+1} and the future aggregate state. Since the household problem is concave, Equ. (16) together with the constraint (15) and a transversality condition (which is guaranteed to hold in a bounded approximation) are both necessary and sufficient for a solution of the household problem.

2.4 Finite Approximation of the Model Equations

2.4.1 Household Productivity, Consumption Function and the Euler Equation

Permanent productivity $\theta_{t,h}$ follows an n_P -state Markov chain and takes on the values $\bar{\theta}_1, \dots, \bar{\theta}_{n_P}$. Transitory productivity $\xi_{t,h}$ has a discrete distribution, taking on the values $\bar{\xi}_l$ with probabilities ω_l^ξ for $l = 1, \dots, n_y$.⁴

At each point in time, and for any value of productivity $\bar{\theta}_j$, the savings behavior is characterized by a critical level χ_{jt} where the borrowing constraint starts binding, and a smooth function for $x > \chi_{jt}$. I approximate the consumption function by a piecewise linear interpolation between the knot points $\bar{x}_{j,i,t}$. Knots are chosen as $\bar{x}_{j,i,t} = \chi_{jt} + \bar{x}_i$, $i = 0, 1, \dots, n_s$ with $0 = \bar{x}_0 < \bar{x}_1 < \dots < \bar{x}_{n_s}$. The consumption function $\hat{\mathcal{C}}_j(x; \mathbf{S}_t)$ can then be represented by $(n_s + 1)n_P$ numbers, giving for each j the critical level χ_{jt} and the level of consumption at $\bar{x}_{j,1,t}, \bar{x}_{j,2,t}, \dots, \bar{x}_{j,n_s,t}$. Notice that consumption at $\bar{x}_{j,0,t}$ equals $\bar{x}_{j,0,t}$ by construction. These numbers are collected into the vector \mathbf{S}_t . The approximated consumption function is then written as $\hat{\mathcal{C}}_j(x; \mathbf{S}_t)$. Again, the time subscript reflects the dependence on aggregate state variables.

For each $\bar{\theta}_j$, the approximation of the saving function has $n_s + 1$ degrees of freedom. We therefore apply a collocation method and require the Euler equation (16) to hold with

⁴In Reiter (2009b), 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 setup.

equality at the knot points $\bar{x}_{j,i,t}$:

$$U' \left(\hat{\mathcal{C}}_j(\bar{x}_{j,i,t-1}; \mathbf{S}_{t-1}) \right) = \beta \sum_{j'=1}^{n_P} \sum_{l=1}^{n_y} \omega_{j,j'}^\theta \omega_l^\xi \left[(\bar{R}_t - \tau_t^k) U' \left(\hat{\mathcal{C}}_{j'}(\bar{x}^{ijlj'}; \mathbf{S}_t) \right) \right] + \eta_{ijt}^c, \\ j = 1, \dots, n_P, \quad i = 0, \dots, n_s \quad (17a)$$

where

$$\bar{x}^{ijlj'} \equiv (\bar{R}_t - \tau_t^k) \left(\bar{x}_{j,i,t-1} - \hat{\mathcal{C}}_j(\bar{x}_{j,i,t-1}; \mathbf{S}_{t-1}) \right) + (\bar{W}_t(1 - \tau_t^l)) \bar{\theta}_{j'} \bar{\xi}_l \quad (17b)$$

Equ. (17a) uses the notation of Sims (2001): the η_{ijt}^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 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 $\underline{k} = \bar{k}_1, \bar{k}_2, \dots, \bar{k}_{n_k} = k^{max}$. 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

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

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

$$K_t = \sum_{j=1}^{n_P} \sum_{i=1}^{n_k} \bar{k}_i \phi_t(j, i) \quad (19)$$

Further define $\phi_t(j)$ as the column vector $(\phi_t(j, 1), \dots, \phi_t(j, n_k))'$, and stack all the $\phi_t(j)$'s

into the column vector

$$\mathbf{\Phi}_t \equiv \begin{bmatrix} \phi_t(1) \\ \vdots \\ \phi_t(n_P) \end{bmatrix} \quad (20)$$

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 $\bar{\theta}_j$, the transition from the distribution $\tilde{\phi}_t(j)$ at the beginning of period t (after the shock to productivity has realized) to the end-of-period distribution $\phi_t(j)$ is given by

$$\phi_t(j) = \Pi_{t,j}^K \tilde{\phi}_t(j) \quad (21)$$

where the elements of the transition matrix $\Pi_{t,j}^K$ are given by

$$\Pi_{t,j}^K(i', i) = \sum_{l=1}^{n_y} \psi(i', R_t \bar{k}_i + W_t \theta_{t,l} \xi_{t,l} - \hat{C}_j(R_t \bar{k}_i + W_t \theta_{t,l} \xi_{t,l}; \mathbf{S}_{t-1})) \quad (22)$$

From the properties of $\psi(i, k)$, each column of $\Pi_{t,l}^K$ has at most $2n_y$ non-zero elements. We can now write the transition from the end-of-period distribution $\mathbf{\Phi}_{t-1}$ to $\mathbf{\Phi}_t$ as the linear dynamic equation

$$\mathbf{\Phi}_t = \Pi_t \mathbf{\Phi}_{t-1} \quad (23a)$$

where

$$\Pi_t = \begin{bmatrix} \Pi_{t,1}^K & 0 & \dots & 0 & 0 \\ 0 & \Pi_{t,2}^K & \dots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \dots & \Pi_{t,n_P-1}^K & 0 \\ 0 & 0 & \dots & 0 & \Pi_{t,n_P}^K \end{bmatrix} (\Pi^P \otimes I_{n_k}) \quad (23b)$$

and Π^P is the Markov transition matrix between the permanent productivity states $\bar{\theta}_j$.

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 numbers⁵ 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 (23). These equations define, for each period t , a system of $n_P(n_s + n_k + 1) + 2$ equations in just as many variables: \mathbf{S}_t , $\mathbf{\Phi}_t$, Z_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).

⁵A law of large numbers for economies with a continuum of agents, using standard analysis and measure theory, is given in Podczeck (2009).

Notice that n_P of those variables are redundant, since $\sum_i \phi_t(l, i) = 1$ for all l . Correspondingly, n_P equations in (23) are linearly dependent since all rows in the $\Pi_{t,j}^K$ add up to one.

2.5 Parameter Values and Functional Forms

I will show numerical results for two different calibrations of the model. In the small calibration, variations in individual productivity are i.i.d., which means that $n_P = 1$. Using $n_k = 1000$ grid points for the distribution of capital and a spline of order $n_s = 100$ for the consumption function, the discrete model has around 1100 variables and can be solved exactly. This has the advantage that we can measure the accuracy of the aggregate models by comparison to the exact solution of the model. The big calibration allows for persistent variations in individual productivity, using $n_P = 31$.

Common to both calibrations are the following parameter values. The frequency of the model is quarterly. Standard values are used for most of the the model parameters: $\beta = 0.99$, $\alpha = 1/3$, $\delta = 0.025$. For the utility function I use CRRA

$$U(c) = \frac{c^{1-\gamma} - 1}{1 - \gamma} \quad (24)$$

with risk aversion parameter $\gamma = 1$ (the accompanying Matlab programs can be used to explore other degrees of risk aversion). For the technology shock I choose $\rho_Z = 0.95$ 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 $\sigma_\tau = 0.01$. In some robustness checks, both the model frequency and σ_τ are varied.

2.5.1 The calibration with i.i.d. shocks

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

2.5.2 The calibration with persistent shocks

The second calibration allows for persistent variations in individual productivity, using $n_P = 31$. From an economic point of view, this calibration is interesting because it leads to a rather realistic distribution of income, consumption and wealth (cf. Table 2.5.2).

This calibration matches three targets. The earnings share of the upper quintile of the distribution is 0.6139. The earnings share of the top 1 percent of the distribution is 0.1476. The conditional annualized variance of persistent income equals 0.15. I assume that from each grid point $\bar{\theta}_j$, the household can only switch to neighboring points $\bar{\theta}_{j-1}, \bar{\theta}_{j+1}$. The probability with which this happens helps to meet the third target. The probability of

The Distributions of Earnings

	Gini	Quintiles					Top Groups (%)		
		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
Model	0.61	1.73	3.46	9.79	23.64	61.39	13.17	15.43	14.76

The Distributions of Wealth (%)

	Gini	Quintiles					Top Groups (%)		
		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
Model	0.74	1.16	2.92	5.25	10.50	80.18	20.26	30.22	13.41

Source for US: Castaneda, Diaz-Gimenez, and Rios-Rull (2003);

Model calibration of Section 2.5.2.

Table 1: Earnings and wealth distribution in the growth model

reaching the highest state is adjusted so as to meet the second target. To match the first target, the income value of the uppermost 20 percent of grid points is modified upwards from a distribution that is equidistant in logs. To avoid that the distribution drifts too much to the boundaries of the state space, the household faces a constant probability of “dying”, that means, being reset to the median value of θ . The inverse of this probability is 50 times the model frequency. Transitory productivity $\xi_{t,i}$ has again two realizations, namely 0.2 and 1.8. This unrealistically high variance of the transitory component helps explaining the lower quintiles of the distribution.

The exact specification of the income process can be seen from the program “caliinc.m”.

2.6 Example 2: An OLG Economy

As a second example, we consider an economy where households have a finite life span, but there is no heterogeneity within cohorts. A household born in period t lives from t to $t + T - 1$. Consumption and labor supply of a household with age i in period t are denoted by $c_{t,i}$ and $l_{t,i}$, respectively. The household supplies one unit of labor in each period before retirement age T^R , therefore $l_{t,i} = 1$ for $i = 0, \dots, T^R - 1$ and $l_{t,i} = 0$ for $i = T^R, \dots, T - 1$. Labor productivity is normalized to $1/T^R$, so that total effective labor supply equals 1.

The household born at t maximizes

$$\mathbb{E}_t \sum_{i=0}^{T-1} \beta^i u(c_{t+i,i}) \quad (25)$$

Asset holdings follow

$$k_{t,i} = R_t k_{t-1,i-1} + \frac{W_t}{T^R} l_{t,i} - c_{t,i} \quad (26)$$

The household starts with no assets, and cannot leave debt at the time of dying. The

budget constraint of a household born at t is therefore

$$k_{t,-1} = 0 \tag{27}$$

$$k_{t+T-1,T-1} \geq 0 \tag{28}$$

Again, there is only one asset (physical capital). The household can save and lend freely in this asset. There is no borrowing or short-sale constraint.

Technological and utility parameters are chosen as in the infinite-life model. To have a model with a similar dimension as before, I choose a monthly frequency and assume 60 years of economic lifetime, so $T = 720$ and $T^R = 480$. For the OLG model, I only consider technology shocks, specified as in (3). There are no tax shocks ($\tau_t^k = \tau_t^l = 0$).

3 Exact and Approximate Aggregation in Linear Models

3.1 Overview of the Solution Method

The method I propose in this paper to compute a numerical solution of heterogeneous agent models (such as the one we have seen in Section 2) involves the following steps:

1. Discretizing the model, as explained in Section 2.4. This usually involves:
 - In a model with a continuum of household, a finite representation of the continuous cross-sectional distribution function of the individual state variables such as household assets, cf. Section 2.4.2.
 - A finite parameterization of the continuous decision functions, for example by spline functions of polynomials.
2. Computing the deterministic steady state of the discrete model. This usually involves:
 - In an outer loop, iteration over steady state prices.
 - In an inner loop, computing the optimal decisions of agents conditional on prices, and then finding the ergodic cross-sectional distribution of the individual state variables.
3. Linearizing the model equations around the steady state. This is done by numeric differentiation, either automatic differentiation or some form of finite differences.
4. If the total number of variables in the linearized discrete model is not too big (1000-2000 on a normal PC), one can compute the solution of the linearized model by any package for linear rational equation systems, for example Sims (2001).
5. If the total number of variables is too big, or if a lower-dimensional approximation is desired, aggregate the model by state and policy reduction, and solve the reduced model. We can distinguish two approaches:

- (a) Identify the minimal state vector that still allows the model to be solved with very high precision, in the range that the floating-point precision of the computer allows. This will be explained in Section 3.4.
- (b) Sacrificing some accuracy, reduce the dimension even further to obtain a small model with reasonable accuracy (in the spirit of Krusell and Smith (1998), Den Haan (2007) and the subsequent literature in macroeconomics). The task then is to find the right set of state variables, in order to obtain (close to) maximal accuracy for a given size of the aggregate model. This will be explained in Section 3.5.

Steps 1–4 have already been described in Reiter (2009b). The topic of the present paper is aggregation (Step 5). As was explained in the introduction, this allows to accurately handle models with a much higher number of variables, and serves as an important input for higher-order solutions.

The linearization in Step 3 gives a solution that is linear in aggregate variables but nonlinear in individual variables. To see what this means, take the growth model as an example. The consumption function, which gives consumption as a nonlinear function of individual assets, is model led as depending linearly on the aggregate state vector, the cross-sectional distribution of capital. If the consumption function is represented by 100 spline coefficients, and the distribution by 1000 histogram values, this linear dependence is described by a 100×1000 -matrix.

I would like to stress that in this paper I do not ask whether the discrete model with a high- but finite-dimensional state vector (obtained in Step 1) is a good approximation to the theoretical model with a continuum of agents. This is an important question, which I leave for future work. I rather focus on the question whether the high-dimensional discrete model can be approximated well by a lower-dimensional model. In the following, I will therefore consider the solution to the discrete model as the “exact solution”.

3.2 The Reduced State Space Model

In this section we review some results of the control literature on state aggregation in linear state space models. From Section 3.3 onwards, we will see how to adapt those techniques to the models that are of interest to economists, namely DSGE models.

In the literature on system and control theory, aggregation is called “model reduction”. We deal with the linear dynamic system

$$x_t = Ax_{t-1} + B\epsilon_t \tag{29a}$$

$$y_t = Cx_t \tag{29b}$$

where A , B and C are given matrices. The vector x contains the state variables, ϵ is the vector of exogenous shocks (“inputs”), which satisfy $E_{t-1} \epsilon_t = 0$ and $E[\epsilon_t \epsilon_t'] = \Sigma_\epsilon$. The vector y is the “output” of the system, the variables that we are ultimately interested in,

and which we observe. The goal then is to approximate (29) by a lower-dimensional model:

$$\hat{x}_t = \hat{A}\hat{x}_{t-1} + \hat{B}\epsilon_t \quad (30a)$$

$$\hat{y}_t = \hat{C}\hat{x}_t \quad (30b)$$

(29) and (30) are driven by the same shock vector ϵ_t , and should generate the same output variables. Therefore, y and \hat{y} have the same dimension. The idea is to choose the matrices \hat{A} , \hat{B} and \hat{C} so that (30) is a good approximation to (29) in the sense that the response of \hat{y} to past ϵ in the reduced model is similar to the response of y to past ϵ in the original model. State aggregation means that the size of the state vector \hat{x} , $\dim(\hat{x})$, is considerable smaller than the size of x , $\dim(x)$.

A natural way to derive the reduced model (30) is by assuming that the reduced state \hat{x} is linearly related to the original state x by

$$\hat{x}_t = Hx_t \quad (31)$$

with appropriately chosen matrix H . We can assume, w.l.o.g., that H is orthogonal:

$$HH' = I \quad (32)$$

Notice that H has more columns than rows, so that $H'H$ is not the identity matrix.

Given H , it is natural to choose \hat{A} , \hat{B} and \hat{C} by least squares. OLS estimates of (30) in the model (29) are given by (cf. Appendix B.1)

$$\hat{A} = H A \Sigma_x H' (H \Sigma_x H')^{-1} \quad (33a)$$

$$\hat{B} = H B \quad (33b)$$

$$\hat{C} = C \Sigma_x H' (H \Sigma_x H')^{-1} \quad (33c)$$

where Σ_x denotes the unconditional covariance matrix of x . It is given by $\mathcal{L}(A, B, \Sigma_\epsilon)$, which is defined as the unique symmetric solution to the discrete Lyapunov equation

$$\mathcal{L}(A, B, \Sigma_\epsilon) = A \mathcal{L}(A, B, \Sigma_\epsilon) A' + B \Sigma_\epsilon B' \quad (34)$$

for any matrices A , B and Σ_ϵ of appropriate size such that Σ_ϵ is non-negative definite and A is asymptotically stable, which means that all eigenvalues are smaller than unity in absolute value. For the special case of $\Sigma_\epsilon = I$, we will also use the short notation $\mathcal{L}(A, B)$.

Appendix B.2 shows that, if A is asymptotically stable, then is \hat{A} . The practical computation of Σ_x , even in cases where the vector x is very large, is dealt with in Appendix B.3. When computing the OLS estimates (33), one has to deal with the fact that the matrix $H \Sigma_x H'$ may be ill-conditioned if H has many rows. This can be handled by Tikhonov-regularization,⁶ cf. Appendix B.4

What remains to be done is to find a suitable way of choosing H . The standard approximation method in control theory is called “balanced reduction”. I will explain it as the combination of two ideas: principal component analysis (PCA), and the “conditional expectations approach” (CEA). The latter idea will be of independent interest for us when we consider aggregation in DSGE models.

⁶I am grateful to Ken Judd for pointing this out to me.

3.2.1 Principal Component Analysis (PCA)

The idea of PCA is to find out in which direction the state vector x varies most strongly. For that, we consider the eigenvalue decomposition of the unconditional covariance matrix of x , Σ_x :

$$\Sigma_x = USU' \quad (35)$$

Here, S is a diagonal matrix with the eigenvalues $\sigma_1, \dots, \sigma_n$, ordered such that $\sigma_i \geq \sigma_{i+1}$. U is the corresponding matrix of eigenvectors of Σ_x . (35) decomposes the variance of x into orthogonal components of decreasing importance (principal components). If the sequence $\sigma_i \geq \sigma_{i+1}$ decays rapidly enough, one can ignore the components $i > k$ for a suitably chosen $k < \dim(x)$. In particular, if σ_i is too small compared to σ_1 , it is numerically indistinguishable from zero. In that case, one can say that the state x lives, up to a very good approximation, in the space spanned by the first k columns of U , which we denote by $U_{:,1:k}$. Then we choose the reduced state \hat{x} as the coefficients of x with respect to this basis: $x_t = U_{:,1:k} \hat{x}_t$. Since the columns of U are orthogonal, we can set $H = U'_{:,1:k}$ and get $Hx_t = U'_{:,1:k} U_{:,1:k} \hat{x}_t = \hat{x}_t$, as in (31).

Having chosen H in this way, it can be easily shown that the OLS estimates (33a) reduce to $\hat{A} = HAH'$ and $\hat{C} = CH'$.

3.2.2 Conditional Expectations Approach (CEA)

If we are only interested in the variables in y , why is it important to know the full state vector x_t ? Because x_t helps to predict future values y_{t+i} . We therefore want the reduced model (30) to give us the correct conditional expectations⁷ of future y 's. Then it is natural to include those conditional expectations in the reduced state vector \hat{x} . Since

$$E_t[y_{t+i}] = E_t[Cx_{t+i}] = CA^i x_t \quad i = 1, 2, \dots \quad (36)$$

this means to set $\hat{x}_{t,i} = (CA^i)x_t$ for $i = 0, \dots, N-1$ for some sufficiently large N . Then the reduced model makes the best possible predictions for y_{t+i} up to $N-1$ periods ahead. Moreover $\hat{x}_{t,i}$ predicts not just y_{t+i} but also $\hat{x}_{t+i-l,l}$ for $l < i$:

$$\hat{x}_{t,i} = E_t[\hat{x}_{t+1,i-1}] = \dots = E_t[\hat{x}_{t+i-1,1}] = E_t[\hat{x}_{t+i,0}] = E_t[Cx_{t+i}], \quad i < N \quad (37)$$

We might therefore set H equal to

$$Q(N) \equiv \begin{bmatrix} C \\ CA \\ CA^2 \\ \dots \\ CA^{N-1} \end{bmatrix} \quad (38)$$

⁷Notice that in the context of linearized solutions, we do not worry about predictions of higher moments of y_{t+i} , because the solution is of the certainty-equivalence type in aggregate variables.

But then there is an obvious problem: to have high precision, we need predictions for many periods ahead, and so it seems we need a large vector \hat{x} , which contradicts the idea of state aggregation. To get state aggregation, we have to pick $k < n \equiv \dim(x)$ rows (or linear combinations of rows) from $Q(N)$. Mathematical theory suggests⁸ to do this via a singular value decomposition (SVD) of $Q(N)$:

$$Q(N) = USV' \quad (40)$$

where $U'U = I$, $V'V = I$, and S is again a diagonal matrix with entries $\sigma_1, \dots, \sigma_n$, ordered such that $\sigma_i \geq \sigma_{i+1}$. In analogy to the procedure of PCA, we set H equal to the first k rows of V' :

$$H = V'_{1:k,:} \quad (41)$$

We can see a strong formal analogy between CEA and PCA if we consider the limiting case $N \rightarrow \infty$:

$$\mathcal{Q} \equiv \sum_{i=0}^{\infty} \left(\frac{\partial E_t y_{t+i}}{\partial x_t} \right)' \left(\frac{\partial E_t y_{t+i}}{\partial x_t} \right) = \sum_{i=0}^{\infty} A^{i'} C' C A^i = \lim_{N \rightarrow \infty} Q(N)' Q(N) = \mathcal{L}(A', C') \quad (42)$$

The matrix \mathcal{Q} is called “observability Gramian”, and measures the sum of squares of the contribution of x to the future y ’s.

If it is more important to predict the y in the not-too-distant future, then it is better to get H from the SVD of $Q(N)$ for large N (below I will choose $N = 1000$, capturing forecasts of the next 250 years) rather than the SVD of \mathcal{Q} .

3.2.3 Balanced Reduction

PCA and CEA implement two different ideas on how to reduce the state vector. PCA asks: “what are the components of x that vary a lot over time”? If a component does not vary, there is no need to include it in the state vector. CEA asks: what are the components of x that help predicting the variables that really matter, namely the future y ? If a component of x varies a lot, but this variation is unrelated to changes in y , it is useless to include it in the state vector. The concept of “Balanced Reduction” combines the two approaches. Define the matrices R , Σ_x , Q , \mathcal{Q} , U , S , V , \tilde{H} as follows:

$$RR' = \Sigma_x \equiv \mathcal{L}(A, B, \Sigma_\epsilon) \quad (43)$$

$$QQ' = \mathcal{Q} \equiv \mathcal{L}(A', C') \quad (44)$$

$$USV' = R'Q \quad (45)$$

$$\tilde{H} = S^{-1/2} V' Q' \quad (46)$$

⁸Given any matrix X with rank n , the rank- k matrix that is closest to X , both in the L_2 and in the Frobenius norm, is given by

$$\tilde{X} = US_k V \quad (39)$$

where USV is the SVD of X , and S_k is obtained from S by zeroing all the singular values σ_i with $i > k$ (Trefethen and Bau III 1997, Theorems 5.8, 5.9).

Here, R is the Cholesky factor of the covariance matrix Σ_x , and Q is the Cholesky factor of the observability Gramian \mathcal{Q} , while U , S and V are the SVD of the matrix $R'Q$ with $U'U = I$, $V'V = I$ and S diagonal with decreasing entries. We take S as the square matrix containing only the non-zero singular values (and drop the columns of U and rows of V corresponding to the zero singular values), so that S is invertible by construction.

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

$$\mathcal{L}(\hat{A}, \hat{B}, \Sigma_\epsilon) = \tilde{H}\mathcal{L}(A, B, \Sigma_\epsilon)\tilde{H}' = \mathcal{L}(\hat{A}', \hat{C}') = (\tilde{H}')^{-1}\mathcal{L}(A', C')\tilde{H}^{-1} = S. \quad (47)$$

Equ. (47) is a remarkable result. It shows that in the new vector \hat{x} the variables are ordered such that \hat{x}_i has both the i -th highest variance, and makes the i -th highest contribution to future values of y . For the reduced model, we pick the first k components of \hat{x} , or the first k rows of \tilde{H} , such that the diagonal elements $S_{i,i}$ are negligible for $i > k$:

$$H = \tilde{H}_{1:k,:} \quad (48)$$

3.2.4 Properties of balanced reduction

Is the reduced model (30) an optimal approximation to (29) in any sense? With the choices of H that we have discussed, it is not a strictly optimal. Nevertheless, balanced reduction has a strong performance guarantee (cf. Antoulas (2005, Theorem 7.10), Antoulas (1999, Section 2.6)):

$$\text{distance}(\text{ExactModel}, \text{ReducedModel}) \leq 2(\sigma_{k+1} + \dots + \sigma_n) \quad (49)$$

Here, the σ 's are the singular values in (45) (called ‘‘Hankel singular values’’) that were omitted in the construction of H in (48). The distance measure in (49) is the Hankel norm, which is defined as the maximum distance in the future response

$$\sqrt{\sum_{i=0}^{\infty} \|y_{t+i} - \hat{y}_{t+i}\|^2} \quad (50)$$

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

$$\sqrt{\sum_{i=0}^{\infty} \|\epsilon_{t-i}\|^2} = 1 \quad (51)$$

In particular, the difference in the usual impulses responses between exact and the reduced models cannot be bigger than the bound (49). This explains why balanced reduction is the standard aggregation method in the control literature.

There exist even better, but more complicated approximations than balanced reduction. The theoretical lower bound on the distance between the two models is σ_{k+1} . This bound can actually be attained (Antoulas 1999, Sections 2.6.1, 3.2). For us, it seems not worthwhile to investigate more complicated methods, because our main problem is how to adapt these aggregation techniques to the type of models that we are interested in, namely linear DSGE models. This is the topic of Sections 3.3 and 3.4.

3.3 The Aggregation Problem in Linear RE Models

The model we want to analyze is given in the form of a linear rational expectations (RE) model, defined through a finite number of equations in a finite number of variables for each model period t . If the original model contains a continuum of state variables, it has first to be discretized. Examples are the discrete models of Section 2.4.3 and A.4. After linearizing this model around the steady state (cf. Section 3.1, Step 3), we assume that the model can be written in the form

$$x_t = Tx_{t-1} + Dd_t + F\epsilon_t \quad (52a)$$

$$d_{t-1} = E_{t-1} [\mathcal{E}_{x1}x_{t-1} + \mathcal{E}_{x0}x_t + \mathcal{E}_{d0}d_t] \quad (52b)$$

$$y_t = Cx_t \quad (52c)$$

with given matrices T , D , F , \mathcal{E}_{x1} , \mathcal{E}_{x0} , \mathcal{E}_{d0} and C . The vector d contains all the decision variables that are determined by the system of expectational equations (52b). In the stochastic growth models, these are the parameters of the consumption function \mathbf{S}_t , which are pinned down by the household Euler equations. The vector x contains all the other variables. For example, in the growth model, this includes the complete information about the cross-sectional distribution of capital, collected into the vector Φ_t , together with Z_t and τ_t^k . Again, y_t collects the aggregate variables of interest. Notice that it is not restrictive in (52c) to assume that y is a known function of x only, because we can always augment x by elements of y and add corresponding equations to (52a).

The solution to this model is a linear decision function

$$d_t = D_X x_{t-1} + D_E \epsilon_t \quad (53)$$

This generates dynamics for the state x of the form (29a), with A and B given by

$$A = T + DD_X \quad (54a)$$

$$B = F + DD_E \quad (54b)$$

We assume that (52) has a unique stable solution.

The aggregation techniques of Section 3.2 do not directly apply to models of the type (52), for the following reasons.

1. The dynamics of the model (52) is not known before we solve it, and in general we first have to deal with the aggregation problem before we can solve the model. In particular, we cannot apply the methods of Section 3.2 that make use of the covariance matrix Σ_x .
2. There is a feedback from the aggregation to the solution of the model. If we replace the state vector x by the reduced vector \hat{x} , we assume, explicitly or implicitly, that the agents in the model base their decision only on \hat{x} , not on x . The way we handle the aggregation problem will therefore have an effect on the solution (53), and therefore the dynamics of the model.

3. The model may not only have a high-dimensional state x , but also a high-dimensional decision vector d . So we have to think not just of state, but also of policy aggregation.
4. Since economic agents are assumed to base their decisions on \hat{x} , we will probably require that \hat{x} contains those aggregate variables that are important for the agents' decision and that we assume are actually observed. So we have to identify those variables and include them into \hat{x} .

To handle the last issue, we assume that the vector of aggregate variables y_t includes all the aggregate variables that economic agents base their decision on. We therefore require that y_t contains at least those variables that enter into equation (52b), and therefore *directly* influence decisions at time $t - 1$ or t . These dependencies are captured in the matrices \mathcal{E}_{x1} and \mathcal{E}_{x0} . While these matrices can be big, it is typically the case that they have very low rank. For example, in the growth model, the state vector x enters the Euler equations (16) only through three variables: the interest rate and the two exogenous states, productivity and capital tax rate. Therefore, the rank of $[\mathcal{E}_{x1} \ \mathcal{E}_{x0}]$ equals three. In the SDP model of Section A, this rank is four. That y contains all the decision-relevant variables then means

$$[\mathcal{E}_{x1} \ \mathcal{E}_{x0}] \in \text{span}(C) \quad (55)$$

where $\text{span}(X)$ denotes the space spanned by the rows of the matrix X . We further require that complete information about y_t is contained in \hat{x}_t . This means that

$$C \in \text{span}(H) \quad (56)$$

Combining (56) and (55) we get the following requirement on H :

$$[\mathcal{E}_{x1} \ \mathcal{E}_{x0}] \in \text{span}(H) \quad (57)$$

3.4 Exact and Almost-Exact Aggregation in Linear RE Models

3.4.1 Exact State Aggregation

Let us first consider the case that exact aggregation of the linear RE model (52) is possible. Denote the dimension of x by n . We want to replace x by a state vector \hat{x} of dimension $k < n$. The key condition for exact aggregation is that there exists a $k \times n$ -matrix H and a $k \times k$ -matrix \hat{T} such that

$$HT = \hat{T}H \quad (58)$$

and H satisfies the spanning requirement (57). We can again normalize H to be orthogonal as in (32). Then we get

$$\hat{T} = HTH' \quad (59)$$

which is similar to a coordinate transformation. Furthermore, (57) and (32) imply that there exist unique matrices $\mathcal{E}_{\hat{x}0}$ and $\mathcal{E}_{\hat{x}1}$ such that

$$\mathcal{E}_{x0} = \mathcal{E}_{\hat{x}0}H \quad (60a)$$

$$\mathcal{E}_{x1} = \mathcal{E}_{\hat{x}1}H \quad (60b)$$

Equ. (56) implies that there is a unique matrix \hat{C} such that

$$C = \hat{C}H \quad (61)$$

Premultiplying (52a) by H and using (31), (58), (60) and (61) we obtain from (52) the reduced model

$$\hat{x}_t = \hat{T}\hat{x}_{t-1} + HDd_t + HF\epsilon_t \quad (62a)$$

$$d_{t-1} = E_{t-1} [\mathcal{E}_{\hat{x}1}\hat{x}_{t-1} + \mathcal{E}_{\hat{x}0}\hat{x}_t + \mathcal{E}_{d0}d_t] \quad (62b)$$

$$y_t = \hat{C}\hat{x}_t \quad (62c)$$

In which sense is (62) equivalent to the original model (52)? Using (58), we have derived (62) from (52). This implies the following. If a stochastic process (x_t, d_t, y_t) is a stable solution of (52), then (\hat{x}_t, d_t, y_t) is a stable solution of (62), because the stability of x_t implies the stability of $\hat{x}_t = Hx_t$. We would like to have a statement saying that (62) has a unique stable solution if and only if (52) has a unique stable solution. But this is not true in general. The reason is that (62) does not tell us anything about those components of x that do not affect future y 's, and therefore do not enter \hat{x} . For example, the original model (52b) may contain an equation $x_{i,t} = 2x_{i,t-1}$ and say that $x_{i,0}$ is predetermined. Then (52b) does not have a stable solution. However, if $x_{i,t}$ does not enter into \hat{x}_t , the reduced model (62) can still have stable solutions. Similarly, (52b) may contain an equation $x_{i,t} = 0.5x_{i,t-1}$ and say that $x_{i,0}$ is not predetermined. Then (52b) has a continuum of stable solutions. If $x_{i,t}$ does not affect \hat{x}_t , then (62) can still have a *unique* stable solution.⁹ The following proposition identifies a sufficient set of conditions such that a unique stable solution of (52) implies a unique stable solution of (62).

Proposition 1. *If*

- (58) holds
- the process (x_t, d_t, y_t) is the unique stable solution of model (52)
- the matrix T is asymptotically stable

then (\hat{x}_t, d_t, y_t) is the unique stable solution of the reduced model (62), with $\hat{x}_t = Hx_t$ always.

Proof. Cf. Appendix C.1. □

What is remarkable about model (62) is that the possibility of aggregation can be judged from the matrix T , which is part of the description of the linear RE model, without knowing the solution of the model. The next section will show how to check for exact and almost-exact aggregation.

⁹Notice that if we say that a model has a stable solution, this means that it has a stable solution for arbitrary initial values of the predetermined variables. For example, we say that the model $x_t = 2x_{t-1}$ with x_0 given does not have a stable solution, although it has one for the special case $x_0 = 0$.

3.4.2 Conditions for exact and almost-exact state aggregation

The last section has shown a very favorable case, namely exact aggregation. To check whether exact aggregation holds, we follow the procedure of Section 3.2.2, and compute the observability matrix

$$Q = \begin{bmatrix} C \\ CT \\ CT^2 \\ \dots \\ CT^{n-1} \end{bmatrix} \quad (63)$$

where $n = \dim(T)$. We have the following

Lemma 1. *If $\text{rank}(Q) = k$, then there is a $k \times n$ -matrix H and a $k \times k$ -matrix \hat{T} satisfying (32), (56) and (58).*

Proof. From the Cayley-Hamilton theorem, there exist real numbers $\lambda_0, \lambda_1, \dots, \lambda_{n-1}$ such that $T^n = \sum_{i=0}^{n-1} \lambda_i T^i$. This implies that CT^n is spanned by the rows of Q , and therefore

$$QT = \Lambda Q \quad (64)$$

with

$$\Lambda \equiv \begin{pmatrix} 0 & I & 0 & 0 & \dots & 0 \\ 0 & 0 & I & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & I \\ \lambda_0 I & \lambda_1 I & \lambda_2 I & \lambda_3 I & \dots & \lambda_{n-1} I \end{pmatrix} \quad (65)$$

If $\text{rank}(Q) = k$, the SVD of Q can be written as

$$Q = [U_1 \ U_2] \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V'_1 \\ V'_2 \end{bmatrix} = U_1 S V'_1 \quad (66)$$

$$S \equiv \text{diag}(\sigma_1, \dots, \sigma_k) \quad (67)$$

Using (66) in (64) we get $U_1 S V'_1 T = \Lambda U_1 S V'_1$. Now set $H = V'_1$, which satisfies (32). We get $HT = S^{-1} U'_1 \Lambda U_1 S H$, since S is invertible by construction. This shows (58) with $\hat{T} = S^{-1} U'_1 \Lambda U_1 S$, which is a $k \times k$ -matrix. To verify (56), notice that $C = [I \ 0 \ \dots \ 0] Q = ([I \ 0 \ \dots \ 0] U_1 S) H$. \square

The interpretation of the lemma is the following. If $\text{rank}(Q) = k$, then there is a vector of k variables, \hat{x}_t , that contains the complete information about the contribution of x_t to $E_t y_{t+i}$. Then we can replace the state x_t in (52) by \hat{x}_t without loss of precision in the prediction of y_{t+i} .

Is this result helpful for the applications that we are interested in? The crucial finding is the following. While the theoretical rank of Q is equal or close to n , it very often turns

out that the *numerical rank* of Q is much lower than n . This means that the the SVD of Q is not given by (66), but by

$$Q = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & E \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} = U_1 S V_1' \quad (68)$$

$$E \equiv \text{diag}(\sigma_{k+1}, \dots, \sigma_n) \quad (69)$$

where $\sigma_{k+1} < \varepsilon \sigma_1$ and ε is machine precision. On a PC, we have $\varepsilon \approx 10^{-16}$. In this case, treating E as zero and proceeding as above we obtain an approximation error that is in the same range as all the errors that result from working on a computer with finite precision.¹⁰ I call this case “almost-exact” aggregation. One could also call it “machine-precision (MP)-exact”, because it is the number of digits of the computer what limits precision here. We will see below that MP-exact aggregation holds for the growth model and the SDP model, but not for the OLG model.

The phenomenon that a large dynamical system allows for almost exact aggregation is not specific to economics. It seems to be a frequent finding in the control literature for systems with few inputs (here: shocks) and few outputs (here: relevant aggregate variables). Antoulas, Zhou, and Zhou (2002) give a partial explanation of this finding, cf. also Appendix C.2.

3.4.3 Exact policy aggregation

The method of Section 3.4.2 allows us to approximate the original model (52) by the model (62) with a much smaller state vector \hat{x} . However, the decision vector d may be just as big as the original state vector. Then we are left with an equation system with few states, but many decision variables. In standard software to solve linear rational expectations models, the computational complexity depends on the total number of variables, not on the number of states. To my knowledge, there is no algorithm available that would exploit the fact that the state vector is very small compared to the decision vector. In this section I will present an iterative method that does this.

The method is based on the following observation. If the model (62) has a unique stable solution, standard arguments show that the decision function can be written in the form

$$d_t = D_{\hat{x}} \hat{x}_{t-1} + D_E \epsilon_t \quad (70)$$

which refers only to \hat{x}_{t-1} , not x_{t-1} . No matter what the dimension of d_t is, we see from (70) that d_t lives in a subspace of dimension at most $\dim(\hat{x}) + \dim(\epsilon)$. This shows that d_t can be approximated as a linear combination of at most $\dim(\hat{x}) + \dim(\epsilon)$ basis vectors. If we collect those into the columns of the matrix G , we can write:

$$d_t = G \hat{d}_t \quad (71)$$

¹⁰Of course, we can be less stringent and consider singular values smaller than, say, $10^{-10} \sigma_1$ as essentially zero, which further reduces the dimension of the approximate model. This will depend on the accuracy requirement in each specific application.

W.l.o.g., we can assume that $G'G = I$, such that $\hat{d}_t = G'd_t$. If we have G or an estimate of it, we can plug (71) into (62), after premultiplying (62b) by G' , and further reduce the model to

$$\hat{x}_t = \hat{T}\hat{x}_{t-1} + HDG\hat{d}_t + HF\epsilon_t \quad (72a)$$

$$\hat{d}_{t-1} = E_{t-1} G' \left[\mathcal{E}_{\hat{x}1}\hat{x}_{t-1} + \mathcal{E}_{\hat{x}0}\hat{x}_t + \mathcal{E}_{d0}G\hat{d}_t \right] \quad (72b)$$

The matrix G spans the space in which the decision vector d_t lives. Standard approximation techniques such as low order polynomials or splines do something similar, and define G implicitly. However, since G is exogenously given, this implies some loss of accuracy. Our aim is to find out the exact subspace in which d_t lives, so that (71) does not reduce the accuracy of the solution. Unfortunately, this is a nonlinear problem, and we need an iterative algorithm to solve it. To do this, assume for a moment that \hat{x} follows

$$\hat{x}_t = \hat{A}\hat{x}_{t-1} + \hat{B}\epsilon_t \quad (73)$$

for given matrices \hat{A} and \hat{B} . Using (73), and writing the optimal decision as (70), we get from (62b) that

$$\begin{aligned} 0 &= E_{t-1} [\mathcal{E}_{\hat{x}1}\hat{x}_{t-1} - (D_{\hat{X}}\hat{x}_{t-2} + D_E\epsilon_{t-1}) + \mathcal{E}_{\hat{x}0}\hat{x}_t + \mathcal{E}_{d0}(D_{\hat{X}}\hat{x}_{t-1} + D_E\epsilon_t)] \\ &= \mathcal{E}_{\hat{x}1} \left(\hat{A}\hat{x}_{t-2} + \hat{B}\epsilon_{t-1} \right) - (D_{\hat{X}}\hat{x}_{t-2} + D_E\epsilon_{t-1}) \\ &\quad + \mathcal{E}_{\hat{x}0} \left(\hat{A}^2\hat{x}_{t-2} + \hat{A}\hat{B}\epsilon_{t-1} \right) + \mathcal{E}_{d0}D_{\hat{X}} \left(\hat{A}\hat{x}_{t-2} + \hat{B}\epsilon_{t-1} \right) \end{aligned} \quad (74)$$

(74) has to be satisfied for all \hat{x}_{t-2} and all ϵ_{t-1} . Collecting terms we get

$$0 = \mathcal{E}_{\hat{x}1}\hat{A} - D_{\hat{X}} + \mathcal{E}_{\hat{x}0}\hat{A}^2 + \mathcal{E}_{d0}D_{\hat{X}}\hat{A} \quad (75)$$

$$0 = \mathcal{E}_{\hat{x}1}\hat{B} - D_E + \mathcal{E}_{\hat{x}0}\hat{A}\hat{B} + \mathcal{E}_{d0}D_{\hat{X}}\hat{B} \quad (76)$$

(75) is a Sylvester equation in $D_{\hat{X}}$, which can be solved quite fast, because state reduction yields a matrix \hat{A} of moderate dimension (bringing \hat{A} into upper triangular form with a Schur decomposition, (75) can be solved column-wise). Then we can use $D_{\hat{X}}$ to get an estimate of the subspace in which the solution lives, and we can iterate on \hat{A} in the following way:

1. Set $\hat{A} = HTH'$.
2. For given \hat{A} , solve (75) for $D_{\hat{X}}$.
3. Choose $G = \text{orth}(D_{\hat{X}})$.
4. Given G , solve the linear model (72). This gives us a solution of the form

$$\hat{d}_t = \hat{D}_{\hat{X}}\hat{x}_{t-1} + \hat{D}_E\epsilon_t \quad (77)$$

Then set $\hat{A} = HTH' + HDG\hat{D}_{\hat{X}}$ and go to 2.

5. Iterate 2– 4 until convergence.

Notice that this algorithm does not just alternate between computing $\hat{D}_{\hat{X}}$, given \hat{A} , and forming \hat{A} from $\hat{D}_{\hat{X}}$. The algorithm iterates not on the decision matrix $\hat{D}_{\hat{X}}$, but on the *subspace* in which the decision vector lives.

There is no guarantee that this algorithm works. First, we have no proof that the iterations converge. A second, more subtle problem is the following. If we have correctly identified G , we know that the exact solution, which we assume to be the unique stable solution of the model (62), is a stable solution of (72). But we have no guarantee that it is the *unique* stable solution of (72). Policy reduction might induce spurious indeterminacy. In particular, even if \mathcal{E}_{d0} is asymptotically stable, the matrix $G'\mathcal{E}_{d0}G$, which is the matrix for iterating (72b) forward, might not be.¹¹ Although there is no proof, the algorithm appears to be very effective and has converged in 2 or 3 iterations in all the applications I have done so far. Getting a deeper understanding of why this works so well must be left for future research.

3.5 Approximate LS Aggregation in Linear RE models

Even if we can achieve almost-exact aggregation by the methods of Section 3.4, we may want to further reduce the dimensionality of the model, at some cost in terms of reduced accuracy. One important reason is that low-dimensional models allow higher-order approximations in aggregate variables, which is important for example in models of asset choice (Reiter 2009a). The great majority of the existing macroeconomic applications, following Krusell and Smith (1998), use approximations with a very low number of state variables, typically one or two. Below we will study approximations with up to 20 state variables. This is the range of variables that one can reasonably handle with nonlinear methods, for example by projection methods with Smolyak polynomials (Malin, Krueger, and Kubler 2007).

The algorithm presented below is a linear version of the Krusell/Smith algorithm. Approximate aggregation in the sense of Krusell/Smith is based on the following ideas:

- Economic agents base their decisions at time t on a reduced set of state variables, \hat{x}_t .
- Agents take as given an aggregate law of motion for the reduced state vector. In the context of linear models, this can be written as

$$\hat{x}_t = \hat{A}\hat{x}_{t-1} + \hat{B}\epsilon_t \quad (78)$$

Agents make their decisions optimally, conditional on the perceived aggregate dynamics (78).

¹¹This problem does not appear in the examples of this paper, but it did happen occasionally in other applications. To solve it, we can try to include more vectors into G (Step 3 of the algorithm). A natural way to do this is to include in G the eigenvectors of the largest (in magnitude) eigenvalues of the matrix \mathcal{E}_{d0} . Since this matrix is sparse, fast algorithms for computing the largest eigenvalues and -vectors are available. Including 6 or 10 eigenvectors has always been sufficient.

- In equilibrium, the dynamics of the model implied by the decisions of agents must be compatible with the aggregate law of motion (78) in a least-squares sense. That means, if agents estimate the model (78) on a long simulation of the disaggregate model implied by the agents' decision function, they obtain the matrices \hat{A} and \hat{B} that they have used in making their decisions.

This can be formalized in the following fixed point problem,

1. Guess \hat{A} and \hat{B} where \hat{A} is asymptotically stable.
2. Solve the system of equations (78) and (72b) to get the matrices $D_{\hat{X}}$ and D_E of the decision rule $\hat{d}_t = \hat{D}_{\hat{X}}\hat{x}_{t-1} + \hat{D}_E\epsilon_t$.
3. Set A and B as

$$A = T + DG\hat{D}_{\hat{X}}H \quad (79a)$$

$$B = F + DG\hat{D}_E \quad (79b)$$

and compute

$$\Sigma_x = \mathcal{L}(A, B, \Sigma_\epsilon) \quad (80)$$

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

$$\hat{A} = H A \Sigma_x H' (H \Sigma_x H')^{-1} \quad (81a)$$

$$\hat{B} = H B \quad (81b)$$

5. Iterate until the results in (81) 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} .

Notice that this algorithm, by exploiting the linearity of the setup, avoids the use of simulation methods, and is therefore not affected by sampling errors, unlike the original (nonlinear) algorithm of Krusell and Smith (1998).

3.6 Evaluating Accuracy

3.6.1 Types of approximation error

Solving the model by the method outlined in Section 3.1 involves three types of approximation error (next to the usual errors from working on a computer with finite precision):

1. error from discretization
2. error from linearization

3. error from aggregation.

The first type of error already affects the steady state solution of the model. It can be estimated by increasing the size of the discrete approximation (grid points in cross-sectional distribution, order of spline approximation of policy function etc.). Some estimates for the growth model are given in Reiter (2009b). The error from the linearization in aggregate variables can be estimated when computing higher order approximations. This is done in Reiter (2009a). Since the topic of the present paper is aggregation, I focus exclusively on the error from aggregation, arising from the reduction of the state and the policy vector.

3.6.2 How to measure the aggregation error

We measure the aggregation error by comparing the solution of the reduced model, represented by the matrices \hat{A} , \hat{B} and \hat{C} from (72), with the solution of the disaggregate model. Often we can compute the exact solution of the (linearized) model, represented by the matrices A , B and C obtained from (54). Then we can measure the quality of an approximate solution directly against the exact solution.

In general, we only have an approximate solution. Then, the key to an accuracy check is the disaggregate model that is implied by the decision functions of the reduced model. The disaggregate model is defined by the matrices A and B given by (79). This is the analogue to the simulation procedure used in Krusell and Smith (1998), Den Haan (2007) and many other papers to check accuracy. In all the examples below, it turns out that checking the approximate model against the disaggregate solution (79) is a good substitute for checking against the exact solution (54).

3.6.3 A variety of accuracy measures

We judge the quality of the the reduced model (30) by its the ability to give correct predictions of the variables of interest y_t , if both models are driven by the same shock. We measure this using several statistics. All statistics refer to the aggregation error.

Accuracy measure 1: unconditional mean squared error

This measure tells us how well a model predicts the relevant aggregate variables (for example capital) on average, where the average is taken over the ergodic distribution of the model solution. To compute it, combine (29a) and (30a) into

$$x_t^* = A^* x_{t-1}^* + B^* \epsilon_t \quad (82)$$

$$x^* \equiv \begin{bmatrix} x_t \\ \hat{x}_t \end{bmatrix}, \quad A^* \equiv \begin{bmatrix} A & 0 \\ 0 & \hat{A} \end{bmatrix}, \quad B^* \equiv \begin{bmatrix} B \\ \hat{B} \end{bmatrix} \quad (83)$$

Since

$$E(y_t - \hat{y}_t) = 0, \quad (84)$$

the unconditional mean squared error is given by

$$E[(y_t - \hat{y}_t)(y_t - \hat{y}_t)'] = [C \quad -\hat{C}] \Sigma_{x^*} [C \quad -\hat{C}]' \quad (85)$$

where $\Sigma_{x^*} = \mathcal{L}(A^*, B^*, \Sigma_\epsilon)$ is the unconditional variance-covariance matrix of x^* . The root mean squared error of the prediction of $y_{i,t}$ is then given by the i -th element on the main diagonal of (85). Below we will report the root mean squared error, normalized by the standard deviation of $y_{i,t}$:

$$\sqrt{\frac{[C_i \quad -\hat{C}_i] \Sigma_{x^*} [C_i \quad -\hat{C}_i]'}{C_i \Sigma_x C_i'}} \quad (86)$$

where C_i is defined as the i -th row of the matrix C . This statistic can be interpreted as the unconditional standard deviation of an infinite-horizon forecast error. It therefore addresses the main concern in Den Haan (2007), who suggests to use long-term forecasts to assess accuracy.

Accuracy measure 2: impulse response error

If the model is driven by several exogenous shocks, the statistic (86) measures the joint effect of all the shocks on the forecast error. To analyze separately how well the reduced model predicts the effects of each type of shocks, one can look at differences in impulse response functions. The impulse response of y_{it} to shock j is given by

$$y_{i,t;j} = C_i A^t B e_j \quad (87)$$

where e_j is the impulse vector which has unit entry at position j and is zero otherwise. I report

$$\frac{\max_t |y_{i,t;j} - \hat{y}_{i,t;j}|}{\max_t |y_{i,t;j}|} \quad (88)$$

where the maximum goes over $t = 1, \dots, 1000$.

Accuracy measure 3: prediction error along transition path

While the first two measures refer to the prediction error in response to shocks, the third measure asks how well the reduced model predicts the dynamics of y if we start out from an arbitrary initial state x_0 . Then $E[y_{i,t}|x_0] = C_i A^t x_0$. From (31) we have $\hat{x}_0 = H x_0$. Then $E[\hat{y}_{i,t}|x_0] = \hat{C}_i \hat{A}^t H x_0$. Then the maximum in the prediction error over all unit vectors x_0 is given by the matrix norm

$$\max_{x_0 \text{ s.t. } \|x_0\|_2=1} \|E[\hat{y}_{i,t}|x_0] - E[y_{i,t}|x_0]\|_2 = \|C_i A^t - \hat{C}_i \hat{A}^t H\|_2 \quad (89)$$

I report

$$\min \left\{ \max_t \beta^t \|C_i A^t - \hat{C}_i \hat{A}^t H\|_2, 1 \right\} \quad (90)$$

where the maximum again goes over $t = 1, \dots, 1000$. Discounting with β gives higher weight to earlier periods.

4 Numerical Results

To relate our findings to the existing literature, we begin the presentation of the results in Section 4.1 with the linearized version of the Krusell/Smith solution. We will confirm earlier findings about the precision of this method in the model with technology shocks, and demonstrate that the model with tax shocks requires an approximation with more than one moment in order to give a reasonably precise approximation. Section 4.2 applies the model reduction techniques of Section 3.2 to all three example economics, and studies how the precision of the reduced model increases with the number of statistics used. Section 4.3 computes the almost-exact aggregation solution of Section 3.4 for all models. Section 4.4 looks again at approximate aggregation with the Krusell/Smith solution, now considering higher-dimensional approximations and different types of state variables. Finally, Section 4.5 discusses whether aggregation in the growth model holds exactly in the theoretical sense, or only up to machine precision in the sense of Section 3.4.2.

4.1 Linearized Krusell/Smith Solution With Few Moments

In a model similar to the one of Section 2.4.3 with $n_P = 1$, the celebrated finding of Krusell and Smith (1998) was that one can obtain a sufficiently precise solution of the model with an approximation where the only endogenous state variable is aggregate capital. To replicate those results, and consider solutions using higher-order moments, define the state vector of the growth model with $n_P = 1$ as $x_t = [\Phi_t \ Z_t \ \tau_t^k]'$, and define the matrix H as

$$H = \begin{pmatrix} \bar{k}_1 & \bar{k}_2 & \dots & \bar{k}_{n_k} & 0 & 0 \\ \bar{k}_1^2 & \bar{k}_2^2 & \dots & \bar{k}_{n_k}^2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \bar{k}_1^m & \bar{k}_2^m & \dots & \bar{k}_{n_k}^m & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix} \quad (91)$$

Then $\hat{x}_t = Hx_t$ contains the first m non-central moments of the distribution of capital, and the two exogenous processes.¹² With this choice of H , we then apply the method of Section 3.5.

Figure 1 presents the results. It shows impulse responses to technology and tax shocks, using approximations with an increasing number of moments. (“Three moments” means that $m = 3$ in (91), etc.) Each panel of Figure 1 compares four different solutions to the model:

¹²A serious numerical implementation needs some modifications. First, non-central moments are integrals over *monomials* in k . If the required number of moments goes beyond four or five, this leads to multicollinearity in the columns of H . One should therefore use integrals over orthogonal polynomials of k , such as Chebyshev polynomials (Judd 1998, Section 6.3). Second, I find it is better to use polynomials in $\log(k)$, not k , except for the first moment, which must exactly be aggregate capital, because of its importance for factor prices. The details for these modifications of H can be found in the Matlab program “hpoly.m”.

1. The exact solution of the discrete model (labeled “Exact”), with A , B defined as in (54).
2. The reduced model, obtained from the exact solution via (33), using the matrix H from (91). This is labeled “ReducSol”.
3. The approximate solution with H from (91), computed by the Krusell/Smith algorithm of Section 3.5 (“ApprAggr”).
4. The disaggregate model (79) obtained from the approximate model (“ApprDisaggr”).¹³ We can interpret this as the outcome of simulating the disaggregate economy, using the household decision rules obtained from the approximate model.

Panel a) of the figure illustrates the famous result that “one moment is enough”. The impulse responses of the four different solutions to a technology shock are visually indistinguishable. The relative error in the impulse response, measured as in (88), equals 0.0029. Expressing this as an R^2 , as is often done in the literature, it would give $R^2 = 1 - 0.0029^2 = 0.999991$.

Panel b) of the figure shows the analogous result for the tax shock. It looks quite different. All the approximations are far off the exact solution. What goes wrong here? A tax shock triggers the following responses. On impulse, the shock is a redistribution from the wealth rich to the poor. Then there is a 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. But this triggers a 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. The approximation error arises because households look at the mean capital stock only. So they 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 dominating more strongly.

The remaining panels of Figure 1 investigate whether the problems can be solved by enlarging the state space. The news is good: the discrepancies between solutions gradually disappear when using more moments, cf. panels c)–f).

There is a second, and perhaps more important piece of good news in Figure 1: 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 approximate model solution (“ApprAggr”) to the disaggregate model under approximate decision functions (“ApprDisaggr”). This is what Krusell and Smith (1998) already did. Those two solutions differ by the same order of magnitude as the exact and approximate solution differ. Applying due diligence, the failure of the approximation does not go unnoticed.

¹³In this section I do not use policy aggregation, so $G = I$ in (79).

4.2 Model Reduction in State Space Form

Even in the simple stochastic growth model, we have seen that in general several moments are necessary to achieve a satisfactory level of accuracy. We now look at these issues in more detail. The questions are: how does the accuracy of the reduced model depend on its dimension? How should the variables in the reduced model be chosen?

In this section, we compute the exact solution of the linearized model, and study whether this solution allows model reduction, using the techniques of Section 3.2. We do this for the models for which we can compute the exact solution, the growth model with $n_P = 1$ (i.i.d. income shocks), the SDP model, and the OLG model. We can then measure accuracy as the distance between the exact solution and the reduced model.

Figure 2 reports results for the growth model. We consider four sets of aggregate statistics: the three approaches of Section 3.2, namely PCA, CEA and balanced reduction, and the traditional macroeconomic approach, namely moments of the cross-sectional distribution, as explained in Section 4.1. In each case, we vary the number of statistics of the cross-sectional distribution between 1 and 100 (again, the exogenous variables Z and τ^k are not counted here, so that the total number of state variables is higher by two). For the approaches based on control theory, the maximum number of statistics is the numerical rank of the covariance matrix or the observability matrix. All the accuracy results refer to the forecast of aggregate capital. This is the relevant variable in this model, because it determines all the prices in the economy. Numbers are decimal logs. All the statistics are truncated above at a relative error of 1. The two top panels of the figure report the impulse response error, defined in (88). The two bottom panels report the RMSE of the infinite-horizon forecast error and the maximum error along a transition, defined in (86) and in (90), respectively.

The following conclusions emerge. First of all, by choosing a state space of about 50 variables, the approximation error can be made very small (relative errors in the range 10^{-8} to 10^{-12}). Given that 10^{-16} is machine precision, we cannot expect more. Balanced reduction performs best when the criterion is to predict the response to shocks, both at short-to-medium horizon (impulse responses) or at infinite horizon (unconditional forecast error). This is what we should expect from the theory reported in Section 3.2.4. For the maximum error in a transition, the CEA approach is best. This is again not surprising, because CEA includes the expected values of (not so distant) future capital stocks in the state vector, so this should work starting from any initial distribution. However, balanced reduction is also doing quite well in this respect, if we consider that the maximum error in a transition is a very tough criterion.

PCA is not competitive, for any of the criteria considered. This shows that the solution contains many components that vary substantially (and are therefore included by PCA) but are not relevant for the prediction of the aggregate capital stock. Changes in the distribution that happen over a range where the consumption function is almost linear will probably fall into this category. It is reassuring to see that the moments-approach, which is an intuitive one, performs reasonably well. It fails only for the transition error. Even putting 100 moments in the state vector, one can still find initial conditions for the

distribution of capital for which the reduced model totally fails to predict the transition path. Only CEA and balanced reduction can fix this problem.

Figure 3 provides the same information for the SDP model, where the prediction error is for the nominal interest rate R . It confirms most of the results obtained from the growth model. There are two significant differences, however. First, moments as state variables (bivariate moments of the joint distribution of prices and firm-specific technology) are not as successful as in the growth model. This underlines the importance of having a model reduction procedure that is independent of the structure of the underlying state space. Another difference is that CEA and balanced reduction perform about equally well with respect to the maximum transition error.

Before you conclude that approximate aggregation always works (as I did believe at some point), it is useful to look at Figure 4, which reports results for the OLG model. The model has 720 state variables (capital held by the 720 cohorts). The moments-based approach has now been replaced by the approach of grouping together nearby cohorts. For example, when using 20 statistics, the state variables are the capital stock held by cohorts 1–60, 61–120, ..., 661–720. The results are strikingly different. Even going up to 120 state variables, errors are still in the range of 10^{-4} to 10^{-5} , and bigger for the maximum transition error. Households belonging to different cohorts behave in a distinctly different way, which does not allow aggregation at a high level of precision. One might feel that a household that has 700 months to live should be almost identical to one that has 701 months to live, but this is not true at the level of precision that we require, and that we have achieved in the models with infinitely-lived agents. Still, the control-theory based techniques of Section 3.2 are useful. Balanced reduction (but not CEA) performs better than simply grouping cohorts.

4.3 Almost-Exact Aggregation

In the last section we have asked whether the exact solution of the model, in cases where we can compute it, is amenable to aggregation. We now turn to the task of applying aggregation directly to the DSGE model. This is the only choice if the model is too big to be solved without aggregation. As we have explained in Section 3.3, the novel element is that now the economic agents themselves are assumed to apply approximate aggregation, so that the aggregation procedure feeds back into the solution of the model. Therefore it is not clear whether the positive results of Section 4.2 go through.

Section 3.4 has shown a way to solve the DSGE model with almost-exact aggregation, if the numerical rank of the observability matrix is smaller than the dimension of the system. Table 2 reports the results from applying this algorithm. The table indicates the variable which should be predicted (capital in the growth model, inflation or nominal interest rate in the SDP model), the frequency of the model (from monthly to annual), the dimensions of state and decision variable of the original model and of the reduced model, then the maximal error in the impulse response of both shocks (technology and shocks in the growth model, technology and interest rate in the SDP model), the maximal error in the transition and the RMSE of the infinite-horizon forecast. For each model, the first line

	V.	Fr.	$d(x)$	$d(d)$	$d(\hat{x})$	$d(\hat{d})$	IR tec.	IR tax	ErrTr.	RMSE
GM	K	4	1003	100	77	27	1.1e-12	4.6e-11	2.5e-11	1.5e-12
							3.4e-12	4.6e-11	2.7e-10	1.5e-8
GM, $\sigma_\tau = 0$	K	4	1003	100	77	27	4.4e-13	3.5e-11	2.0e-11	2.1e-8
							4.8e-12	6.4e-11	3.9e-10	3.0e-8
GM	K	1	1003	100	102	26	6.2e-13	1.8e-12	3.3e-11	8.0e-13
							1.2e-12	4.1e-11	5.2e-11	2.0e-8
GM	K	12	1003	100	65	23	5.1e-12	7.2e-10	3.1e-10	3.9e-8
							3.8e-12	7.2e-10	2.9e-10	3.5e-8
GM, $n_P = 31$	K	4	31003	3100	202	31	2.5e-11	4.9e-11	7.8e-8	3.7e-11
SDP	π	4	744	734	94	32	1.3e-12	2.2e-12	2.3e-12	1.6e-12
							7.5e-11	1.7e-10	3.0e-10	2.5e-8
SDP	R	4	744	734	94	32	1.7e-12	5.1e-13	2.3e-12	1.3e-12
							7.0e-11	6.1e-11	3.0e-10	3.4e-8

Table 2: Results Almost-Exact Aggregation

reports these statistics for the difference between the the reduced model and the implied simulated model (“ApprAggr” and “ApprDisaggr” in the terminology of Section 4.1). The second line reports the difference between the the reduced model and the exact solution, in those cases where it is available (“ApprAggr” and “Exact”). The results are a clear success. Relative errors are mostly in the range of 10^{-10} or better. On a machine that does basic operations with 16 digits accuracy, one cannot expect more. The RMSE of the infinite-horizon forecast is difficult to compute, and it is not clear whether differences below 10^{-8} are significant or just measurement errors.

Notice that the difference between solutions “ApprAggr” and “ApprDisaggr”, which can always be computed, is a very good indicator for the difference between solutions “ApprAggr” and “Exact”.

The dimension of \hat{x} is chosen by the algorithm as the numerical rank of the observability matrix (63). Of course, we can reduce the dimension of \hat{x} quite a bit if we are willing to give up some precision. For example, in the benchmark growth model, if we choose $\dim(\hat{x}) = 40$ rather than $\dim(\hat{x}) = 77$, we lose about two or three digits of precision, which would still be sufficient for any practical application.

4.4 Approximate Aggregation

We now return to the problem of approximate aggregation with a small number of state variables, which we had already touched in Section 4.1. We will test the accuracy of the Krusell/Smith algorithm of Section 3.5, applied to the growth model with $n_P = 1$. Now we vary the number of endogenous state variables between 1 and 20, and compare the use of moments to the use of CEA and balanced reduction. PCA was dropped from the list, because it already proved to be non-competitive in Section 4.2.

The results in Figure 5 should be compared to Figure 2. Accuracy is measured as the distance between the approximate solution and the exact solution of the linearized model. Results for more than 12 moments are not reported because the algorithm did not converge. There is a slight difference in the choice of state variables for the case of CEA and of balanced reduction, compared to Section 4.2. The first statistic is now always the aggregate capital stock, because it is what determines prices, and the reduced model assumes that the agents know prices, so this information has to be contained in the state vector.

The main finding is that the accuracy of the KS method is quite similar to the accuracy of directly aggregating the solution of the model, as we did in Section 4.2. That the aggregation feeds back into the model solution, because agents base their decision on the aggregate model, does not pose a serious problem for the accuracy of the approximate model. The exception to this finding is that the maximum error along a transition is now considerably higher, although accuracy is still very good for the CEA approach. The KS approach selects the dynamics of \hat{x} to be optimal in the OLS sense, which does not care about the transition error from arbitrary initial conditions. If the error along a transition path is the most important criterion, one could certainly find modifications of the algorithm that improve in this direction.

Results for the growth model with $n_P = 31$ and for the SDP, which are not reported here, confirm these conclusions.

4.5 Is Exact Aggregation Possible?

In Section 4.3 we have seen that a model with a very reduced state vector is able to provide an approximation basically up to machine precision. We then have to ask whether the theoretical model allows exact aggregation. This means that the observability matrix Q in (63) has exact, not just numerical rank, smaller than n . With purely numerical methods, it is not easy to give a definite answer to this question, but the Popov-Belevitch-Hautus test is helpful in this respect: the observability matrix of the pair (T, C) has rank equal to $\dim(T)$ if and only if there exists no right eigenvector of T orthogonal to the rows of C (Williams and Lawrence 2007, Theorem 4.7). More generally, the number of eigenvectors of T not orthogonal to C is a good indicator of the rank of the observability matrix. This is illustrated, for the case that T can be diagonalized, in Appendix C.2. Figure 6 reports this test for the growth model with $n_P = 1$. Here, C here is the vector that computes the mean of the cross-sectional distribution of capital, K . The upper panel plots the absolute eigenvalues of the matrix T in (52a), in ascending order. They are quite evenly spread between 0 and 0.9995. The middle panel plots the decimal log of the inner product of C and the eigenvectors corresponding to those eigenvalues. For the about 50 eigenvalues closest to unity, this inner product is large, indicating that the components of those eigenvalues are important for the dynamics of aggregate capital. The lower panel plots the inner product again, but now in ascending order. We see that for some eigenvectors, the inner product is so small that we cannot be absolutely sure that it is different from zero, given that the machine precision is 16 digits. However, only a few values are close to zero by this

standard, and for more than 90 percent of the eigenvectors, the inner product is definitely nonzero. We can therefore say that the very strong reduction in the number of states in the calculations of Section 4.3 is not a theoretical, just a numerical property of the model.

5 Conclusions

The main result of the paper is that linearized DSGE models with a high number of state variables (more than 30000 in one example) can be solved with very high precision, using methods of approximate state and policy reduction. The paper has shown this with two prominent examples of heterogeneous agent models, a stochastic growth model and a state-dependent pricing model. Approximations with 50 to 200 state variables solve the model basically up to machine precision. Even fewer states are sufficient if the approximation is only required to capture the effects of stochastic shocks, not to predict the transition from arbitrary initial conditions. The paper has shown how to choose the state variables so as to get almost optimal accuracy for a given level of aggregation.

Obviously, not all high-dimensional economic models can be solved accurately. The experience so far suggests that the following considerations are important:

- For reasons of computational efficiency, the transition matrix (matrix T in (52)) should be sparse. With more than 10000 state variables, a dense T might not even fit into computer memory. Economically this means that, from any given individual state today (a given level of capital, for example), there is only a small set of states tomorrow that the agent can reach with positive probability. The level of sparsity is usually a function of the time period. A model at monthly frequency will probably be sparser, and therefore easier to handle, than a model at annual frequency.
- It is useful if the cross-sectional distribution affects individual decisions only through a small set of variables. For example, the distribution of capital in the growth model affects households only through factor prices, which are a function of the aggregate capital stock. Knowing the whole cross-sectional distribution of capital is important for households only insofar it helps to predict *future* factor prices.
- Discrete differences between types of agents may not allow almost-exact aggregation. For example, an OLG model with many cohorts seems to require as many variables as cohorts, if a very high level of accuracy is to be achieved.

Interestingly, the number of exogenous driving forces seems to play no essential role for aggregation. The theory of almost-exact aggregation in Section 3.4 was based on the matrix T which describes the dependence of the current state on last period's state, not on the shocks.

The most important issue that is left for future research is to investigate under what conditions the discretized model is a good approximation to the theoretical model with continuous state space, and whether the discrete model converges in a suitable sense to

the continuous model, if the dimension of the state vector goes to infinity. Another interesting avenue for future research is to compute approximate solutions of heterogeneous agent models that can be interpreted as solutions under bounded rationality or limited information. This will impose restrictions on the state variables that the economic agents are supposed to consider when making their decisions. For example, those variables should be measured and observed in reality. In finding the optimal state vector in this paper, we have not considered any such restrictions.

A Example 3: A Model of Heterogeneous Firms and State-Dependent Pricing (SDP)

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.

A.1 Firms

There is a continuum of firms and each of them is the monopolistically competitive producer of a differentiated good. At the beginning of a period, firm h is characterized by two individual state variables, its own price inherited from the last period, $P_{t-1,h}$, and its idiosyncratic productivity level $\theta_{t,h}$. Each firm produces with the Cobb-Douglas production function

$$y_{t,h} = Z_t \theta_{t,h} l_{t,h} \quad (92)$$

The aggregate level of technology, Z_t , is assumed to be given by the following process

$$\ln Z_t \equiv z_t = \rho_z z_{t-1} + e_{z,t}, \quad (93)$$

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

Utility-maximization on the part of households implies that demand for good i is given by

$$y_{t,h} = \left(\frac{P_{t,h}}{P_t} \right)^{-\epsilon} Y_t^d, \quad (94)$$

where Y_t^d is aggregate demand.

In order to change its price the firm must pay a fixed cost, denoted by $\kappa_{t,h}$, which is measured in units of the aggregate good and given by

$$\kappa_{t,h} = \begin{cases} 0 & \text{if } P_{t,h} = P_{t-1,h}, \\ \kappa & \text{otherwise,} \end{cases} \quad (95)$$

The price adjustment cost κ is stochastic. It is independent across time and across firms. It is realized each period before the firm set its price, and before production takes place.

Each firm chooses prices $p_{t,h}$ and labor input $l_{t,h}$ so as to maximize its market value

$$E_0 \sum_{t=0}^{\infty} Q_{0,t}^R \left\{ \frac{p_{t,h}}{P_t} y_{t,h} - W_t l_{t,h} - \kappa_{t,h} \right\}, \quad (96)$$

subject to the constraints (92)–(95). Here, W_t is the real wage, and $Q_{0,t}^R$ is the real stochastic discount factor that will be defined in Section A.2.

I use a discretization that is very similar to that of Section 2.4.2, with the cross-sectional distribution of prices replacing the distribution of household capital, and the distribution of firm productivity replacing household productivity. Productivity $\theta_{t,h}$ follows an n_P -state Markov chain and takes on the values $\bar{\theta}_1, \dots, \bar{\theta}_{n_P}$. $\Pi_{j,j'}^P$ denotes the probability of switching from state j to j' . The level of prices is discretized in an analogous fashion to the discretization of capital in the growth model. I approximate the distribution of prices by a finite number of mass points at a predefined grid $\bar{p}_1, \bar{p}_2, \dots, \bar{p}_{n_p}$. Again, the boundaries \bar{p}_1 and \bar{p}_{n_p} must be chosen such that in equilibrium very few firms are close to it. For the dynamics of the price level I make the same assumption that I used for capital in the growth model. The firm chooses the price continuously, and this price is valid for the present period, but at the end of the period the price jumps randomly to the neighboring grid points such that the expected value is maintained. So we define the function $\psi(i, p)$ analogously to the function $\psi(i, k)$ in Section 2.4.2. Define $\phi_t(j, i)$ as the fraction of firms at time t that have productivity level $\bar{\theta}_j$ and price \bar{p}_i . Stack those into a column vector Φ_t exactly as in (20). Then we arrive at the dynamic equation (23) in the same way as in Section 2.4.

Since the firm problem is non-convex, optimality has to be characterized by the Bellman equation, not by Euler equations. The variables in the model that characterize behavior are therefore the $V_t(j, i)$, defined as the value function of a firm with productivity level $\bar{\theta}_j$ and price \bar{p}_i . The Bellman equation can be written as

$$V_{t-1}(j, i) = Q_{t-1,t}^R \max_{p,l} \left\{ \frac{p}{P_t} y_{t,h} - W_t l - \kappa_{t,h} + \sum_{i'=1}^{n_p} \sum_{j'=1}^{n_P} \psi(i', p) \Pi_{j,j'}^P V_t(j', i') \right\} + \eta_{i,j,t}^c, \\ j = 1, \dots, n_P, \quad i = 0, \dots, n_p \quad (97)$$

again subject to (92)–(95). Equ. (97) is the analogue to (17a) in the growth model. Similar to Φ_t , we stack all the values $V_t(j', i')$ into the column vector \mathbf{V}_t .

A.2 Households

The representative household maximizes expected discounted utility

$$E_0 \sum_{t=0}^{\infty} \beta^t U(C_t, L_t),$$

where β is the subjective discount factor, L_t is the number of hours worked and C_t denotes a Dixit-Stiglitz consumption aggregate. Defining $C_t(j, i)$ as the amount of consumption

bought from firms with price level \bar{p}_j and productivity level $\bar{\theta}_i$, this aggregate can be written as

$$C_t \equiv \left(\sum_{j=1}^{n_P} \sum_{i=1}^{n_p} C_t(j, i)^{\frac{\epsilon-1}{\epsilon}} \phi_t(j, i) \right)^{\frac{\epsilon}{\epsilon-1}} \quad (98)$$

with corresponding price index

$$P_t \equiv \left(\sum_{j=1}^{n_P} \sum_{i=1}^{n_p} \bar{p}_j^{1-\epsilon} \phi_t(j, i) \right)^{\frac{1}{1-\epsilon}} \quad (99)$$

where ϵ is the elasticity of substitution between different varieties of consumption goods. 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

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

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

$$P_t C_t + E_t \{Q_{t,t+1} D_{t+1}\} \leq D_t + P_t W_t L_t + T_t, \quad (100)$$

where $Q_{t,s}$ denotes the price in period t for nominal payments in period s , and D_{t+1} gives the nominal payoff associated with the portfolio held at the end of period t . T_t is nominal dividend income resulting from ownership of firms. Note that the stochastic discount factor $Q_{t,s}$ is unambiguous in equilibrium since households are identical. Denoting the real stochastic discount factor by $Q_{t,s}^R \equiv Q_{t,s} \left(\frac{P_s}{P_t} \right)$, the household Euler equation is given by

$$Q_{t,t+1}^R U_C(C_t, L_t) = \beta E_t U_C(C_{t+1}, L_{t+1}) \quad (101)$$

The first order condition for labor supply is

$$\frac{\eta C_t}{1 - L_t} = W_t, \quad (102)$$

A.3 Market Clearing and Monetary Policy

The goods and labor market clearing conditions are

$$y_{t,h} = Y_{t,h}^d \text{ for all } i. \quad (103)$$

$$L_t = \sum_{j=1}^{n_P} \sum_{i=1}^{n_p} L_t(j, i) \phi_t(j, i) \quad (104)$$

$$Y_t^d = C_t + \kappa_t \quad (105)$$

$$\kappa_t = \sum_{j=1}^{n_P} \sum_{i=1}^{n_p} \kappa_t(j, i) \phi_t(j, i) \quad (106)$$

Monetary policy takes the form of a simple interest rate rule

$$R_t = R_{t-1}^{\phi_r} \left(\beta^{-1} \left(\frac{P_t}{P_{t-1}} \right)^{\phi_\pi} \right)^{1-\phi_r} e^{e_{r,t}}, \quad (107)$$

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.

A.4 The Discrete Model

For each t , we have the equations (23), (93), (97)–(99), (101), (102) and (104)–(107) in the variables Φ_t , Z_t , R_t , \mathbf{V}_t , Y_t^d , L_t , C_t , W_t , P_t and κ_t .

A.5 Parameter values

I use standard parameter values: $\eta = 1.5$, $\epsilon = 7$, $\beta = 1/1.01$, $\phi_\pi = 1.5$ and $\phi_r = 0.9$. The steady state inflation rate is 0.5 percent quarterly. Firm productivity levels are equidistant in logs; from each point, only neighboring points can be reached, the transition probability is symmetric. The correlation of the technology shock is 0.95, that of the monetary shock is zero. The cost function for price adjustment is tent-shaped, with support $[0, 2E\kappa]$, and $E\kappa$ is calibrated so as to give a frequency of price adjustment of 1/3 per quarter.

A.6 Choice of grid

$n_p = 501$ grid points are used in p , and $n_P = 9$ grid points in productivity.

- The aggregate grid has $n_p \cdot n_P$ points. However, the ergodic set will be much smaller than that. To compute the solution of the discretized model, we only have to consider the $\phi_t(j, i)$ and $V_t(j', i')$ on the ergodic set.
- The grid \bar{p}_j is defined relative to the aggregate price level. Therefore, the price \bar{p}_j at time $t + 1$ stands for the price \bar{p}_j at time t , multiplied by the inflation rate P_{t+1}/P_t .
- The grid points \bar{p}_j have to be chosen such that the distance between grid points is not a multiple of the steady state inflation rate. This makes that a firm that does not change its price will end up next period between two grid points. In this way, it is guaranteed that the the solution with (infinitesimally small) aggregate shocks stays on the ergodic set.

B Asymptotic Estimation of Reduced Model

B.1 Estimation Formula

Assume the true model is

$$x_t = Ax_{t-1} + B\epsilon_t \quad (108)$$

where the ϵ is a vector of i.i.d. shocks. We think of x as being a high-dimensional state vector, and want to estimate a VAR in a lower-dimensional vector of statistics \hat{x} , which is related to x by

$$\hat{x}_t = Hx_t \quad (109)$$

with known matrix H . The estimated model is

$$\hat{x}_t = \hat{A}\hat{x}_{t-1} + \hat{B}\epsilon_t + u_t \quad (110)$$

The error term u_t is supposed to capture the aggregation error. The normal equations for the OLS estimation of this model are

$$\frac{1}{T} \sum_{i=1}^T \hat{x}_t [\hat{x}'_{t-1} \quad \epsilon'_t] = \frac{1}{T} \sum_{i=1}^T [\hat{A} \quad \hat{B}] \begin{bmatrix} \hat{x}_{t-1} \\ \epsilon_t \end{bmatrix} [\hat{x}'_{t-1} \quad \epsilon'_t] \quad (111)$$

Using (109) and (108) we can write (111) as

$$\frac{1}{T} \sum_{i=1}^T H(Ax_{t-1} + B\epsilon_t) [x'_{t-1}H' \quad \epsilon'_t] = \frac{1}{T} \sum_{i=1}^T [\hat{A} \quad \hat{B}] \begin{bmatrix} Hx_{t-1} \\ \epsilon_t \end{bmatrix} [x'_{t-1}H' \quad \epsilon'_t] \quad (112)$$

In the limit $T \rightarrow \infty$, the means converge to their unconditional expectations, and we get

$$[H\mathcal{L}(A, B, \Sigma_\epsilon)H' \quad HB\Sigma_\epsilon] = [\hat{A} \quad \hat{B}] \begin{bmatrix} H\mathcal{L}(A, B, \Sigma_\epsilon)H' & 0 \\ 0 & \Sigma_\epsilon \end{bmatrix} \quad (113)$$

which gives (33).

B.2 Stability of Reduced Model

Define the spectral radius of a square matrix X by $\varrho(X)$.

Lemma 2. *If $\varrho(A) < 1$ and*

$$\Sigma = A\Sigma A' + BB' \quad (114)$$

for some matrix $B \neq 0$, then $\varrho(HA\Sigma H'[H\Sigma H']^{-1}) < 1$ for any H such that $H\Sigma H'$ is regular.

Proof. Assume first that Σ has full rank. Because it is positive definite, it can be written as $\Sigma = SS'$ with S regular. Define $Q = S^{-1}$. From (114) we get

$$I = Q\Sigma Q' = QAQ^{-1}Q\Sigma Q'(Q')^{-1}A'Q' + QBB'Q' = \tilde{A}\tilde{A}' + \tilde{B}\tilde{B}' \quad (115)$$

where $\tilde{A} \equiv QAQ^{-1}$ and $\tilde{B} \equiv QB$. It follows from (115) that $\|\tilde{A}\|_2 < 1$. [For any x with $x'x = 1$, we get $x'Ix = 1 = x'\tilde{A}\tilde{A}'x + x'\tilde{B}\tilde{B}'x$, therefore $x'\tilde{A}\tilde{A}'x < 1$ and therefore $\|\tilde{A}'\|_2^2 < 1$ and $\|\tilde{A}'\|_2 < 1$]. Define $\hat{A} \equiv H\Sigma H'[H\Sigma H']^{-1}$. Then $\hat{A} = \tilde{H}\tilde{A}\tilde{H}'[\tilde{H}\tilde{H}']^{-1}$ where $\tilde{H} \equiv HS$. Then $\hat{A}^k = \tilde{H} \left(\tilde{A}^{k-1} \tilde{H}' [\tilde{H}\tilde{H}']^{-1} \tilde{H} \right)^{k-1} \tilde{A} \tilde{H}' [\tilde{H}\tilde{H}']^{-1}$. Since $\tilde{H}' [\tilde{H}\tilde{H}']^{-1} \tilde{H}$ is an orthogonal projection, $\|\tilde{H}' [\tilde{H}\tilde{H}']^{-1} \tilde{H}\|_2 = 1$. Then $\|\hat{A}^k\|_2 \leq \|\tilde{H}\|_2 \|\tilde{A}\|_2^k \|\tilde{H}' [\tilde{H}\tilde{H}']^{-1}\|_2$. Then (using Gelfand's formula of the spectral radius)

$$\varrho(\hat{A}) = \lim_{k \rightarrow \infty} \left(\|\hat{A}^k\|_2 \right)^{1/k} \leq \lim_{k \rightarrow \infty} \left(\|\tilde{H}\|_2 \|\tilde{A}\|_2^k \|\tilde{H}' [\tilde{H}\tilde{H}']^{-1}\|_2 \right)^{1/k} = \|\tilde{A}\|_2 < 1 \quad (116)$$

If Σ is singular, for any $\epsilon > 0$ there is a regular $\tilde{\Sigma}$ with $\|\tilde{\Sigma} - \Sigma\|_2 < \epsilon$, for which the above argument applies. Then ϵ can be chosen such that $H\tilde{\Sigma}H'$ is regular. The claim then follows from the continuity of the spectral radius. \square

B.3 Solving Discrete-Time Lyapunov Equations

Taking as given the asymptotically stable $n \times n$ -matrix A and the $n \times m$ -matrix B , we want to compute \mathcal{R} , defined as the unique solution of the Lyapunov equation

$$\mathcal{R} = A\mathcal{R}A' + BB' \quad (117)$$

If n is not too large, there are several ways of computing \mathcal{R} . A simple and robust one is the squared Smith iteration (Benner, Quintana-Ort, and Quintana-Ort 2002), also called “doubling algorithm”:

- Set $\mathcal{R} := BB'$.
- Repeat m times the following:

$$\begin{aligned} \mathcal{R} &:= \mathcal{R} + A\mathcal{R}A' \\ A &:= A^2; \end{aligned}$$

After m steps, \mathcal{R} has the same value as obtained from the simple scheme $\mathcal{R}_{j+1} := A\mathcal{R}_jA' + BB'$ after 2^m iterations. If \mathcal{R} is a variance, this would be the variance of the process after 2^m periods, if it was started from a constant value. This algorithm is implemented in the Matlab file `doubling.m`. Numerically even more robust, but computationally more costly

is to compute the Cholesky factor of \mathcal{R} . There is a version of the doubling algorithm that does this (Benner, Quintana-Ort, and Quintana-Ort 2002), implemented in the Matlab file `doubling_chol.m`.

On a PC, this works well as long as n is not much bigger than, say, 1000. For larger n , we can still compute a precise approximation to \mathcal{R} if the numerical rank of \mathcal{R} is not too large. If \mathcal{R} has rank k , it can be written as

$$\mathcal{R} = USU', \quad U'U = I \quad (118)$$

for some symmetric $k \times k$ -matrix S . From (117) we get

$$S = U'\mathcal{R}U = \hat{A}S\hat{A}' + UBB'U' \quad (119)$$

with $\hat{A} \equiv U'AU$. Then (119) can be solved, for example, by the doubling method described above. It remains to find such an U . Define

$$R(N) = [B \quad AB \quad A^2B \quad \dots \quad A^{N-1}B] \quad (120)$$

Then $\mathcal{R} = \lim_{N \rightarrow \infty} R(N)R(N)'$. We therefore want U to span the column space of $R(N)$ for sufficiently large N . This is done by the following algorithm, which is implemented in the Matlab file “`gramianspace.m`”.

1. Set $h := B$ and $R := h$.
2. For $i = 1, \dots, N$, do the following. Set

$$\begin{aligned} h &:= Ah \\ R &:= [R \quad h] \end{aligned}$$

What is crucial is that we store R not in its original form, because R would grow prohibitively large for large N , but rather we maintain and update the QR-decomposition $R = qr$ where q is orthogonal and r is upper triangular. Updating this decomposition when a new column is added is straightforward, because the QR-decomposition of a matrix can be computed by householder reflections on each column of the matrix (Golub and Van Loan 1983, Section 5.2). What is important is that we update q and r only if the new column is not spanned, up to some specified precision, by the earlier columns. In that way, the size of q and r does not grow beyond the numerical rank of R .

3. Compute the SVD $r = u\sigma v'$ and set $U = qu$. Since both q and u are orthogonal, so is U .

If $\text{rank}(\mathcal{R}) = k$, the matrix U obtained is of size $n \times k$. Then we get the $k \times k$ -matrix S from (119). Notice that, in general, we cannot compute $\mathcal{R} = USU'$, because it has dimension $n \times n$ and might not fit into the computer memory. But we can compute, for any H not having too many rows, $H\mathcal{R}H'$ as $(HU)S(U'H)'$.

B.4 Tikhonov regularization

When H has many columns (reduced model with many state variables), the matrix $H\Sigma_x H'$ may be seriously ill-conditioned. This is similar to a multi-collinearity problem in a regression problem of the form $Y = X\beta + u$. A common approach to solve this problem is Tikhonov regularization (called “ridge regression” in econometrics). The idea is to add a small positive constant ϵ on the diagonal of $X'X$. The identification problem is then solved (artificially). Estimates are biased towards zero, but have a smaller RMSE for sufficiently small ϵ (Hoerl and Kennard 1970).

I prefer to work directly on the singular values of X (rather than the eigenvalues of $X'X$). To do this, calculate the SVD of X as $USV' = X$. Then replace diagonal element σ_k in S by $\max(\sigma_k, \epsilon\sigma_1)$ for some small ϵ (I have used $\epsilon = 1e - 14$). Call the resulting matrix \tilde{S} . Then compute $\beta = V\tilde{S}^{-1}U'Y$. To apply this to the OLS problem in Section B.1, set $X = (HL^{-1/2})'$ and $Y = (HAL^{-1/2})'$, where $L^{1/2}$ is a square root of $H\Sigma_x H'$.

C Almost-exact aggregation

C.1 Proof of Proposition 1)

The discussion before Proposition 1) has already shown that the stable solution of (52) gives a stable solution of (62) with $\hat{x}_t = Hx_t$. It remains to be shown that (62) has no other stable solution.

Choose an $(n - k) \times n$ -matrix \tilde{H} such that $H\tilde{H}' = 0$, and $\begin{bmatrix} H \\ \tilde{H} \end{bmatrix}$ is regular. Since H satisfies (32) by construction, such a matrix can always be found. Then

$$\begin{bmatrix} H \\ \tilde{H} \end{bmatrix} \begin{bmatrix} H' & \tilde{H}' \end{bmatrix} = I \quad (121)$$

Next we show that $\tilde{H}T\tilde{H}'$ is asymptotically stable. Define

$$T^* \equiv \begin{bmatrix} H \\ \tilde{H} \end{bmatrix} T \begin{bmatrix} H' & \tilde{H}' \end{bmatrix} = \begin{bmatrix} \hat{T} & 0 \\ \tilde{H}TH' & \tilde{H}T\tilde{H}' \end{bmatrix} \quad (122)$$

From (121) it follows that (122) is a similarity transformation, so that T^* has the same eigenvalues as T . Since T^* is block-triangular, it follows that the eigenvalues of $\tilde{H}T\tilde{H}'$ are eigenvalues of T^* , and therefore of T . This shows that $\tilde{H}T\tilde{H}'$ is asymptotically stable.

Take any stable solution (\hat{x}_t, d_t) of (62). Define $\tilde{x}_t = \tilde{H}x_t$. Premultiplying (52a) by \tilde{H} we get

$$\tilde{x}_t = \tilde{H}T\tilde{H}'\tilde{x}_{t-1} + \tilde{H}TH'\hat{x}_{t-1} + \tilde{H}Dd_t + \tilde{H}F\epsilon_t \quad (123)$$

This together with the stability of $\tilde{H}T\tilde{H}'$ shows that \tilde{x}_t is stable. From (121), one can recover x_t from \hat{x}_t and \tilde{x}_t by $x_t = H'\hat{x}_t + \tilde{H}'\tilde{x}_t$. The stability of \hat{x} and of \tilde{x} then implies

the stability of x_t . Furthermore, (x_t, d_t) satisfies (52), because it satisfies (123) and (62), and $\begin{bmatrix} H \\ \tilde{H} \end{bmatrix}$ has full rank.

Therefore, every stable solution \hat{x} of (62) generates a (different) stable solution of (52). Since the latter system has a unique stable solution by assumption, so does (62).

C.2 The Numerical Rank of the Observability Matrix

To understand better the conditions for almost-exact aggregation, consider the special case where C has only one row, and where the matrix T has distinct eigenvalues, so that it can be diagonalized and written as

$$T = G\Lambda G^{-1} \quad (124)$$

Here G is the matrix of eigenvectors of T and Λ is the diagonal matrix holdings the eigenvalues λ_i . Then the observability matrix is given by

$$Q = \begin{bmatrix} CG\Lambda^0 G^{-1} \\ CG\Lambda^1 G^{-1} \\ \vdots \\ CG\Lambda^{n-1} G^{-1} \end{bmatrix} = \begin{bmatrix} CG \\ CG\Lambda \\ \vdots \\ CG\Lambda^{n-1} \end{bmatrix} G^{-1} \quad (125)$$

Since G^{-1} has full rank, the rank of Q equals the rank of

$$\tilde{Q} = \begin{bmatrix} CG \\ CG\Lambda \\ \vdots \\ CG\Lambda^{n-1} \end{bmatrix} = \begin{bmatrix} \tilde{c}_1 & \tilde{c}_2 & \dots & \tilde{c}_{n-1} & \tilde{c}_n \\ \tilde{c}_1\lambda_1 & \tilde{c}_2\lambda_2 & \dots & \tilde{c}_{n-1}\lambda_{n-1} & \tilde{c}_n\lambda_n \\ \tilde{c}_1\lambda_1^2 & \tilde{c}_2\lambda_2^2 & \dots & \tilde{c}_{n-1}\lambda_{n-1}^2 & \tilde{c}_n\lambda_n^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \tilde{c}_1\lambda_1^{n-1} & \tilde{c}_2\lambda_2^{n-1} & \dots & \tilde{c}_{n-1}\lambda_{n-1}^{n-1} & \tilde{c}_n\lambda_n^{n-1} \end{bmatrix} \quad (126)$$

where \tilde{c}_i is the i -th element of the column vector CG . Since the λ_i are distinct by assumption, the columns of the matrix

$$\begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_{n-1} & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \dots & \lambda_{n-1}^2 & \lambda_n^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \dots & \lambda_{n-1}^{n-1} & \lambda_n^{n-1} \end{bmatrix} \quad (127)$$

are independent. Then (126) has as many independent columns as there are non-zero elements \tilde{c}_i . The Popov-Belevitch-Hautus test of Section 4.5 is a consequence of this.

From those considerations, there are at least two reasons why the numerical rank of Q may be smaller than n . First, two elements λ_i and λ_{i+1} may be equal up to machine precision. And second, if an λ_i is sufficiently smaller in absolute value than λ_1 , then the powers λ_i^k become insignificant for high enough k . For example, assume that $\lambda_i \leq 0.5\lambda_1$. Then $\lambda_i^k \leq 10^{-16}\lambda_1^k$ for all $k > 53$. That means that from the 55-th row of Q onwards, the entries in columns i and higher are numerically negligible compared to the entry in the first column.

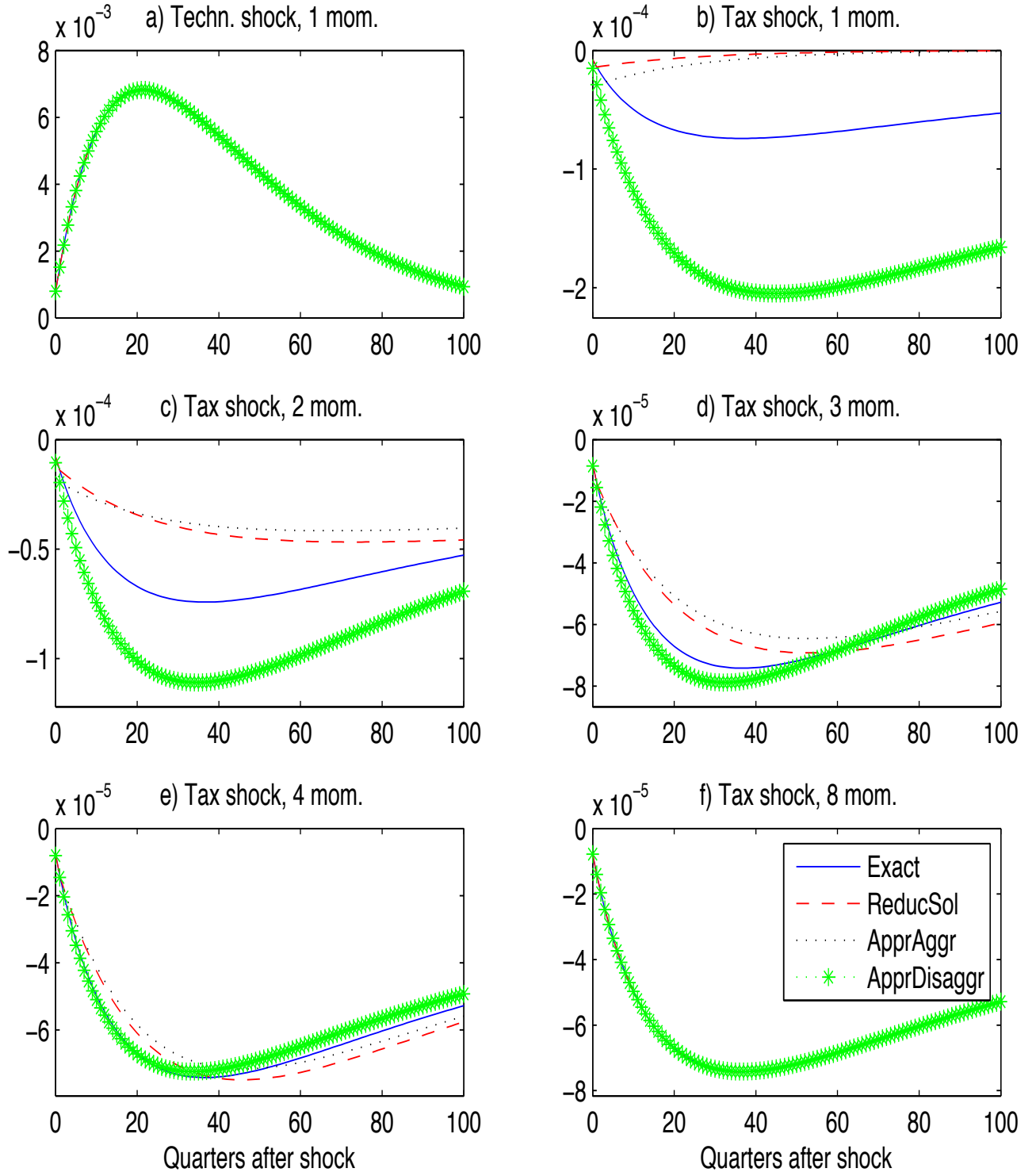


Figure 1: Impulse response functions for K , stochastic growth model, $n_P = 1$

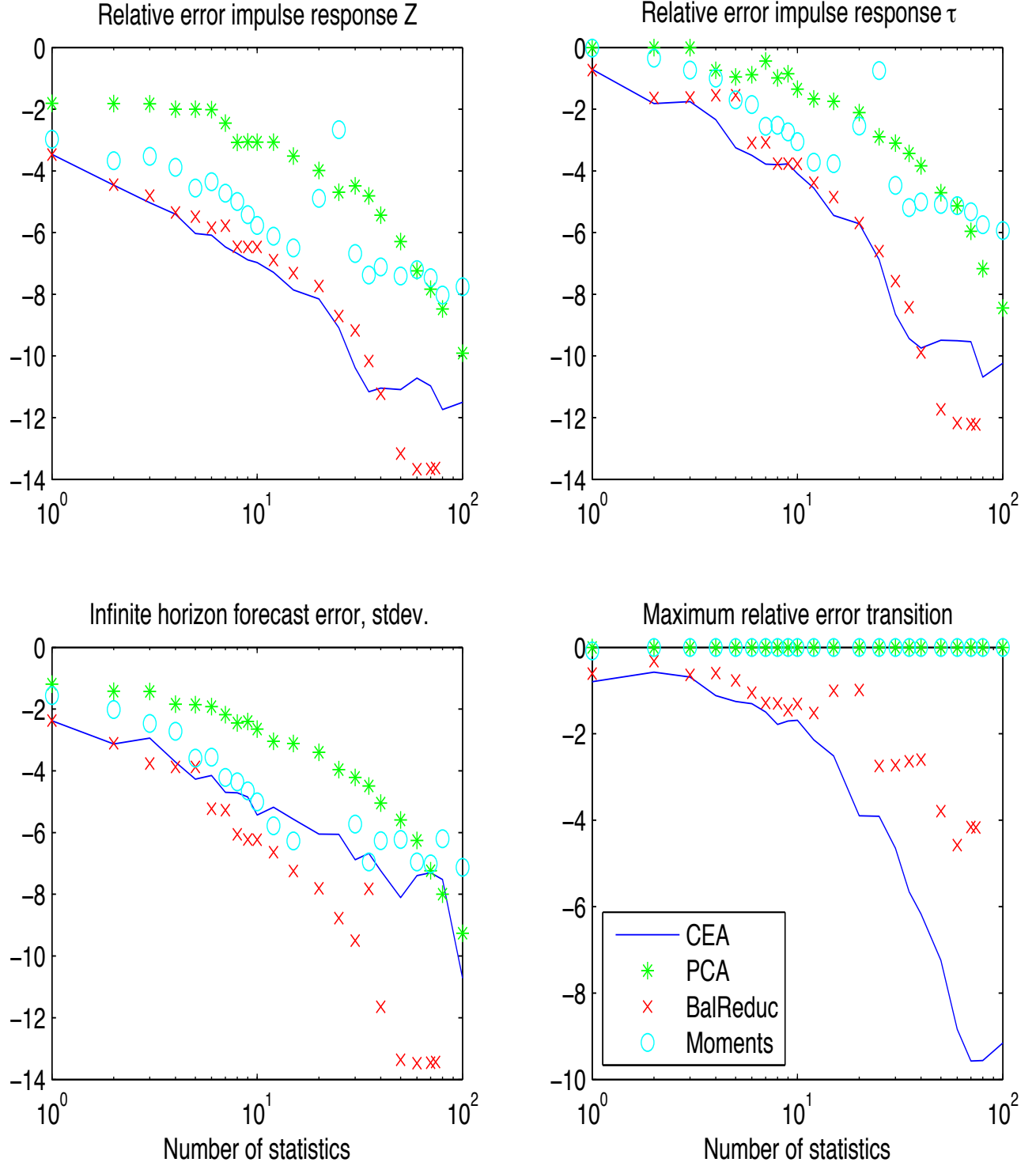


Figure 2: log10 of approximation errors in K , stochastic growth model with model reduction, $n_P = 1$

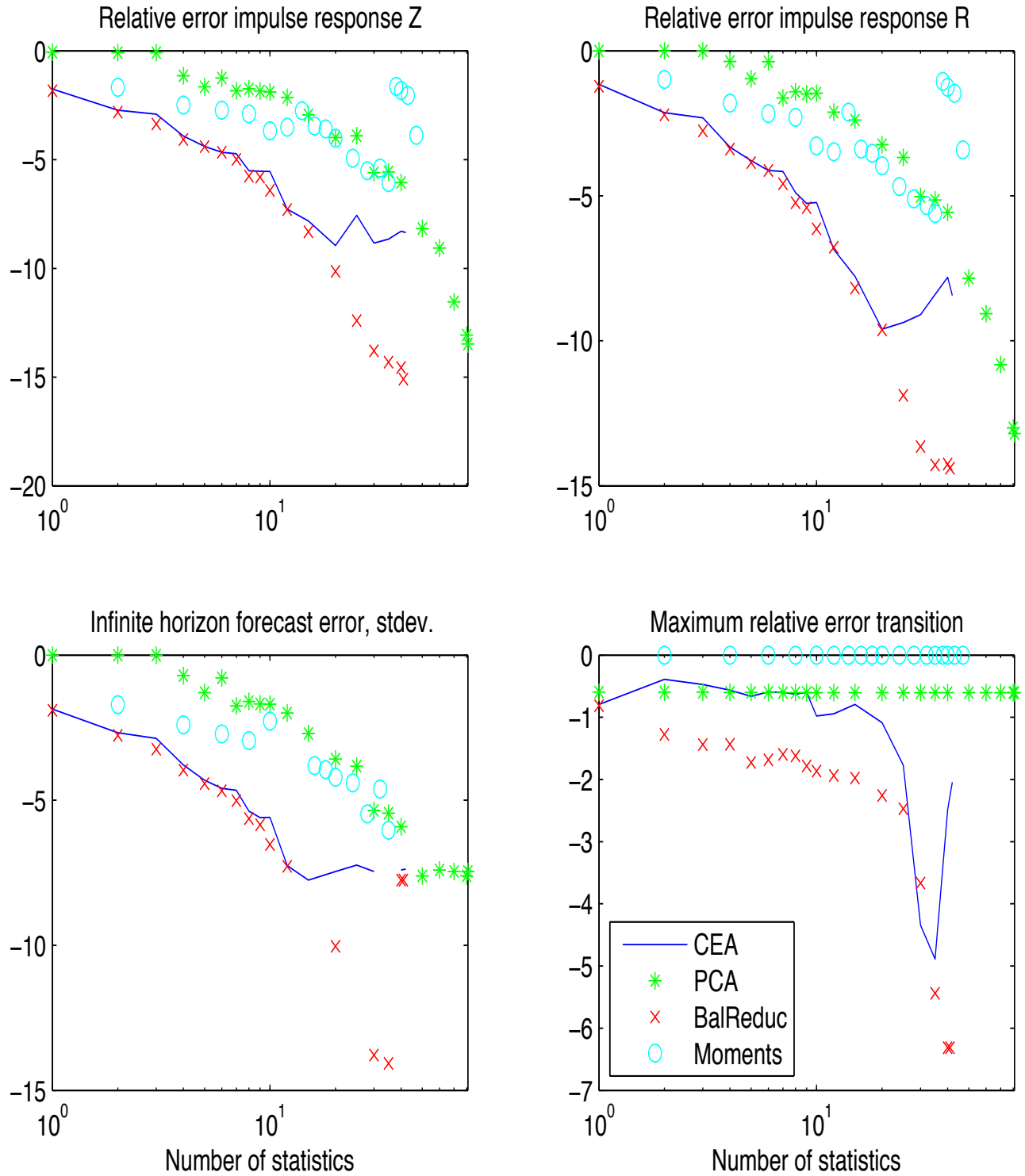


Figure 3: log10 of approximation errors in R , state-dependent pricing model with model reduction

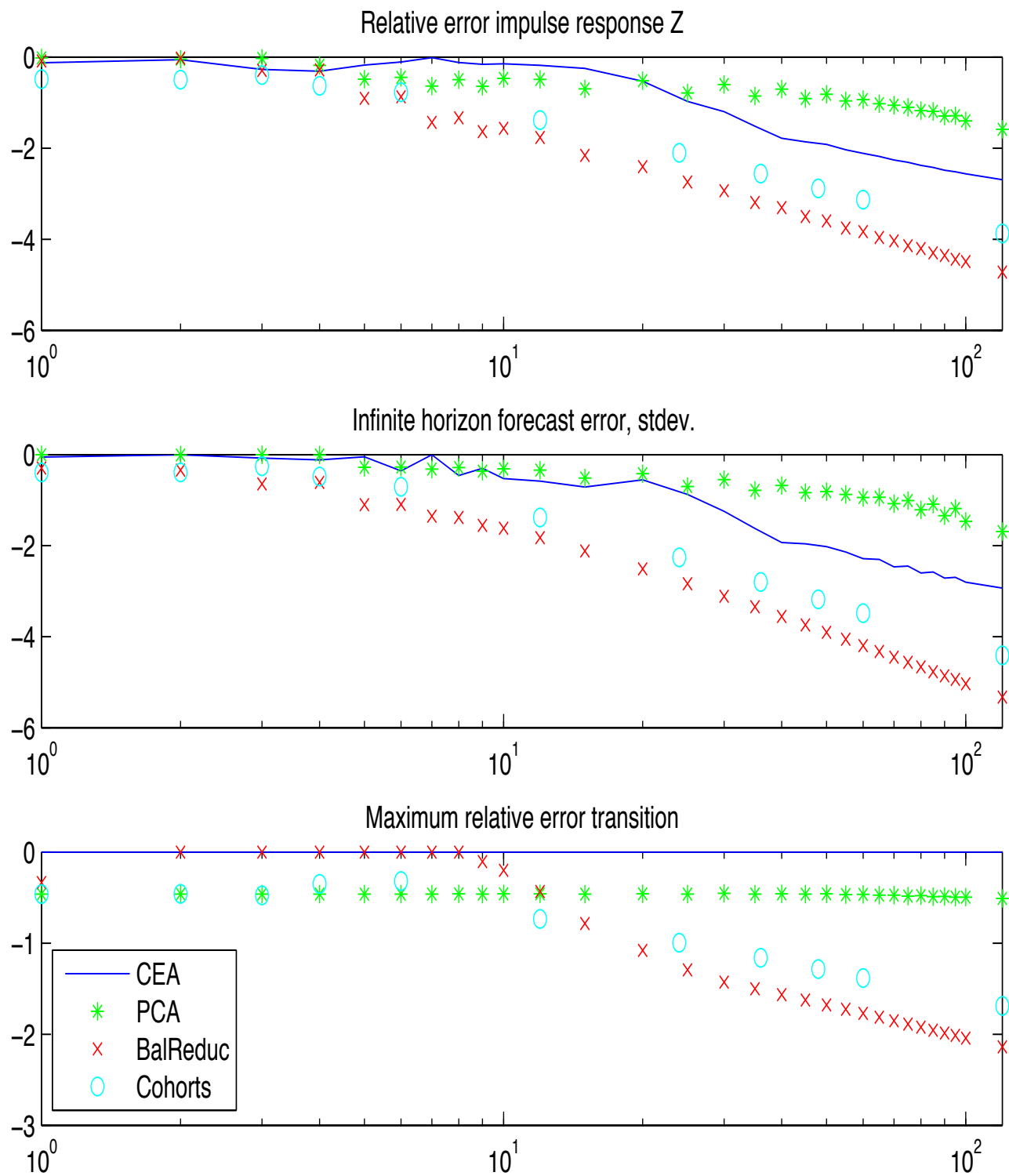


Figure 4: log10 of approximation errors in K , OLG model with model reduction

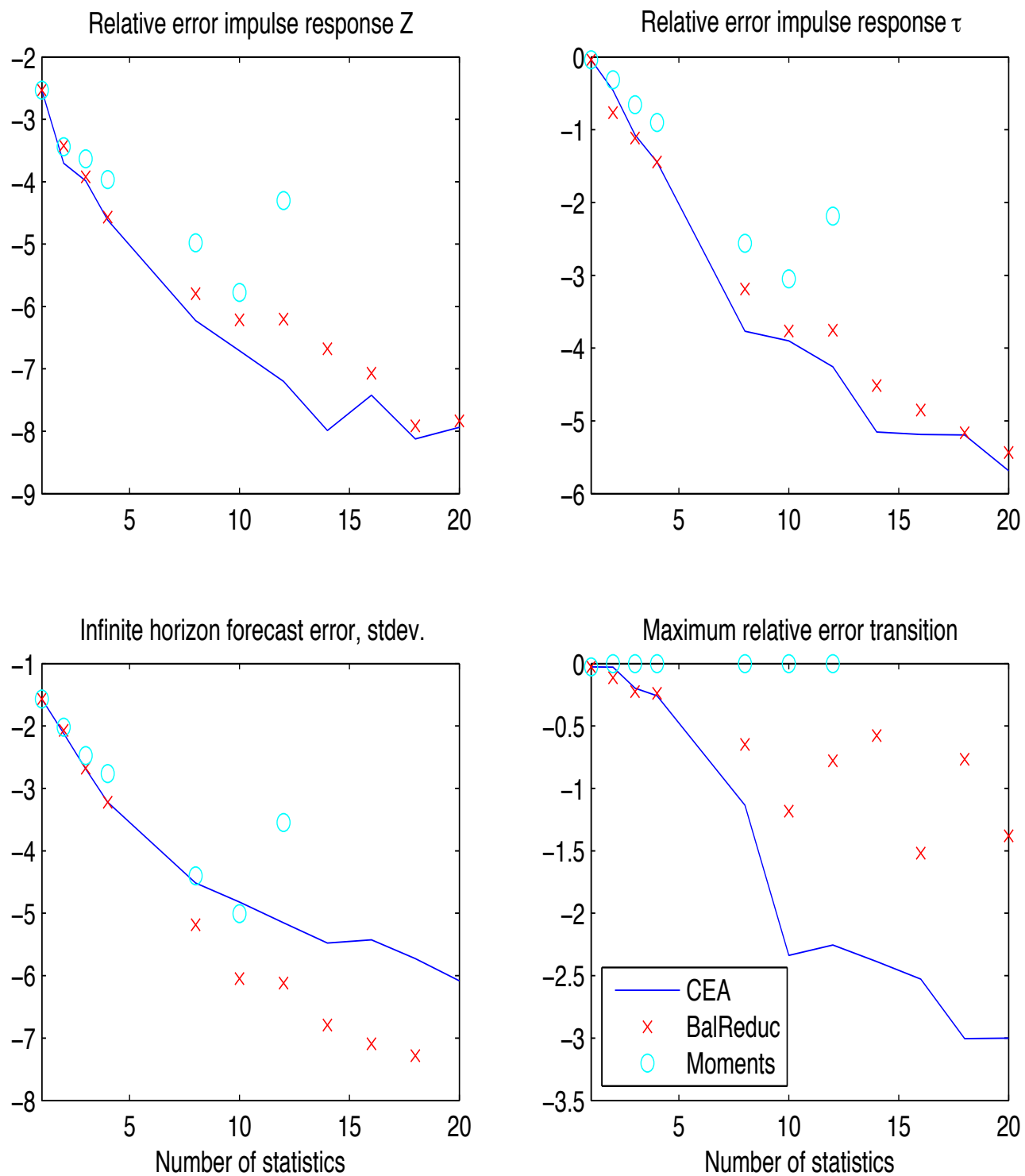


Figure 5: \log_{10} of approximation errors in K , stochastic growth model with approximate aggregation

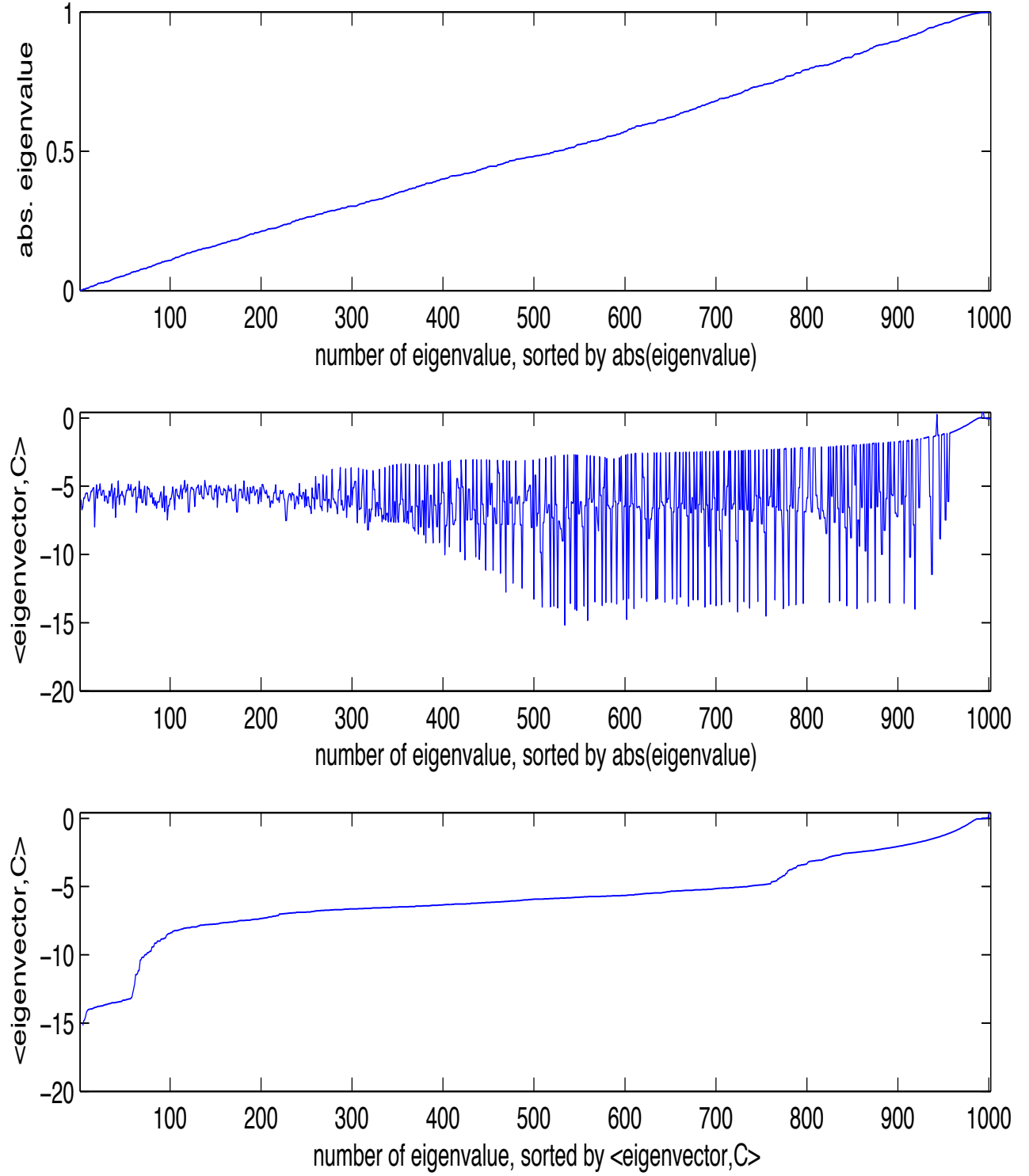


Figure 6: Popov-Belevitch-Hautus test for exact aggregation, growth model, $n_P = 1$

References

- Antoulas, A. C. (1999). Approximation of linear dynamical systems. In J. Webster (Ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*, volume 11. Wiley.
- Antoulas, A. C. (2005). *Approximation of Large-Scale Dynamical Systems*. SIAM.
- Antoulas, A. C., S. Y. Zhou, and Y. Zhou (2002). On the decay rate of hankel singular values and related issues. *Systems and Control Letters* 46, 323–342.
- Benner, P., E. Quintana-Ort, and G. Quintana-Ort (2002). Numerical solution of discrete stable linear matrix equations on multicomputers. *Parallel Algorithms and Applications* 17(1), 127–146.
- 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.
- Costain, J. and A. Nákov (2008). Dynamics of the price distribution in a general model of state-dependent pricing. Banco de España WP 831.
- 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. (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.
- Den Haan, W. J. and P. Rendahl (2009). Solving the incomplete markets model with aggregate uncertainty using explicit aggregation. Manuscript.
- Golub, G. H. and C. F. Van Loan (1983). *Matrix Computations*. Baltimore: Johns Hopkins Press.
- Güvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics* 12(1), 58–79.
- Haefke, C. and M. Reiter (2009). Endogenous labor market participation and the business cycle. Manuscript.
- Hoerl, A. and R. Kennard (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12, 55–67.
- Judd, K. L. (1998). *Numerical Methods in Economics*. Cambridge and London: MIT Press.
- Krusell, P. and A. A. Smith (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy* 106(5), 867–96.
- Krusell, P. and A. 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.

- 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.
- Midrigan, V. (2009). Menu costs, multi-product firms, and aggregate fluctuations. 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). Nonlinear approximate aggregation in heterogeneous agent models. Draft.
- Reiter, M. (2009b). Solving heterogeneous agent models by projection and perturbation. *Journal of Economic Dynamics and Control* 33(3), 649–665.
- Reiter, M., T. Sveen, and L. Weinke (2009). Lumpy investment and state-dependent pricing in general equilibrium. Norges Bank Working Paper 2009/05.
- Sims, C. A. (2001). Solving linear rational expectations models. *Computational Economics* 20(1-2), 1–20.
- Trefethen, L. N. and D. Bau III (1997). *Numerical Linear Algebra*. SIAM.
- Williams, R. L. and D. A. Lawrence (2007). *Linear state-space control systems*. John Wiley and Sons.