State Reduction and Second-order Perturbations of Heterogeneous Agent Models

Michael Reiter, IHS Vienna and NYU Abu Dhabi

Numerical Methods in Macroeconomics, Goethe University Frankfurt, 23.10.2024

Table of Content

Outline

Linearization in aggregate states: basic idea

1. Stochastic household problem solved non-linearly as a function of individual state
   - Example: consumption as a function of cash on hand, represented as a cubic spline on 100 knot points
   - 100 variables: consumption at each knot point
   - 100 equations: Euler equation on each knot point
2. High-dimensional representation of cross-sectional distribution of states
   - Distribution of capital as a histogram with 1000 bins
   - 1000 variables: fraction of HH in each bin
   - 1000 equations: Kolmogorov equation for each bin
3. Linearization around StSt without aggregate shocks: 100×1000 matrix: consumption at each knot point as a function of the histogram
4. Solved by standard techniques for linearized model; Blanchard/Kahn condition: 100 eigenvalues bigger than 1.

Methods for HA models

- Krusell and Smith (1998): small aggregate state, nonlinear, consistency in OLS sense
- Reiter (2009): large aggregate state, linear in aggregates
  Further developments:
  - loss-less model reduction:Reiter (2010a)
  - continuous time: Ahn, Kaplan, Moll, Winberry, and Wolf (2018)
  - linear superposition of impulse response functions: Boppart, Krusell, and Mitman (2018); efficient implementation: Auclert, Bardóczy, Rognlie, and Straub (2021)
- Linear and quadratic perturbation in continuous time: Bilal (2023),
- General method perturbation in discrete time: Bhandari, Bourany, Evans, and Golosov (2023)
- Old and new alternatives: (Den Haan 1997; Reiter 2010b; Mertens and Judd 2017; Winberry 2018; Bhandari, Evans, Golosov, and Sargent 2021; Grand and Ragot 2022)

New method: second-order perturbation

Advantages:

- higher accuracy
- precautionary behavior w.r.t. aggregate shocks, risk premia
- approximation of welfare

Challenges:

1. state reduction
2. smoothness
3. efficient computation

Discrete versus continuous time

1. Continuous time: Bilal (2023)
   - First differentiate, then discretize
   - Object to approximate: gradient of value function w.r.t. distribution
   - High-dimensional, sparse
   - Relatively fast to compute solution
   - Relatively slow to simulate model
2. Discrete time: this paper
   - First discretize, then differentiate
   - Reduction to low-dimensional state vector
   - Some overhead in determining state vector
   - Very fast to simulate aggregates (including cross-sectional statistics of distribution)
   - Low dimension useful for estimation

General Model

$$
\begin{aligned}
\text{Exogenous dyn.} \quad & z_t = \mathcal{Z}(z_{t-1}, \varepsilon_t) & \text{(1a)} \\
\text{Endogenous dyn.:} \quad & q_t = \mathcal{Q}(S_{t-1}, \varepsilon_t, a_t) & \text{(1b)} \\
\text{Aggr. equilibrium:} \quad & 0 = \mathcal{A}(S_{t-1}, \varepsilon_t, a_t, \mathrm{E}_t\, a_{t+1}, \mathrm{E}_t\, V_{t+1}) & \text{(1c)} \\
\text{Bellman equ.:} \quad & V_t = \mathcal{F}(a_t, \mathrm{E}_t\, a_{t+1}, \mathrm{E}_t\, V_{t+1}) & \text{(1d)} \\
\text{Distribution dyn.:} \quad & D_t = \Pi(a_t, \mathrm{E}_t\, a_{t+1}, \mathrm{E}_t\, V_{t+1})\, D_{t-1} & \text{(1e)}
\end{aligned}
$$

where $S_{t-1} \equiv (D_{t-1}, q_{t-1}, z_{t-1})$.
Compact notation:

$$\mathcal{M}(S_{t-1}, \varepsilon_t, \Theta_t, \mathrm{E}_t\, \Theta_{t+1}) = 0 \qquad (2)$$

Implicit, optimal policy: $P_t = \mathcal{P}(S_{t-1}, \varepsilon_t, a_t, \mathrm{E}_t(V_{t+1}))$

Outline

Loss-less state reduction in linearized model

Intuition: Assume that we want to predict y_{t+i} where

$$y_t = CS_t, \qquad S_t = AS_{t-1} + B\varepsilon_t, \qquad \mathrm{E}_{t-1}\,\varepsilon_t = 0 \tag{3}$$

Then

$$\mathrm{E}_t\, y_{t+i} = CA^i S_t \tag{4}$$

Conditional expectations of future y's are QS_t with

$$Q \equiv \begin{bmatrix} C \\ CA \\ \dots \\ CA^{n-1} \end{bmatrix} \tag{5}$$

SVD of Q:

$$Q = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} = U_1 S V_1', \qquad S \text{ diagonal} \tag{6}$$

U_1 spans the linear combination of the states that contains the relevant information about the expected y's.

I call it "Conditional Expectations Approach (CEA)" (Reiter 2010a).

Reduction to few states

Idea: use first columns of U_1 for state reduction.

- How many are needed for good approximation?
- Are these states useful for nonlinear approximation? For perturbation, yes!

States and proxy distributions

- Replace the distribution D_t by a vector of "moments" m_t, linear in D_t:

$$m_t = HD_t \tag{7}$$

- Use the idea of a "proxy distribution" D_t^{pd} (Reiter 2010b). In deviations from the steady state:

$$D_t^{pd} = D^* + \left[\Sigma_D H'(H\Sigma_D H')^{-1}\right](m_t - m_t^*) \tag{8}$$

  D_t^{pd} is the expectation of D_t conditional on $HD_t = m_t$.
- Replace distribution dynamics by

$$m_t = H\Pi(s_{t-1}, \varepsilon_t, a_t, \mathrm{E}_t\, V_{t+1})\Phi^{pd} m_{t-1} \tag{9}$$

Selecting state variables

- Minimal state vector:
  - Aggregate capital (necessary for factor prices)
  - Exogenous processes: lagged values and current shocks
- Different ways to add more states:
  - MOM: additional moments of the cross-sectional distribution (beyond the first moment)
  - COH: capital owned by adjacent cohorts (OLG model)
  - PCA: principal component analysis;
    components of distribution that fluctuate strongly
  - CEA: conditional-expectations approach (Reiter 2010a): first elements in SVD of space that spans conditional expectations of all future variables in linearized model

In all cases, orthogonalize the states so that covariance matrix in linear model is diagonal.

Outline

Smoothness

- Perturbation builds on smoothness of involved functions (ideally analytic)
- Discretized methods (histograms etc.) non-smooth
- To make problem of agent smoother:
  - approximate policy or value function by cubic spline (not piecewise linear)
  - make transition probabilities smooth function of decisions
  - include smooth i.i.d. shock ξ (one for each endogenous continuous state) into the model (similar to Childers (2018))
  - at each state, identify critical values of ξ where regime changes
    - inequality constraint (on state or policy) starts binding
    - discrete choice switches

Distribution dynamics

- set of grid points κ_i, $i = 1, \ldots, n_k$.
- Interval boundaries $\bar{\kappa}_j = (\kappa_j + \kappa_{j+1})/2$ for $j = 1, \ldots, n_k - 1$.
- "saving function" $K'(\kappa_i, \xi; \Omega)$
- Transition probabilities between grid points i and j:

$$\Pi_{i,j}(x) = prob[K'(\kappa_i, \xi; \Omega) \in (\bar{\kappa}_{j-1}, \bar{\kappa}_j)] \tag{10}$$

$$= cdf(\Xi(\bar{\kappa}_j, \kappa_i; \Omega)) - cdf(\Xi(\underline{\kappa}_j, \kappa_i; \Omega)) \tag{11}$$

with

$$\Xi(k, \kappa_i; \Omega) \equiv \begin{cases} \underline{\xi} & \text{if } k \leq K'(\kappa_i, \underline{\xi}; \Omega) \\ \bar{\xi} & \text{if } k \geq K'(\kappa_i, \bar{\xi}; \Omega) \\ \xi \text{ s.t. } K'(\kappa_i, \xi; \Omega) = k & \text{else} \end{cases} \tag{12}$$

Threshold points

We consider three types of threshold points:

1. The constraint on the continuous state starts binding.
2. The constraint on some other choice variable starts binding.
3. The optimal discrete choice switches.

They will be determined by implicit differentiation of a system of three equations.

Handling regime changes

Task: differentiate the integral of a function w.r.t. parameters α and β that has a kink or discountinuity at a threshold point $\hat{\xi}$:

$$\frac{\partial^2 \int g(\xi)\phi(\xi)\,d\xi}{\partial\alpha\partial\beta} = \int_{-\bar{\xi}}^{\hat{\xi}} \frac{\partial^2 g_1(\xi)}{\partial\alpha\partial\beta}\phi(\xi)\,d\xi + \int_{\hat{\xi}}^{\bar{\xi}} \frac{\partial^2 g_2(\xi)}{\partial\alpha\partial\beta}\phi(\xi)\,d\xi$$
$$+ \frac{\partial\hat{\xi}}{\partial\alpha}\frac{\partial\hat{\xi}}{\partial\beta}\cdot(g_1'(\hat{\xi}) - g_2'(\hat{\xi}))$$
$$+ \frac{\partial\hat{\xi}}{\partial\beta}\cdot\left(\frac{\partial g_1(\hat{\xi})}{\partial\alpha} - \frac{\partial g_2(\hat{\xi})}{\partial\alpha}\right) + \frac{\partial\hat{\xi}}{\partial\alpha}\cdot\left(\frac{\partial g_1(\hat{\xi})}{\partial\beta} - \frac{\partial g_2(\hat{\xi})}{\partial\beta}\right)$$
$$+ \frac{\partial^2\hat{\xi}}{\partial\alpha\partial\beta}\cdot(g_1(\hat{\xi}) - g_2(\hat{\xi})) \quad (13)$$

Conclusion: you have to identify and differentiate the threshold point as well.

Outline

Notation

- Greek letters α and β run over the elements of the time-t state vector $\tilde{x}_t = (\tilde{s}_{t-1}, \varepsilon_t)$.
- Greek letters γ and δ run over the elements of the predetermined states $\tilde{s}_t$ at the end of period t.
- Greek letters λ and μ run over the future shocks ε_{t+1}.
- Roman letters i and j, run over all time-t variables in $\tilde{\Theta}_t = [\tilde{s}_t; \tilde{y}_t; \tilde{V}_t]$
- The uppercase Roman letters I and J run over the elements of future variables $\tilde{\Theta}_{t+1}$.

Quadratic perturbation in states

$$R^k_{\alpha\beta} + \mathcal{M}^k_i H^i_{\alpha\beta} + \mathcal{M}^k_l \left[G^l_\gamma H^\gamma_{\alpha\beta} + \frac{1}{2} \hat{H}^l_{\gamma\delta} (G^\gamma_\beta G^\delta_\alpha + G^\gamma_\alpha G^\delta_\beta) \right] = 0$$

where

- $R^k_{\alpha\beta}$ only contains linear terms
- G, H: linear and quadratic coefficients
- $\mathcal{M}^k_\alpha \equiv \frac{\partial \mathcal{M}^k}{\partial x^\alpha_t}$

Fast iteration: lagged update of ALM

1. Set $\hat{H}^i_{\alpha\beta} = 0$ for all i, α and β.
2. Given $\hat{H}$, and separately for each pair (α, β), solve for $H^i_{\alpha\beta}$:

$$R^k_{\alpha\beta} + \mathcal{M}^k_i H^i_{\alpha\beta} + \mathcal{M}^k_l \left[G^l_\gamma \hat{H}^\gamma_{\alpha\beta} + \hat{H}^l_{\alpha\beta} (G^\alpha_\beta G^\beta_\alpha + G^\alpha_\alpha G^\beta_\beta)/2 \right] = 0$$

3. Set $\hat{H}^i_{\alpha\beta} = H^i_{\alpha\beta}$ for all i, α and β.
4. Iterate 2. and 3. until convergence.

Effect of aggregate uncertainty

Precaution w.r.t. aggregate shocks $H^i_{\sigma\sigma}$:

$$\mathcal{M}^k_i H^i_{\sigma\sigma} + \mathcal{M}^k_l \left[G^l_\gamma H^\gamma_{\sigma\sigma} + H^l_{\lambda\mu} \Sigma_{\lambda\mu} + H^l_{\sigma\sigma} \right] = 0 \qquad (14)$$

Outline

Models: technology

- Cobb-Douglas production function:

$$Y_t = Z_t K_{t-1}^{\alpha} L_t^{1-\alpha} \tag{15}$$

- Total factor productivity:

$$Z_t = 1 + \rho_z \cdot (Z_{t-1} - 1) + \epsilon_{z,t}, \qquad \epsilon_{z,t} \sim (0, \sigma_z) \tag{16}$$

- Capital dynamics:

$$K_t = I_t + (1 - \delta_t) K_{t-1} \tag{17}$$

- Competitive factor markets
- Aggregate resource constraint $Y_t = C_t + I_t$

Infinite horizon models

1. Divisible labor, utility $\log(c) + \eta \log(1 - L)$, $L \geq 0$.
2. Indivisible-labor: almost identical to Chang and Kim (2007); non-convex consumer problem
3. Idiosyncratic household productivity follows Markov chain

A model of stochastic aging (OLG)

- Worker households: 12 age groups, groups 9-12 retired, groups 1-8 work,
- idiosyncratic labor productivity shock
- HH moves to higher age group with positive probability
- Groups 11 and 12 have positive probability of dying, are followed by young agent
- Accidental bequests
- 3 aggregate shock processes: TFP, age-specific technology, depreciation

Outline

Accuracy checks

1. Deterministic solution (without precautionary effect): compare IR function of quadratic approximation (w.r.t. states) to nonlinear perfect foresight solution (response to one-time shock).
2. Stochastic solution: ”Euler equation errors” along simulated paths (consistency check, not a comparison to exact solution; in the paper).

No precise error estimate of precautionary terms, but they are simple function of quadratic terms w.r.t. states.

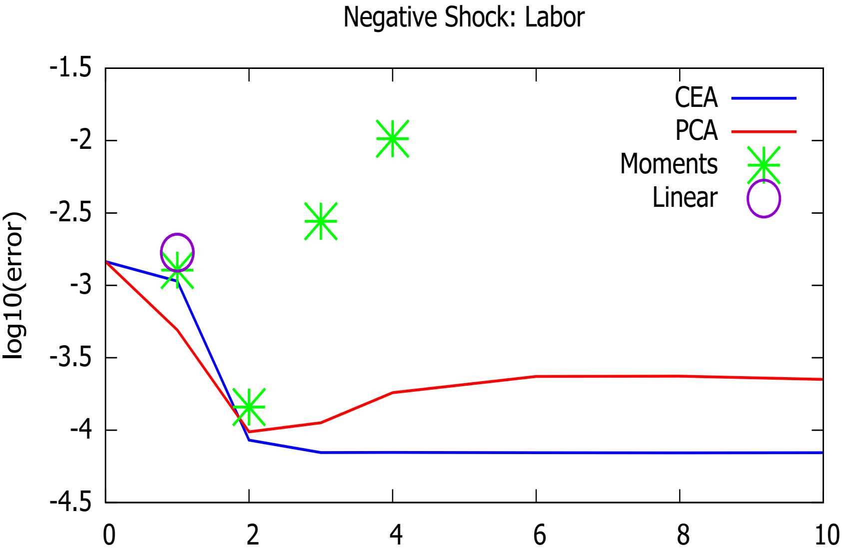
Approximation errors from aggregation and linearization

- Aggregation error is linear in shock size (standard deviation)
- Linearization error is quadratic in shock size
- Error from quadratic approximation is third order in shock size, etc.

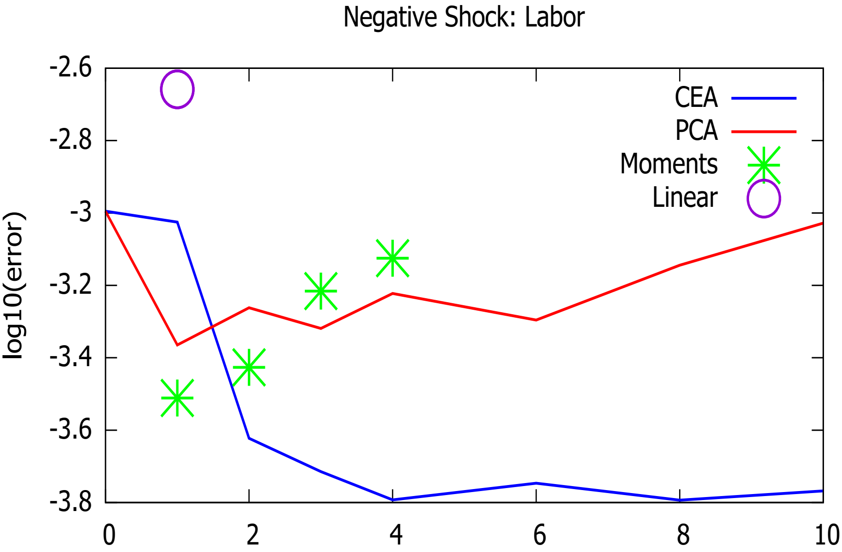
Accuracy Divisible Labor Model

σ	Type	#s	Labor Neg	Labor NegPos	Investment Neg	Investment NegPos	Ca Neg
1			(3.7e-3)	(1.7e-5)	(5.8e-4)	(5.9e-7)	(4.9e-3)
	LIN		1.69e-5	1.67e-5	5.85e-7	5.88e-7	6.88e-6
	-	0	1.28e-4	2.60e-6	4.93e-6	1.58e-7	2.66e-5
	MOM	2	7.87e-6	2.96e-6	2.19e-7	1.47e-7	3.74e-6
	PCA	2	6.47e-6	2.61e-6	4.67e-7	1.47e-7	6.14e-6
	CEA	2	2.35e-6	1.75e-6	2.11e-7	1.50e-7	2.85e-6
10			(3.9e-2)	(1.6e-3)	(5.8e-3)	(4.7e-5)	(4.8e-2)
	LIN		1.68e-3	1.62e-3	4.71e-5	4.67e-5	6.19e-4
	-	0	1.46e-3	1.35e-4	5.04e-5	6.20e-6	2.81e-4
	MOM	2	1.44e-4	1.43e-4	4.42e-6	4.75e-6	7.48e-5
	PCA	2	9.73e-5	9.54e-5	7.58e-6	3.28e-6	9.72e-5
	CEA	2	8.52e-5	1.49e-5	1.74e-6	5.57e-7	9.76e-6

Divisible-labor model, one-time shock 10 stdevs



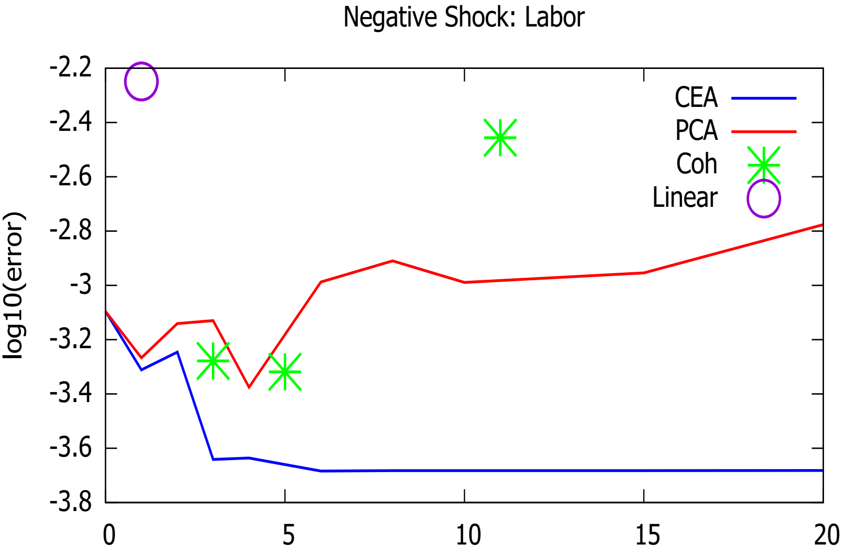
Indivisible-labor model, one-time shock 10 stdevs



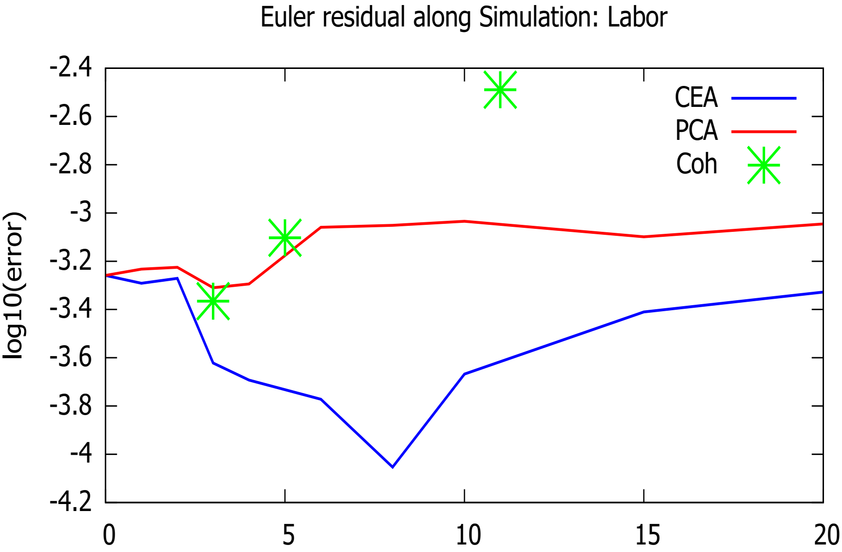
Accuracy OLG Model

σ	Type	#s	Labor Neg	Labor NegPos	Investment Neg	Investment NegPos	Ca[p] Neg
3			(1.3e-2)	(5.6e-3)	(1.3e-2)	(2.9e-4)	(2.1e-2)
	LIN		5.64e-3	5.57e-3	2.91e-4	2.89e-4	1.45e-3
	-	0	8.03e-4	3.26e-4	9.31e-5	1.44e-5	8.61e-4
	COH	5	4.80e-4	3.60e-4	2.23e-5	1.09e-5	1.14e-4
	PCA	4	4.21e-4	3.80e-4	6.37e-5	5.39e-5	1.41e-3
	CEA	4	2.31e-4	1.40e-4	1.53e-5	1.13e-5	6.52e-5
10			(1.4e-2)	(6.5e-2)	(4.6e-2)	(3.2e-3)	(7.7e-2)
	LIN		6.55e-2	6.49e-2	3.25e-3	3.24e-3	1.61e-2
	-	0	5.72e-3	4.87e-3	2.84e-4	2.58e-4	3.02e-3
	COH	5	4.60e-3	3.96e-3	1.39e-4	1.73e-4	1.39e-3
	PCA	4	3.70e-3	2.42e-3	7.19e-4	6.81e-4	1.74e-2
	CEA	4	5.14e-3	4.50e-3	1.43e-4	2.23e-4	2.90e-4

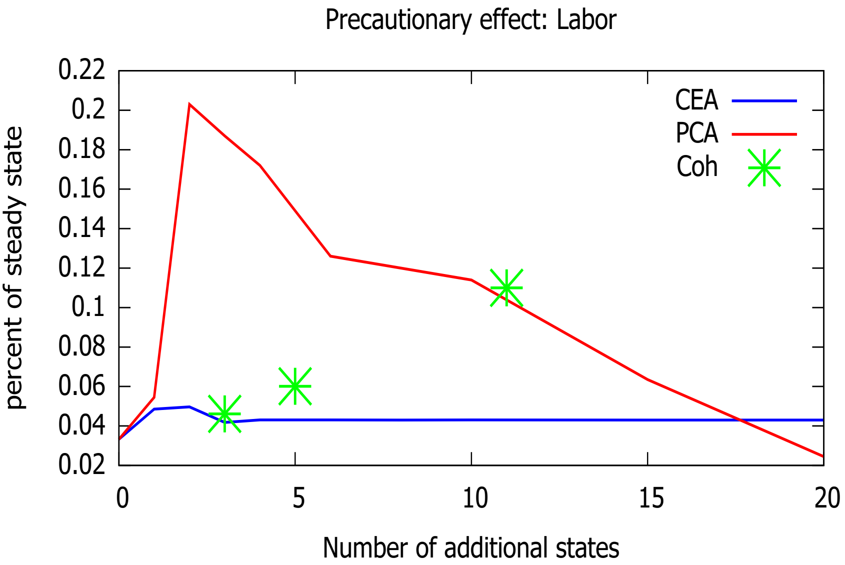
OLG model, one-time shock 3 stdevs



OLG model, residual along simulation path



OLG model, precautionary effect



Computation times

- Written in Julia
- Computations done on a Windows desktop with an AMD Ryzen 7-3700X 8-core CPU.
- Timing somewhat stochastic (garbage collection)

	Divisible Labor		Indiv. Labor		OLG	
Solution	#states	seconds	#st.	sec.	#st.	sec.
SteadyState	1401	1.6	8501	17.7	16803	25.0
Linear	1401	3.0	8501	35.8	16803	57.2
Quadratic	5	2.7	5	29.5	7	85.6
Quadratic	10	4.7	10	47.2	10	113.3
Quadratic	15	6.9	15	70.0	15	141.5
Quadratic	20	10.3	20	102.3	20	188.5

Conclusions

- Second-order perturbation fast to compute
- Accuracy at least one order of magnitude better than in linear solution
- High accuracy even for large shocks
- Few additional states are sufficient, in all models considered; adding more states can be detrimental
- CEA approach works best in all example models.
- Precautionary effect w.r.t. aggregate variables of similar magnitude as in RA models.

Summary

- State selection by CEA works well for second-order perturbation.
- Improvement in accuracy over linear solution by at least an order of magnitude.
- In example models (no strong aggregate nonlinearity) accuracy high even for very large shocks.

Outline

Ideas

Neural networks very flexible, but gradient descent can be very slow, therefore

- Set up neural network as a generalization of linear solution with state reduction
- Start from perturbation solution; don't learn everything from scratch
- Combine with nonlinear transformation of variables (Judd 2002)
- Solve for value and policy functions by backward iteration (Reiter 2010b).
- Use stochastic gradient descent for other stuff, (adding further states if necessary, nonlinear least squares etc.)

Linear solution as network

$$
\begin{aligned}
S &\to s && \text{optimal state aggregation in linear model} \\
s &\to (y, v) && \text{linearized solution} \\
v &\to V && \text{optimal value function aggregation in linear model} \\
(y, V) &\to P && \text{solve for optimal policy point-wise} \\
P &\to \Pi && \text{policy determines state transition matrix} \\
S' &= \Pi(S)
\end{aligned}
$$

Enlarging the network, example

$S \to s$	optimal state aggregation in linear model
$s \to s^2$	quadratic functions of states
$s \to z = BC(s)$	nonlinear transformations (Box-Cox)
$z \to Sig(z)$	map into bounded range (Sigmoid)
$s \to (y, v)$	linearized solution
$v \to V$	value function aggregation in linear model
$(y, V) \to P$	solve for optimal policy point-wise
$P \to \Pi$	policy determines state transition matrix
$S' = \Pi(S)$	

Important: scale nonlinear transformations so as to leave derivatives at steady state unchanged.

Outline of solution method: ”shallow learning”

- Preserve information from perturbation solution
- Temporarily fix grid by simulation, solve value function by backward induction
- Find useful nonlinear transformations
- Approximate value function pointwise by nonlinear LS
- Use gradient descent on some subproblems, not on the one big problem

Ahn, S., G. Kaplan, B. Moll, T. Winberry, and C. Wolf (2018).
When Inequality Matters for Macro and Macro Matters for Inequality.
NBER Macroeconomics Annual 32(1), 1–75.

Auclert, A., B. Bardóczy, M. Rognlie, and L. Straub (2021, September).
Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models.
Econometrica 89(5), 2375–2408.

Bhandari, A., T. Bourany, D. Evans, and M. Golosov (2023, September).
A perturbational approach for approximating heterogeneous agent models.
NBER Working Papers 31744, National Bureau of Economic Research, Inc.

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2021).

Inequality, Business Cycles, and Monetary-Fiscal Policy.

Econometrica 89(6), 2559–2599.

Bilal, A. (2023).
Solving heterogeneous agent models with the master equation.
Manuscript,Harvard.

Boppart, T., P. Krusell, and K. Mitman (2018).
Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative.
Journal of Economic Dynamics and Control 89(C), 68–92.

Chang, Y. and S.-B. Kim (2007).
Heterogeneity and aggregation: Implications for labor-market fluctuations.
American Economic Review 97(5), 1939–1956.

Childers, D. (2018).
Solution of Rational Expectations Models with Function-Valued States.
Working Paper.

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

Grand, F. L. and X. Ragot (2022).
Optimal policies with heterogeneous agents: Truncation and transitions.
Manuscript.

Judd, K. L. (2002).
Perturbation methods with nonlinear changes of variables.
Manuscript,Hoover Institution.

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

Mertens, T. M. and K. L. Judd (2017).
Solving an incomplete markets model with a large cross-section of agents.

?

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

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

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

Winberry, T. (2018).
A toolbox for solving and estimating heterogeneous agent macro models.
Quantitative Economics, forthcoming.