Solving and Estimating Heterogeneous Agent Models with Aggregate Uncertainty by Perturbation Methods

Lecture

C. Bayer, U Bonn, CEPR & IZA

B. Born, FS &CEPR

R. Luetticke, UCL & CEPR

November 2020

Heterogeneous agents models with aggregate uncertainty

These models are computational demanding to solve

- The original Krusell and Smith (1997, 1998) algorithm is notoriously slow
- Therefore, many papers use MIT shocks
- or are restricted to relatively simple household decisions

Heterogeneous agents models with aggregate uncertainty

These models are computational demanding to solve

- The original Krusell and Smith (1997, 1998) algorithm is notoriously slow
- Therefore, many papers use MIT shocks
- or are restricted to relatively simple household decisions

- We depart from the Reiter (2002, 2009) perturbation method
- And (try to) provide an accessible algorithm that can deal with high-dimensional heterogeneity

Reiter (2002): Solve by perturbation

The heterogeneous agent model:

- Can be written as a non-linear difference equation:
  $\mathbb{E}F(X_t, X_{t+1}, Y_t, Y_{t+1}, \varepsilon_{t+1}) = 0$
- that is function valued and
- needs to be linearized around the stationary equilibrium (StE)

Reiter (2002): Solve by perturbation

The heterogeneous agent model:

- Can be written as a non-linear difference equation:
  $\mathbb{E}F(X_t, X_{t+1}, Y_t, Y_{t+1}, \varepsilon_{t+1}) = 0$
- that is function valued and
- needs to be linearized around the stationary equilibrium (StE)
- Functions need to be approximated by finite dimensional objects (e.g. coefficients of polynomials, splines, etc.)
- We show how to do this in a smart efficient way

Heterogeneous agents models with aggregate uncertainty

Content

- General model setup
- Perturbation solution
- Our reduction method
- Application: Solving the Krusell&Smith model
- Application: Estimation of HANK models

Heterogeneous agents models with aggregate uncertainty

Available codes:

- Perturbation vs MIT shock for KS model (Matlab)
  `https://github.com/ralphluet/KS_Perturbation_vs_MIT`
- Perturbation with our reduction for KS and HANK models (Matlab)
  `https://github.com/ralphluet/perturbation_codes`
- Perturbation with our reduction for HANK models (Python)
  `https://github.com/econ-ark/BayerLuetticke`
- Perturbation with our reduction for estimating HANK models (Julia)
  `https://github.com/BenjaminBorn/HANK_BusinessCycleAndInequality`

Model Setup

Recursive Dynamic Planning Problem

Consider a household problem in presence of aggregate and idiosyncratic risk

- S_t is an (exogenous) aggregate state
- s_{it} is a partly endogenous idiosyncratic state
- μ_t is the distribution over s

- Bellman equation:

$$\nu(s_{it}, S_t, \mu_t) = \max_{x \in \Gamma(s_{it}, P_t)} u(s_{it}, x) + \beta \mathbb{E} \nu(s_{it+1}(x, s_{it}), S_{t+1}, \mu_{t+1})$$

Recursive Dynamic Planning Problem

Consider a household problem in presence of aggregate and idiosyncratic risk

- S_t is an (exogenous) aggregate state
- s_{it} is a partly endogenous idiosyncratic state
- μ_t is the distribution over s

- Euler equation:

$$u'[x(s_{it}, S_t, \mu_t)] = \beta R(S_t, \mu_t)\mathbb{E}u'[x(s_{it+1}, S_{t+1}, \mu_{t+1})],$$

No aggregate risk

Recall how to solve for a StE

- Discretize the state space (vectorized)
- Optimal policy $\bar{\mathrm{h}}(\mathrm{s}_{\mathrm{it}};\mathrm{P})$ induces flow utility $\bar{\mathrm{u}}_{\bar{\mathrm{h}}}$ and transition probability matrix $\Pi_{\bar{\mathrm{h}}}$
- Discretized Bellman equation

$$\bar{v} = \bar{\mathrm{u}}_{\bar{\mathrm{h}}} + \beta \Pi_{\bar{\mathrm{h}}} \bar{v} \tag{1}$$

holds for optimal policy (assuming a linear interpolant for the continuation value)

No aggregate risk

- and for the law of motion for the distribution (histograms)

$$\mathrm{d}\bar{\mu} = \mathrm{d}\bar{\mu}\Pi_{\bar{\mathrm{h}}} \tag{2}$$

No aggregate risk

Equilibrium requires

- $\bar{\mathrm{h}}$ is the optimal policy given P and ν (being a linear interpolant)
- $\bar{\nu}$ and $\mathrm{d}\bar{\mu}$ solve (1) and (2)
- Markets clear (some joint requirement on $\bar{\mathrm{h}}, \mu, \mathrm{P}$, denoted as $\Phi(\bar{\mathrm{h}}, \mu, \mathrm{P}) = 0$)

This can be solved for efficiently

- $\mathrm{d}\bar{\mu}$ is vector corresponding to the unit-eigenvalue of $\Pi_{\bar{\mathrm{h}}}$
- Using fast solution techniques for the DP, e.g. EGM
- Using a root-finder to solve for P

Introducing aggregate risk

With aggregate risk

- Prices and distributions change over time

Yet, for the household

- Only prices and continuation values matter
- Distributions do not influence the decisions directly

Redefining equilibrium (Reiter, 2002)

A **sequential equilibrium with *recursive individual planning***

- A sequence of discretized Bellman equation, such that

$$\nu_t = \bar{u}_{P_t} + \beta \Pi_{h_t} \nu_{t+1} \tag{3}$$

  holds for optimal policy, h_t (which results from ν_{t+1} and P_t)
- and a sequence of histograms, such that

$$d\mu_{t+1} = d\mu_t \Pi_{h_t} \tag{4}$$

  holds given the optimal policy
- (Policy functions, h_t, that are optimal given P_t, ν_{t+1})
- Prices, distributions and policies lead to market clearing

Compact notation (Schmitt-Grohé and Uribe, 2004)

The equilibrium conditions as a non-linear difference equation

- Controls: $Y_t = \begin{bmatrix} \nu_t & P_t & Z_t^Y \end{bmatrix}$ and
- States: $X_t = \begin{bmatrix} \mu_t & S_t & Z_t^X \end{bmatrix}$
  where Z_t are purely aggregate states/controls
- Define

$$F\left(d\mu_t, S_t, d\mu_{t+1}, S_{t+1}, \nu_t, P_t, \nu_{t+1}, P_{t+1}, \varepsilon_{t+1}\right) \tag{5}$$

$$= \begin{bmatrix} d\mu_{t+1} - d\mu_t \Pi_{h_t} \\ \nu_t - \left(\bar{u}_{h_t} + \beta \Pi_{h_t} \nu_{t+1}\right) \\ S_{t+1} - H(S_t, d\mu_t, \varepsilon_{t+1}) \\ \Phi(h_t, d\mu_t, P_t, S_t) \\ \varepsilon_{t+1} \end{bmatrix}$$

s.t.

$$h_t(s) = \arg \max_{x \in \Gamma(s, P_t)} u(s, x) + \beta \mathbb{E} \nu_{t+1}(s') \tag{6}$$

Compact notation (Schmitt-Grohé and Uribe, 2004)

In words

- First set of equations: Difference of one forward iteration of the distribution to assumed value.
- Second set of equations: Difference of one backward iteration of the value function (or policy functions in EGM) to assumed value.
- Last two sets of equations: Macro model.

Compact notation (Schmitt-Grohé and Uribe, 2004)

The equilibrium conditions as a non-linear difference equation

- Function-valued difference equation $\mathbb{E}F(X_t, X_{t+1}, Y_t, Y_{t+1}, \varepsilon_{t+1}) = 0$
- turns real-valued when we replace the functions by their discretized counterparts
- Standard techniques to solve by perturbation (Dynare etc)

Perturbation References: General

- General:
  1. A First Look at Perturbation Theory by James G. Simmonds and James E. Mann Jr.
  2. Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory by Carl M. Bender, Steven A. Orszag.

- This lecture:
  1. "Perturbation Methods for General Dynamic Stochastic Models" by Hehui Jin and Kenneth Judd.
  2. "Computational Methods for Economists" by Jesus Fernandez-Villaverde.
  3. "Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function" by Martin Uribe and Stephanie Schmitt-Grohe.

Non-linear difference equation

- A large class of economic models can be written as a set of non-linear difference equations of the form
$$\mathrm{E_t f(s_{t+1}, s_t, c_{t+1}, c_t) = 0}$$
where s are all state and c are all control variables now.

Perturbation methods

- More generally, functional equations of the form:

$$\mathcal{H}(\mathrm{d}) = 0$$

  for an unknown decision rule d.
- Perturbation solves the problem by specifying:

$$\mathrm{d}^{\mathrm{n}}(\mathrm{x}, \theta) = \sum_{\mathrm{i}=0}^{\mathrm{n}} \theta_{\mathrm{i}} (\mathrm{x} - \mathrm{x}_0)^{\mathrm{i}}$$

- We use implicit-function theorems to find coefficients θ_{i}'s
- Inherently local approximation.

Motivation

- Many complicated mathematical problems have:
  - either a particular case
  - or a related problem

  that is easy to solve.

- Often, we can use the solution of the simpler problem as a building block of the general solution.

- Sometimes perturbation is known as asymptotic methods.

A simple example

- Imagine we want to compute $\sqrt{26}$ by hand
- Note that:
$$\sqrt{26} = \sqrt{25 * 1.04} = 5 * \sqrt{25 * 1.04} \approx 5 * 1.02 = 5.1$$
- Exact solution: $\sqrt{26} = 5.09902$
- More generally:
$$\sqrt{\mathrm{x}} = \sqrt{\mathrm{y}^2 * (1 + \epsilon)} = \mathrm{y} * \sqrt{(1 + \epsilon)} \approx \mathrm{y} * (1 + \epsilon)$$
- Accuracy depends on how big ϵ is

Applications in economics

- Judd and Guu (1993) showed how to apply it to economic problems
- Recently, perturbation methods have been gaining much popularity
- In particular, second- and third-order approximations are easy to compute and notably improve accuracy
- Perturbation theory is the generalization of the well-known linearization strategy
- Hence, we can use much of what we already know about linearization

Regular versus singular perturbations

- Regular perturbation: a small change in the problem induces a small change in the solution.
- Singular perturbation: a small change in the problem induces a large change in the solution.

- Example: excess demand function.
- Most problems in economics involve regular perturbations.
- Sometimes, however, we can have singularities.
  Example: introducing a new asset in an incomplete market model.

Perturbation methods

- Back to our economic model cast in the following form:

$$\mathrm{E_t f(s_{t+1}, s_t, c_{t+1}, c_t) = 0}$$

where s are state and c are control variables.

- Rewrite f introducing a parameter for uncertainty :

$$\mathrm{E_t f(s_{t+1}, s_t, c_{t+1}, c_t; \sigma) = 0}$$

.

Evolution of states

- Dynamic stochastic general equilibrium models in addition have a structure where a subset of state variables s_t^1 are predetermined and endogenous while the remainder are exogenously driven as

$$\mathrm{s}_{\mathrm{t}+1}^2 = \mathrm{H}_2(\mathrm{s}_\mathrm{t}^2, \sigma) + \sigma\Sigma\epsilon_{\mathrm{t}+1}$$

  where $\epsilon_{\mathrm{t}+1}$ are i.i.d. with zero mean, unit covariance and bounded support.
- Stacking all state variables, we can write

$$\mathrm{s}_{\mathrm{t}+1} = [\mathrm{H}_1(\mathrm{s}_\mathrm{t}, \sigma); \mathrm{H}_2(\mathrm{s}_\mathrm{t}, \sigma)] + \sigma\eta\epsilon_{\mathrm{t}+1}$$

- with $\eta = [0; \Sigma]$

Evolution of controls

- That c_t is a control means that there is a function $c_t = G(s_t, \sigma)$
- The goal is to solve for the unknown H_1 and G.

Local approximation

- Take a Taylor series approximation of G and H:

$$
\begin{aligned}
\mathrm{G}(\mathrm{s},\sigma) = {} & \mathrm{G}(\mathrm{s}^*,\sigma^*) + \mathrm{G}_{\mathrm{s}}(\mathrm{s}^*,\sigma^*)(\mathrm{s}-\mathrm{s}^*) + \mathrm{G}_{\sigma}(\mathrm{s}^*,\sigma^*)(\sigma-\sigma^*) \\
& + 1/2\mathrm{G}_{\mathrm{ss}}(\mathrm{s}^*,\sigma^*)(\mathrm{s}-\mathrm{s}^*)^2 + \mathrm{G}_{\mathrm{s}\sigma}(\mathrm{s}^*,\sigma^*)(\mathrm{s}-\mathrm{s}^*)(\sigma-\sigma^*) \\
& + 1/2\mathrm{G}_{\sigma\sigma}(\mathrm{s}^*,\sigma^*)(\sigma-\sigma^*)^2 + \ldots
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{H}(\mathrm{s},\sigma) = {} & \mathrm{H}(\mathrm{s}^*,\sigma^*) + \mathrm{H}_{\mathrm{s}}(\mathrm{s}^*,\sigma^*)(\mathrm{s}-\mathrm{s}^*) + \mathrm{H}_{\sigma}(\mathrm{s}^*,\sigma^*)(\sigma-\sigma^*) \\
& + 1/2\mathrm{H}_{\mathrm{ss}}(\mathrm{s}^*,\sigma^*)(\mathrm{s}-\mathrm{s}^*)^2 + \mathrm{H}_{\mathrm{s}\sigma}(\mathrm{s}^*,\sigma^*)(\mathrm{s}-\mathrm{s}^*)(\sigma-\sigma^*) \\
& + 1/2\mathrm{H}_{\sigma\sigma}(\mathrm{s}^*,\sigma^*)(\sigma-\sigma^*)^2 + \ldots
\end{aligned}
$$

Local approximation

- Replace s and c in F:

$$\begin{aligned}\mathrm{F}(\mathrm{s},\sigma) &\equiv \mathrm{E_t f}(\mathrm{H}(\mathrm{s_t},\sigma)+\sigma\eta\epsilon_{\mathrm{t}+1},\mathrm{s_t},\mathrm{G}[\mathrm{H}(\mathrm{s_t},\sigma)+\sigma\eta\epsilon_{\mathrm{t}+1},\sigma],\mathrm{G}(\mathrm{s_t},\sigma))\\ &= 0\end{aligned}$$

- The goal is to solve for the unknown $\mathrm{H_1}$ and G.
- Local approximation means that we solve for $\mathrm{H_1},\mathrm{G}$ by taking a Taylor expansion of F around the non-stochastic steady state s^*, for which $\sigma^* = 0$.
- Typically we can solve for this steady state without solving a recursive problem.

Local approximation

- Define non-stochastic steady state as vectors (s^*, c^*) :

$$f(s^*, s^*, c^*, c^*) = 0$$

- $c^* = G(s^*, 0)$ and $s^* = H(s^*, 0)$

- Note that if $\sigma = 0$, then $E_t f = f$

Local approximation

- Approximation of G and H around the point $(\mathrm{s},\sigma)=(\mathrm{s}^*,0)$

$$\mathrm{G}(\mathrm{s},0)=\mathrm{G}(\mathrm{s}^*,0)+\mathrm{G}_{\mathrm{s}}(\mathrm{s}^*,0)(\mathrm{s}-\mathrm{s}^*)+\mathrm{G}_{\sigma}(\mathrm{s}^*,0)\sigma$$
$$\mathrm{H}(\mathrm{s},0)=\mathrm{H}(\mathrm{s}^*,0)+\mathrm{H}_{\mathrm{s}}(\mathrm{s}^*,0)(\mathrm{s}-\mathrm{s}^*)+\mathrm{H}_{\sigma}(\mathrm{s}^*,0)\sigma$$

- $\mathrm{G}(\mathrm{s}^*,0),\mathrm{H}(\mathrm{s}^*,0)$ identified by steady state values
- Remaining coefficients are identified by:

$$\mathrm{F}_{\mathrm{s}}(\mathrm{s}^*,0)=0$$
$$\mathrm{F}_{\sigma}(\mathrm{s}^*,0)=0$$

Local approximation

- Take derivative of F w.r.t. uncertainty :

$$\begin{aligned}\mathrm{F}_\sigma(\mathrm{s}^*,0) &= \mathrm{E_t f_{s'}}(\mathrm{H}_\sigma + \eta\epsilon') + \mathrm{f_{c'}}[\mathrm{G_s}(\mathrm{H}_\sigma + \eta\epsilon') + \mathrm{G}_\sigma] + \mathrm{f_c G}_\sigma \\ &= \mathrm{f_{s'}H}_\sigma + \mathrm{f_{c'}}[\mathrm{G_s H}_\sigma + \mathrm{G}_\sigma] + \mathrm{f_c G}_\sigma \\ &= 0\end{aligned}$$

- This is a system of n equations:

$$\begin{pmatrix}\mathrm{f_{s'}} + \mathrm{f_{c'}G_s} & \mathrm{f_{c'}} + \mathrm{f_c}\end{pmatrix}\begin{pmatrix}\mathrm{H}_\sigma \\ \mathrm{G}_\sigma\end{pmatrix} = 0$$

- This equation is linear and homogeneous in $\mathrm{H}_\sigma, \mathrm{G}_\sigma$. Thus we have that $\mathrm{H}_\sigma = 0$ and $\mathrm{G}_\sigma = 0$.

Local approximation

Important theoretical result:

- In words, up to first order, we do not need to adjust the steady state solution when changing aggregate risk σ.
- Expected values of $\mathrm{s_t}$ and $\mathrm{c_t}$ are equal to their non-stochastic steady-state values.
- In a first order approximation the certainty equivalence principle holds, i.e., the policy function is independent of the variance-covariance matrix of ϵ.
- Interpretation: no precautionary behavior.

Local approximation

- Differentiation w.r.t s yields:

$$\mathrm{F_s(s^*,0) = f_{s'}H_s + f_s + f_{c'}G_sH_s + f_cG_s = 0}$$

- In matrix form:

$$\begin{pmatrix} \mathrm{f_{s'}} & \mathrm{f_{c'}} \end{pmatrix} \begin{pmatrix} \mathrm{I} \\ \mathrm{G_s} \end{pmatrix} \mathrm{H_s} = -\begin{pmatrix} \mathrm{f_s} & \mathrm{f_c} \end{pmatrix} \begin{pmatrix} \mathrm{I} \\ \mathrm{G_s} \end{pmatrix}$$

Local approximation

- Let $A = [f_{s'} \quad f_{c'}]$ and $B = [f_s \quad f_c]$
- Let $\hat{s}_t \equiv s_t - s^*$, then postmultiply:

$$A \begin{pmatrix} I \\ G_s \end{pmatrix} H_s \hat{s}_t = -B \begin{pmatrix} I \\ G_s \end{pmatrix} \hat{s}_t$$

- Consider a perfect foresight equilibrium. In this case, $H_s \hat{s}_t = \hat{s}_{t+1}$

$$A \begin{pmatrix} I \\ G_s \end{pmatrix} \hat{s}_{t+1} = -B \begin{pmatrix} I \\ G_s \end{pmatrix} \hat{s}_t$$

Local approximation

- This leaves us with a system of quadratic equations that we need to solve for $\mathrm{H_s}, \mathrm{G_s}$.
- Procedures to solve rational expectations models:
  1. Blanchard and Kahn (1980).
  2. Uhlig (1999).
  3. Sims (2000).
  4. Klein (2000).

Local properties of the solution I

- Perturbation is a local method.
- It approximates the solution around the deterministic steady state of the problem.
- It is valid within a radius of convergence.

Local properties of the solution II

- What is the radius of convergence of a power series around x? An $r \in \mathcal{R}_{+}^{\infty}$ such that $\forall x', |x' - z'| < r$, the power series of x will converge.
- A Remarkable Result from Complex Analysis:
  The radius of convergence is always equal to the distance from the center to the nearest point where the decision rule has a (non-removable) singularity. If no such point exists then the radius of convergence is infinite.
- Singularity here refers to poles, fractional powers, and other branch powers or discontinuities of the functional or its derivatives.

Solution

- Using an eigenvalue decomposition of $H_s = P\Lambda P^{-1}$ we obtain,

$$\begin{aligned} AZ\Lambda &= BZ & Z &= [I; G_s]P \\ A &= [\partial F_{s'}, \partial F_{c'}] & B &= -[F_s, F_c] \end{aligned}$$

which implies that the solution corresponds to a subset of the solutions to the generalized eigenvalue problem

$$AXD = BX$$

- But which?
- If we have a stable system, then $\lim_{t\to\infty} H_s^t = 0$. Therefore, we are searching for the exactly n_s eigenvalues smaller than unity.

Solution

- Splitting all solutions to the eigenvalue problem above and below eigenvalues of 1, we obtain

$$\mathrm{A}\begin{bmatrix} \mathrm{X}_{11} & \mathrm{X}_{21} \\ \mathrm{X}_{12} & \mathrm{X}_{22} \end{bmatrix}\begin{bmatrix} \mathrm{D}_1 & 0 \\ 0 & \mathrm{D}_2 \end{bmatrix} = \mathrm{B}\begin{bmatrix} \mathrm{X}_{11} & \mathrm{X}_{21} \\ \mathrm{X}_{12} & \mathrm{X}_{22} \end{bmatrix}$$

and therefore

$$\mathrm{H}_\mathrm{s} = \mathrm{X}_{11}\mathrm{D}_1\mathrm{X}_{11}^{-1}$$

and

$$\mathrm{G}_\mathrm{s} = \mathrm{X}_{12}\mathrm{X}_{11}^{-1}$$

Alternative Solution: Time Iteration

- Rendahl (2018) extends linear time iteration
- Intuitive, robust, and easy to implement algorithm. Now x is vector of all controls and states. $\mathrm{F_n}$ is transition matrix of states and controls.
- Rewrite difference equation in "end of last period" notation

$$\mathrm{Ax_{t-1} + Bx_t + Cx_{t+1} = 0}$$

- Let $\mathrm{F_n}$ be a candidate solution:

$$\mathrm{Ax_{t-1} + Bx_t + CF_n x_t = 0}$$
$$\mathrm{x_t = -(B + CF_n)^{-1} A x_{t-1}}$$

- Thus update guess $\mathrm{F_{n+1}}$ as

$$\mathrm{F_{n+1} = -(B + CF_n)^{-1} A}$$

Higher order approximations

- Obtaining higher-order approximations to the solution of the non-linear system is a sequential procedure.
- The coefficients of the ith term of the jth-order approximation are given by the coefficients of the ith term of the ith order approximation, for $j > 1$ and $i < j$.
- More importantly, obtaining the coefficients of the ith order terms of the approximate solution given all lower-order coefficients involves solving a linear system of equations.

Higher order approximations

- We can iterate this procedure as many times as we want.
- We can obtain n-th order approximations.
- Levintal (2017) uses tensor-unfolding to work with higher-order derivatives

- Problems:
  1. Existence of higher order derivatives.
  2. Numerical instabilities.
  3. Computational costs.

Quality of approximation

- David Childers (2018), "Automated Solution of Heterogeneous Agent Models":
- Under some regularity conditions the solution algorithm is guaranteed to converge to the first derivative of the true infinite dimensional solution as the discretization is refined.
- Convergence rates for the approximation are provided as well, depending on the choices of interpolation method including polynomials, splines, histograms, and wavelets.

So, is all solved?

The dimensionality of the system F is still an issue

- With high dimensional idiosyncratic states, discretized value functions and distributions become large objects
- For example:
  4 income states (grid points)
  $\times$ 100 illiquid asset states
  $\times$ 100 liquid asset states
  $\implies \geq 40,000$ control variables in F
- Same number of state variables

Our Reduction Method

Reiter (2009): Reduce dimensionality ex ante

Problem:

- Dimensionality of the difference equation is large

Proposal:

- Reduce dimensionality ex ante (before solving the StE)
- e.g. (sparse) splines to represent policy functions
- Then linearize

Winberry (2016) extends this to the distribution functions

What we do

Proposal:

- Reduce dimensionality after StE, but before linearization
- Extract from the StE the *important* basis functions to represent individual policies (akin to image compression)
- Perturb only those basis functions but use the StE as "reference frame" for the policies (akin to video compression)
- Similarly for distributions (details later)

Our idea

1.) Apply compression techniques as in video encoding

- Apply a discrete cosine transformation to all value/policy functions (Chebycheff polynomials on roots grid)
- Define as reference "frame": the StE value/policy function
- Write fluctuations as differences from this reference frame
- Assume all coefficients of the DCT from the StE close to zero do not change after shock

Our idea

2.) Transform joint-distribution μ into copula and marginals

- Calculate the Copula, $\bar{\Xi}$ of μ in the StE
- Perturb the marginal distributions
- Approximate changes in the Copula (via DCT) or use fixed Copula to calculate an approximate joint distribution from marginals

- Idea follows Krusell and Smith (1998) in that some moments of the distribution do not matter for aggregate dynamics

Copula

A distribution of probabilities

A *Copula* is a joint distribution function of univariate marginal probabilities for a multivariate stochastic variable. It maps $[0,1]^n \to [0,1]$

Sklar's theorem

Every distribution function F can be represented by the marginal distribution functions F_i and a *Copula*, Ξ, with $F(x_1,\ldots,x_n) = \Xi\left[F_1(x_1),\ldots,F_n(x_n)\right]$.

Details

1.) Apply compression techniques as in video encoding

- Let $\bar{\Theta} = \mathrm{dct}(\bar{v})$ be the coefficients obtained from the DCT of the value function in StE
- A DCT expresses a finite sequence of data points in terms of sum of cosine functions at different frequencies
- Linear, invertible function $f = \Re^N - > \Re^N$ (equivalently: an invertible NxN matrix)
- x_n is transformed to X_n according to:

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\pi/N(n + 1/2)k\right], k = 0, ..., N-1$$

Details

1.) Apply compression techniques as in video encoding

- DCT yields the coefficients of the fitted (multi-dimensional) Chebyshev polynomial, where the polynomial is constructed such that the tensor grid for s is mapped to the Chebyshev knots.
  See Ahmed et al. (1974) for the seminal contribution.
- Importantly, the absolute value of the coefficients has an interpretation in terms of the R^2 contribution of the corresponding polynomial in fitting the data.
- This allows us to order and select the polynomial terms based on their importance.

Excursus: Global polynomial

- Express a function by the coefficients ψ of a polynomial

$$\hat{f}(x) = \sum_{j=1}^{n} \psi_j c_j(x)$$

where $c_i(x)$ are known basis functions such as $c_j(x) = x^j$.
- Better than ordinary polynomials are usually Chebyshev polynomials of which the baseline functions are

$$c_j(x) = \cos(j \ \arccos x)$$

- These are orthogonal on [-1,1], i.e.

$$\int_{-1}^{1} c_i(x) c_j(x) \frac{1}{\sqrt{1-x^2}} dx = 0 \forall i \neq j$$

Excursus: Global polynomial

- Since the evaluation points $\mathrm{x_i}$ are known ("grid"), we can compute $\mathrm{C} = [\mathrm{c_j}(\mathrm{x_i})]_{\mathrm{i}=1\ldots\mathrm{M},\mathrm{j}=1\ldots\mathrm{n}}$
- The vector of function values $\widehat{\mathrm{f}} = [\hat{\mathrm{f}}(\mathrm{x_i})]_{\mathrm{i}=1\ldots\mathrm{M}}$ is then given by

$$\widehat{\mathrm{f}} = \mathrm{C}\psi$$

- Therefore, we can obtain an optimal (minimal MSE) as

$$\psi^* = (\mathrm{C}'\mathrm{C})^{-1}\mathrm{C}'\ \mathrm{f}$$

- The big **advantage** of polynomials is that they can be integrated analytically and that they are differentiable of any order.

Excursus: Global polynomial: issues

- **Runge's Phenomenon:** Since polynomials tend to infinity as $x \to \infty$ it is not true that the overall fit of a global polynomial gets better, if more grid points and higher order polynomials are used (oscillating behavior).

- Choosing **Chebyshev polynomials** as basis functions and
- grid points as the **roots** $x_i = \cos(\frac{2i-1}{2N})$ for $i = 1 \ldots N$ of these polynomials minimizes approximation error.

Excursus: Discrete Cosine Transforms

A first observation

- Suppose Chebychev root grid-points are not suitable for our problem.
- Then, we can write $f(x) = f(g(y))$ and
- generate the grid x_i by applying g to the Chebyshev nodes y_i,
- with basis functions $c_j(x) = \cos\left(j \ \arccos g^{-1}(x)\right)$

Discrete Cosine Transform (DCT) and lossy compression

- In particular, if we do not intend to evaluate off-grid, we do not need to know g but just the nodes $y_i = \cos\left(\frac{2i-1}{2N}\pi\right)$ and grid values x_i
- and obtain an equivalent representation of f_i in terms of coefficients.
- Shrinking ≈ 0-coefficients to 0 leaves $\hat{f}_i$ close to unchanged.
- In addition $C'C = I$.

Details

1.) Apply compression techniques as in video encoding

- Define an index set $\mathcal{I}$ that contains the x percent largest (i.e. most important) elements from $\bar{\Theta}$
- Let θ be a sparse vector with non-zero entries only for elements $\mathrm{i} \in \mathcal{I}$
- Define $\tilde{\Theta}(\theta_t) = \begin{cases} \bar{\Theta}(\mathrm{i}) + \theta_t(\mathrm{i}) & \mathrm{i} \in \mathcal{I} \\ \bar{\Theta}(\mathrm{i}) & \text{else} \end{cases}$

Details

Decoding

- Now we reconstruct $\nu_t = \nu(\theta_t) = \mathrm{idct}(\tilde{\Theta}(\theta_t))$
- This means that in the StE the reduction step adds no additional approximation error as $\nu(0) = \bar{\nu}$ by construction
- Yet, it allows to reduce the number of derivatives that need to be calculated from the outset

Details

2) Analogously for the distribution function

- Define μ_t as $\Xi_t(\bar{\mu}_t^1, \ldots, \bar{\mu}_t^n)$ for n being the dimensionality of the idiosyncratic states
- The StE distribution is obtained when $\mu = \bar{\Xi}(\bar{\mu}^1, \ldots, \bar{\mu}^n)$

- We can treat the copula as an interpolant defined on the grid of steady-state marginal distributions, and also approximate Ξ_t as a sparse expansion around the steady-state copula $\bar{\Xi}$.
- The most extreme variant of this is to treat the copula as time fixed.

Details

2) Analogously for the distribution function

- Typically prices are only influenced through the marginal distributions
- The approach ensures that changes in the mass of one dimension, say wealth, are distributed in a sensible way across the other dimensions
- The implied distributions look "similar" to the StE one

Obtaining the copula function of the StE

To obtain an estimate of the Copula of the StE:

1. Accumulate the histogram along every dimension to obtain CDF estimate, M.
2. Add a leading zero to the CDF matrix, M, along every dimension.
3. Calculate marginal distributions, $\mathrm{m_i}$, from the CDF (summing out other dims)
4. Obtain the Copula estimate as an interpolant of M on $\{\mathrm{m_1}, \ldots, \mathrm{m_n}\}$

$$\bar{\Xi} = \textsc{griddedInterpolant}(\{\mathrm{m_1}, \ldots, \mathrm{m_n}\}, \mathrm{M})$$

.

Details

Too many equations

- The system

$$
\begin{aligned}
&\mathrm{F}\Big(\{\mathrm{d}\mu_t^1,\dots,\mathrm{d}\mu_t^n\},\mathrm{S}_t,\{\mathrm{d}\mu_{t+1}^1,\dots,\mathrm{d}\mu_{t+1}^n\},\mathrm{S}_{t+1},\\
&\theta_t,\mathrm{P}_t,\theta_{t+1},\mathrm{P}_{t+1}\Big)=\\
&\begin{bmatrix}
\mathrm{d}\bar{\Xi}(\bar{\mu}_t^1,\dots,\bar{\mu}_t^n)-\mathrm{d}\bar{\Xi}(\bar{\mu}_t^1,\dots,\bar{\mu}_t^n)\Pi_{h_t}\\
\mathrm{dct}\left[\mathrm{idct}(\tilde{\Theta}(\theta_t))-\left(\bar{u}_{h_t}+\beta\Pi_{h_t}\mathrm{idct}(\tilde{\Theta}(\theta_{t+1}))\right)\right]\\
\mathrm{S}_{t+1}-\mathrm{H}(\mathrm{S}_t,\mathrm{d}\mu_t)\\
\Phi(h_t,\mathrm{d}\mu_t,\mathrm{P}_t,\mathrm{S}_t)
\end{bmatrix}
\end{aligned}
\tag{7}
$$

  has too many equations

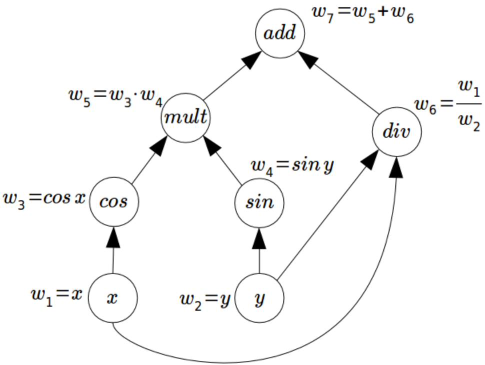
- Use only difference in marginals and the differences on $\mathcal{I}$

Excursus: Automatic differentiation

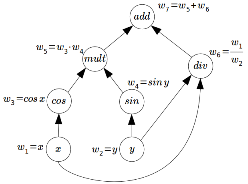
- Modern computer languages like Julia offer easy to implement automatic differentiation libraries.
- Automatic differentiation is neither:
  - Symbolic differentiation
    - Inefficient code
  - nor Numerical differentiation (the method of finite differences)
    - If you make h too small, then your accuracy gets killed by floating point roundoff
    - If h is too big, then approximation errors start ballooning
- AD avoids these problems: it calculates exact derivatives, so your accuracy is only limited by floating point error.

Excursus: Automatic differentiation

- AD applies the chain rule to your function
- Any complicated function f can be rewritten as the composition of a sequence of primitive functions
- Let $f(x, y) = \cos x \sin y + \frac{x}{y}$



Excursus: Automatic differentiation



$$\mathrm{Df} = \mathrm{Dw}_7 = \mathrm{D}(\mathrm{w}_5 + \mathrm{w}_6) = \mathrm{Dw}_5 + \mathrm{Dw}_6$$

$$\mathrm{Dw}_6 = \mathrm{D}\frac{\mathrm{w}_1}{\mathrm{w}_2} = \frac{\mathrm{w}_1\mathrm{Dw}_2 - \mathrm{w}_2\mathrm{Dw}_1}{\mathrm{w}_2^2}$$

$$\mathrm{Dw}_5 = \mathrm{Dw}_3\mathrm{w}_4 = \mathrm{w}_3\mathrm{Dw}_4 + \mathrm{w}_4\mathrm{Dw}_3$$

$$\mathrm{Dw}_4 = \mathrm{D}\sin \mathrm{w}_2 = \cos \mathrm{w}_2 \cdot \mathrm{Dw}_2$$

$$\mathrm{Dw}_3 = \mathrm{D}\cos \mathrm{w}_1 = -\sin \mathrm{w}_1 \cdot \mathrm{Dw}_1$$

$$\mathrm{Dw}_2 = \mathrm{Dy}$$

$$\mathrm{Dw}_1 = \mathrm{Dx}$$

Excursus: Automatic differentiation

- AD is implemented by a nonstandard interpretation of the program in which real numbers are replaced by dual numbers and the numeric primitives are lifted to operate on dual numbers.
- Dual numbers: Replace every number x with the number $\mathrm{x} + \mathrm{x}'\varepsilon$, where x' is a real number, but ε is an abstract number with the property $\varepsilon^2 = 0$

- Julia does this for you!
- ForwardDiff Package: `www.juliadiff.org/`
- Examples: `www.juliadiff.org/ForwardDiff.jl/stable/user/advanced.html`

Excursus: Julia

- Julia combines three key features for highly intensive computing tasks as perhaps no other contemporary programming language does: it is fast, easy to learn and use, and open source.

- Introduction by Fernandez-VillaVerde:
  `www.sas.upenn.edu/~jesusfv/Chapter_HPC_8_Julia.pdf`
- Introduction by QuantEcon:
  `https://lectures.quantecon.org/jl/`

Application: Krusell-Smith model

A simple KS economy

Incomplete Markets and TFP

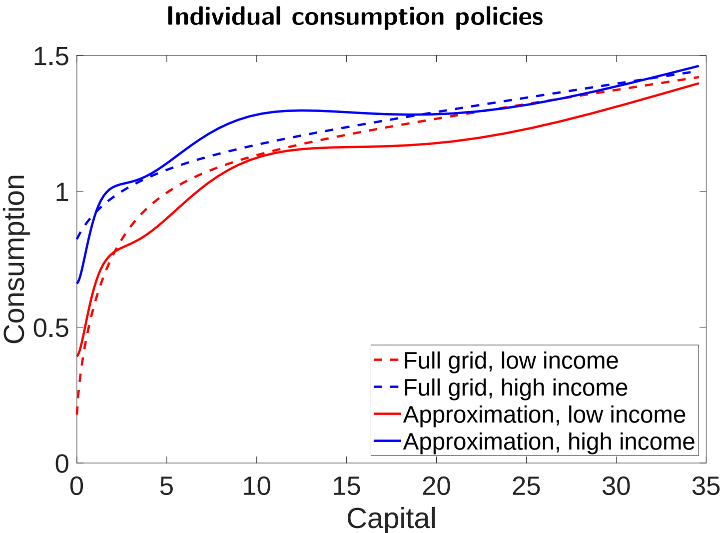
- Household productivity can be high or low
- No contingent claims
- Households save in capital goods (which they rent out)
- Households supply labor (disutility) and consume (utility)
- Aggregate productivity (TFP) follows a log AR-1 process
- Cobb-Douglas production function

A simple KS economy

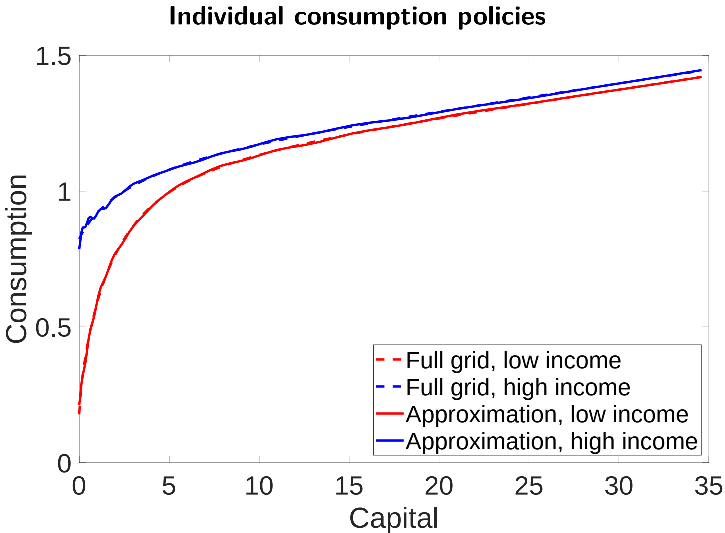
Numerical setup

- Asset grid has 100 points ($\Longrightarrow$ a total grid size of 200)
- Policies solved by EGM (instead of VFI)

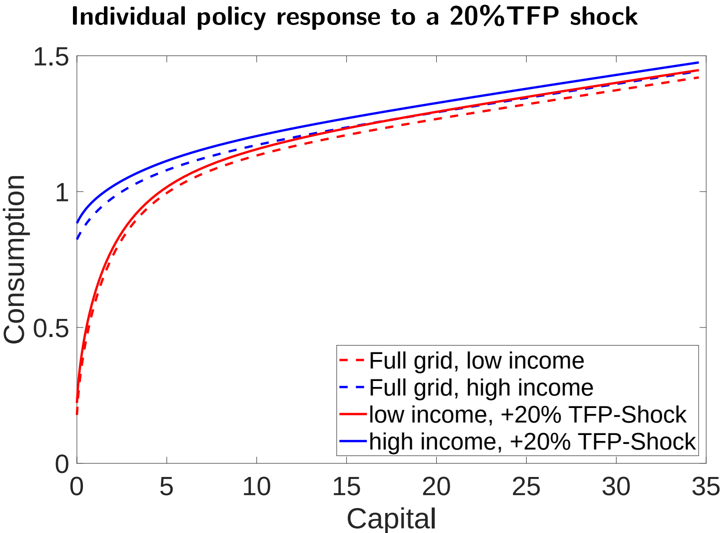
Different levels of "compression"



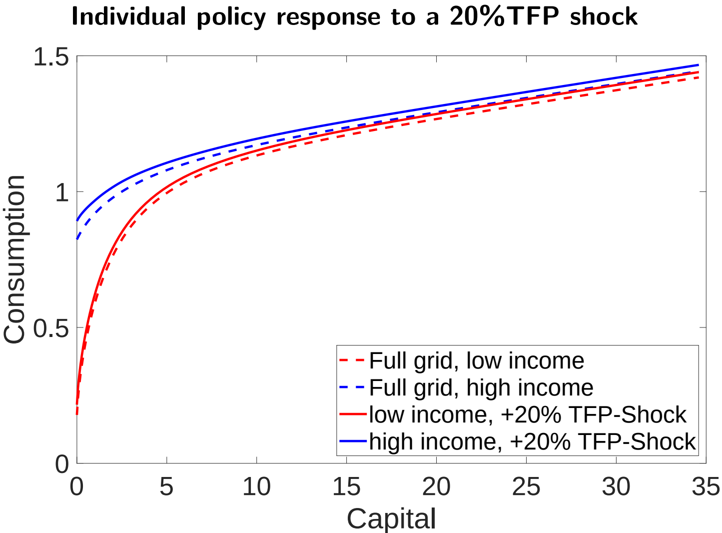
Different levels of “compression”



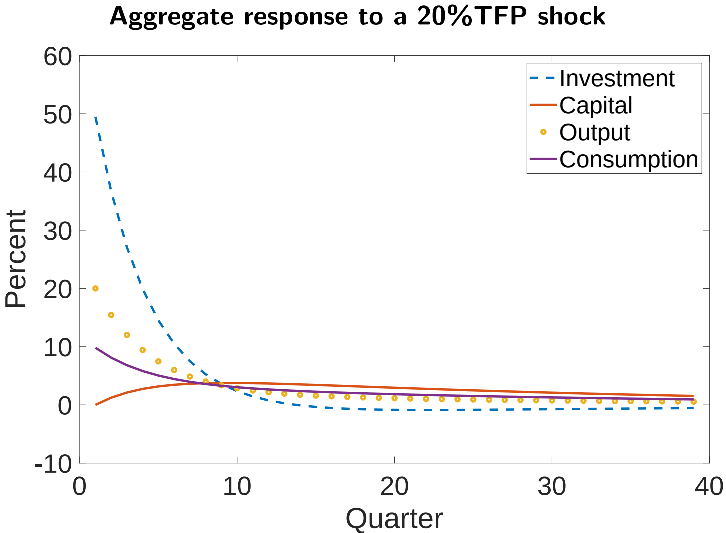
Different levels of “compression”



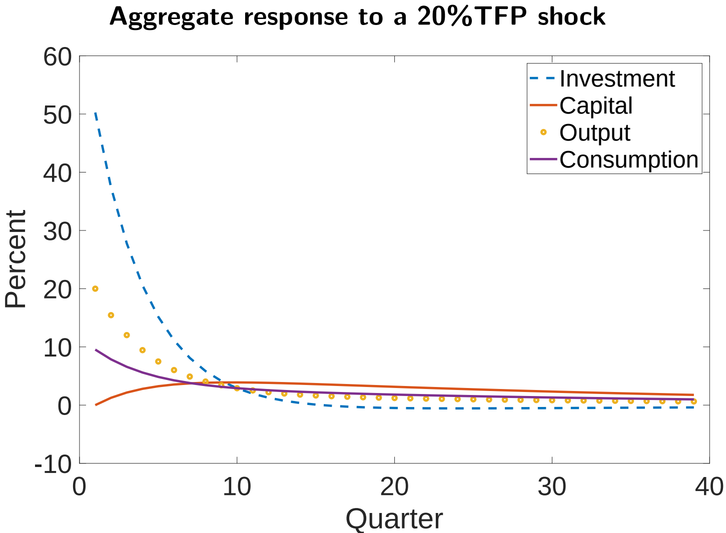
Different levels of "compression"



Different levels of “compression”



Different levels of “compression”

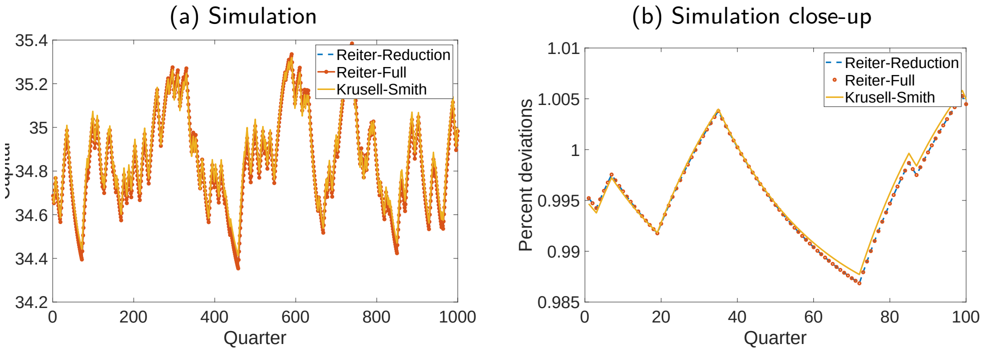


Taking stock

- When looking only at the StE policy function one concludes that roughly 50% of the information is needed to reconstruct the policies well
- This is roughly level of state reduction Reiter (2009) approach would achieve
- Using the StE as reference one can achieve much higher reduction
- For the aggregate dynamics maintaining only 3-6% of the basis functions suffices

Simulation performance

Figure: Simulations of Krusell & Smith model



Notes: Both panels show simulations of the Krusell & Smith (1998) model with TFP shocks solved with (1) the Reiter method with our proposed state-space reduction, (2) the original Reiter

Error Statistics

Table: Den Haan errors

	Absolute error (in %) for capital K_t		
	Reiter-Reduction	Reiter-Full	K-S
Mean	0.0119	0.0119	0.1237
Max	0.0152	0.0152	0.3491

Notes: Differences in percent between the simulation of the linearized solutions of the model and simulations in which we solve for the intratemporal equilibrium prices in every period and track the full histogram over time for $t = \{1, ..., 1000\}$; see Den Haan (2010)

Computing time

Table: Run time for Krusell & Smith model

	StE	K & S	Reiter-Reduction	Reiter-Full
in seconds	6.28	49.85	0.43	0.91

Notes: Run time in seconds on a Dell laptop with an Intel i7-7500U CPU @ 2.70GHz x 4. Model calibration and number of grid points as in Den Haan et al. (2010). Code in Matlab.

Application: Estimating HANK models

Bayer, Born, Luetticke (2020): Shocks, Frictions, and Inequality in US Business Cycles

What we do

- Fuse two-asset HANK model with a Smets-Wouters-type medium scale DSGE model
- Estimate the model using (Bayesian) full-information approach
- IRF analysis and variance decompositions

- Research Question:
  What shocks and frictions drive the US business cycle and US inequality?

Overview of the model

Workers		Production Sector	Government
Trade Assets	**Obtain Income**	**Produce and Differentiate Consumption Goods**	**Monetary Authority, Fiscal Authority**
Bonds, b>B; and Illiquid Assets (trading friction)	Wages set by unions s.t. Rotemberg wage adjustment costs (Idiosyncratic Income Risk) Interest Dividends Profits	**Intermediate goods producers** Rent capital & labor Competitive Market for Intermediate Goods **Entrepreneurs** Monopolistic resellers s.t. Rotemberg price adjustment costs	Policy Rules: • Monetary authority sets nominal interest rate -> Taylor rule • Fiscal authority supplies government debt, consumes goods, taxes labor income and profits -> Spending rule

Introducing more macro structure

Linearization techniques easily allow for more structure

- Say, we want to add price stickiness, monetary, and fiscal policy.
- This requires additional extra state variables.
- This is numerical cheap when linearizing.

Fusing incomplete markets and price stickiness

A HANK model

- Introduce price stickiness (NKPC),
- introduce a central bank (Taylor rule),
- introduce a fiscal authority (Expenditure rule),
- introduce capital goods producers (quadratic adjustment costs).

Issues:

- We need to deal with assigning monopoly rents.
- We need to deal with the portfolio problem (Bonds vs. Capital).
- We need to make labor supply endogenous

A HANK model

See Bayer et al., 2019

Extending the household sector

1. Assume GHH preferences (for business cycles reasonable)

$$u(c,n) = \frac{\left(c - h\frac{n^{1+\gamma}}{1+\gamma}\right)^{1-\xi}}{1-\xi}$$

   Scaling with productivity h allows for easy aggregation w.l.o.g. if taxes are linear.

2. Assign profits to either to (a) a group of households, (b) the government, or (c) a profit-asset.

A HANK model

Modeling portfolio choice: easy version

- All households hold the same bonds-to-capital ratio.
- All assets can be traded without any friction.
- Choice is over total wealth.
- For first order approximation: Returns must equal in expectations, i.e. define a safe return on bonds R_t, prices of capital goods q_t and rental rates of capital r_t, then

$$\mathbb{E}_t \frac{r_{t+1} + q_{t+1}}{q_t} = R_{t+1}$$

Equilibrium conditions (idiosyncratic part)

This leaves us with the following equilibrium conditions:
(A) idiosyncratic part, using linear interpolations in micro problem:

1. Recursive planning. For the vectors of marginal utilities $u_{c,t}$:

$$u_{c,t} = \underbrace{\beta R_{t+1}\Gamma_t(u_{c,t+1} + \lambda_{t+1})}_{\text{one EGM backwards step}}$$

   with Γ_t induced by optimal policies,
   - given future marginal utils $u_{c,t+1}$, and expected returns R_{t+1}
   - and current incomes determined through wages w_t, dividends r_t, profits π_t, and capital prices q_t.
2. Law of motion for distribution of capital

$$\mu_{t+1} = \mu_t\Gamma_t$$

Equilibrium conditions (summary variables)

(B) summary variables, model free:

1. It is useful to introduce an aggregate control that summarize μ_t: $K_t := \sum_j k_j \mu_t^j$ where k_j is the capital grid.
2. Let $\phi_t := \frac{B_t}{K_t}$ be the bonds-to-capital ratio entering period t.
3. For any unit of capital households hold, they have $r_t + q_t + \phi_t R_t$ resources for consumption.
4. Every unit of capital for next period sells at $q_t + \phi_{t+1}$

Equilibrium conditions (macro model)

(C) prices:

1. Factor prices as controls from FOCs of firms

$$w_t = (1-\alpha) mc_t Z_t \left(\frac{K_t}{N_t}\right)^{\alpha}, \quad r_t = \alpha mc_t Z_t \left(\frac{N_t}{K_t}\right)^{1-\alpha} - \delta,$$

prices of undifferentiated goods, mc_t, and total profits accordingly

$$\hat{\pi}_t = \beta \mathbb{E}_t \hat{\pi}_{t+1} + \kappa \left(mc_t^{-1} - \bar{\mu} \right)$$

$$\Pi_t = (1 - mc_t) Y_t - \text{adjustment costs/profits}$$

2. Returns on government bonds from Taylor rule (state variable)

$$R_{t+1} = R_t^{\rho_R} \hat{\pi}_t^{(1-\rho_R)\theta_\pi} \hat{Y}_t^{(1-\rho_R)\theta_Y}$$

Observe that adjustment costs are zero up to first order around stationary equilibrium.

Equilibrium conditions (macro model)

(D) aggregate quantities:

1. Labor supply
$$(1-\tau)w_t = N_t^{\gamma}$$
2. Production of capital goods (ignore externality)
$$q_t = 1 + \phi \frac{K_{t+1} - K_t}{K_t}$$
3. Total output and components
$$Y_t = Z_t K_t^{\alpha} N_t^{1-\alpha}, \quad C_t = Y_t - G_t - K_{t+1} + (1-\delta)K_t$$
4. A fiscal rule (spending adjusts, B_t is a state, G_t a control)
$$\hat{G}_t = (\hat{B}_t \hat{R}_t / \hat{\pi}_t)^{\rho_B} \hat{\pi}_t^{\gamma_\pi} \hat{Y}_t^{\gamma_Y}, \quad B_{t+1} = G_t + B_t R_t / \pi_t - \tau w_t N_t$$
5. Goods market clearing is residual.

2 asset models

HANK models have more action with more assets

- The literature has highlighted the role of wealthy hand-to-mouth consumers (Kaplan et al., 2014).
- HANK models with more assets feature asset substitution as in the older Keynesian literature (see e.g. Tobin, 1969), which is supported by the data (see Bayer et al., 2019; Luetticke, 2020).

A tractable structure

- For many applications it suffices to assume that capital can only be traded from time to time randomly (Calvo shock).

2 asset models

2 marginal values of assets

- Marginal value of liquid assets results from usual envelop condition

$$V_b(h, b, k) = Ru_c$$

- Marginal value of illiquid assets results from usual envelop condition

$$V_k(h, b, k|\text{adjust}) = (q + r)u_c^a$$

when trade is possible and from the marginal value of the dividend payment plus discounted marginal value if no trade is possible

$$V_k(h, b, k|\text{not}) = ru_c^n + \beta \mathbb{E} V_k'(h', b', k)$$

- Thus, $V_k(h, b, k) = \lambda(q + r)u_c^a + (1 - \lambda)(ru_c^n + \beta \mathbb{E} V_k'(h', b', k))$
- Optimal asset choices require $q\mathbb{E}V_b(h', b', k') = \mathbb{E}V_k(h', b', k')$ which allows us to trace out potentially optimal $(b', k')(h)$ pairs

Computational aspects of HANK-2

- requires both assets (100 points each)
- 22 income states
- uses both value functions and consumption policies as controls
- Full set would have $> 600,000$ variables
- Reduction to 219 distribution states and ca 700 controls for value functions and policies

Observations

Changing the aggregate macro structure is easy

- As long as a change in the model does not affect what income is composed of and which choices households can make given prices and incomes, but only how prices are formed, we can change the aggregate part of the model without touching the micro part.

Changing parameters does not mean we have to update everything

- This means that writing $F(\dots) = [F_1, F_2]'$,

$$F_1 = \begin{bmatrix} d\bar{\Xi}(\bar{\mu}_t^1, \dots, \bar{\mu}_t^n) - d\bar{\Xi}(\bar{\mu}_t^1, \dots, \bar{\mu}_t^n)\Gamma_t \\ dct\left[idct(\tilde{\Theta}(\theta_t)) - \left(\bar{u}_{h_t} + \beta\Gamma_t idct(\tilde{\Theta}(\theta_{t+1}))\right)\right] \end{bmatrix}$$

$$F_2 = \begin{bmatrix} S_{t+1} - \Xi(S_t, P_t) \\ \Phi(h_t, \mu_t, P_t, S_t) \end{bmatrix}$$

implies we only need to update the Jacobian of F_2 when changing model parameters and $\Phi(h_t, d\mu_t, P_t, S_t) = \bar{\Phi}(P_t, S_t)$ except for summary equation.

Sources of Fluctuations

Standard in complete markets model

- total factor productivity
- gov. bond spread (a.k.a. “risk premium”)
- price markup
- wage markup
- monetary policy
- government spending

Sources of Fluctuations

Standard in complete markets model

- total factor productivity
- gov. bond spread (a.k.a. "risk premium")
- price markup
- wage markup
- monetary policy
- government spending

New in the incomplete markets model

- idiosyncratic income risk
- tax progressivity

Solution and Estimation

Solution method

- The distribution Θ over $\mathrm{b}, \mathrm{k}, \mathrm{h}$ is a state variable
- First-order perturbation of the non-linear difference equation $\mathrm{EF}(\mathrm{x_t}, \mathrm{x_{t+1}}, \epsilon_\mathrm{t}) = 0$ around the stationary equilibrium to obtain a local approximation to the solution
- We approximate the policy functions as sparse polynomials around their stationary equilibrium values and approximate the distribution functions by histograms of their marginals and a time-varying Copula as in Bayer and Luetticke (2020).

Estimation: Numerical details

- i.e., method linearizes the resulting non-linear difference equation.
- Write as $\mathrm{A}\mathrm{x}_{\mathrm{t}} = \mathrm{B}\mathrm{x}_{\mathrm{t}+1}$ and solve using Klein's method to obtain G and H.
- Since estimated parameters do not directly show up in household problem, they change only small parts of A, B during estimation.
- Need to calculate only few derivatives of F in each iteration of the estimation.
- Convenient also in terms of coding: Only dynare-like macro model block needs to be edited to change the macro model.

Estimation: Bayesian approach

Similar to An and Schorfheide (2007) and Fernández-Villaverde (2010)

1. Kalman filter to obtain the likelihood from the state-space representation of the model solution.
2. Advantage of State-Space: Deal with mixed frequency and missing observations.
3. Roughly one evaluation of the Kalman filter every other second.
4. Maximize posterior likelihood
5. Random Walk Metropolis Hastings algorithm to draw from posterior

Gaussian State Space

- Solution to linearized model takes state space form, which can be written as

$$x_{t+1} = Gx_t + w_{t+1},\ w_{t+1} \overset{iid}{\sim} \mathcal{N}(0, Q) \tag{8}$$

$$y_t = Hx_t + \nu_t,\ \nu_t \overset{iid}{\sim} \mathcal{N}(0, R) \tag{9}$$

- G, H, Q, R are functions of the model parameters θ
- x_t is an $n_x \times 1$ vector of states
- y_t is an $n_y \times 1$ vector of observables
- w_t is a $p \times 1$ vector of structural errors
- ν_t a vector of measurement errors
- Assumption: w_t and ν_t are orthogonal

$$E_t(w_{t+1}\nu_s) = 0 \quad \forall\ t+1 \text{ and } s \geq 0$$

Fundamental Problem: Unobserved States

- This implies that

$$y_t = H\left(Gx_{t-1} + w_t\right) + \nu_t \tag{10}$$

- Thus, y_t is normally distributed:

$$y_t \sim \mathcal{N}(HGx_{t-1}, HQH + R) \tag{11}$$

- If all states were observed, we could directly construct the likelihood $f(y_T, \dots, y_1|\theta)$
- We could then run optimizer over our estimated parameter set $\tilde{\theta} \subseteq \theta$ to get ML estimate of $\tilde{\theta}$
- Problem: we have unobserved states and cannot use equation (11)
- Solution: turn to Kalman filter to back out states from the observed data $\rightarrow$ Filtering problem

Kalman Filter: Summary

At time t, given $\hat{x}_{t|t-1}, \Sigma_{t|t-1}$ and observing y_t

1. Compute the forecast error in the observations using

$$a_t = y_t - H\hat{x}_{t|t-1} \tag{12}$$

2. Compute the Kalman Gain K_t using

$$K_t = G\Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1} \tag{13}$$

3. Compute the state forecast for next period given today's information

$$\hat{x}_{t+1|t} = G\hat{x}_{t|t-1} + K_t\left(y_t - H\hat{x}_{t|t-1}\right) = G\hat{x}_{t|t-1} + K_t a_t \tag{14}$$

4. Update the covariance matrix

$$\Sigma_{t+1|t} = (G - K_t H)\,\Sigma_{t|t-1}\,(G - K_t H)' + Q + K_t R K_t' \tag{15}$$

Kalman Filter: Initialization

- How to initialize filter at $t = 0$ where no observations are available?

  $\rightarrow$ start with unconditional mean $E(x)$ and Variance Σ

- Given covariance stationarity, the unconditional mean is

$$E(x) = Ex_{t+1} = E(Gx_t + w_{t+1}) = GE(x) \Rightarrow (I - G)E(x) = 0$$

  hence, $E(x) = 0$

- For the covariance matrix, we have

$$\begin{aligned} \Sigma &= E\left[(Gx_t + w_t) (Gx_t + w_t)' \right] \\ &= E\left[Gx_t {x_t}' G' + w_t {w_t}' \right] \\ &= G \Sigma G' + Q \end{aligned} \tag{16}$$

  $\rightarrow$ so-called Lyapunov-equation

Metropolis Hastings-Algorithm

- Start with a vector θ_0
- Repeat for $\mathrm{j} = 1, \ldots, \mathrm{N}$
  - Generate $\tilde{\theta}$ from $\mathrm{q}(\theta_{\mathrm{j}-1}, \cdot)$ and u from $\mathcal{U}(0,1)$
  - If $\tilde{\theta}$ is valid parameter draw (steady state exists, Blanchard-Kahn conditions satisfied etc.) and $\mathrm{u} < \alpha(\theta^{\mathrm{j}-1}, \theta^{\mathrm{j}})$ set $\theta_{\mathrm{j}} = \tilde{\theta}$
  - Otherwise, set $\theta_{\mathrm{j}} = \theta_{\mathrm{j}-1}$ (implies setting $\pi(\tilde{\theta}) = 0$ if draw invalid)
- Return the values $\{\theta_0, \ldots, \theta_{\mathrm{N}}\}$
- After the chain has passed the transient stage and the effect of the starting values has subsided, the subsequent draws can be considered draws from the posterior

$\Rightarrow$ burnin required that assures remaining chain has converged

The Random-Walk Metropolis Hastings Algorithm

- As long as the regularity conditions are satisfied, any proposal density will ultimately lead to convergence to the invariant distribution
- However: speed of convergence may differ significantly
- In practice, people often use the Random-Walk Metropolis Hastings algorithm where

$$\mathrm{q}\left(\theta, \tilde{\theta}\right) = \mathrm{q}_{\mathrm{RW}}\left(\tilde{\theta} - \theta\right) \tag{17}$$

  and q_{RW} is a multivariate density
- The candidate $\tilde{\theta}$ is thus given by the old value θ plus a random variable increment

$$\tilde{\theta} = \theta + \mathrm{z}, \mathrm{z} \sim \mathrm{q}_{\mathrm{RW}} \tag{18}$$

Estimation: Two-step procedure

- First, we calibrate or fix all parameters that affect the steady state of the model.
- Second, we estimate by full-information methods all parameters that only matter for the dynamics of the model, i.e., the aggregate shocks, frictions, and policy rules.
- We set the priors for shocks, frictions, and policy rules to standard values from the representative agent literature

Observables

Quarterly US data from 1954Q1 – 2019Q4

In first-differences

- GDP, Consumption, Investment
- the real wage

In log-levels

- GDP deflator based inflation rates
- Hours worked per capita
- the (shadow) federal funds rate

All demeaned and without measurement error.

Observables

Further data non-quarterly availability

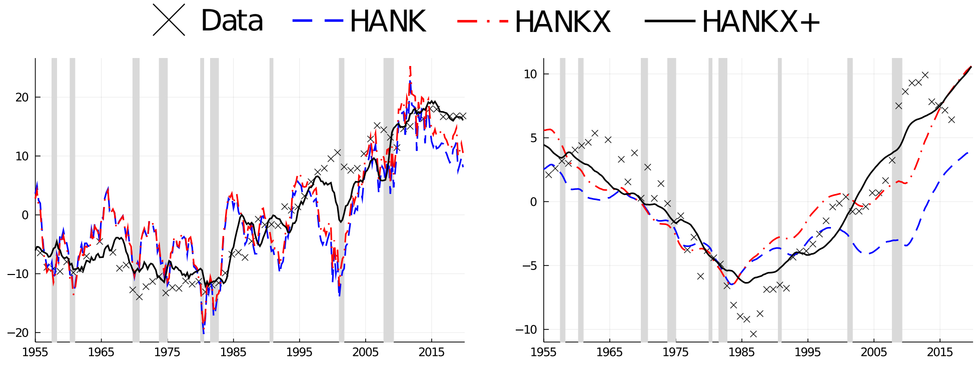
- Measures of inequality:
  - Wealth share of the top 10% (Piketty-Saez WID) (1954 – 2019)
  - Income share of the top 10% (Piketty-Saez WID) (1954 – 2019)

All in log-levels, demeaned and with measurement error.

Estimated model variants

Data / Shocks	Aggregate Data	+ Cross-sectional Data
Aggregate Shocks	HANK (vs RANK)	HANKX
+ Cross-sectional Shocks		HANKX+

Wealth Inequality in the US



(a) Top 10% income share (b) Top 10% wealth share

Shocks and Frictions in US Business Cycles

Variance decomposition: GDP and components

US business cycles: Summary

HANK and RANK models give only a somewhat different view

Estimation results

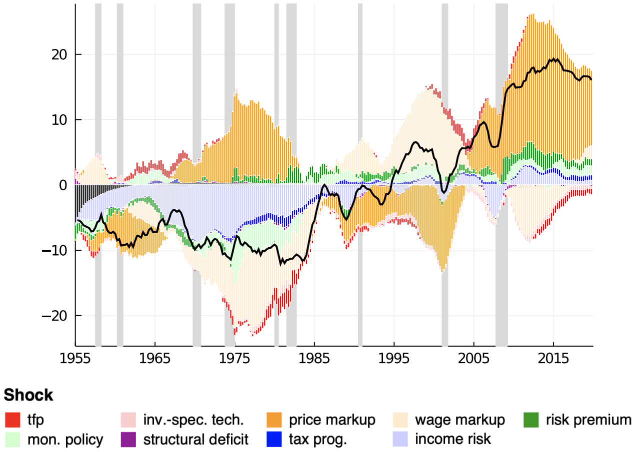
- Key is the estimation that makes the dynamics of both models more similar
- Estimated HANK model features less nominal and real frictions than RANK

Decomposition results

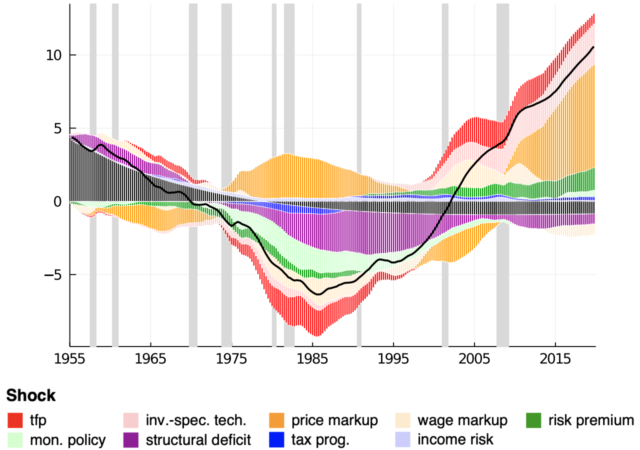
- Investment-specific technology becomes less important because it induces wealth effects on consumption via asset prices
- Risk premium, monetary, and wage markup shocks become more important
- Income risk shocks can partly replace risk premium shocks

Shocks and Frictions in US Inequality

Shock decomposition: Income share of top 10%



Shock decomposition: Wealth share of top 10%



Contribution of shocks to US inequality 1985-2019

Shock	Top 10% Income	Top 10% Wealth
TFP, ϵ^{Z}	-0.38	2.63
Inv.-spec. tech., ϵ^{Ψ}	-0.17	3.26
Price markup, $\epsilon^{\mu Y}$	11.69	4.3
Wage markup, $\epsilon^{\mu W}$	5.82	0.87
Risk premium, ϵ^{A}	-0.62	2.07
Income risk, ϵ^{σ}	2.57	-0.14
Monetary policy, ϵ^{R}	1.30	1.98
Structural deficit, ϵ^{G}	-0.05	1.60
Tax progressivity, τ^{P}	1.54	0.67
Sum of shocks	21.55	16.79

US inequality: Summary

Business cycle shocks are important drivers of inequality dynamics

Income inequality

- Price and wage markups explain two-third of the increase since 1985
- Rising income risk and falling tax progressivity explain the remaining one-third

Wealth inequality

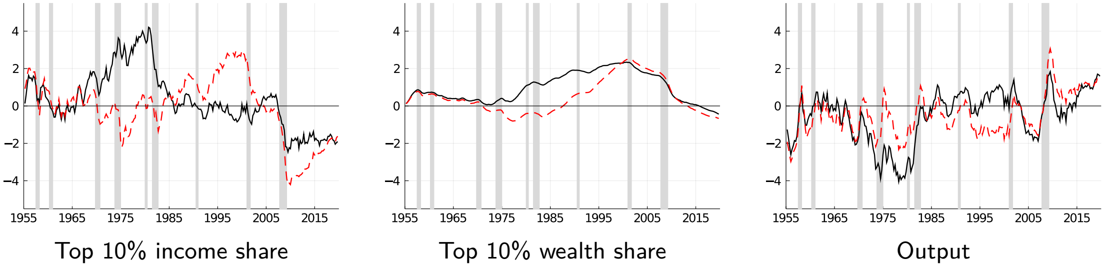
- Technology shocks via their effect on asset prices explain most of the increase since 1985
- The two markup shocks explain only one-third of this increase
- Monetary policy and fiscal deficit shocks are important as well

Policy Counterfactuals

Policy counterfactual: Inequality

- How important are the estimated policy coefficients for the evolution of inequality?
- Run estimated shock sequence with counterfactually set policy parameters
  - Hawkish monetary policy (double inflation response, θ_{π})
  - Dovish monetary policy (double output response, θ_{Y})

Counterfactual evolution of inequality: Monetary policy



Log point deviations from baseline. Black: Hawkish; Red: Dovish

Policy counterfactual: Summary

Effect of monetary policy depends on supply vs demand shocks

Hawkish monetary policy (triple θ_{π})

- Higher inequality in the 70s as markup (cost-push) shocks are important

Dovish monetary policy (triple θ_{Y})

- Lower inequality in the 70s and aftermath of the Great Recession

Very persistent effect on wealth inequality.

Summary: Bayer et al. (2020)

Our HANK model can jointly explain the US business cycle and inequality

US business cycle

- Not a radically different view on the US business cycle
- HANK models stress the importance of portfolio choice for the transmission of aggregate shocks

US inequality

- Business cycles are important to understand the evolution of US inequality.
- Business cycle shocks and policy responses can account for most of the increase in US inequality since the 1980s.

Conclusion

No excuse!

- Even when heterogeneity is high dimensional,
- our algorithm is an easy approach to these models
- It is a fast and simple to code

Conclusion

No excuse!

- It requires knowledge of only two standard tools of macro:
  1. Solving a recursive het. agent model for a StE
  2. Linearizing a rep. agent model
  3. (and a little twist in between)
- The fixed design for dimensionality reduction allows to employ the method to estimate models with standard techniques

Bibliography I

Ahmed, N., T. Natarajan, and K. R. Rao (1974). "Discrete cosine transform". *IEEE Transactions on Computers C-23* (1), 90–93.

An, Sungbae and Frank Schorfheide (2007). "Bayesian analysis of DSGE models". *Econometric Reviews 26* (2-4), 113–172.

Bayer, Christian, Benjamin Born, and Ralph Luetticke (2020). "Shocks, Frictions, and Inequality in US Business Cycles". CEPR Discussion Paper 14364.

Bayer, Christian and Ralph Luetticke (2020). "Solving heterogeneous agent models in discrete time with many idiosyncratic states by perturbation methods". *Quantitative Economics forthcoming*.

Bayer, Christian, Ralph Luetticke, Lien Pham-Dao, and Volker Tjaden (2019). "Precautionary savings, illiquid assets, and the aggregate consequences of shocks to household income risk". *Econometrica 87* (1), 255–290.

Den Haan, Wouter J (2010). "Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents". *Journal of Economic Dynamics and Control 34* (1), 79–99.

Bibliography II

- Den Haan, Wouter J, Kenneth L Judd, and Michel Juillard (2010). "Computational suite of models with heterogeneous agents: incomplete markets and aggregate uncertainty". *Journal of Economic Dynamics and Control 34* (1), 1–3.
- Fernández-Villaverde, Jesús (2010). "The econometrics of DSGE models". *SERIEs 1* (1), 3–49.
- Kaplan, Greg, Giovanni L Violante, and Justin Weidner (2014). "The wealthy hand-to-mouth". *NBER Working Paper No. 20073*.
- Krusell, Per and Anthony A. Smith (1998). "Income and wealth heterogeneity in the macroeconomy". *Journal of Political Economy 106* (5), 867–896.
- ——— (1997). "Income and wealth heterogeneity, portfolio choice, and equilibrium asset returns". *Macroeconomic Dynamics 1* (02), 387–422.
- Luetticke, Ralph (2020). "Transmission of monetary policy with heterogeneity in household portfolios". *American Economic Journal: Macroeconomics forthcoming*.
- Reiter, Michael (2002). "Recursive computation of heterogeneous agent models". *mimeo, Universitat Pompeu Fabra*.

Bibliography III

——— (2009). "Solving heterogeneous-agent models by projection and perturbation". *Journal of Economic Dynamics and Control 33* (3), 649–665.

Schmitt-Grohé, Stephanie and Martın Uribe (2004). "Solving dynamic general equilibrium models using a second-order approximation to the policy function". *Journal of Economic Dynamics and Control 28* (4), 755–775.

Tobin, James (1969). "A general equilibrium approach to monetary theory". *Journal of Money, Credit and Banking 1* (1), 15–29.

Winberry, Thomas (2016). "A toolbox for solving and estimating heterogeneous agent macro models". *mimeo, Chicago Booth.*