Estimating HANK:
Macro Time Series and Micro Moments

Sushant Acharya, William Chen, Marco Del Negro,
Keshav Dogra, Ethan Matlin, Reca Sarfati

Bank of Canada, Federal Reserve Bank of New York, Harvard, MIT

ASSA 2021 Econometric Society Session on
"Estimation of Macroeconomic Models"

Goal: Using HANK models for policy analysis

- Discuss some of the computational and conceptual (fitting both macro and micro data) challenges associated with **estimating HANK models**, and try to address some of these challenges

How to estimate HANK?

- From a **computational** point of view, both the *solution* and the *likelihood* of HANK models are much more expensive than those of RANK models
  - Standard approaches/packages are no longer adequate, and we need to come up with alternative ones

- Two requirements

  1. Fast likelihood computation for a large state vector $\rightarrow$ Chandrasekhar recursion

  2. Parallelization $\rightarrow$ **Sequential Monte Carlo**

Fitting macro *and* micro data

- From a **methodological** point of view, we want the model parameters to fit both the macro time series *and* (at least some) moments of the cross sectional distributions

$\Rightarrow$ propose an approach that combines time series likelihood-based methods with methods applied to micro data (e.g., method of moments)

$\Rightarrow$ Use this approach to ask:

Is there a trade-off between fitting macro and micro data?

- Does fitting macro time series requires high or low average MPCs (marginal propensities to consume) in the HA model?
    - HANK model with low average MPCs (most agents are permanent income agents) are pretty close to RANK models
- Does inequality matter for macro?

HA models estimation: the literature

- Challe et al., 2017: Bayesian estimation of HANK using a trick to simplify the wealth distribution
- Winberry, 2018: parametric reduction of state space, Bayesian estimation of shock process parameters in a heterogeneous firms model
- Cho, 2018: Reiter, 2009, approximation & dimension reduction, maximum likelihood estimation of non-steady state parameters in HANK
- Auclert et al., 2019: estimate HANK using MA(∞) representation; see also Hagedorn et al., 2019
- Bayer et al., 2019: Bayesian estimation of selected non-steady state parameters in HANK
- Fernandez-Villaverde et al., 2019: estimation of non linear HA model
- Mongey and Williams, 2017
- Liu and Plagborg-Møller, 2019

The model

- **Heterogeneous agents** version of **Smets & Wouters**
- Huggett economy where agents trade *liquid* (government) *bonds*
- There is *capital* in the economy (we want a stochastic growth model)
  - Capital is owned by a mutual fund (whose shares are *not* traded) that makes investment decisions and remits profits to agents (Auclert et al., 2018)

\+ (most of) Smets & Wouters' bells and whistles (as in Bayer et al., 2019):

real rigidities	*nominal rigidities*
investment adjustment costs	price stickiness
variable capital utilization	wage stickiness
	partial indexation to lagged inflation

NO habit persistence

Households

- Households solve

$$\max_{\{C_{i,t},B_{i,t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \left(\prod_{k=0}^{t} \beta_k \right) \left\{ \ln C_{i,t} - \varphi \frac{H_t^{1+\phi_h}}{1+\phi_h} \right\}$$

s.t.

$$C_{i,t} + \frac{1}{1+r_t} B_{i,t+1} = W_t s_{i,t} e_{i,t} H_t + B_{i,t} + T_t, \qquad B_{i,t+1} \geq 0$$

- $s_{i,t}$ *persistent* (2-state Markov chain w/ transition prob p_{ij}), $e_{i,t}$ *i.i.d.* independent across households
- T_t are lump sum transfers to households (from monopolistic competitors, mutual fund) minus taxes

Mutual fund, production, and price rigidities

- *Mutual fund* invests s.t. adjustment costs, rent out capital to firms, pay dividends to households (note: dividends can be negative if mutual fund decides to invest)

- *Firms*: Monpolistically competitive intermediate goods producers hire labor and capital, pay dividends to households, produce

$$Y_t(j) = K_t(j)^{\alpha} \left(Z_t L_t(j)\right)^{1-\alpha}$$

- Firms adjust prices s.t. Rotemberg cost $\rightarrow$ *Phillips curve*

- ... and sell to final good producers who combine varieties into CES aggregate Y_t

Wage rigidities

- Monopolistically competitive unions hire labor H_t from households (common across households) and sell to competitive employment agency
- Unions choose nominal wage s.t. Rotemberg adjustment cost to max average household utility (Auclert et al., 2018)

Monetary and fiscal policy, market clearing

- *Monetary policy* follows Taylor rule:

$$\left(\frac{1+i_t}{1+i}\right) = \left(\frac{1+i_{t-1}}{1+i}\right)^{\rho_R} \left[\left(\frac{\Pi_t}{\Pi^*}\right)^{\psi_\pi} \left(\frac{Y_t}{Y_{t-1}} e^{-\gamma}\right)^{\psi_y}\right]^{1-\rho_R} e^{\varepsilon_t^R}$$

- *Fiscal policy* matters (non-Ricardian):

$$\frac{B_{t+1}^g}{1+r_t} = B_t^g + G_t - T_t^g \text{ where } T_t^g = \delta_b \left[B_t^g + G_t - \frac{b^g Z_t}{1+r_t}\right] + (1-\delta_b) T^g Z_t$$

- Market clearing:

$$\int C_{i,t} di + X_t + G_t = Y_t$$

Shocks

- Technology growth is AR(1) $\rightarrow$ common stochastic trend in consumption, output, ...

$$\ln\left(\frac{Z_t}{Z_{t-1}}\right) = z_t = (1-\rho_z)\,\gamma + \rho_z z_{t-1} + \sigma_z \epsilon_{z,t}$$

- 6 other shocks: β_t, G_t/Y_t, marginal efficiency of investment, price markup, wage markup, monetary policy — all follow AR(1) processes

Solution algorithm

- (Large!) set of equilibrium conditions $\mathbb{E}_t F(x_{t+1}, y_{t+1}, x_t, y_t) = 0$ where the x_t's are the states (including the wealth distribution), and the y_t's are the jumps (including the individual decision variables)

- Steps (Reiter, 2009; Ahn et al., 2018; **Childers, 2018**; ...):

1. Solve numerically the (*non-linear*) steady state $F(x, y, x, y) = 0$ (set aggregate shocks to zero)
2. (Log-)linearize F (analytical differentiation), yielding system $A\mathbb{E}_t \begin{bmatrix} \hat{x}_{t+1} \\ \hat{y}_{t+1} \end{bmatrix} = B \begin{bmatrix} \hat{x}_t \\ \hat{y}_t \end{bmatrix}$
   - Note: linearization *only* with respect to *aggregate* variables $\rightarrow$ heterogeneity in the response to the shock across the distribution is preserved
3. *Reduction* (Reiter, 2010): instead of tracking the (very large) n-vectors summarizing the distribution/individual policy functions, "aggregate" across adjacent bins — so we only need to track an n/k-vectors
4. Solve reduced system using standard Klein algorithm

Estimation: Macro data

- With HANK models we have two datasets that we want the model to "fit":
  - Macro time series (Y^m) and micro distributions (Y^d)

Y^m: *macro time series* ($Y^m = y^m_{1:T}$)

- in this application, same observables as Smets and Wouters (2007): output, consumption, investment, and wage growth, total hours worked, inflation, and the federal funds rate, for the period 1964Q1-2007Q4.

$\Rightarrow$ Likelihood of macro time series (as in standard DSGE models estimation) :

$$p(y^m_{1:T}|\theta)$$

Estimation: Micro data

Y^d: *micro data/distributions*

- in this application, a few *moments*/targets from *steady state* distribution (vector $\bar{m}(Y^d)$), taken from Kaplan et al., 2018:

(Liquid) Wealth Distribution	$\bar{m}(Y^d)$	(σ_d)	*Income Distribution*	$\bar{m}(Y^d)$	(σ_d)
Fraction zero-liquid wealth agents	0.10	(10%)	Cross-sectional variance:	0.70	(5%)
Average MPC	0.16	(20%)	Average variance of 1-year income change	0.23	(5%)

$\Rightarrow$ Penalty function for micro moments (quasi-likelihood of micro data):

$$log\ \ p(Y^d|\theta) = -\frac{1}{2}\left(log\,\bar{m}(\theta) - log\,\bar{m}(Y^d)\right)'\ \Sigma_d^{-1}\ \left(log\,\bar{m}(\theta) - log\,\bar{m}(Y^d)\right)$$

where $\bar{m}(\theta)$ are the model implied moments, and Σ_d is a diagonal matrix with σ_d^2 on the diagonal

Estimation: Combining macro and micro Data

- Posterior:

$$p(\theta|Y^m, Y^d) \propto \underbrace{p(y^m_{1:T}|\theta)}_{\text{Likelihood}} \; \underbrace{p(Y^d|\theta)\; p(\theta)}_{\text{Prior}}$$

- $p(Y^d|\theta)$ is viewed as a prior (Del Negro and Schorfheide, 2008, "Forming priors for DSGE models"): $p(Y^d|\theta) \propto p(\theta|Y^d)$ implicitly generates a prior for all parameters affecting the steady state

- $p(\theta)$ is the "standard" prior (generally same as in Smets and Wouters) for those parameters that do not affect the steady state, or that enter the RANK version of the model

  - For the parameters that do not enter RANK (e.g., income process) $p(\theta)$ is uninformative (uniform within bounds), given that the prior is implicitly defined by $p(Y^d|\theta)$

- Similar to Mongey and Williams, 2017, but done *jointly*

Alternative interpretation

- **Composite likelihood** approach (Canova and Matthes, 2018):

$$p(\theta|Y^m, Y^d) \;\propto\; \underbrace{p(y^m_{1:T}|\theta)\; p(Y^d|\theta)}_{\text{Composite Likelihood}} \;\; \underbrace{p(\theta)}_{\text{Prior}}$$

  where the $p(Y^d|\theta)$ recalls the "likelihood" used in micro studies following Chernozhukov and Hong, 2003 (e.g., Jarosch 2015; Lise et al, 2015)

- Note, in this paper we are not trying to fit the *time series* of the *entire* cross sectional distribution (Liu and Plagborg-Møller, 2019) — only some *moments* of the *steady state* distribution

Controlling the weight of the micro data prior

- In order to formally investigate whether there is a trade-off between fitting macro and micro data we can introduce a parameter Υ_d that controls the weight on the prior $p(Y^d|\theta)$:

$$p(\theta|Y^m, Y^d) \propto \underbrace{p(y^m_{1:T}|\theta)}_{\text{Likelihood}} \underbrace{p(Y^d|\theta)^{\Upsilon_d}\, p(\theta)}_{\text{Prior}}$$

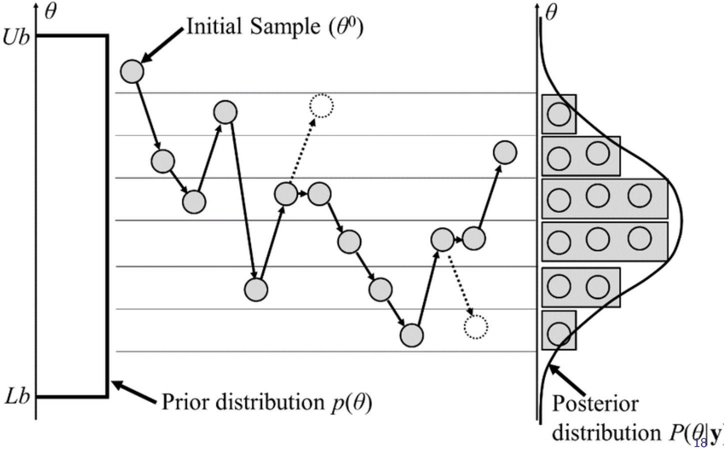
- As $\Upsilon_d \to \infty$ we force the model to meet the micro targets Y^d (equivalent to $\Sigma_d \to 0$ in the penalty function)

Making estimation feasible: Monte Carlo methods

- The posterior $p(\theta|Y^m, Y^d)$ does not have a known form $\rightarrow$ Monte Carlo methods
- Standard approach to obtaining draws from the *posterior distribution* in DSGE estimation: *Markov Chain* Monte Carlo (Random Walk Metropolis Hastings; e.g., Dynare)

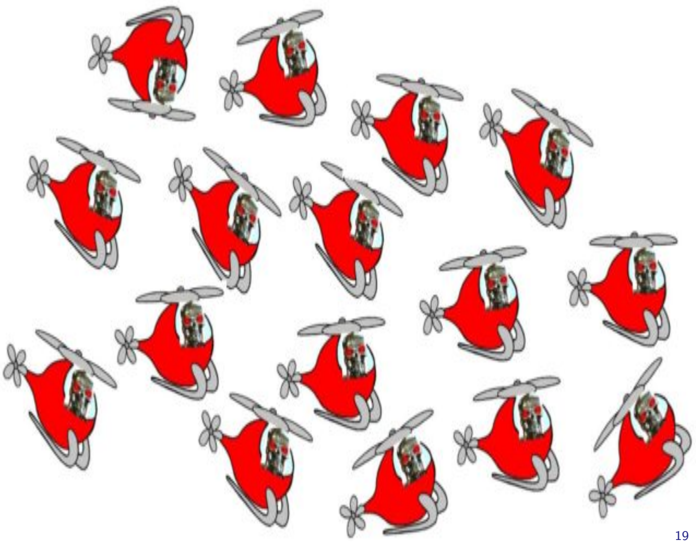
- Start with one particle θ and let it travel the posterior distribution (always accept moves "up" and only sometimes accept moves "down")

- Problem for HANK: *It cannot be parallelized* (it's Markov!)



"New" approach: Sequential Monte Carlo

- "New" to the DSGE estimation literature (Creal, 2007; Herbst and Schorfheide, 2014, 2015); old for the statistics literature (Gordon et al., 1993 Chopin, 2002, ...)
- Start with a swarm of particles

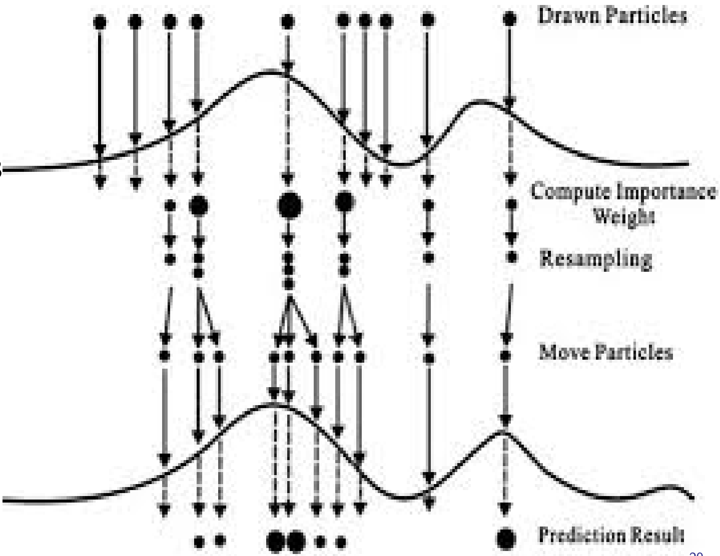


"New" approach: Sequential Monte Carlo

- "New" to the DSGE estimation literature (Creal, 2007, Herbst and Schorfheide, 2014, 2015); old for the statistics literature (Gordon et al., 1993, Chopin, 2002, ...)

- Start with a swarm of particles

- ... and let them all travel and "adapt" to the posterior

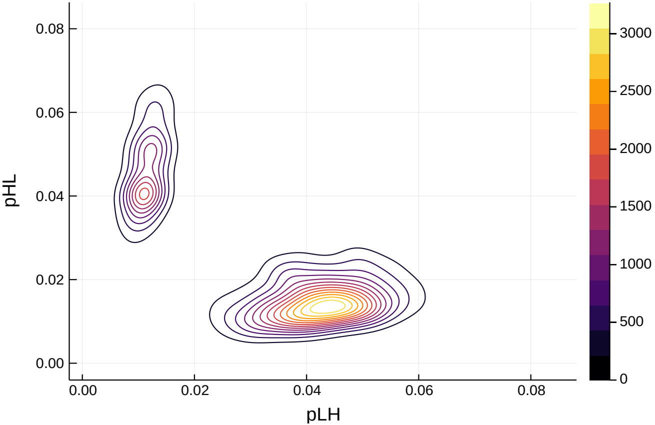


Why SMC?

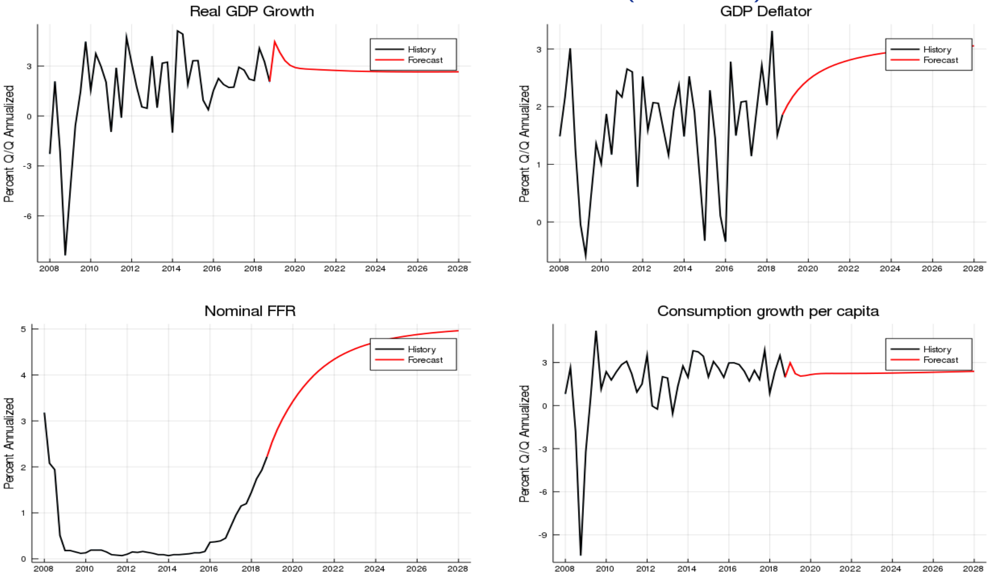
- Reasons to use SMC for HA models in particular, and in general models whose likelihood is costly to evaluate

  1. *It can be parallelized*
  2. *Swarm of particles can be re-used* as a bridge for new estimations ("online" estimation)
     - "Online estimation of DSGE models" Cai, Del Negro, Herbst, Matlin, Sarfati, Schorfheide, 2019; see also our blog and our Julia SMC package on GitHub
     - new data $\rightarrow$ **routine forecasting** (and forecasting evaluation exercises) becomes feasible
     - better approximation $\rightarrow$ start with a coarse (but fast to compute) approximation/reduction, and refine progressively
  3. *Robust* to multimodality

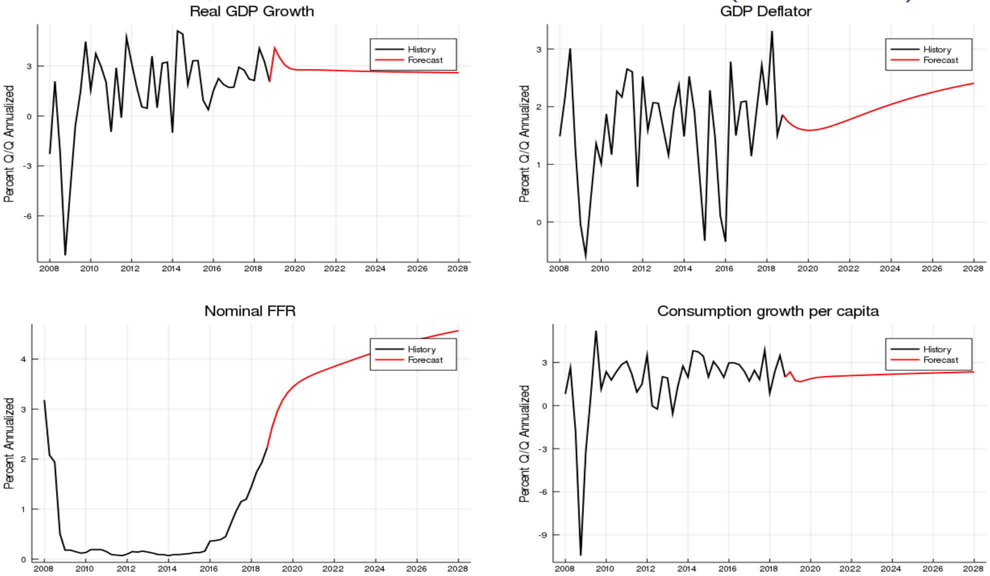
Example of multimodality



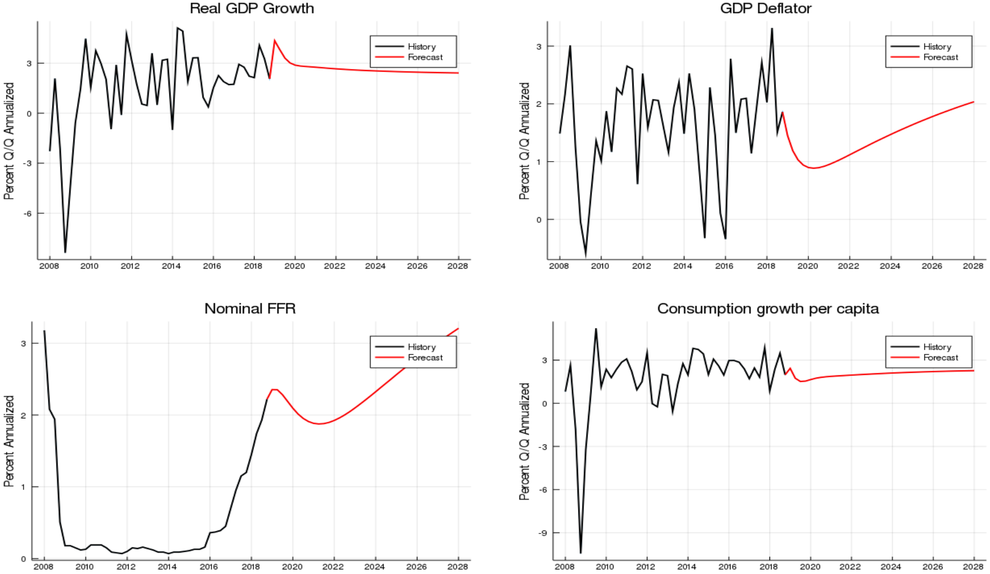
Forecasts from HANK! ($\Upsilon_d = 0$)



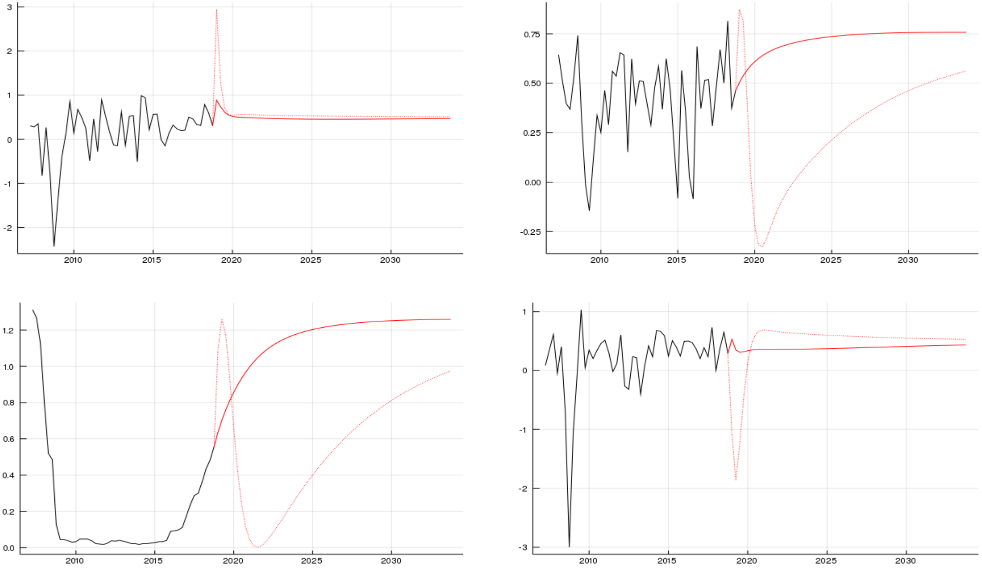
World premiere: forecasts from HANK! ($\Upsilon_d = 0.25$)



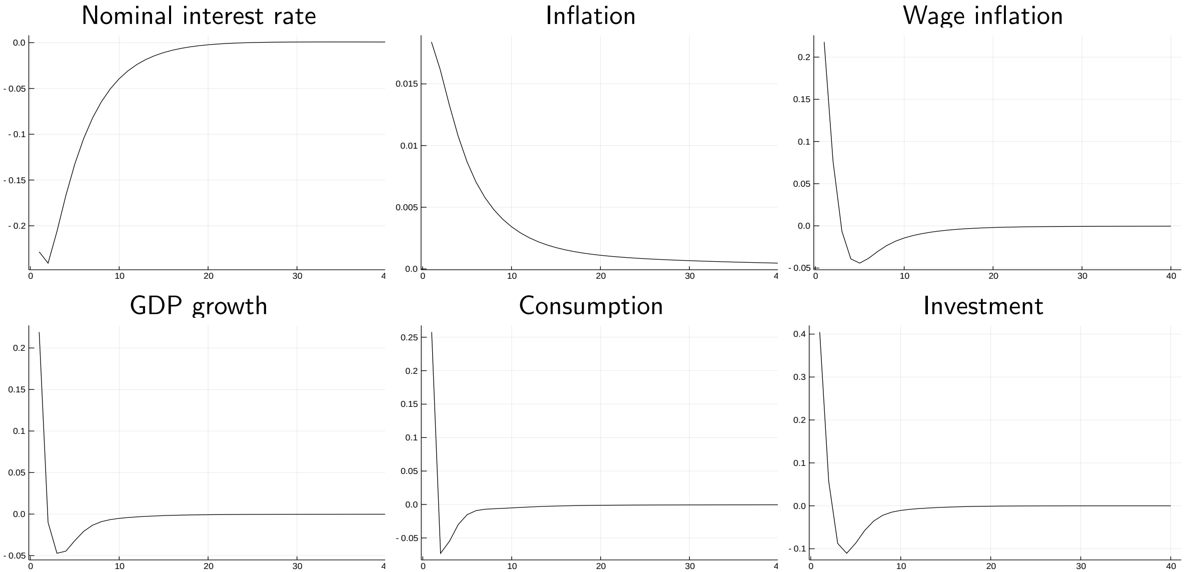
World premiere: forecasts from HANK! ($\Upsilon_d = 1$)



... vs calibrated parameters

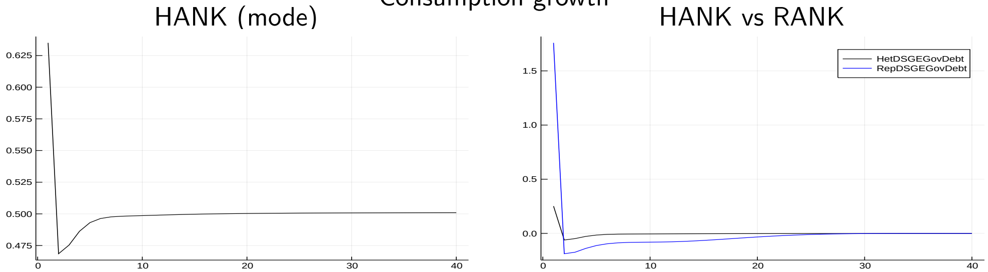


Impulse responses to a monetary policy shock

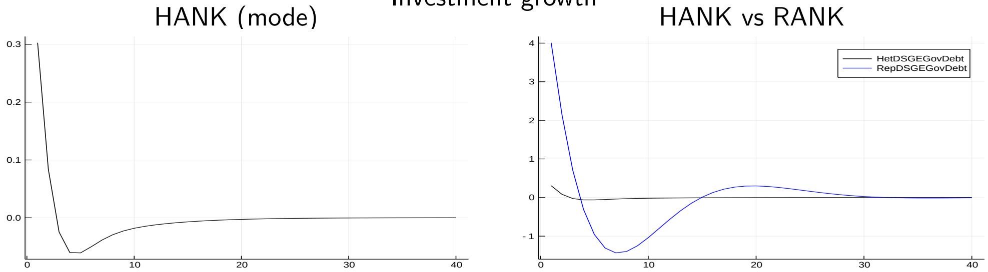


HANK vs RANK (same parameters)

Consumption growth



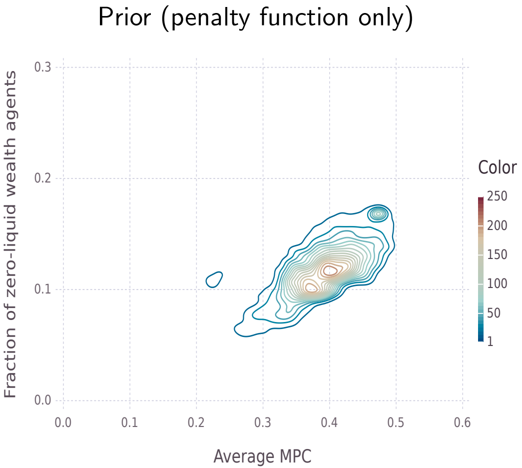
Investment growth

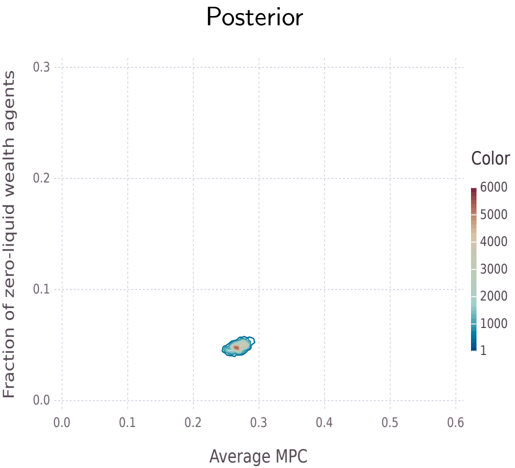


Selected parameter estimates

param	prior	prior_mean	mode	mean	std	5%	95%
			nominal rigidities				
κ_p	*Gamma*	0.500	0.001	0.002	0.001	0.001	0.003
κ_w	*Gamma*	0.500	0.387	0.342	0.090	0.203	0.503
			policy rule				
ψ_π	*Normal*	1.500	1.422	1.392	0.149	1.150	1.632
ψ_y	*Normal*	0.120	0.204	0.204	0.041	0.136	0.271
ρ_R	*Beta*	0.750	0.817	0.804	0.029	0.753	0.849
			fiscal rule				
δ_b	*Uniform*	0.500	0.988	0.873	0.104	0.670	0.992

Prior and posterior distribution for micro targets





Conclusions

- Estimating HANK models presents both conceptual (fitting both macro and micro data) and computational (desperate need for parallelization) challenges
- We try to address some of these challenges
- We apply our approach to a discrete time HANK model
- Preliminary results show that
    - We can use HANK to produce reasonable forecasts
    - The macro data want an average MPC of about 0.23, significantly above zero

Prior and posterior distribution for micro targets ($\Upsilon_d = 1$)