New Frontiers in Heterogeneous Agent Macroeconomics

Benjamin Moll
London School of Economics

Cowles Lecture
2026 North American Summer Meeting of the Econometric Society
http://benjaminmoll.com/cowles_lecture/


Tiphaine Wibault
Yucheng Yang
Christian Wolf
Clarisse Wibault
Andreas Schaab
Peter Maxted
Juan Duque
Matthieu Gomez
Xavier Gabaix
Felipe Alves
Lenya Ryzhik
SeHyoun Ahn
Paco Buera
David Laibson
Jean-Michel Lasry
Bob Lucas
Lukasz Rachel
Yves Achdou
Jakob Foerster
Zhiyu Fu
Abhijit Banerjee
Gisle Natvik
Greg Kaplan
Gianluca Violante
Andreas Fagereng
Yongseok Shin
Tom Winberry
Oleg Itskhoki
Victor Zhorin
Maike Osborne
Sebastian Towers
Mark Aguiar
Johannes Forkel
Chiyuan Wang
George Whittle
Martin Holm
Pierre-Louis Lions
Galo Nuño
Pascual Restrepo
Rob Townsend
Florian Scheuer
Jiequn Han
Emilien Gouin-
Bonenfant

Plan

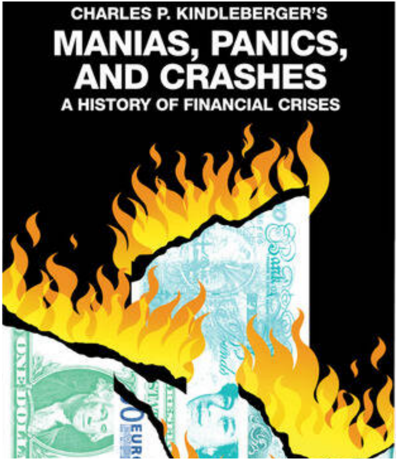
Introduction to Heterogeneous Agent Macroeconomics

- some typical models (HANK, ...)
- some central findings
- a key difficulty

New Frontiers

1. Why not just much simpler two-type models (TANK)?
2. Reinforcement learning for HA macro... as an equilibrium computation device
3. Reinforcement learning for HA macro... as a model of human learning

Long-run goal: HA models with proper aggregate risk, non-linearities



... as in macro-finance but with heterogeneity, ABM but with forward-looking agents

Talk is based on four papers of mine

Structural Reinforcement Learning for Heterogeneous Agent Macroeconomics

with Yucheng Yang, Chiyuan Wang & Andreas Schaab, *Working Paper (2025)*

https://arxiv.org/abs/2512.18892

Recurrent Structural Policy Gradient for Partially Observable Mean Field Games

with Clarisse Wibault, Johannes Forkel, Sebastian Towers, Tiphaine Wibault, Juan Duque, George Whittle, Andreas Schaab, Yucheng Yang, Chiyuan Wang, Maike Osborne & Jakob Foerster

Proceedings of the 43rd International Conference on Machine Learning (ICML 2026 spotlight paper)

https://arxiv.org/abs/2602.20141

The Trouble with Rational Expectations in Heterogeneous Agent Models: A Challenge for Macroeconomics

The Economic Journal (2026)

https://benjaminmoll.com/challenge/

Mean Field Games without Rational Expectations

with Lenya Ryzhik, *Communications in Contemporary Mathematics (2026)*

https://benjaminmoll.com/MFGRatx/

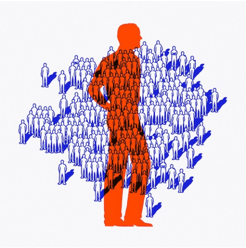
Other References

- Auclert, Rognlie and Straub (2025) “Fiscal and Monetary Policy with Heterogeneous Agents”
- Kaplan, Moll and Violante (2018) “Monetary Policy According to HANK”
- Kaplan, Moll and Violante (2023) “The Very Model of Modern Monetary Policy”
- New York Times (2023) “The ‘Representative Agent’ Is Always Rational. The Rest of Us Are Not.”
- Financial Times (2024) “Macroeconomists, meet HANK: how new models are changing the discipline”
- Bianchi and Kaplan (2026) “How Small is Small? Non-linearities in Heterogeneous Agent Models”
- Angeletos, Lian and Wolf (2024) “Can Deficits Finance Themselves?”
- Auclert, Rognlie and Straub (2024) “The Intertemporal Keynesian Cross”
- Rognlie (2025) “Comment on Debortoli-Galí”
- Bardoczy, Sim & Tischbirek (2025) “Macroeconomic Effects of Excess Savings”

Introduction to HA macroeconomics

Heterogeneous agent macroeconomics

- Approach: study macro questions in terms of distributions of micro variables rather than just aggregates
  - typical example: distributions of income and wealth
- Attractive for two reasons
  - conceptually: integrated approach to macro and distribution
  - empirically: integrated approach to micro and macro data



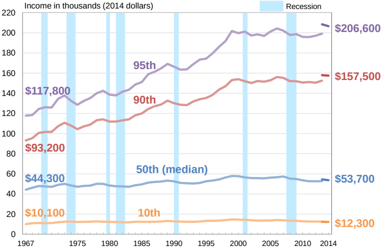
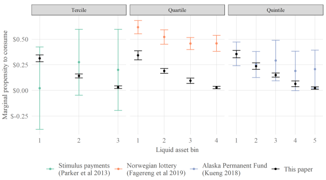


Figure 6: Marginal Propensity to Consume by Asset Buffer



Note: This figure compares the estimates of heterogeneity by assets in the passthrough of income shocks to consumption. Parker et al. (2013), Fagereng, Holm and Natvik (2018) and Kueng (2018) use terciles, quartiles, and quintiles respectively. To enable comparability with these prior papers, we calculate the marginal propensity to consume (instead of the elasticity of consumption to income) using their respective bin cutoffs. Our paper, Parker et al. (2013), and Kueng (2018) measure the MPC on nondurables. Fagereng, Holm and Natvik (2018) measures the MPC on total consumption. See Section 3.5 for details.

Today: HA models for business-cycle macro

- Monetary and fiscal policy
- Booms and busts
- $\Rightarrow$ HA models with aggregate risk
- Three example economies
  1. Huggett (1993) model with aggregate risk
  2. Krusell and Smith (1998) = Real Business Cycle model with heterogeneity
  3. HANK = Heterogeneous Agent New Keynesian model

Simplest textbook model: Huggett (1993) with agg. risk

- Continuum of agents i, heterog. in $(b_{i,t}, y_{i,t})$, $y_{i,t}$ = idios. risk, agg. shock z_t
- State of the economy: distribution $G_t(b, y)$ and agg. shock z_t
- Households choose consumption $c_{i,t}$ to maximize

$$v_{i,0} = \max_{\{c_{i,t}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_{i,t}) \quad \text{subject to}$$

$$b_{i,t+1} = (1 + r_t)(b_{i,t} + y_{i,t} z_t - c_{i,t}), \quad y_{i,t+1} \sim \mathcal{T}_y(\cdot | y_{i,t}), \quad b_{i,t+1} \geq \underline{b}$$

- Save ($b_{i,t} > 0$) or borrow ($b_{i,t} < 0$) at interest rate r_t

Simplest textbook model: Huggett (1993) with agg. risk

- Continuum of agents i, heterog. in $(b_{i,t}, y_{i,t})$, $y_{i,t}$ = idios. risk, agg. shock z_t
- State of the economy: distribution $G_t(b, y)$ and agg. shock z_t
- Households choose consumption $c_{i,t}$ to maximize

$$v_{i,0} = \max_{\{c_{i,t}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_{i,t}) \quad \text{subject to}$$

$$b_{i,t+1} = (1 + r_t)(b_{i,t} + y_{i,t} z_t - c_{i,t}), \quad y_{i,t+1} \sim \mathcal{T}_y(\cdot | y_{i,t}), \quad b_{i,t+1} \geq \underline{b}$$

- Save ($b_{i,t} > 0$) or borrow ($b_{i,t} < 0$) at interest rate r_t

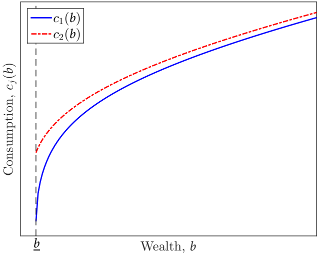
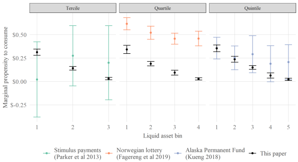


Figure 6: Marginal Propensity to Consume by Asset Buffer



Note: This figure compares the estimates of heterogeneity by assets in the passthrough of income shocks to consumption. Parker et al. (2013), Fagereng, Holm and Natvik (2018) and Kueng (2018) use terciles, quartiles, and quintiles respectively. To enable comparability with these prior papers, we calculate the marginal propensity to consume (instead of the elasticity of consumption to income) using their respective bin cutoffs. Our paper, Parker et al. (2013), and Kueng (2018) measure the MPC on nondurables. Fagereng, Holm and Natvik (2018) measures the MPC on total consumption. See Section 3.5 for details.

Simplest textbook model: Huggett (1993) with agg. risk

- Continuum of agents i, heterog. in $(b_{i,t}, y_{i,t})$, $y_{i,t}$ = idios. risk, agg. shock z_t
- State of the economy: distribution $G_t(b, y)$ and agg. shock z_t
- Households choose consumption $c_{i,t}$ to maximize

$$v_{i,0} = \max_{\{c_{i,t}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_{i,t}) \quad \text{subject to}$$

$$b_{i,t+1} = (1 + r_t)(b_{i,t} + y_{i,t} z_t - c_{i,t}), \quad y_{i,t+1} \sim \mathcal{T}_y(\cdot | y_{i,t}), \quad b_{i,t+1} \geq \underline{b}$$

- Save ($b_{i,t} > 0$) or borrow ($b_{i,t} < 0$) at interest rate r_t
- Market clearing: interest rate r_t such that

$$\int b'_t(b, y, z_t)\, dG_t(b, y) = 0, \quad \text{all } t$$

Note: agent problem depends on G_t only through low-dimensional price r_t

More compact notation: individual states s, prices p

- Continuum of agents i, heterogeneous in $s = (b, y)$
- State of the economy: distribution $G_t(s)$ and agg. shock z_t
- Price vector p_t, here only one price $p_t = r_t$
- Households choose consumption $c_{i,t}$ to maximize

$$v_{i,0} = \max_{\{c_{i,t}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_{i,t}) \quad \text{subject to}$$

$$s_{i,t+1} \sim \mathcal{T}_s(\cdot | s_{i,t}, c_{i,t}, z_t, p_t) = \text{budget constraint + income process}$$

- Market clearing: price p_t (interest rate) such that

$$\int b'_t(s, z_t)\, dG_t(s) = 0, \quad \text{all } t$$

Note: agent problem depends on G_t only through low-dimensional price p_t

Even more compact notation: equilibrium price functional

- Continuum of agents i, heterogeneous in $s = (b, y)$
- State of the economy: distribution $G_t(s)$ and agg. shock z_t
- Households choose consumption $c_{i,t}$ to maximize

$$v_{i,0} = \max_{\{c_{i,t}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_{i,t}) \quad \text{subject to}$$

$$s_{i,t+1} \sim \mathcal{T}_s(\cdot | s_{i,t}, c_{i,t}, z_t, p_t)$$

- Low-dimensional equilibrium price functional

$$p_t = P^*(G_t, z_t), \qquad z_{t+1} \sim \mathcal{T}_z(\cdot | z_t)$$

Note: agent problem depends on G_t only through low-dim. price functionals

Even more compact notation: equilibrium price functional

- Continuum of agents i, heterogeneous in $s = (b, y)$
- State of the economy: distribution $G_t(s)$ and agg. shock z_t
- Households choose consumption $c_{i,t}$ to maximize

$$v_{i,0} = \max_{\{c_{i,t}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_{i,t}) \quad \text{subject to}$$

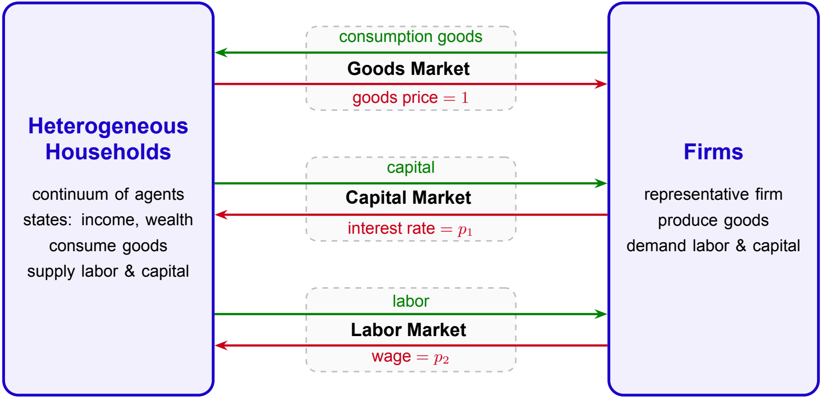
$$s_{i,t+1} \sim \mathcal{T}_s(\cdot | s_{i,t}, c_{i,t}, z_t, p_t)$$

- Low-dimensional equilibrium price functional

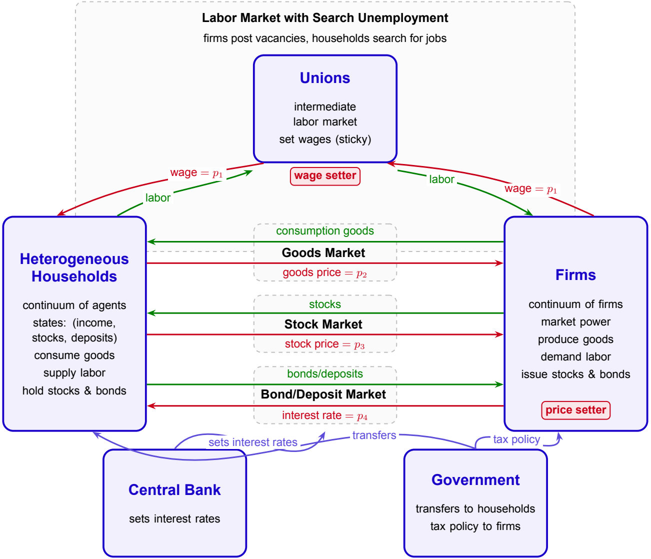
$$p_t = P^*(G_t, z_t), \qquad z_{t+1} \sim \mathcal{T}_z(\cdot | z_t)$$

Generalizes to $s \in \mathbb{R}^{n_s}$, $z \in \mathbb{R}^{n_z}$, $p \in \mathbb{R}^{n_p}$, reward function $R(s, a, z, p)$, a = actions

Adding production: Krusell and Smith (1998)



HANK model with sticky wages & prices, portfolio choice



Central findings

Can’t do justice here but see these review articles

ANNUAL REVIEWS

Annual Review of Economics

Fiscal and Monetary Policy with Heterogeneous Agents

Adrien Auclert,[1,5,6] Matthew Rognlie,[2,3,5] and Ludwig Straub[4,5,6]

Annu. Rev. Econ. 2025. 17:539–62

First published as a Review in Advance on May 7, 2025

The *Annual Review of Economics* is online at economics.annualreviews.org

https://doi.org/10.1146/annurev-economics-091624-044646

JEL codes: E21, E24, E32, E52, E62

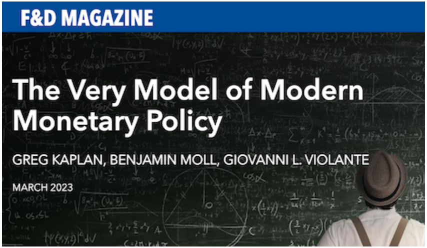
OPEN ACCESS CC BY

Keywords

fiscal policy, monetary policy, heterogeneous agents, HANK, fiscal multiplier, sequence space

Abstract

In the past decade, a new paradigm for fiscal and monetary policy analysis has emerged, combining the canonical macro model of income and wealth inequality with the New Keynesian model. These heterogeneous-agent New Keynesian (HANK) models feature new transmission channels and allow for the joint study of aggregate and distributional effects. We review key developments in this literature through the lens of a canonical HANK model. Monetary and balanced-budget fiscal policy have similar aggregate effects as in the standard New Keynesian model, while deficit-financed fiscal policy is much more expansionary. We discuss the split between direct and indirect effects of policy as well as the implications of cyclical income risk, maturity structure, nominal assets, behavioral frictions, and many other extensions to the model. Throughout, we highlight the benefits of using sequence-space methods to solve and analyze this class of models.

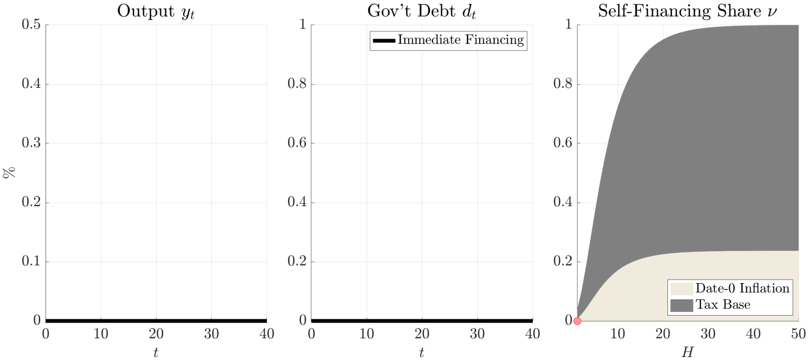


HANK models impart new lessons about redistribution and the heterogeneous effects of monetary policy and shed new light on traditional central bank objectives of inflation control and output stabilization. Here are four broad lessons, and some preliminary thoughts, on how HANK models may illuminate our current high-inflation environment.

Lesson 1: Predicting indirect policy impacts

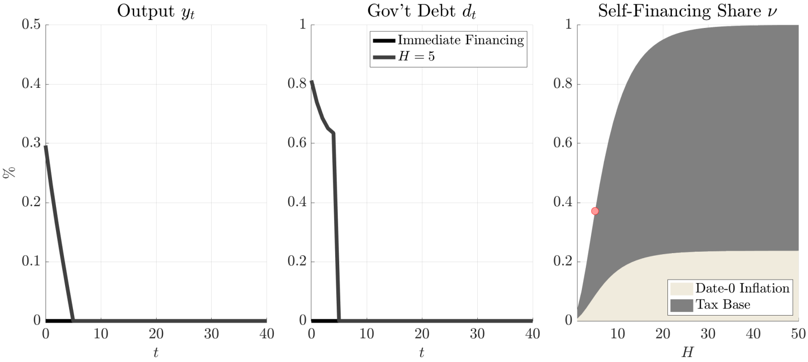
One central finding: fiscal stimulus is partly self-financing

Angeletos, Lian, and Wolf (2024) "Can Deficits Finance Themselves?"



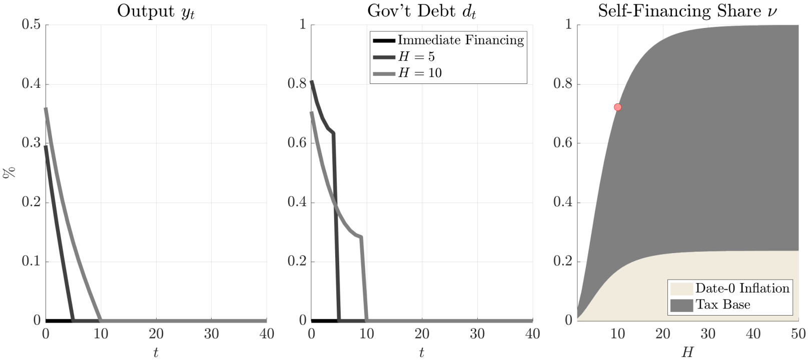
One central finding: fiscal stimulus is partly self-financing

Angeletos, Lian, and Wolf (2024) "Can Deficits Finance Themselves?"



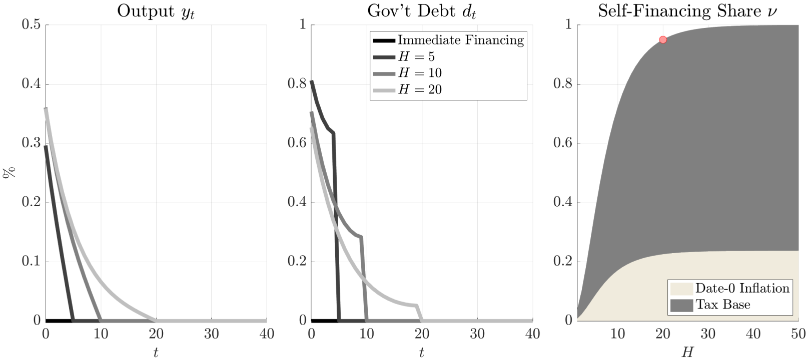
One central finding: fiscal stimulus is partly self-financing

Angeletos, Lian, and Wolf (2024) "Can Deficits Finance Themselves?"



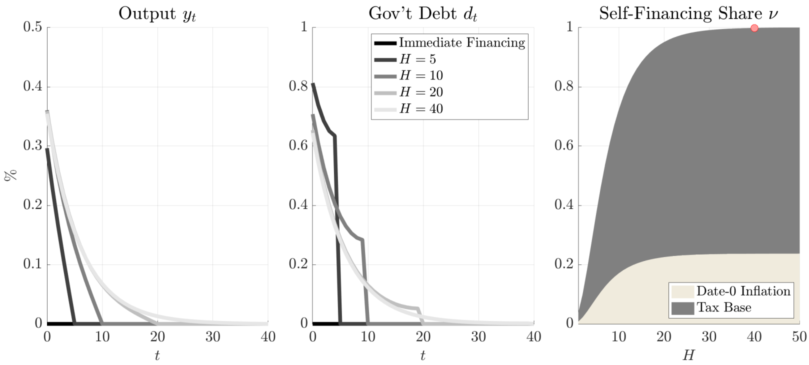
One central finding: fiscal stimulus is partly self-financing

Angeletos, Lian, and Wolf (2024) "Can Deficits Finance Themselves?"



One central finding: fiscal stimulus is partly self-financing

Angeletos, Lian, and Wolf (2024) "Can Deficits Finance Themselves?"



Paper's quantitative exercise: in practice, stimulus is **40-95%** self-financing

A key difficulty in

HA models with aggregate risk

A key difficulty in HA models with aggregate risk

- Key problem: rational expectations + general equilibrium
  $\Rightarrow$ distribution = state variable in Bellman equation ("Master equation")
  - true even though households/firms only care about prices
  - intuition: equilibrium prices are not Markov, only the distribution is
    $\Rightarrow$ forecast prices by forecasting distributions
- Despite recent impressive advances, still no general, efficient global solution method for HA models with aggregate risk
- This still really holds back HA literature, e.g. ~~non-linearities, crises~~
- Bianchi and Kaplan (2026): substantial non-linearities even for small shocks

Key difficulty: equilibrium prices are not Markov

- Discretize individual state $s \in \{s_1, ..., s_J\}$ with $J = J_1 \times ... \times J_n$
- Value function, distribution, etc are J-dimensional vectors

$$\mathbf{v}_t = \begin{bmatrix} v_t(s_1) \\ \vdots \\ v_t(s_J) \end{bmatrix}, \qquad \mathbf{g}_t = \begin{bmatrix} g_t(s_1) \\ \vdots \\ g_t(s_J) \end{bmatrix}$$

- Consumption policy $c = \pi(s, \cdot) \Rightarrow J \times J$ transition matrix for s

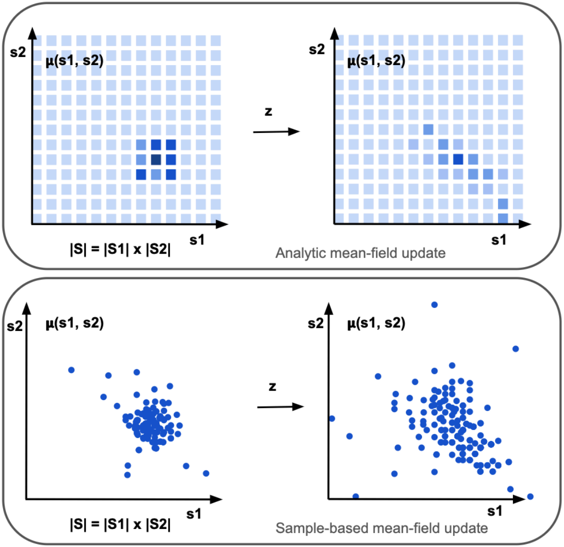
$$\mathbf{A}_{\pi,z_t} \quad \text{with entries} \quad \Pr(s_{i,t+1} = s_{j'} | s_{i,t} = s_j) = \mathcal{T}_s(s_{j'} | s_j, \pi(s_j, \cdot), z_t, p_t)$$

- Law of motion for distribution $\mathbf{g}_t$ (Chapman-Kolmogorov equation)

$$\mathbf{g}_{t+1} = \mathbf{A}_{\pi,z_t}^{\top} \mathbf{g}_t, \quad z_{t+1} \sim \mathcal{T}_z(\cdot | z_t)$$

- Note: high-dimensional state $(\mathbf{g}_t, z_t)$ is Markov

Tracking the distribution



Key difficulty: equilibrium prices are not Markov

- Equilibrium prices satisfy

$$p_t = P^*(\mathbf{g}_t, z_t)$$
$$\mathbf{g}_{t+1} = \mathbf{A}_{\pi,z_t}^\top \mathbf{g}_t$$
$$z_{t+1} \sim \mathcal{T}_z(\cdot|z_t)$$

- Difficulty: while $(\mathbf{g}_t, z_t)$ is Markov, low-dimensional p_t is not Markov!

- Dynamic programming can only handle Markov states $\Rightarrow$ Master equation

$$V(s, \mathbf{g}, z) = \max_c u(c) + \beta \mathbb{E}\left[V(s', \mathbf{g}', z')|s, \mathbf{g}, z\right] \quad \text{s.t.} \quad s' \sim \mathcal{T}_s(\cdot|s, c, z, P^*(\mathbf{g}, z))$$

- Without Markov transition prob's: cannot even write Bellman equation!

- But what if there was a way to approximate value and policy functions with p_t process for which there are **no Markov transition probabilities?**

New Frontiers

New Frontiers in HA macroeconomics

1. Why not just much simpler two-type models (TANK)?

2. Reinforcement Learning for HA macro... as an equilibrium computation device

3. Reinforcement Learning for HA macro... as a model of human learning

Why not just much simpler two-type models (TANK)?

TANK = Two-Agent New Keynesian model à la Campbell-Mankiw

- fraction λ are myopic "spenders" with MPC $= 1$
- fraction $1 - \lambda$ are permanent-income "savers" with MPC ≈ 0
- calibrate λ to match average MPCs, say $\lambda = 30\%$

$\Rightarrow$ ~~Ricardian equivalence~~, just like HANK

... and these TANK models are, of course, much more tractable and transparent

Debortoli-Galí critique: TANK $\approx$ HANK for aggregate outcomes

- "A suitably specified and calibrated TANK model captures reasonably well [HANK's] implications for aggregate output and the main channels through which aggregate shocks are transmitted."

  https://www.journals.uchicago.edu/doi/full/10.1086/735272

Until recently: HANK literature did not have a very good reply to this critique

Why the Debortoli-Galí critique overreaches

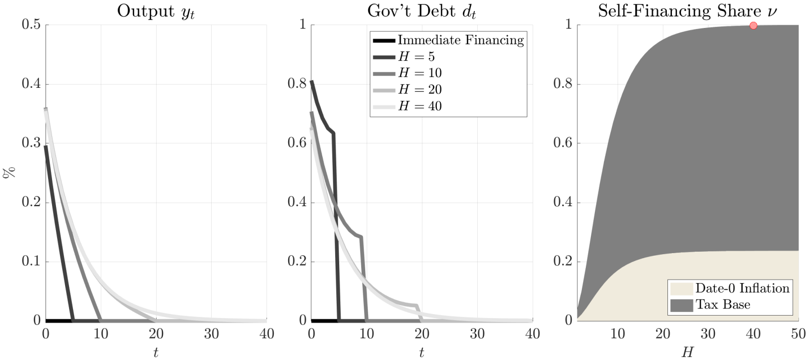
Not my own arguments but those of:

- Angeletos, Lian and Wolf (2024) "Can Deficits Finance Themselves?"
- Auclert, Rognlie and Straub (2024) "The Intertemporal Keynesian Cross"
- Rognlie (2025) "Comment on Debortoli-Galí"
- Bardoczy, Sim and Tischbirek (2025) "The Macroeconomic Effects of Excess Savings"

In a nutshell: **TANK = bad way of breaking Ricardian equivalence that imports many RANK pathologies via permanent-income savers**

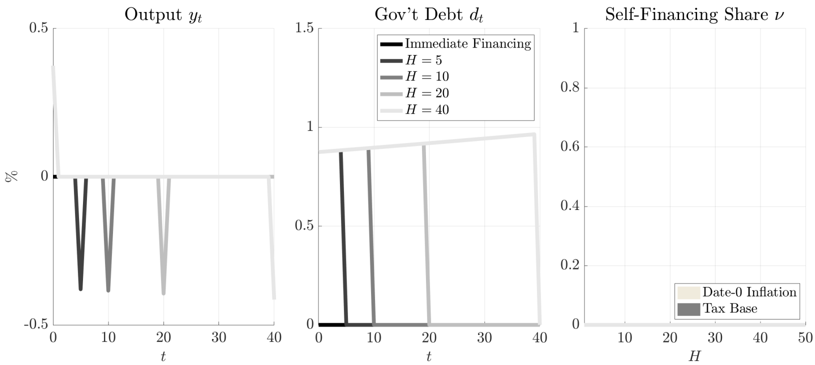
Recall HANK finding: fiscal stimulus is partly self-financing

Angeletos, Lian, and Wolf (2024) "Can Deficits Finance Themselves?"



Paper's quantitative exercise: in practice, stimulus is **40-95%** self-financing

Angeletos et al: TANK looks completely different, self-financing = 0



Rognlie's discussion of Debortoli-Galí: *"I think [DG] overreaches – and that, indeed, TANK is a far weaker substitute for HANK than this paper indicates. [...] As I will show, reasonable alternative shocks – such as government spending shocks, financed by debt that is not paid off too quickly – open up a large gap between TANK and HANK."*

What's going on? Intertemporal MPCs

$$\mathbf{M} = \begin{pmatrix} M_{00} & M_{01} & M_{02} & M_{03} & \cdots \\ M_{10} & M_{11} & M_{12} & M_{13} & \cdots \\ M_{20} & M_{21} & M_{22} & M_{23} & \cdots \\ M_{30} & M_{31} & M_{32} & M_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Intertemporal MPCs

Response to income increase at $s = 0$

$$\mathbf{M} = \begin{pmatrix} M_{00} & M_{01} & M_{02} & M_{03} & \cdots \\ M_{10} & M_{11} & M_{12} & M_{13} & \cdots \\ M_{20} & M_{21} & M_{22} & M_{23} & \cdots \\ M_{30} & M_{31} & M_{32} & M_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Intertemporal MPCs

Response to income increase at $s = 1$

$$\mathbf{M} = \begin{pmatrix} M_{00} & M_{01} & M_{02} & M_{03} & \cdots \\ M_{10} & M_{11} & M_{12} & M_{13} & \cdots \\ M_{20} & M_{21} & M_{22} & M_{23} & \cdots \\ M_{30} & M_{31} & M_{32} & M_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Intertemporal MPCs

Response to income increase at $s = 2$

$$\mathbf{M} = \begin{pmatrix} M_{00} & M_{01} & M_{02} & M_{03} & \cdots \\ M_{10} & M_{11} & M_{12} & M_{13} & \cdots \\ M_{20} & M_{21} & M_{22} & M_{23} & \cdots \\ M_{30} & M_{31} & M_{32} & M_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

iMPCs for RANK and TANK with spender share λ

$$\mathbf{M}^{\text{RA}} = \begin{pmatrix} (1-\beta) & (1-\beta)\beta & (1-\beta)\beta^2 & \cdots \\ (1-\beta) & (1-\beta)\beta & (1-\beta)\beta^2 & \\ (1-\beta) & (1-\beta)\beta & (1-\beta)\beta^2 & \\ \vdots & & & \ddots \end{pmatrix}, \qquad \mathbf{M}^{\text{TA}} = (1-\lambda)\mathbf{M}^{\text{RA}} + \lambda\mathbf{I}$$

- Note: $\mathbf{M}^{\text{RA}}$ has two weird properties:
  1. columns don't converge to zero as $t \to \infty$: spend income very slowly
  2. rows converge to zero very slowly (β^s): responsive to infinite future
- **$\mathbf{M}^{\text{TA}}$ inherits these properties** $\Rightarrow$ reason for zero self-financing!
- $\Rightarrow$ **TANK = bad way of breaking Ricardian equivalence that imports many RANK pathologies via permanent-income savers**
- Auclert, Rognlie and Straub (2024), Bardoczy, Sim and Tischbirek (2025): similar findings for cumulative multipliers and depletion of excess savings

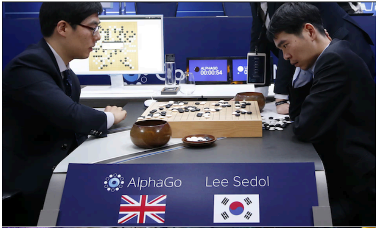
Reinforcement learning for HA macro

Reinforcement learning for HA macro

1. RL as an equilibrium computation device

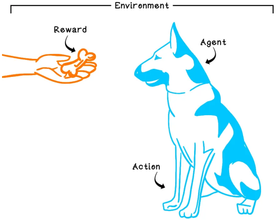
2. RL as a model of human learning

RL: learning value & policy functions in incompletely-known Markov decision processes from Monte Carlo simulation (a.k.a. “approximate DP”)



Playing Atari with Deep Reinforcement Learning

Volodymyr Mnih Koray Kavukcuoglu David Silver Alex Graves Ioannis Antonoglou
Daan Wierstra Martin Riedmiller
DeepMind Technologies
{vlad,koray,david,alex.graves,ioannis,daan,martin.riedmiller} @ deepmind.com



Sutton and Barto (2018) “Reinforcement Learning: An Introduction”
Zhao (2025) “Mathematical Foundations of Reinforcement Learning”

Computing an expected value

Random variable x

How compute expected value $\mathbb{E}[x]$? Two approaches:

1. **Exact:** know probability distribution $f(x) \Rightarrow$ calculate

$$\mathbb{E}[x] = \int x f(x) dx$$

2. **Monte Carlo:** don't know f but can sample $\{x_1, x_2, \ldots, x_N\}$

$$\mathbb{E}[x] \approx \bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

   Or update incrementally (stochastic approximation method):

$$\bar{x}_k = \frac{1}{k} \sum_{n=1}^{k} x_n \quad \text{satisfies} \quad \bar{x}_k = \bar{x}_{k-1} + \frac{1}{k} \left(x_k - \bar{x}_{k-1} \right), \quad \frac{1}{k} = \text{“learning rate”}$$

Computing a value function

For now: eliminate actions and individual states

$$v(p_0) = \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(p_t)\right], \quad p_t = \text{exogenous stochastic process}$$

Two approaches:

1. **Dynamic programming:** p_t Markov and know $f(p'|p)$

$$v(p) = u(p) + \beta \int v(p') f(p'|p) dp'$$

Computing a value function

For now: eliminate actions and individual states

$$v(p_0) = \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(p_t)\right], \quad p_t = \text{exogenous stochastic process}$$

Two approaches:

1. **Dynamic programming:** p_t Markov and know $f(p'|p)$

$$v(p) = u(p) + \beta \int v(p') f(p'|p) dp'$$

2. **Monte Carlo:** don't know f but sample N trajectories $\{p_t^n\}_{t=0}^T$

$$v(p_0) \approx \widehat{v}(p_0) = \frac{1}{N} \sum_{n=1}^{N} \sum_{t=0}^{T} \beta^t u(p_t^n)$$

Computing a value function

For now: eliminate actions and individual states

$$v(p_0) = \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(p_t)\right], \quad p_t = \text{exogenous stochastic process}$$

Two approaches:

1. **Dynamic programming:** p_t Markov and know $f(p'|p)$

$$v(p) = u(p) + \beta \int v(p') f(p'|p) dp'$$

2. **Monte Carlo:** don't know f but sample N trajectories $\{p_t^n\}_{t=0}^T$

$$v(p_0) \approx \widehat{v}(p_0) = \frac{1}{N} \sum_{n=1}^{N} \sum_{t=0}^{T} \beta^t u(p_t^n)$$

Can also extend to compute optimal policy (Howard): policy gradient method

Computing a value function

For now: eliminate actions and individual states

$$v(p_0) = \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(p_t)\right], \quad p_t = \text{exogenous stochastic process}$$

Two approaches:

1. **Dynamic programming:** p_t Markov and know $f(p'|p)$

$$v(p) = u(p) + \beta \int v(p') f(p'|p) dp'$$

2. **Temporal difference learning:** update v incrementally at each t

$$\widehat{v}_{t+1}^k(p_t^k) = \widehat{v}_t^k(p_t^k) + \alpha \left[\left(r(p_t^k) + \beta \widehat{v}_t^k(p_{t+1}^k)\right) - \widehat{v}_t^k(p_t^k)\right]$$

Can also extend to compute optimal policy (Howard): policy gradient method

RL as an equilibrium computation device

Two papers and an open-source library

Structural Reinforcement Learning for Heterogeneous Agent Macroeconomics

Yucheng Yang[*,1] Chiyuan Wang[*,2] Andreas Schaab[3] Benjamin Moll[4]

https://arxiv.org/abs/2512.18892

Recurrent Structural Policy Gradient for Partially Observable Mean Field Games

---

Clarisse Wibault[1 2] Johannes Forkel[1] Sebastian Towers[1] Tiphaine Wibault[3] Juan Duque[4] George Whittle[2] Andreas Schaab[5] Yucheng Yang[6] Chiyuan Wang[7] Maike Osborne[2] Benjamin Moll[† 8] Jakob Foerster[† 1]

https://arxiv.org/abs/2602.20141

⚡ MFAX: Mean-Field Games in JAX ⚡

https://clarisse-wibault.github.io/rspg/

Google Colab notebook for macroeconomics environment

Sidestepping the Master Equation via RL

"Structural RL": hybrid w RL about equilibrium prices but DP for individual states

Outcome: efficient & flexible global solution method for HA models with agg risk

- solves problems traditional methods struggle with:

1. non-trivial market clearing (Huggett with agg. risk) ≈ 1 min on Google Colab
2. portfolio choice ≈ 1 min
3. HANK with forward-looking price/wage Phillips curve ≈ 4 min
4. ...

How does it work?

- in contrast to dynamic programming, RL can handle non-Markov states
- replace distribution with low-dim. prices in state space, grid-based not DNNs
- efficient market clearing using policy functions (= demand curves)

Sidestepping the Master Equation via RL

Recall: states $s = (b, y)$, price $p = r$, agents choose $c_{i,t}$ to maximize

$$v_{i,0} = \max_{\{c_{i,t}\}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(c_{i,t})\right] \text{ s.t. } s_{i,t+1} \sim \mathcal{T}_s(\cdot | s_{i,t}, c_{i,t}, z_t, p_t), \quad p_t = P^*(G_t, z_t)$$

Assumption 1: agents observe prices p_t but not distribution $G_t(s)$ ▸ Wold representation

Assumption 2: consumption policy π does not condition on price histories

$$c_{i,t} = \pi(s_{i,t}, z_t, p_t)$$

Extension: keep track of price histories (h lags or RNN) $\Rightarrow$ similar results

Similarity to standard RL: don't know transition probabilities of p_t but can sample

Difference to standard RL: agents know individual dynamics

- know $u, \mathcal{T}_s$: utility function & budget constraint, RE about income process
- want hybrid method that takes advantage of this structural knowledge

RL approach: approximate $\mathbb{E}$ by sampling N model-generated p trajectories

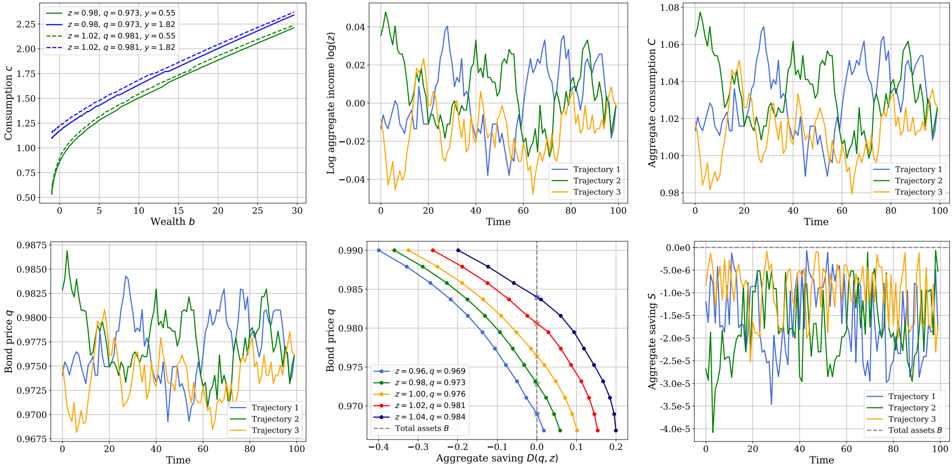
Computational experiments

- Efficient implementation in JAX for GPUs, run on Google Colab
- Algorithm = stochastic (Monte Carlo) $\Rightarrow$ present averages over multiple runs

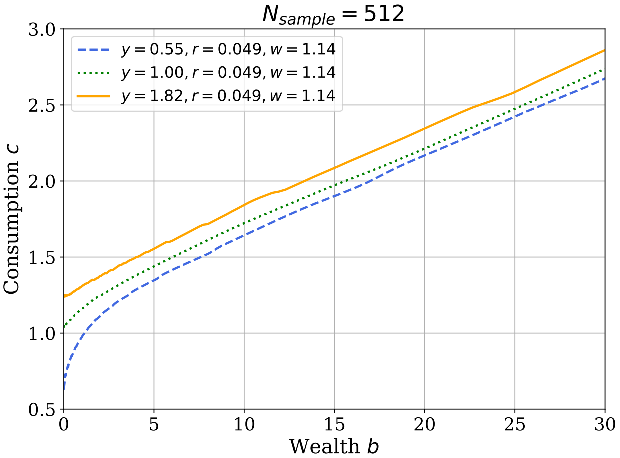
Model	Average converge epoch	# Runs	Average Runtime (sec)
Krusell-Smith	462.3	10	36.77
Huggett with agg. shocks	573.0	10	45.91
Portfolio choice	637.5	10	84.3
HANK with agg. shocks	707.5	10	246.49
Partial Equilibrium (Huggett)	355.8	10	24.50

Note: all experiments were implemented on the A100 GPU on Google Colab

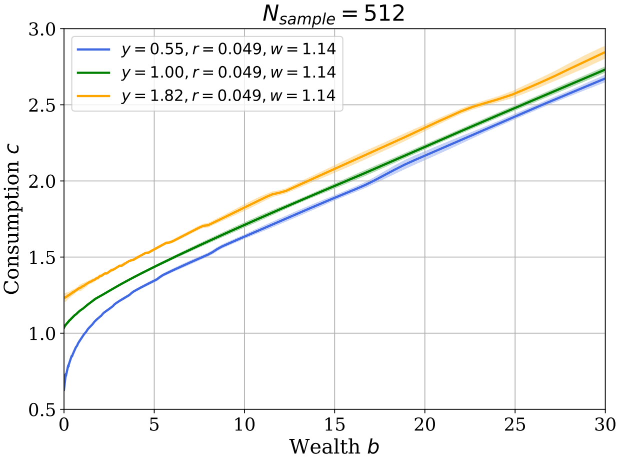
Some simulated trajectories under the optimal policy



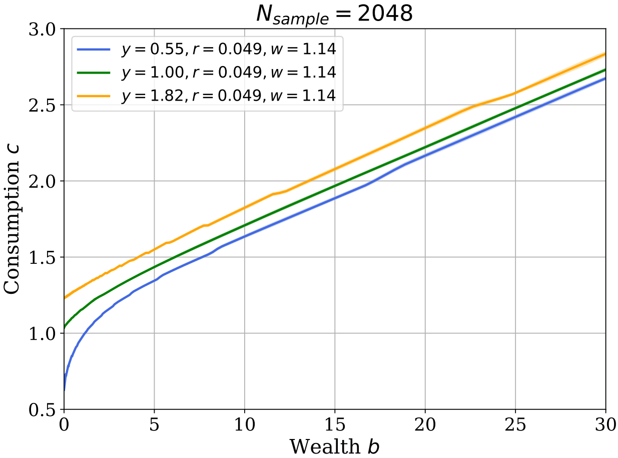
Consumption policy function: single run



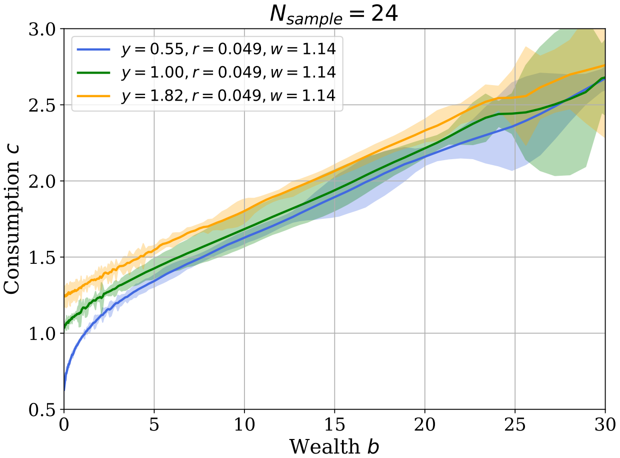
Consumption policy function: multiple runs



Larger sample size $N \Rightarrow$ more precise estimate



Smaller sample size $N \Rightarrow$ noisier estimate



RL as a model of human learning

Interesting question: Could RL form the basis for an empirically realistic theory of expectations formation?

SRL & RSPG papers: RL by the computational economist, not the model agents

But can these approaches perhaps be tweaked to model human learning and expectation formation? Discuss this in conclusion of SRL paper:

- RL has attractive idea at its core: learn by sampling

Interesting question: Could RL form the basis for an empirically realistic theory of expectations formation?

SRL & RSPG papers: RL by the computational economist, not the model agents

But can these approaches perhaps be tweaked to model human learning and expectation formation? Discuss this in conclusion of SRL paper:

- RL has attractive idea at its core: learn by sampling

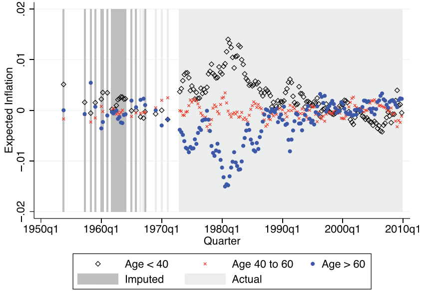


FIGURE I

Inflation Expectations by Age Group Relative to Cross-Sectional Mean

Interesting question: Could RL form the basis for an empirically realistic theory of expectations formation?

SRL & RSPG papers: RL by the computational economist, not the model agents

But can these approaches perhaps be tweaked to model human learning and expectation formation? Discuss this in conclusion of SRL paper:

- RL has attractive idea at its core: learn by sampling
- Literature: RL underpins substantial share of human and animal learning
  see e.g. Niv, Glimcher, Caplin-Dean, Gershman-Daw, Baberis-Jin,...

Interesting question: Could RL form the basis for an empirically realistic theory of expectations formation?

SRL & RSPG papers: RL by the computational economist, not the model agents

But can these approaches perhaps be tweaked to model human learning and expectation formation? Discuss this in conclusion of SRL paper:

- RL has attractive idea at its core: learn by sampling
- Literature: RL underpins substantial share of human and animal learning
  see e.g. Niv, Glimcher, Caplin-Dean, Gershman-Daw, Baberis-Jin,...

Required tweaks would likely be substantial:

1. convert to fully online, incremental RL algorithm, with agents updating policies and value estimates continuously while interacting with their environment
2. incorporate biased or experience-weighted sampling
3. not just model-free RL

Summary

Introduction to Heterogeneous Agent Macroeconomics

- some typical models (HANK, ...)
- some central findings
- a key difficulty: prices aren't Markov

New Frontiers

1. TANK $\neq$ HANK for some important questions
2. RL as an equilibrium computation device
3. RL as a model of human learning