Deep Learning Solutions to Master Equations for Continuous Time Heterogeneous Agent Macroeconomic Models

Jonathan PAYNE

Princeton

joint work with Zhouzhou GU, *Mathieu* LAURIÈRE *and Sebastian* MERKEL

DSE 2023

August 24th 2023

Introduction

- Great progress in using deep learning to solve discrete time, heterogeneous agent economies.
- We focus on continuous time, heterogeneous agent economies ("mean field games"):
  - Consider economies with aggregate shocks, long-term assets, portfolio choice, illiquidity.
  - Consider different finite dimensional approximations to the distribution (finite agents, projection).
  - Solve the resulting high dimensional PDE(s) using neural network approximations.
- *How do we test it?* Compare to solutions for canonical economic models.
  (e.g. Aiyagari (1994), Krusell and Smith (1998), Basak and Cuoco (1998), and extensions).
- *What economic question are we answering?* The impact of housing policy on inequality.
  (in preliminary follow-up paper Gu & Payne (2023) "Housing Policy and Inequality").

Outline

Outline

Environment

- Continuous time, infinite horizon economy.
- Populated by $I = [0, 1]$ households who consume goods, supply labor, and save wealth.
- Representative firm rents capital and labor to produce goods by $Y_t = e^{z_t} K_t^{\alpha} L_t^{1-\alpha}$, where
  - K_t is capital hired, L_t is labor hired,
  - z_t is productivity (the exogenous aggregate state variable): follows $dz_t = \eta(\bar{z} - z_t)dt + \sigma dB_t^0$
  - B_t^0 is a common Brownian motion process; it generates filtration $\mathcal{F}_t^0$.
- Competitive markets for goods (numeraire), capital rental (return r_t), and labor (wage w_t).

Household Problem

- Household i has idiosyncratic state $x_t^i = (a_t^i, n_t^i)$, where a_t^i is wealth, n_t^i is labor endowment.
- Given belief about the price processes, household chooses consumption $c = \{c_t^i\}_{t\geq 0}$ to solve:

$$\begin{aligned} &\max_{\{c_t^i\}_{t\geq 0}} \mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} u(c_t^i) dt\right] \\ &s.t. \quad da_t^i = (\tilde{w}_t n_t^i + \tilde{r}_t a_t^i - c_t^i) dt =: \mu_t^a dt, \quad a_t^i \geq \underline{a} \\ &\qquad\quad n_t^i \in \{n_1, n_2\}, \text{ switches at idiosyncratic Poisson rate } \lambda(n_t^i) \end{aligned} \tag{1}$$

    - $u(c) = c^{1-\gamma}/(1-\gamma)$: utility function, ρ: discount rate,
    - $(\tilde{r}, \tilde{w}) = \{\tilde{r}_t, \tilde{w}_t\}_{t\geq 0}$ are agent beliefs about prices processes,
    - $\underline{a}$: borrowing limit.

- Let $G_t = \mathcal{L}(x_t^i | \mathcal{F}_t^0)$ and g_t be population distribution and density of $(x_t^i)_{i\in I}$, given history $\mathcal{F}_t^0$
    - Non degenerate because households get uninsurable idiosyncratic labor endowment shocks.

Equilibrium

Definition: Given an initial density g_0, an equilibrium for this economy consists of a collection of $\mathcal{F}_t^0$-adapted stochastic process, $\{c_t^i, g_t, z_t, q_t := [r_t, w_t] : t \geq 0, i \in I\}$, such that:

1. Given belief that price process is $\tilde{q}$, household consumption process, c_t^i, solves problem (3),
2. Given belief that price process is $\tilde{q}$, firm choose capital and labor optimally:

$$r_t = e^{z_t} \partial_K F(K_t, L) - \delta, \qquad w_t = e^{z_t} \partial_L F(K_t, L)$$

3. The price vector $q_t = [r_t, w_t]$ satisfies market clearing conditions:

$$K_t = \sum_{j \in \{1,2\}} \int a g_t(a, n_j) da, \qquad L = \sum_{j \in \{1,2\}} \int n_j g_t(a, n_j) da$$

4. Agent beliefs about the price process are consistent: $\tilde{q} = q$

Equilibrium (Combining Equations For Prices)

Definition: Given an initial density g_0, an equilibrium for this economy consists of a collection of $\mathcal{F}_t^0$-adapted stochastic process, $\{c_t^i, g_t, q_t := [r_t, w_t], z_t : t \geq 0, i \in I\}$, such that:

1. Given belief that price process is $\tilde{q}$, household consumption process, c_t^i, solves problem (3),
2. The price vector $q_t = [r_t, w_t]$ satisfies:

$$q_t = \begin{bmatrix} r_t \\ w_t \end{bmatrix} = \begin{bmatrix} e^{z_t}\partial_K F(K_t, L) - \delta \\ e^{z_t}\partial_L F(K_t, L) \end{bmatrix} =: Q(z_t, g_t), \text{ where } K_t = \sum_{j\in\{1,2\}} \int a g_t(a, n_j) da$$

3. Agent beliefs about the price process are consistent: $\tilde{q} = q$

Having a closed form expression for prices in terms of (z_t, g_t) makes problem very tractable.

Recursive Representation of Equilibrium

- Aggregate states: $x = (z, g)$, individual states: (a, n), household value fn: $V(a, n, z, g)$.
- Given a belief $dg_t(x) = \tilde{\mu}_g(z_t, g_t)dt$, household at $x = (a, n)$ choose c to solve HJBE:

$$0 = \max_c \Big\{ -\rho V(a,n,z,g) + u(c) + \partial_a V(a,n,z,g)(w(z,g)n + r(z,g)a - c) \\ +\lambda(n)\left(V(a,\check{n},z,g) - V(a,n,z,g)\right) + \partial_z V(a,n,z,g)\mu^z(z) + 0.5\left(\sigma^z\right)^2 \partial_{zz} V(a,n,z,g) \\ + \int_{\mathcal{X}} \frac{\partial V}{\partial g}(y,z,g)\tilde{\mu}^g(y,z,g)dy \Big\}, \quad s.t. \quad \text{BC: } \frac{\partial V}{\partial a}\Big|_{a=\underline{a}} \geq u'(wn + r\underline{a})$$

where $\check{n}$ is complement of n.

- For optimal policy rule, $c^*(a, n, z, g; \tilde{\mu}^g)$, for z_t, population density, g, evolves by KFE:

$$dg_t(a,n) = \underbrace{\left[-\partial_a[(w(z,g)n + r(z,g)a - c^*)g_t(a,n)] - \lambda(n)g_t(a,n) + \lambda(\check{n})g_t(a,\check{n})\right]}_{=:\mu^g(a_t,n_t,z_t,g_t;\tilde{\mu}^g)}dt$$

- In equilibrium $\tilde{\mu}^g = \mu^g$.

Recursive Representation of Equilibrium (Soft Borrowing Constraint)

- Aggregate states: $x = (z, g)$, individual states: (a, n), household value fn: $V(a, n, z, g)$.
- Given a belief $dg_t(x) = \tilde{\mu}_g(z_t, g_t)dt$, household at $x = (a, n)$ choose c to solve HJBE:

$$
\begin{aligned}
0 = \max_c \Big\{ &-\rho V(a,n,z,g) + u(c) - \mathbf{1}_{a_t \le \underline{a}} \psi(a_t) + \partial_a V(a,n,z,g)(w(z,g)n + r(z,g)a - c) \\
&+ \lambda(n)\left(V(a,\check{n},z,g) - V(a,n,z,g)\right) + \partial_z V(a,n,z,g)\mu^z(z) + 0.5\left(\sigma^z\right)^2 \partial_{zz} V(a,n,z,g) \\
&+ \int_{\mathcal{X}} \frac{\partial V}{\partial g}(y,z,g)\tilde{\mu}^g(y,z,g)dy \Big\}, \quad s.t. \quad \cancel{\text{BC: } \frac{\partial V}{\partial a}\Big|_{a=\underline{a}} \ge u'(\cdot)}, \; \psi(a) = -\frac{1}{2}\kappa(a - \underline{a})^2
\end{aligned}
$$

where $\check{n}$ is complement of n.

- For optimal policy rule, $c^*(a, n, z, g; \tilde{\mu}^g)$, for z_t, population density, g, evolves by KFE:

$$
dg_t(a,n) = \underbrace{\left[-\partial_a\left[(w(z,g)n + r(z,g)a - c^*)g_t(a,n)\right] - \lambda(n)g_t(a,n) + \lambda(\check{n})g_t(a,\check{n})\right]}_{=:\mu_t^g(c_t^*, a_t, n_t, z_t, g_t; \tilde{\mu}^g)} dt
$$

- In equilibrium $\tilde{\mu}^g = \mu^g$.

“Master Equation” Representation of Equilibrium

- “Master equation” substitutes KFE, market clearing, and belief consistency into HJBE.
- Equilibrium value function $V(a, n, z, g)$ characterized by one PDE:

$$\begin{aligned} 0 = & -\rho V(a,n,z,g) + u(c^*(a,n,z,g)) + \mathbf{1}_{a_t \le \underline{a}} \psi(a_t) \\ & + \partial_a V(a,n,z,g)(w(z,g)n + r(z,g)a - c^*(a,n,z,g)) \\ & + \lambda(x)\left(V(a,\check{n},z,g) - V(a,n,z,g)\right) + \partial_z V(a,n,z,g)\mu^z(z) + 0.5\left(\sigma^z\right)^2 \partial_{zz} V(a,n,z,g) \\ & + \int_{\mathcal{X}} \frac{\partial V}{\partial g}(y,z,g)\mu^g(c^*(y,z,g),y,z,g)dy =: \mathcal{L}V \end{aligned}$$

where the optimal control c^* is characterised by:

$$u'(c^*(a,n,z,g)) = \partial_a V(a,n,z,g).$$

Solution Outline

- Goal: solve Master equation numerically
- Problem: Master equation contains an infinite dimensional derivative.
- Solution: three main ingredients:
  1. High but finite dimensional approximation to distribution and Master equation:
     (i). Replace continuum of agents by a finite population of agents, or
     (ii). Project distribution onto a finite dimensional set of basis functions (e.g. indicator functions, eigenfunctions, Chebyshev polynomials ...).
  2. Parameterize V by neural network, and
  3. Train the parameters to minimize the (approximate) master equation residual.

Ingredient 1: Comparing Finite Population and Projection

	Finite Population More	**Projection** More
Distribution approx.	Finite collection of agents $\hat{g} \approx \{(a_t^i, n_t^i) : i \leq N\}$	Finite projection coefficients $\hat{g}_t(x) \approx \sum_{i=1}^N \alpha_t^i h^i(x)$
KFE approximation	Evolution of other agents' states	Evolution of projection coefficients α_t^i (complex)
Capital market	Sum up agent positions $K_t = \sum_i^N a_t^i$	Approximate the integral $K_t = \sum_j \int a\hat{g}_t(a, n_j)da$

Finite agent approach introduces small sample error in aggregates.
Projections have more complicated KFE approximations.

Ingredients 2 & 3: The Algorithm

Approximate value function by neural network $V(x, z, \hat{g}) \approx V(x, z, \hat{g}; \theta)$ with parameters θ.

Starting with an initial θ^0. At iteration n with guess θ^n:

1. Randomly sample $S^n = \{(x_m, z_m, \hat{g}_m)\}_{m \leq M}$ from the state space.

2. Calculate the weighted average error:
$$\mathcal{E}(\theta^n, S^n) = \kappa^e \mathcal{E}^e(\theta^n, S^n) + \kappa^f \mathcal{E}^f(\theta^n, S^n), \quad \text{where}$$
   - $\mathcal{E}^e(\theta^n, S^n) := \frac{1}{M} \sum_{m \leq M} |\hat{\mathcal{L}}(x_m, z_m, \hat{g}_m)|$ is error in Master equation $\hat{\mathcal{L}}$
   - $\mathcal{E}^f(\theta^n, S^n)$ is penalty for "wrong" shape (e.g. penalty for non-concavity of V)

3. Update the NN parameters using "stochastic" gradient descent:
$$\theta^{n+1} = \theta^n - \alpha_n D_\theta \mathcal{E}(\theta^n, S^n)$$

4. Repeat until $\mathcal{E}(\theta^n, S^n) \leq \epsilon$ where ϵ is a precision threshold.

Neural Network Q & A

- ***Q.*** *What type of network architecture do we use?*
  - For finite population approximation, typically feed forward, 5 layers, 64 neurons
  - For projection, used "LSTM" neural network (following the "Deep Galerkin" architecture)
- ***Q.*** *What are the main differences to discrete time?*
  - Need to calculate derivatives rather than expectations (we do this with automatic differentiation)
  - Need to choose where to sample rather than always simulating economy
    (We follow [Gopalakrishna, 2021] and increase sampling where error in master equation is large)
- ***Q.*** *Why do we need shape constraints?*
  - Neural network can find "bad" approximate solutions,
    (E.g. consumption-saving problem has approximate solution $V \approx 0$ for high γ.) More
  - Option: penalize shape that correspond to known "bad" solutions.
  - Option: train ϕ satisfying $V(a, n, z, g) = \phi(a, n, z, g; \theta)(a - \underline{a})^{1-\gamma}$ instead of training V

Neural Network Q & A

- ***Q.*** *What about slowing down the updating?*
  - For projection methods, we use "Howard improvement algorithm" to slow down the rate of updating (fix policy rule for some iterations and just update V).
  - [Duarte, 2018] and [Gopalakrishna, 2021] suggest introducing a "false" time step but so far we have not found this necessary (or found a way to implement at high scale).
  - We use shape constraints as a replacement.
- ***Q.*** *What about imposing symmetry and/or dimension reduction?*
  - [Han et al., 2021] and [Kahou et al., 2021] suggest feeding the distribution through a preliminary neural network that reduces the dimension and imposes symmetry.
  - We find we can solve the problem with and without this approach.
- ***Q.*** *What if we have boundary conditions?*
  - Then we sample separately from boundary and add a loss for the boundary condition.
  - But, we have found replacing inequality boundary conditions with penalties is helpful.

Related Literature

Machine learning for macro-economic models:

- Discrete time (e.g. [Azinovic et al., 2022], [Han et al., 2021], [Maliar et al., 2021], [Kahou et al., 2021], [Bretscher et al., 2022], [Wagner, 2023])
- Discrete time approximation to forward and backward differential stochastic equations (e.g. [Han et al., 2018], [Huang, 2022])
- Continuous time (e.g. [Duarte, 2018], [Gopalakrishna, 2021], [Fernandez-Villaverde et al., 2020])
- Portfolio choice and housing (e.g. [Azinovic and Žemlička, 2023], [Gaegauf et al., 2023])
- *This paper:* solve analytical formulation of continuous time model with distributions.

Machine learning for physics and mean field games:

- Controls or value functions in MFGs (e.g. [Perrin et al., 2022, Germain et al., 2022, Laurière, 2021], [Laurière, 2021, Carmona and Laurière, 2022, Hu and Lauriere, 2022])
- We build on the Deep Galerkin Method (DGM) and Physics Informed Neural Networks (PINNs) (e.g. [Sirignano and Spiliopoulos, 2018], [Raissi et al., 2017], [Li et al., 2022])
- *This paper:* integrates market clearing conditions in DGM and PINN

Testing the Algorithm

- Test version of model with fixed aggregate productivity (Aiyagari (1994)):

	Finite Agent NN	Projection NN
Master equation MSE	3.135×10^{-5}	2.548×10^{-4}

  - Neural network solutions match finite difference solution at steady state and on transition paths.
  - Example plots: comparison to [Ahn et al., 2018] Plots

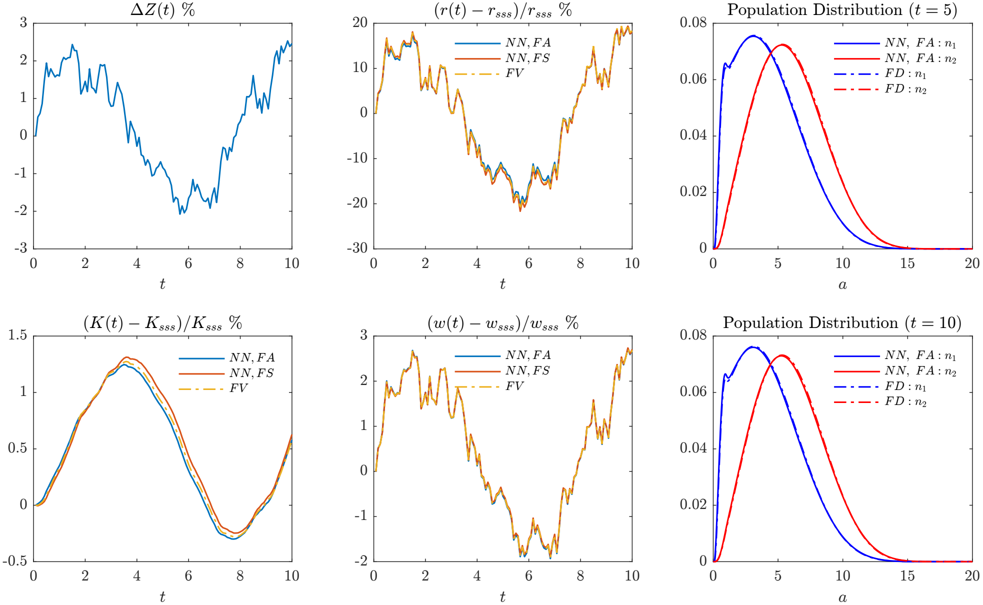
- Test version with stochastic aggregate productivity (Krusell-Smith (1998)):

	Finite Agent NN	Projection NN
Master equation MSE	3.037×10^{-5}	9.639×10^{-5}

  - Neural network solutions generate similar output to traditional methods.
  - Example plots: comparison to [Fernández-Villaverde et al., 2018]

KS: Numerical Results

More Plots



Outline

Environment

- Krusell-Smith has a very simple market clearing condition.
- We now consider how to work with more complicated market clearing.
- As an example, we use our paper Gu-Payne (2023), which studies how housing policy impacts inequality in a model with aggregate risk.
- Gu-Payne (2023) makes a number of changes to the previous model:
  - Introduces an illiquid asset, housing, and
  - Introduces long-term equity.

Environment

- Continuous time, infinite horizon economy.
- Consumption good produced by a "Lucas tree" according to stochastic process:

$$dy_t = \eta(\bar{y} - y_t)dt + \sigma dB_t^0, \tag{2}$$

- Assets: short term bonds in zero net supply, equity in Lucas tree, housing in fixed supply H:
  - "Liquid" competitive markets for goods, bonds (at interest rate r_t), and equity (at price q_t).
  - "Illiquid" housing; trading housing at rate $\iota_{i,t}$ incurs transaction cost: $\Psi(\iota_{i,t}, h_{i,t}) = \frac{1}{2}\psi\iota_{i,t}^2/h_{i,t}$ (price of housing is p_t).
- Population approximated by I of agents (start with finite agent approximation):
  - Get flow utility $u(c_t^i)$ from consuming c_t^i goods and $\zeta_{i,t}\nu(h_{i,t}, a_{i,t})$ from housing $h_{i,t}$, where
  - $\zeta_{i,t} \in \{n_1, n_2\}$ is idiosyncratic housing need ("life-stage") , which switches at rate $\lambda(\zeta_t^i)$.
  - Face collateral borrowing constraint: $a_t \geq -\kappa p_t h_{i,t}$

Agent Problem

- Idiosyncratic states: $x_t^i = [a_t^i, h_t^i, \zeta_t^i]$, a_t^i is liquid wealth, h_t^i is housing, ζ_t^i is housing need.
- Given their beliefs, agent i chooses (c_i, b_i, ι_i) to maximise utility s.t. state evolution:

$$V(x_0^i, z_0) = \max_{c^i, b^i, \iota^i} \mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} (u(c_t^i) + \zeta_{i,t} \nu(h_{i,t}, a_{i,t}) + \mathbf{1}_{a_t \leq -\kappa p_t h} \phi(a_t, h_t)) dt \right], \tag{3}$$

$$s.t. \quad dz_t = \ldots, \quad dx_t^i = \ldots, \quad \Psi(a, h) := -0.5\psi(a + \kappa p h)^2$$

Full HJBE

- FOCs give choices in terms of value function and derivatives:

$$\begin{aligned}
[c_t^i]: \quad & c_t^i = (u')^{-1} \left(\partial V_i / \partial a_t^i \right) \\
[b_t^i]: \quad & b_t^i = \left[\frac{\partial^2 V}{\partial (a_t^i)^2} (\sigma_t^q)^2 \right]^{-1} \left[-\frac{\partial V}{\partial a_i} \left(r_r - \mu_t^q - \frac{y}{q_t} \right) + \frac{\partial^2 V}{\partial (a_t^i)^2} a_t^i (\sigma_t^q)^2 \right. \\
& \left. + \frac{\partial^2 V}{\partial a_i \partial y} \sigma_t^q \sigma_t^y + \sum_j \frac{\partial^2 V}{\partial a_t^i \partial a_t^j} \sigma_t^q \tilde{\sigma}_t^{a_t^j} \right] \qquad (4) \\
[\iota_i^i]: \quad & \iota_i = \frac{h_i}{\psi} \left(\frac{\partial V_i / \partial h_t^i}{\partial V_i / \partial a_t^i} - p_t \right)
\end{aligned}$$

Equilibrium

Definition: Given an initial distribution, g_0, an equilibrium for this economy consists a collection of $\mathcal{F}_t^0$ adapted stochastic processes $\{c_{i,t}, b_{i,t}, \iota_{i,t}, a_{i,t}, g_t, r_t, q_t, p_t, y_t : \ t \geq 0, i \in \mathcal{I}\}$ such that:

1. Given their belief about the price processes $(\tilde{r}, \tilde{q}, \tilde{p})$, individual i's consumption decision $c_{i,t}$, bond holdings $b_{i,t}$, and rate of housing purchase, $\iota_{i,t}$ solve optimization problem,
2. Market clearing conditions are satisfied: (i) goods market $\sum_i c_{i,t} = y_t$, (ii) stock market $\sum_i (a_{i,t} - b_{i,t}) = q_t$, (iii) bond market $\sum_i b_{i,t} = 0$, and (iv) housing market $\sum_i \iota_{i,t} = 0$.
3. Agent beliefs about the price process are consistent with the optimal behaviour of other agents in the sense that $(\tilde{r}, \tilde{q}_t, \tilde{p}_t) = (r_t, q_t, p_t)$.

Master Equation Formulation

- Define $\xi_a := \partial_a V_i(x,z,g)$ and $\xi_h := \partial_h V_i(x,z,g)$.
- Using market clearing and agent optimization:
  - We can get (r,p) in closed form in terms of (ξ_a, ξ_h, z, g) Derivation

$$r - \left(\mu^q + \frac{y}{q}\right) = \frac{q + \mathbf{1}\cdot(\mathbf{M}^{-1}\boldsymbol{\xi}_y)\sigma_q\sigma_y}{\mathbf{1}\cdot(\mathbf{M}^{-1}\boldsymbol{\xi})}, \qquad p = \frac{1}{H}\left(\sum_i \frac{\xi_{h,i}}{\xi_{a,i}} h_i\right), \qquad \mathbf{M}_{ij} := \sigma_q^2 \xi_{i,a_j}$$

  - But we only have $q_t = q(z,g)$ implicitly; only know it must satisfy Ito's lemma. (Non-trivial market clearing condition implies a PDE for q)
- Substituting the equilibrium KFE, market clearing, and the (implict and explicit) pricing expressions, we are left with the following "implicit" master equations:

$$0 = -\rho + \frac{y}{q} + \mu^q + \mu^{\xi_a} + j^{\xi_a} + \sigma^{\xi_a}\sigma^q + \frac{1}{\xi_a}\frac{\partial \phi}{\partial a}, \tag{5}$$

$$0 = -\rho + \frac{\partial \Psi}{\partial h} + \mu^{\xi_h} + j^{\xi_h} + \frac{1}{\xi_h}\frac{\partial \phi}{\partial h}, \qquad \mu^x, \sigma^x, j^x = \text{drift, volatility, jump in } x \,/\, \xi^x \tag{6}$$

Modified Algorithm (new parts in red)

Approximate the value function derivatives (ξ_a, ξ_h) and price triplet (q, μ^q, σ^q) by neural networks with collective parameters θ_ξ, θ_q respectively.

Starting with an initial θ^0. At iteration n with guess θ^n:

1. Sample $S^n = \{(x_m, z_m, \hat{g}_m)\}_{m \leq M}$ from the state space.
2. Update θ_ξ using loss on master equation and then update θ_q using consistency conditions.
3. Repeat until $\mathcal{E}(\theta^n, S^n) \leq \epsilon$ where ϵ is a precision threshold.

Detail on Step 2: Updating θ_ξ and θ_q

2 i. **KFE Block**: Calculate the evolution of the distribution of liquid wealth and housing. (Use change of variable to evolution of wealth and housing shares $\{\eta_i = a_i/A, \varphi_i = h_i/H\}$

2 ii. **Master Equation Block**: Using KFE, evaluate the weighted average error:

$$\mathcal{E}^\xi(\theta_\xi^n, S^{ne}) = \frac{w^a}{M} \sum_{m \le M} |\mathcal{L}^{hm}/\xi_h^m| + \frac{w^h}{M} \sum_{m \le M} |\mathcal{L}^{am}/\xi_a^m|. \tag{7}$$

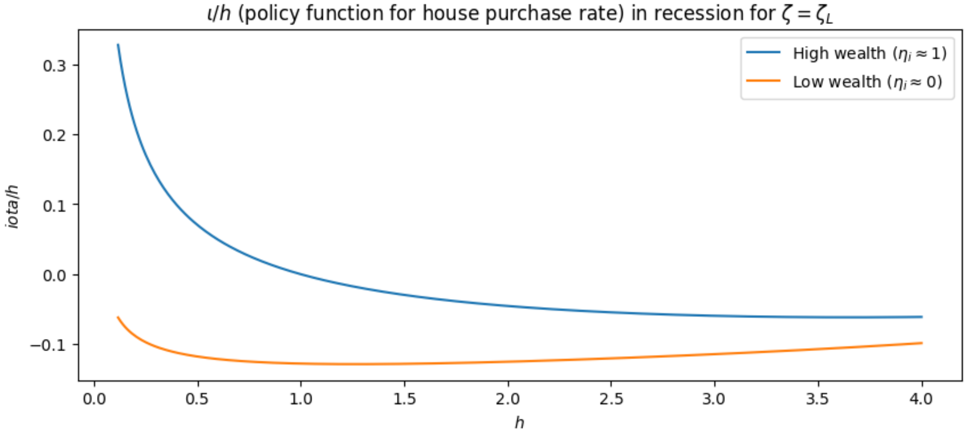
where $\mathcal{L}^{am}$ and $\mathcal{L}^{hm}$ are the error in sample m for the pdes for ξ_a and ξ_h respectively. Update parameters θ_ξ^n using stochastic gradient descent.

2 iii. **Equilibrium consistency**: Using KFE, evaluate the weighted average error by goods market clearing condition and Ito's lemma applied to $q(z, g)$:

$$\mathcal{E}^q(\theta_\xi^n, S^{ne}) = \frac{1}{M} \left(\sum_{m \le M} \epsilon_c \left| \sum_i c_i - y \right| + \epsilon_\mu |\mathcal{L}^{\mu m}| + \epsilon_\sigma |\mathcal{L}^{\sigma m}| \right). \tag{8}$$

where $\mathcal{L}^{\mu m}$ and $\mathcal{L}^{\sigma m}$ are the errors in sample m in the consistency equation by Itô's Lemma. Update θ_q^n using stochastic gradient decent.

Housing Purchases by Low/High Wealth Agents (High ζ, Recession)



Poor agents who need housing are forced to sell it recessions and buy it back in expansions

$\Rightarrow$ Average return on housing is low for poorer agents (high for wealthier agents)

$\Rightarrow$ Subsidies to encourage home ownership have complicated impact on inequality

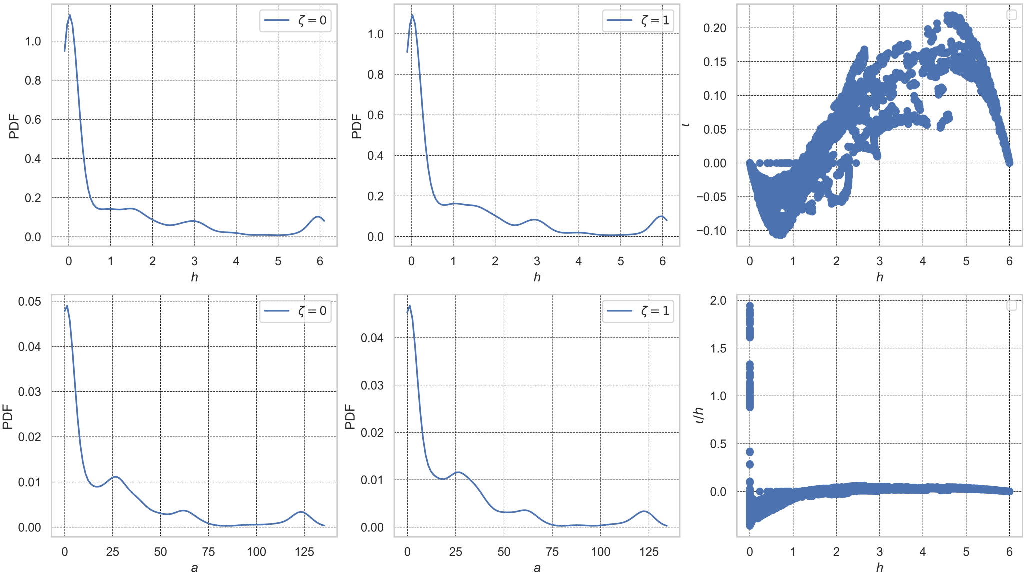
Intuition: Revisiting the Housing First Order condition

- Revisit:

$$\iota_i = \frac{h_i}{\psi}\left(\frac{\partial V_i/\partial h}{\partial V_i/\partial a} - p\right)$$

- Decreasing utility gain from housing: negative ι for rich households
- Binded financially constraint: negative ι for poor households
- Unconstrained but lacking houses: positive ι for “mid-classes”

Results: Ergodic Distribution



Outline

Conclusion

- *This talk:* showed how we use neural networks to solve continuous time, heterogeneous agent models with long-term and illiquid assets.
- *How are we using the tools:* Evaluating historical housing policy.
- *Practical Lessons:* for continuous time deep learning
  1. Working out the correct sampling approach is very important.
  2. Neural networks have difficulty dealing with inequality constraints.
  3. Enforcing shape constraints and/or rescaling functions is important.
  4. Need tighter tolerance than finite difference.

Thank You!

References I

[Achdou et al., 2022] Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2022).
Income and wealth distribution in macroeconomics: A continuous-time approach.
The Review of Economic Studies, 89(1):45–86.

[Ahn et al., 2018] Ahn, S., Kaplan, G., Moll, B., Winberry, T., and Wolf, C. (2018).
When inequality matters for macro and macro matters for inequality.
NBER macroeconomics annual, 32(1):1–75.

[Azinovic et al., 2022] Azinovic, M., Gaegauf, L., and Scheidegger, S. (2022).
Deep equilibrium nets.
International Economic Review, 63(4):1471–1525.

[Azinovic and Žemlička, 2023] Azinovic, M. and Žemlička, J. (2023).
Economics-inspired neural networks with stabilizing homotopies.
arXiv preprint arXiv:2303.14802.

[Bretscher et al., 2022] Bretscher, L., Fernández-Villaverde, J., and Scheidegger, S. (2022).
Ricardian business cycles.
Available at SSRN.

[Cardaliaguet et al., 2015] Cardaliaguet, P., Delarue, F., Lasry, J.-M., and Lions, P.-L. (2015).
The master equation and the convergence problem in mean field games.
arXiv.

References II

[Carmona, 2020] Carmona, R. (2020).
Applications of mean field games in financial engineering and economic theory.
To appear in: Machine Learning and Data Sciences for Financial Markets, Cambridge University Press.

[Carmona and Laurière, 2022] Carmona, R. and Laurière, M. (2022).
Deep learning for mean field games and mean field control with applications to finance.
Machine Learning in Financial Markets: A guide to contemporary practises, editors: A. Capponi and C.-A. Lehalle, Cambridge University Press.

[Duarte, 2018] Duarte, V. (2018).
Machine learning for continuous-time economics.
Available at SSRN 3012602.

[Fernández-Villaverde et al., 2018] Fernández-Villaverde, J., Hurtado, S., and Nuño, G. (2018).
Financial Frictions and the Wealth Distribution.
Working Paper, pages 1–51.

[Fernandez-Villaverde et al., 2020] Fernandez-Villaverde, J., Nuno, G., Sorg-Langhans, G., and Vogler, M. (2020).
Solving high-dimensional dynamic programming problems using deep learning.
Unpublished working paper.

References III

[Gaegauf et al., 2023] Gaegauf, L., Scheidegger, S., and Trojani, F. (2023).
A comprehensive machine learning framework for dynamic portfolio choice with transaction costs.
Available at SSRN 4543794.

[Germain et al., 2022] Germain, M., Laurière, M., Pham, H., and Warin, X. (2022).
DeepSets and their derivative networks for solving symmetric PDEs.
Journal of Scientific Computing, 91(2):63.

[Gopalakrishna, 2021] Gopalakrishna, G. (2021).
Aliens and continuous time economies.
Swiss Finance Institute Research Paper, (21-34).

[Han et al., 2018] Han, J., Jentzen, A., and E, W. (2018).
Solving high-dimensional partial differential equations using deep learning.
Proceedings of the National Academy of Sciences, 115(34):8505–8510.

[Han et al., 2021] Han, J., Yang, Y., and E, W. (2021).
DeepHAM: A global solution method for heterogeneous agent models with aggregate shocks.
arXiv preprint arXiv:2112.14377.

[Hu and Lauriere, 2022] Hu, R. and Lauriere, M. (2022).
Recent developments in machine learning methods for stochastic control and games.
ssrn.4096569.

References IV

[Huang, 2022] Huang, J. (2022).
A probabilistic solution to high-dimensional continuous-time macro-finance models.
Available at SSRN 4122454.

[Kahou et al., 2021] Kahou, M. E., Fernández-Villaverde, J., Perla, J., and Sood, A. (2021).
Exploiting symmetry in high-dimensional dynamic programming.
Technical report, National Bureau of Economic Research.

[Laurière, 2021] Laurière, M. (2021).
Numerical methods for mean field games and mean field type control.
Mean Field Games, 78:221.

[Li et al., 2022] Li, J., Yue, J., Zhang, W., and Duan, W. (2022).
The deep learning Galerkin method for the general stokes equations.
Journal of Scientific Computing, 93(1):1–20.

[Maliar et al., 2021] Maliar, L., Maliar, S., and Winant, P. (2021).
Deep learning for solving dynamic economic models.
Journal of Monetary Economics, 122:76–101.

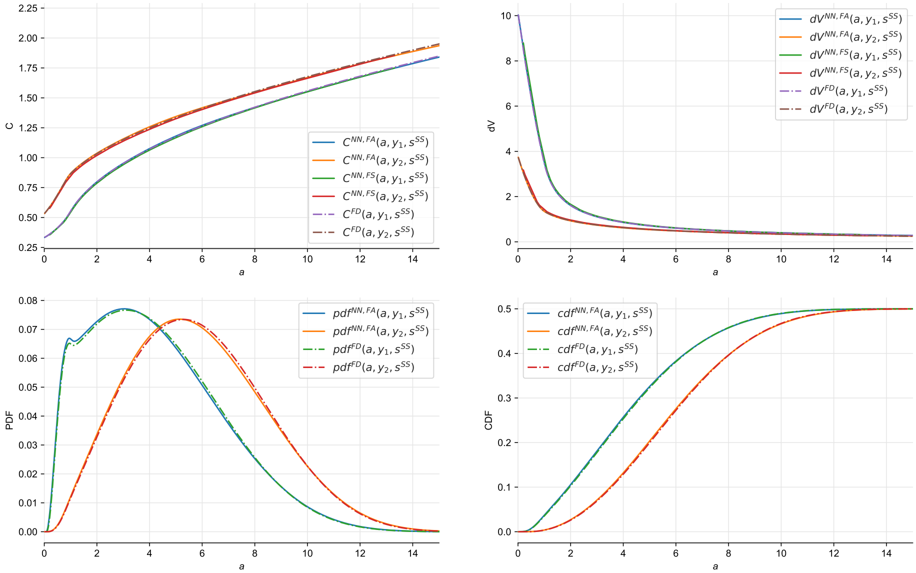
[Perrin et al., 2022] Perrin, S., Laurière, M., Pérolat, J., Élie, R., Geist, M., and Pietquin, O. (2022).
Generalization in mean field games by learning master policies.
In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 9413–9421.

References V

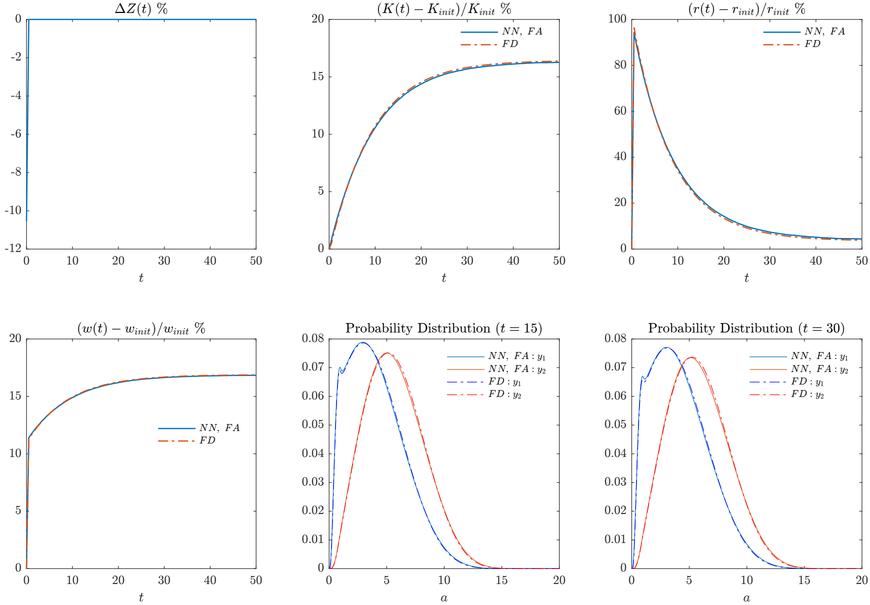
[Raissi et al., 2017] Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2017).
Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations.
arXiv preprint arXiv:1711.10561.

[Sirignano and Spiliopoulos, 2018] Sirignano, J. and Spiliopoulos, K. (2018).
DGM: A deep learning algorithm for solving partial differential equations.
Journal of computational physics, 375:1339–1364.

ABH: Numerical Results Back



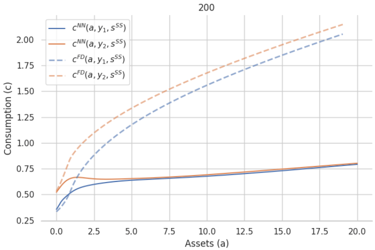
ABH: Numerical Results Back

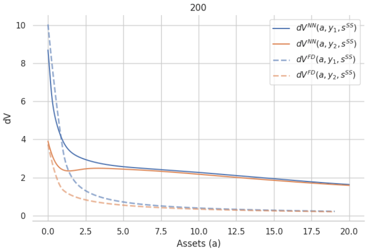


ABH: Numerical Results

Back

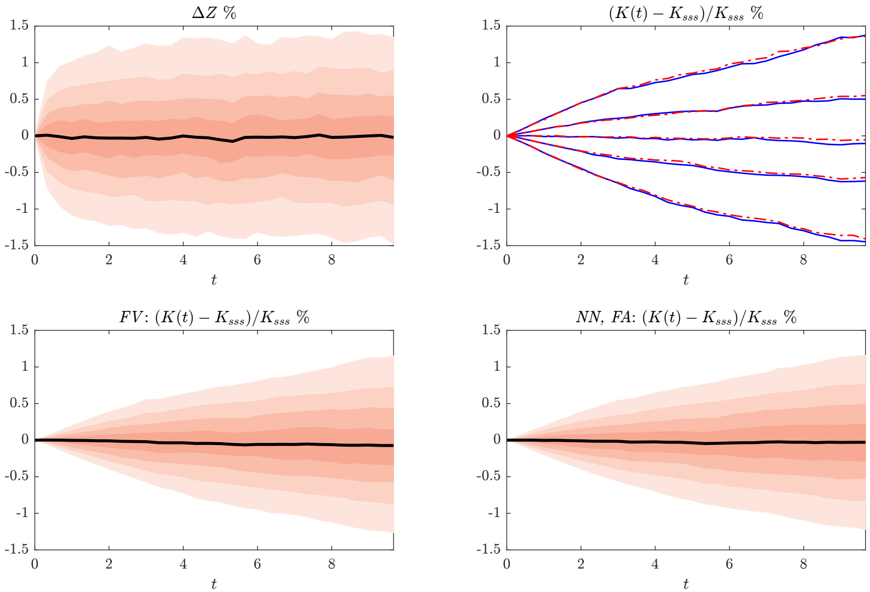
Training of the neural network (FA approach):





Play/Pause

KS: Numerical Results Back



Household's HJB Equation in Housing Model

$$
\begin{aligned}
\rho V_i(x_i) = \max_{b_i, c_i, \iota_i,} \; & u(c_i) \\
& + \zeta_{i,t}\nu(h_{i,t}, a_{i,t})\frac{\partial V_i}{\partial a_i}\mu_{a_i}(b_i, c_i, \iota_i, \cdot) + \frac{\partial V_i}{\partial y}\mu^y + \lambda(\zeta_i)(V_i(a_i, h_i, \tilde{\zeta}_i, \cdot) - V_i(a_i, h_i, \zeta_i, \cdot)) \\
& + \frac{1}{2}\frac{\partial^2 V_i}{\partial a_i^2}\sigma_{a_i}^2(b_i, \cdot) + \frac{1}{2}\frac{\partial^2 V_i}{\partial y^2}\sigma_y^2 + \frac{\partial^2 V_i}{\partial a_i \partial y}\sigma_{a_i}(b_i, \cdot)\sigma_y \\
& + \sum_{j \neq i}\frac{\partial^2 V_i}{\partial a_i \partial a_j}\sigma_{a_i}(b_i, \cdot)\hat{\sigma}_{a_j}(\cdot) + \sum_{j \neq i}\frac{\partial V_i}{\partial a_j}\hat{\mu}_{a_j}(\cdot) + \sum_{j \neq i}\frac{\partial^2 V_i}{\partial a_j \partial y}\hat{\sigma}_{a_j}(\cdot)\sigma_y \\
& + \sum_{j \neq i}\lambda(\zeta_j)(V_i(a_i, h_i, \zeta_i, \tilde{\zeta}_j \cdot) - V_i(a_i, h_i, \zeta_i, \zeta_j, \cdot)) \\
& + \frac{1}{2}\sum_{j \neq i, j' \neq i}\frac{\partial^2 V_i}{\partial a_j \partial a_{j'}}\hat{\sigma}_{a_j}(\cdot)\hat{\sigma}_{a_j'}(\cdot) + \phi(a_i, h_i, \kappa_i)
\end{aligned} \tag{9}
$$

Back

Derivations

The first order condition of optimal portfolio choice condition in (4) can be further written into a matrix form:

$$\mathbf{M}(\boldsymbol{a}-\boldsymbol{b})=\boldsymbol{n} \tag{10}$$

By multiplying both sides with $\mathbf{M}^{-1}$, the risky asset holding can be written as:

$$\boldsymbol{a}-\boldsymbol{b}=\mathbf{M}^{-1}\boldsymbol{n} \tag{11}$$

Further, the bond market clearing condition can be essentially written as: $\boldsymbol{\iota}\cdot\boldsymbol{b}=0$, we have:

$$\boldsymbol{\iota}\cdot(\boldsymbol{a}-\boldsymbol{b})=\boldsymbol{\iota}\cdot(\mathbf{M}^{-1}\boldsymbol{n})=q. \tag{12}$$

Plug in the expression for $\boldsymbol{n}$, then we can get the expression for risk-premium.

Closed form solution for housing price p_t with quadratic transaction cost: $\Psi(h_{i,t},\iota_{i,t})=\frac{1}{2}\kappa\frac{\iota_{i,t}^2}{h_{i,t}}$

$$p+\kappa\frac{\iota_i}{h_i}=\frac{\partial V_i/\partial h_i}{\partial V_i/\partial a_i}\rightarrow p=\frac{1}{H}\left(\int_i\frac{\xi_{h,i}}{\xi_{a,i}}h_i di\right) \tag{13}$$

back

Results: Losses

Master Equation ξ_a	Master Equation ξ_h	Goods Market	q-Drift	q-Volatility
1.02×10^{-2}	2.31×10^{-3}	3.12×10^{-4}	8.10×10^{-4}	5.57×10^{-3}

Recursive Representation (Appendix)

- Assume equilibrium exists that is recursive in aggregate states: $\{z, g\}$.
- Given a belief $dg_t(x) = \hat{\mu}_g(z_t, g_t)dt + \hat{\sigma}(z_t, g_t)dB_t^0$, household choose c to solve HJBE:

$$
\begin{aligned}
0 = \max_{c \in \mathcal{C}(x,z,g)} \Big\{ & \rho V(x,z,g) + u(c) + D_x V(x,z,g)\mu^x(c,x,z,Q(z,g)) \\
& + \lambda(x)\left(V(x+\gamma^x(x), z, g) - V(x,z,g)\right) + \text{higher order terms} \\
& + \partial_z V(x,z,g)\mu^z(z) + 0.5\left(\sigma^z(z)\right)^2 \partial_{zz} V(x,z,g) \\
& + \int_{\mathcal{X}} \left(\hat{\mu}_g(z_t, g_t) \frac{\partial V}{\partial g}(x,z,g)(y) + \text{higher order terms} \right) dy \Big\} \\
& s.t. \quad \text{BCs} \quad \Psi(V)(x) = 0, \Phi(V)(x) \geq 0
\end{aligned}
$$

- For optimal policy rule, $c^*(x,z,g)$, for z_t, population density, g, evolves by KFE:

$$
dg_t(x) = \mu_g(c^*(x,z,g), z_t, g_t)dt - \operatorname{div}[\sigma^x(c,s,z,q)g_t(x)]dB_t^0. \tag{14}
$$

- In equilibrium $\hat{\mu}_g = \mu_g$.

Back

ABH: Master Equation

- The Master equation is:

$$0 = (\mathcal{L}V)(a,y,g) = (\mathcal{L}^h V)(a,y,g) + (\mathcal{L}^g V)(a,y,g)$$

- where, in this ABH model, the operators are defined by:

$$\begin{aligned}
(\mathcal{L}^h V)(a,y,g) :=& -\rho V(a,y,g) + u(c^*(a,y,g)) + \mathbf{1}_{a\leq \underline{a}}\psi(a) \\
&+ \partial_a V(a,y,g) s(a,y,c^*(a,y,g), r(\bar{g}), w(\bar{g})) \\
&+ \lambda(y)(V(a,\tilde{y},g) - V(a,y,g)) \\
(\mathcal{L}^g V)(a,y,g) :=& \int_{\mathbb{R}} \frac{\partial V}{\partial g}(a,y,g)(b)\,(\lambda(\tilde{y})g(b,\tilde{y}) - \lambda(y)g(b,y))\,db \\
&+ \int_{\mathbb{R}} \partial_b \frac{\partial V}{\partial g}(a,y,g)(b) s\,(a,y,c^*(a,y,g), r(\bar{g}), w(\bar{g}))\, g(b,y)db
\end{aligned}$$

- and the optimal control satisfies the following FOC:

$$\partial_a V(a,y,g) = u'(c^*(a,y,g))$$

Back

ABH Model: Equilibrium

Equilibrium: We say that $q^* = [r^*, w^*]$ and c^* form an equilibrium if:

1. Given their belief q^*, the optimal control of a representative agent is c^*
2. Their belief is consistent with c^*:
$$q_t^* = Q(\bar{g}_t^*) = [r(\bar{g}_t^*), w(\bar{g}_t^*)]$$
where g^* is the distribution generated if everyone uses c^*

Value function:

- Agents want to know the equilibrium c^* for "any" possible distribution
- Value function of an agent depends on the current g_t

ABH: Derivative of Master Equation

- We will rather approximate $W(a,y,g) = \partial_a V(a,y,g)$, which solves the PDE:

$$0 = (\mathfrak{L}W)(a,y,g) = (\mathfrak{L}^h W)(a,y,g) + (\mathfrak{L}^g W)(a,y,g)$$

- where the operators $\mathfrak{L}^h$ and $\mathfrak{L}^g$ are defined by:

$$\begin{aligned}
(\mathfrak{L}^h W)(x,g) := & \, (r(\bar{g}) - \rho) W(a,y,g) + \mathbf{1}_{a \leq \underline{a}} \psi'(a) \\
& + \partial_a W(a,y,g) s(a,y,c^*(a,y,g), r(\bar{g}), w(\bar{g})) \\
& + \lambda(y)(W(a,\tilde{y},g) - W(a,y,g)) \\
(\mathfrak{L}^g W)(x,g) := & \int_{\mathbb{R}} \frac{\partial W}{\partial g}(a,y,g)(b) \left(\lambda(\tilde{y}) g(b,\tilde{y}) - \lambda(y) g(b,y)\right) db \\
& + \int_{\mathbb{R}} \partial_b \frac{\partial W}{\partial g}(a,y,g)(b) s\left(a,y,c^*(a,y,g), r(\bar{g}), w(\bar{g})\right) g(b,y) db
\end{aligned}$$

- with the FOC:

$$W(a,y,g) = u'(c^*(a,y,g)).$$

- We apply the algorithm to this PDE for W.

Approach A: Finite Population

- Replace distribution g_t by finite number of agent $\hat{g}_t := \{x_t^i : i \leq I\}$.
  - Agent $i \leq I$ behaves as if their individual actions do not influence prices.
  - So, their belief is: $\hat{q}_t = \hat{Q}(z_t, \hat{g}_t^{-i})$, where $\hat{g}_t^{-i} := \{x_t^j : j \neq i\}$
- $V(x^i, z, \hat{g})$ solves $(\hat{\mathcal{L}}V)(x^i, z, \hat{g}) = 0$ subject to BCs, where $\hat{\mathcal{L}} := \hat{\mathcal{L}}^h + \hat{\mathcal{L}}^g$

$$(\hat{\mathcal{L}}^h V)(x^i, z, \hat{g}) := (\mathcal{L}^h V)(x^i, z, \hat{g})$$

$$(\hat{\mathcal{L}}^g V)(x^i, z, \hat{g}) := \sum_{j \leq I} \frac{\partial V}{\partial x^j}(x^i, z, \hat{g}) \mu^x(c^*(x^j, z, \hat{g}), x^j, z, \hat{Q}(z, \hat{g}^{-j}))$$
$$+ \sum_{j \leq I} \lambda(x^j) \left(V(x^i, z, \{x^j + \gamma^x(x^j), \hat{g}^{-j}\}) - V(x^i, z, \hat{g}^{-i}) \right)$$

  - $\hat{\mathcal{L}}^h$ stays the same; $\hat{\mathcal{L}}^g$ changes to capture impact of changes in other agent positions
  - Converges to original model as $I \to \infty$ (see [Carmona, 2020])

Back

Approach B: Projection Onto Basis

- Approximate the distribution $g_t(x)$ by $\sum_{i=1}^{N} \alpha_t^i h^i(x)$, where:
  - α_t^i is a time varying coefficient, $h^i(x)$ is basis function, and
  - Example bases: Indicator Functions, Chebyshev polynomial, Eigenfunctions, ...
  - Distribution characterized by coefficients: $\hat{g}_t := \{\alpha_t^1, ...\alpha_t^N\}$.
  - Substituting $\sum_{i=1}^{N} \alpha_t^i h^i(x)$ into KFE implicitly gives the law of motion for the coefficients:

$$d\alpha_t^i = \hat{\mu}_\alpha^i(z,\hat{g})dt, \quad \text{where } \hat{\mu}_\alpha^i(z,\hat{g}) \text{ solve } \sum_{i=1}^{N} \hat{\mu}_\alpha^i(z,\hat{g}) h^i(x) = \hat{\mathcal{L}}^k \left[\sum_{i=1}^{N} \alpha_i(t) h^i(x) \right]$$

- $V(x^i, z, \hat{g})$ solves $(\hat{\mathcal{L}}V)(x^i, z, \hat{g}) = 0$ subject to BCs, where $\hat{\mathcal{L}} := \hat{\mathcal{L}}^h + \hat{\mathcal{L}}^g$:

$$(\hat{\mathcal{L}}^h V)(x,z,\hat{g}) := (\mathcal{L}^h V)(x,z,\hat{g}), \quad (\hat{\mathcal{L}}^g V)(x,z,\hat{g}) := \sum_{i=1}^{N} \hat{\mu}_\alpha^i(z,\hat{g}) \frac{\partial V}{\partial \alpha_i}(x,z,\hat{g})$$

Discrete state space Eigenvectors

Back

Approach B.1: Project Onto Discrete State Space

- We approximate the distribution by a histogram:
  - Basis is a collection of N^x points: $x_1, \ldots, x_{N^x}$, in $\mathcal{X}$.
  - We approximate g_t by a vector $\alpha_t \in \mathbb{R}^{N^x}$ of mass points at $x_1, \ldots, x_{N^x}$.
- Law of motion of the mass points is the finite difference approximation to the KFE.

Back

Approach B.2: Projection Onto Eigenfunctions

- Let $\{e_i\}_{i\geq 1}$ be eigenfunctions of KFE operator $\hat{\mathcal{L}}^k_{z=0}$ without aggregate shocks:

$$\mathcal{L}^k_{z=0} e_i = \lambda_i e_i, \quad \text{where } \lambda_i \text{ are eigenvalues}$$

  Use finite subset of eigenfunctions of $\hat{\mathcal{L}}^k_{z=0}$ as basis:

$$g_t(x) \approx \sum_{i\leq I} \alpha_t^i e^i(x), \quad \text{so distribution characterized by } \hat{g} = \{\alpha_t^1, \ldots, \alpha_t^I\}$$

- Drifts of the coefficients $\{\hat{\mu}_\alpha^i\}_{i\leq I}$ satisfy a collection of equations for $i \leq I$:

$$\sum_{i\leq I} \hat{\mu}_\alpha^i \langle e_i, e_j\rangle = \underbrace{\sum_{i\leq I} \alpha_t^i \lambda_i^A \langle e_i, e_j\rangle}_{\text{Weighted } \mathcal{L}^k_{z=0} \text{ eigenvalues}} + \underbrace{\int_{\mathcal{X}} e_j(x) \Big(\big(\mathcal{L}^k_z - \mathcal{L}^k_{z=0}\big) \Big(\sum_{i\leq I} \alpha_t^i e_i \Big) \Big)(x) dx}_{\text{Weighted difference between } \mathcal{L}^k_z - \mathcal{L}^k_{z=0}}$$

- **Remark:** We approximate operator difference $\hat{\mathcal{L}}^k_z - \hat{\mathcal{L}}^k_{z=0}$:
  - Many papers perturb z or $g(x)$ (e.g. [Cardaliaguet et al., 2015], [Alverez (2023)], [Bilal (2023)])
  - We "perturb" $\hat{\mathcal{L}}^k_z$ in the operator space.

Back

Approximate Solutions

- Consider HJB equation for the Merton problem (consumption and portfolio choice):

$$\rho V(a) = \max_{c,\theta} u(c) + V'(a)((r + (\bar{R} - r)\theta)a - c) + \frac{1}{2}\sigma^2\theta^2 a^2 V''(a) \tag{15}$$

- Suppose V_0 is the exact solution of Merton's problem, we plug in a scaled solution $k^{-\gamma}V_0$:

$$\rho k^{-\gamma}V_0 = \frac{c^{1-\gamma}k^{1-\gamma}}{1-\gamma} + k^{-\gamma}V_0'((r + (\bar{R} - r)\theta)a - kc) + \frac{1}{2}\sigma^2\theta^2 a^2 k^{-\gamma}V_0''(a) \tag{16}$$

  Which implies that the loss function (with no loss of generality, we use L1 loss here) will be:

$$Loss = \left|\left(k^{1-\gamma} - k^{-\gamma}\right)\right| \cdot \underbrace{\left|-cV_0' + \frac{c^{1-\gamma}}{1-\gamma}\right|}_{\text{Finite value}}$$

  - $\gamma < 1$, no problem because $k^{1-\gamma}$ will explode while $k^{-\gamma}$ vanishes as $k \to \infty$.
  - $\gamma > 1$, a very large k can be problematic because both $k^{1-\gamma}$ and $k^{-\gamma}$ vanish as $k \to \infty$.

- Hence, in the economically relevant case $\gamma > 1$, computer is very good at finding "cheat solution" by simply push value function to be very close to zero.

Back

Example: Projection of Distribution on Chebyshev Polynomials

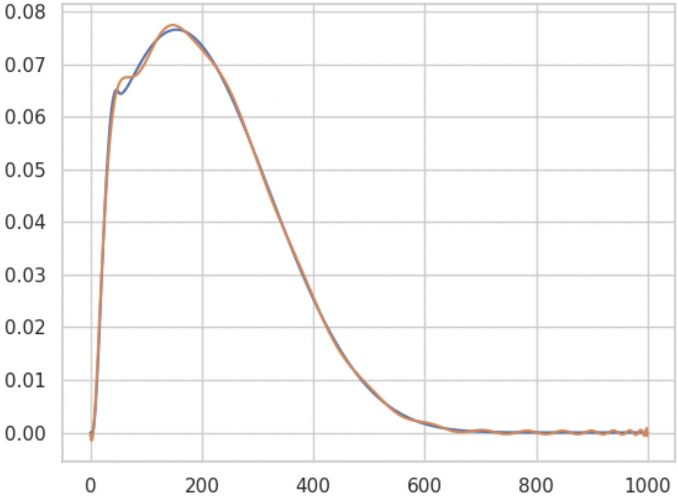


Figure: Capital to Labor Ratio vs borrowing constraint a_{lb}

Example: Projection of Distribution on Chebyshev Polynomials

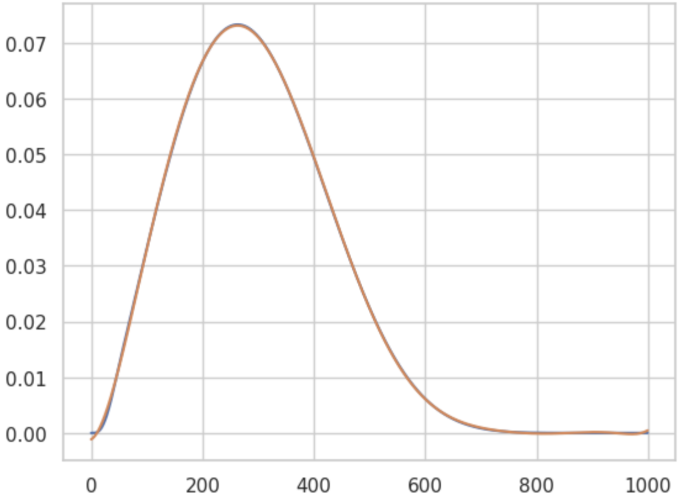


Figure: Capital to Labor Ratio vs borrowing constraint a_{lb}

ABH: Numerical Results with Projection

Results with projection technique based on 7 eigenvectors