Discussion of "Estimating Nonlinear Heterogeneous Agent Models with Neural Networks " by Hanno Kase, Leonardo Melosi and Matthias Rottner

Discussion by Vytautas Valaitis

University of Surrey

Bank of Spain Annual Research Conference, 2025

Overview

Economists seek to solve and estimate nonlinear models with household heterogeneity and aggregate risk, building on Krusell and Smith (1998).

Overview

Economists seek to solve and estimate nonlinear models with household heterogeneity and aggregate risk, building on Krusell and Smith (1998).

Current progress:

- Estimate by *linearizing* around aggregate states: Reiter (2009), Bayer et al. (2024), Winberry (2018),

Overview

Economists seek to solve and estimate nonlinear models with household heterogeneity and aggregate risk, building on Krusell and Smith (1998).

Current progress:

- Estimate by *linearizing* around aggregate states: Reiter (2009), Bayer et al. (2024), Winberry (2018),
- Use ANN to capture *nonlinear* aggregate dynamics: Fernández-Villaverde et al. (2023), Azinovic et al. (2022),

Overview

Economists seek to solve and estimate nonlinear models with household heterogeneity and aggregate risk, building on Krusell and Smith (1998).

Current progress:

- Estimate by *linearizing* around aggregate states: Reiter (2009), Bayer et al. (2024), Winberry (2018),
- Use ANN to capture *nonlinear* aggregate dynamics: Fernández-Villaverde et al. (2023), Azinovic et al. (2022),
- Keeping track of the wealth distribution : Bayer et al. (2019), Algan et al. (2010), Maliar et al. (2021).

Overview

Economists seek to solve and estimate nonlinear models with household heterogeneity and aggregate risk, building on Krusell and Smith (1998).

Current progress:

- Estimate by *linearizing* around aggregate states: Reiter (2009), Bayer et al. (2024), Winberry (2018),
- Use ANN to capture *nonlinear* aggregate dynamics: Fernández-Villaverde et al. (2023), Azinovic et al. (2022),
- Keeping track of the wealth distribution : Bayer et al. (2019), Algan et al. (2010), Maliar et al. (2021).

This paper estimates a nonlinear HANK model, while keeping track of the wealth distribution . Illustrate the method in three settings:

Overview

Economists seek to solve and estimate nonlinear models with household heterogeneity and aggregate risk, building on Krusell and Smith (1998).

Current progress:

- Estimate by *linearizing* around aggregate states: Reiter (2009), Bayer et al. (2024), Winberry (2018),
- Use ANN to capture *nonlinear* aggregate dynamics: Fernández-Villaverde et al. (2023), Azinovic et al. (2022),
- Keeping track of the wealth distribution : Bayer et al. (2019), Algan et al. (2010), Maliar et al. (2021).

This paper estimates a nonlinear HANK model, while keeping track of the wealth distribution . Illustrate the method in three settings:

- Linearized small-scale DSGE model - can compare to the analytical solution.
- A RANK model with a ZLB - compare with other nonlinear estimation methods.
- A nonlinear HANK model (ZLB + aggregate risk) - **main contribution**.

Overview: Model Solution

- Treat the estimable parameters $\tilde{\Omega}$ as pseudo-state variables.
- Build two networks

$$\psi_t^i = \psi^I(\mathbb{S}_t^i, \mathbb{S}_t, \tilde{\Omega}|\bar{\Omega}) \qquad \psi_t^A = \psi^A(\mathbb{S}_t, \tilde{\Omega}|\bar{\Omega})$$

- Where the state vector $\mathbb{S} \in \{\mathbb{S}^A, \{\mathbb{S}^i\}_{i=1}^L\}$ includes the states of all agents L.
- This discretization allows to keep track of the whole distribution !

Overview: Model Solution

- Treat the estimable parameters $\tilde{\Omega}$ as pseudo-state variables.
- Build two networks

$$\psi_t^i = \psi^I(\mathbb{S}_t^i, \mathbb{S}_t, \tilde{\Omega}|\bar{\Omega}) \qquad \psi_t^A = \psi^A(\mathbb{S}_t, \tilde{\Omega}|\bar{\Omega})$$

- Where the state vector $\mathbb{S} \in \{\mathbb{S}^A, \{\mathbb{S}^i\}_{i=1}^L\}$ includes the states of all agents L.
- This discretization allows to keep track of the whole distribution !
- Curse of dimensionality
  - Make draws from the parameter space *instead* of solving on the grid .
  - The relevant share of the state space shrinks as the number of state variables $\uparrow$.
  - Make draws from the simulated model.

Overview: Model Estimation

$$\mathcal{L}(\mathbb{Y}_{1:T}|\tilde{\Omega}) = \prod_{t=1}^{T} p_{\tilde{\Omega}}\left(y_t|y_{t-1},...,y_1,\tilde{\Omega}\right)$$

When the model is nonlinear:

- Need to use MC filters, which are time-consuming and less accurate.

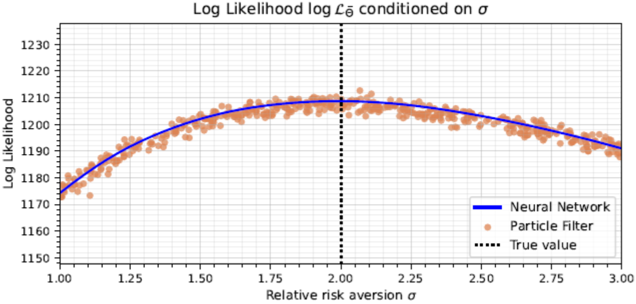
Overview: Model Estimation

$$\mathcal{L}(\mathbb{Y}_{1:T}|\tilde{\Omega}) = \prod_{t=1}^{T} p_{\tilde{\Omega}}\left(y_t|y_{t-1},...,y_1,\tilde{\Omega}\right)$$

When the model is nonlinear:

- Need to use MC filters, which are time-consuming and less accurate.

Instead: Use the ANN to approximate the mapping between $\mathcal{L}(\mathbb{Y}_{1:T}|\tilde{\Omega})$ and $\tilde{\Omega}$, dubbed as the ANN *particle filter*. This allows for fast and precise likelihood evaluation.



How to choose the network architecture?

Current literature offers little guidance on how to pick the network hyper-parameters.

How to choose the network architecture?

Current literature offers little guidance on how to pick the network hyper-parameters.

We want a network that approximates well BUT allows for a *fast* convergence: 1. trains fast, 2. stable across iterations.

Stability especially relevant because:

- In the context of model solution, data is *endogenous* to the network weights.
- Every iteration there is a *new* sample draw.
- Any analogies to parameter dampening in the regression?

How to choose the network architecture?

Current literature offers little guidance on how to pick the network hyper-parameters.

We want a network that approximates well BUT allows for a *fast* convergence: 1. trains fast, 2. stable across iterations.

Stability especially relevant because:

- In the context of model solution, data is *endogenous* to the network weights.
- Every iteration there is a *new* sample draw.
- Any analogies to parameter dampening in the regression?

Wide or Narrow?

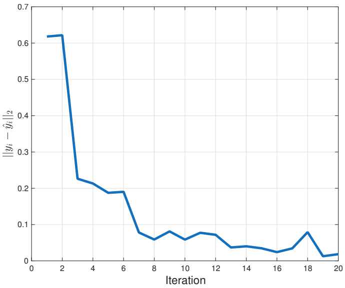
- Deep networks generalize better, making them more stable when the new data arrives. Wide networks memorize and may be less stable to new data.
- Training narrow networks is more efficient .

Wide or Narrow?

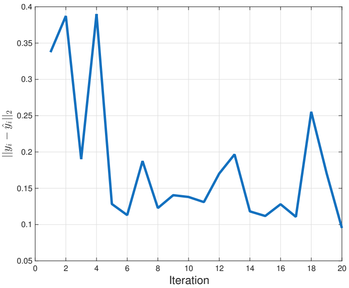
- Deep networks generalize better, making them more stable when the new data arrives. Wide networks memorize and may be less stable to new data.
- Training narrow networks is more efficient .

Experiment:

- Consider a eide shallow network (27 neurons, 1 layer) and a deep narrow network (9 layers 3 neurons each).
- Each iteration: learn the mapping $y_i = f(X_i)$ with the new data draws.



(a) Narrow; Av. training time 23s.



(b) Wide; Av. training time 44s.

Discretization or Moments?

Discretization

Currently, the model is solved by assuming that the economy consists of a discrete number L agents:

- Allows to keep track of all the features of the wealth distribution,
- Requires a large L for a precise solution when the wealth distribution matters,
- Large L complicates the ANN stability and the algorithm convergence .

Discretization or Moments?

Discretization

Currently, the model is solved by assuming that the economy consists of a discrete number L agents:

- Allows to keep track of all the features of the wealth distribution,
- Requires a large L for a precise solution when the wealth distribution matters,
- Large L complicates the ANN stability and the algorithm convergence .

Distribution Moments

Alternatively, one can use selected moments of the wealth distribution in $\mathbb{S}$.

- Simulate a histogram (Young, 2010) and compute selected moments as elements of $\mathbb{S}$,
- Potentially easier to train the network and converge the algorithm and NOT more costly to simulate,
- Which moments to use?
- OR one could use the histogram as part of $\mathbb{S}$.

Discretized distribution when individual nonlinearity is relevant

This paper makes significant progress in handling aggregate nonlinearity.
Alternatively, it can be exploited to solve and estimate models with individual nonlinearity and aggregate risk:

Discretized distribution when individual nonlinearity is relevant

This paper makes significant progress in handling aggregate nonlinearity. Alternatively, it can be exploited to solve and estimate models with individual nonlinearity and aggregate risk:

Relevant examples include:

- 2-asset HANK with aggregate uncertainty - price of illiquid asset determined by the constrained agents.

Discretized distribution when individual nonlinearity is relevant

This paper makes significant progress in handling aggregate nonlinearity. Alternatively, it can be exploited to solve and estimate models with individual nonlinearity and aggregate risk:

Relevant examples include:

- 2-asset HANK with aggregate uncertainty - price of illiquid asset determined by the constrained agents.
- Asset pricing with heterogeneous agents - asset prices more sensitive to prediction errors in the aggregate law of motion.

Discretized distribution when individual nonlinearity is relevant

This paper makes significant progress in handling aggregate nonlinearity. Alternatively, it can be exploited to solve and estimate models with individual nonlinearity and aggregate risk:

Relevant examples include:

- 2-asset HANK with aggregate uncertainty - price of illiquid asset determined by the constrained agents.
- Asset pricing with heterogeneous agents - asset prices more sensitive to prediction errors in the aggregate law of motion.
- Firm dynamics with capital illiquidity - capital prices depend on financial frictions.

Discretized distribution when individual nonlinearity is relevant

This paper makes significant progress in handling aggregate nonlinearity. Alternatively, it can be exploited to solve and estimate models with individual nonlinearity and aggregate risk:

Relevant examples include:

- 2-asset HANK with aggregate uncertainty - price of illiquid asset determined by the constrained agents.
- Asset pricing with heterogeneous agents - asset prices more sensitive to prediction errors in the aggregate law of motion.
- Firm dynamics with capital illiquidity - capital prices depend on financial frictions.

Additionally:

- Ramsey problems with heterogeneous agents - one needs to keep track of the whole distribution, discrete L does the trick!

Thank You!

References I

Yann Algan, Olivier Allais, and Wouter J. Den Haan. Solving the incomplete markets model with aggregate uncertainty using parameterized cross-sectional distributions. *Journal of Economic Dynamics and Control*, 34(1):59–68, 2010.

Marlon Azinovic, Luca Gaegauf, and Simon Scheidegger. Deep equilibrium nets. *International Economic Review*, 63(4):1471–1525, 2022.

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

Jesús Fernández-Villaverde, Hurtado Samuel, and Galo Nuño. Financial frictions and the wealth distribution. *Econometrica*, 91(3):869–901, 2023.

Per Krusell and Anthony A. Smith, Jr. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896, 1998.

Lilia Maliar, Serguei Maliar, and Pablo Winant. Deep learning for solving dynamic economic models. *Journal of Monetary Economics*, 122:76–101, 2021.

Eric R. Young. Solving the incomplete markets model with aggregate uncertainty using the krusell–smith algorithm and non-stochastic simulations. *Journal of Economic Dynamics and Control*, 34(1):36–41, 2010.