"Inequality and Zero Lower Bound"

by Jesús Fernández Villaverde, Joel Marbet, Galo Nuño, Omar Rachedi

Discussion: Lilia Maliar

November 4, 2021

What do they do in this paper?

Excellent and novel paper:

- extensively analyzes a heterogeneous-agent version of the basic new Keynesian (HANK) model with occasionally binding constraints;
- uses deep-learning techniques to solve Krusell and Smith's (KS) (1998) type of model;
- presents novel economic findings.

This paper

HANK:

- Aggregate and idiosyncratic labor-income risk;
- One asset, bonds (no capital);
- One borrowing constraint on bonds;
- ZLB;
- Progressive labor earnings taxes.

Main numerical findings:

- Frequency of ZLB in HANK (10%) is higher than in RANK (8%);
- Long-run real interest rate depends on the interaction of the inflation target and wealth inequality;
- ZLB leads to larger increases in income inequality during severe recessions.

Neural networks

- *Idea of the standard KS algorithm*:
  – represent an infinite dimensional object – wealth distribution – with a polynomial of a finite set of moments (usually mean).
- *FVMNR use a combination of the following algorithms:*
  – a stochastic simulation algorithm of Maliar et al. (2010);
  – histograms of Young (2010) to represent the distribution;
  – a neural-network algorithm of Fernandez-Villaverde et al. (FVHN) (2020).
- *Idea of the FVHN algorithm:*
  – approximate aggregate laws of motion with a general function of the distributional moments;
  – instead of polynomial, use a neural network.
- Why neural networks?
  – Very flexible – allow for kinks, high degrees of non-linearities, well-behaved outside their training areas
  – Universal approximators.

Comments

1. Choices made in the numerical analysis.

2. Solution procedure.

3. Related literature.

Comment 1a: IRF are not very informative

- For aggregate shocks, the paper reports impulse response functions (IRF).
  – *Q1: Given that there is both aggregate and idiosyncratic risk, why the IRF are so smooth?*
  – It's a common practice in the literature to use the methodology of Koop, Pesaran and Potter (KPP) (1996) to produce IRF with multiple sources of risk, and they are usually not so smooth.
- For idiosyncratic shocks, the paper reports individual groups' responses to three specific realizations.
  – *Q2: How can a specific realization be representative of general tendencies?*
  – Need to report standard IRF as suggested by KPP.

Comment 1b: Calibration

- Inflation target affects the real rate at too high levels of wealth inequality – wealth Gini coefficient assumed is 0.96–0.98.
  – *Q3: Are such values of wealth Gini realistic? Will the effects be quantitatively significant under realistic inequality?*
  – In the U.S. economy, the Gini is equal to 0.76.
- The model predicts that the probability of ZLB events is 10% with a 2% inflation target.
  – *Q4: Is this probability realistic?* Average frequency of ZLB in the US economy where the inflation target is 2% was 5% (Reifschneider and Williams, 1999).
- Precautionary savings are the key to understanding the frequency of ZLB.
  – *Q5: Why no sensitivity results with respect to the level of prudence determining the size of precautionary savings?*
  – They assume GHH preferences with risk aversion of 1 and Frisch elasticity of 1.

Comment 1c: Other statistics should be reported

- Running times.
- Distributive implications of the model:
  - Gini coefficients of wealth, income, consumption;
  - Wealth and income shares of distribution percentiles;
  - Fraction of households on the borrowing constraint.
- Aggregate model's implications:
  - Standard deviations of and correlations between aggregate variables;
  - Correlations between aggregate variables and shocks;
  - Correlations between Gini coefficients and shocks.
- $\Longrightarrow$ *Otherwise, we do not know whether the model does well in reproducing empirical observations.*

Comment 2: Deep learning techniques are not used to a full potential

- The present paper:
  - Only aggregate law of motion is approximated by a neural network.
  - The other objects (consumption function, value function) are approximated using standard approximation techniques (exogenous grids).
- *Maliar, Maliar and Winant (MMW) (2018, 2020) "Deep Learning for Solving Dynamic Economic Models". Journal of Monetary Economics 122.*
  - Do not invent a new computational method for economic models but show how to reformulate the models themselves to fit into the existing state-of-the-art computational methods
  - MMW adapt deep learning framework built on multilayer neural networks with stochastic optimization to solving dynamic economic models.
  - All unknown functions are found from the same optimization problem.

Comment 2: Deep learning techniques are not used to a full potential

1. **Model**: $\begin{cases} E_\epsilon\left[f_1\left(X\left(s\right),\epsilon\right)\right]=0 \\ \quad\quad\ldots \\ E_\epsilon\left[f_n\left(X\left(s\right),\epsilon\right)\right]=0 \end{cases}$

   s = state, $X\left(s\right)$ = decision function, ϵ = innovations.
2. Parameterize $X\left(s\right)\simeq\varphi\left(s;\theta\right)$ with a **deep neural network**.
3. Construct **objective function** for deep learning training

   $$\min_\theta\left(E_\epsilon\left[f_1\left(\varphi\left(s;\theta\right),\epsilon\right)\right]\right)^2+\ldots+\left(E_\epsilon\left[f_n\left(\varphi\left(s;\theta\right),\epsilon\right)\right]\right)^2\to 0$$
4. **All-in-one expectation** operator is a critical novelty:

   $$\left(E_\epsilon\left[f_j\left(\varphi\left(s;\theta\right),\epsilon\right)\right]\right)^2=E_{(\epsilon_1,\epsilon_2)}\left[f_j\left(\varphi\left(s;\theta\right),\epsilon_1\right)\cdot f_j\left(\varphi\left(s;\theta\right),\epsilon_2\right)\right]$$

   with ϵ_1,ϵ_2 = two independent draws.
5. **Stochastic gradient descent** for training (random grids)
6. Google **TensorFlow** platform – software that leads to break-ground applications (image, speech recognition, etc).

Comment 2: Deep learning techniques are not used to a full potential

- **KS** (1998) use a reduced state space:
  X_i (*variables of agent* i, *aggregate moments*)
  $\Rightarrow$ few state variables
- **MMW (2018)** uses the true state space:
  X_i (*variables of all agents, distributions*)
  $\Rightarrow$ hundreds of state variables
  How do we deal with such a large state space?

1. Neural network automatically performs the model reduction
   – it learns to summarize information from many inputs into a smaller set of hidden layers.
2. Neural network deals with ill conditioning
   – it learns to ignore collinear and redundant variables.

Comment 3: Related papers

Other papers that introduce aggregate uncertainty and non-linear dynamics within a HANK economy:

- **Paper # 1:** Maliar and Maliar (2020). "Deep Learning: Solving HANC and HANK Models in the Absence of Krusell-Smith Aggregation", SSRN WP.
- **Paper # 2:** Gorodnichenko, Maliar, Maliar and Naubert (2020) “Household Savings and Monetary Policy under Individual and Aggregate Stochastic Volatility“, CEPR working paper 15614.

Comment 3: Related paper #1: Maliar and Maliar (2020)

- The same HANK model as in FVMNR:
  - bonds are the only asset;
  - borrowing constraint on bonds;
  - ZLB.
- Neural networks are used for solving not just for aggregate law of motion but the entire model is formed as an objective function of the deep learning approximation.
- Work with actual state space: 1,000 of agents (2,000 state variables)
- The key distinctive feature of neural network is model reduction:
  - 2,000 of state variables are reduced to 32 or 64 neurons in the hidden layer.

Comment 3: Related paper #2: Gorodnichenko et al. (2020)

- Households
  - Three types of assets: bonds (liquid), shares and machines (illiquid)
  - Three borrowing constraints, one per each asset
  - Idiosyncratic shocks to productivity level and volatility
  - Heterogenous labor
- Firms
  - CRS with machines and labor
  - Aggregate shocks to TFP level and volatility
  - Sticky prices (Rotemberg)
- Government
  - Fiscal policy (progressive income taxation)
  - Monetary policy (Taylor rule with ZLB)

Comment 3: Related paper #2: Gorodnichenko et al. (2020)

- Household productivity

$$\begin{aligned} \text{risk:} \quad & \eta_{\ell,t}(j) = \rho^{\ell}\eta_{\ell,t-1}(j) + \exp(\sigma_{\ell,t-1})\,\varepsilon_{\ell,t}(j) \\ \text{uncertainty:} \quad & \sigma_{\ell,t} = \rho^{\sigma_l}\sigma_{\ell,t-1} + \sigma_{\sigma_\ell}\varepsilon_{\sigma_\ell,t} \end{aligned}$$

- Aggregate TFP

$$\begin{aligned} \text{risk:} \quad & \eta_{\theta,t} = \rho^{\theta}\eta_{\theta,t-1} + \exp(\sigma_{\theta,t-1})\,\varepsilon_{\theta,t} \\ \text{uncertainty:} \quad & \sigma_{\theta,t} = \rho^{\sigma_\theta}\sigma_{\theta,t-1} + \sigma_{\sigma_\theta}\varepsilon_{\sigma_\theta,t} \end{aligned}$$

  where $\varepsilon_{\ell,t}$, $\varepsilon_{\sigma_\ell,t}$, $\varepsilon_{\theta,t}$, $\varepsilon_{\sigma_\theta,t} \sim \mathcal{N}(0,1)$.
- Again, we work with the true state space of the model – the whole distribution is included.
- We study
  - how risk and uncertainty affect inequality;
  - what is the role of non-linearities (including ZLB) in the model's predictions.

Comment 3: Grain of salt about deep learning technology

- Neural network is a promising approximator but has a large number of parameters and is highly non-linear.
- There are some analytic results on local convergence of neural networks but convergence is not guaranteed.
- Stochastic optimization is magical but its convergence rate is lower and not guaranteed.

Thank you!