Discussion of Kase, Rottner and Stohler: "Generative Economic Modeling"

Michael Reiter

Institute for Advanced Studies, Vienna

March 23, 2026

Motivation

Growing literature on deep learning for HA models.

- What is the problem?
  ”we provide an alternative to standard deep learning-based approaches, which **often require extensive fine-tuning** and **can suffer from instability** due to their endogenous feedback loops.”
- How to solve it:
  ”our methodology **ensures stability and scalability by using precomputed solutions from conventional methods**, allowing for efficient training and accurate approximations of the full economic model.”

Abstract model

- Model

$$\mathbb{S}_t = f\left(\mathbb{S}_{t-1}, \nu_t, \psi\left(S_{t-1}, \nu_t\right), |\Theta\right) \qquad \text{State transition}$$

$$\mathbb{F}(\psi)(\mathbb{S}) = 0 \qquad \text{Euler equation}$$

  where
  - $\mathbb{S}_t$: state vector
  - ν_t: shock vector
  - ψ: policy function
  - Θ: parameters of economic model
- Approximate solution: $\hat{\psi}\left(\mathbb{S}|\Psi\right)$ with parameters Ψ.
- Standard way to find Ψ: choose it so that $\mathbb{F}(\psi)(\mathbb{S}_i) \approx 0$ on some set $\mathbb{S}_i$ (or minimize sum of squared residuals)

Why is this difficult?

- Make $\hat{\psi}(\mathbb{S}|\Psi)$ flexible enough to handle big model (large Ψ)
- How to solve nonlinear system $\mathbb{F}\big(\hat{\psi}(\mathbb{S}|\Psi)\big)(\mathbb{S}_i) = 0$?
  A complicated equation system; sum of squared residuals is difficult to differentiate through in reverse mode.
- Feedback loop: grid $\mathbb{S}_i$ is endogenous, usually simulated from intermediate solutions; an issue especially when $\mathbb{S}$ contains a distribution

How can ”generative modelling” be stable?

- Solving for general equilibrium is much harder than
  - learning the properties of a given sample
  - solving a dynamic optimization problem (either by time iteration or by reinforcement learning)
- Traditional methods for smaller problems often get around it:
  - Backward induction from a reasonable starting point
  - Homotopy (continuation) methods
- What does ”generative modelling” do to solve a large model?
  1. Define simpler subproblems of the full problem, which sufficiently overlap
  2. Solve each subproblem by traditional solution method, which is supposed to be robust
  3. Train a large NN from the simulated solutions of subproblems; much more robust than solving the set of Euler equations

The title: what does ”generative” mean?

- page 2: ”Our method belongs to the class of generative artificial intelligence because we employ a neural network to generate results for the complete model that includes all features and states, something we have not used for the training process. ”
- page 10: ”For this reason, we want to use deep learning to learn the underlying dynamics of the full model from the set of submodels, leveraging deep learning's generative capacity.”

Does generative modeling generate something new?

- or just summarize the results of the submodels?
  Cautious remarks by the authors:
  - ”... some higher-order interaction terms remain unobserved during training because each submodel includes only a subset of features ... ”
  - the surrogate approximation will contain some residual error ...”
  - but ”in many economic models, the contribution of higher-order interaction terms—such as those arising from multiple shocks—tends to diminish.”
- If submodels are solved by traditional methods, does final outcome inherit the limitations of the traditional model (such as functional form of ALM in Krusell/Smith)?

Simple analogy

Problem: learn joint distribution from marginal distributions

- Use additional assumptions, for example fixed copula (Bayer/Luetticke)
- Here:
  - solution described by NN with given structure
  - parameters obtained by stochastic gradient descent
- Question:
  If generative solution is successful, is it because the model does not have significant ”higher-order interaction”, or because the solution captures it (somewhat miraculously, like ”double-descent” phenomenon, Belkin et al., 2019)

Can we test for it?

- Approximation error in full model compared to approximation errors in submodels:
  1. Submodel with traditional method vs. submodel with NN
     If there is no overfitting (error in validation sample not higher than in training sample), does the NN (almost) exactly replicate the KS solution on submodel? Does it improve on it?
  2. Submodel with NN vs. full model with NN
- More precise comparison of approximation error in full model compared to generative approach
- More precise analysis of solution in generative approach; does it correctly capture aspects of higher-order interaction? (where full solution is available)

How general is the method?

All examples work through shutting down aggregate shocks

- In HA models, could it work with shutting down individual shocks? (more substantial simplification)
- Other types of submodels, without shutting down shocks?

Can you go further?

Is your solution a good starting point for improving the NN directly?

Summary

- Presents an interesting novel idea to solve complex DSGE models, including heterogeneous agent models
- Ex ante it is not obvious that it works well (do the subproblems together effectively sample the relevant state space of the complete model?)
- Paper shows it works on some interesting examples
- Further explorations welcome
- More work on understanding (interpreting) the method welcome