Spooky Boundaries at a Distance: Inductive Bias, Dynamic Models, and Behavioral Macro

Presenter: **Mahdi Ebrahimi Kahou;** Jesus Fernandez-Villaverde; Sebastian Gomez-Cardona; Jesse Perla; Jan Rosa Discussion: **Lilia Maliar**

October 18, 2024

"Spooky boundaries"

(1) 

[1] *.jpg*

A timely paper to discuss before Halloween!

O que este artigo faz?

Excellent, creative, and ambitious paper:

- The authors demonstrate how deep learning (DL) can be applied to solve dynamic economic models.
- DL can help satisfy long-run boundary conditions, such as transversality conditions, without explicitly imposing them.
- The machine learning (ML) concepts that are shown to be relevant for this result are inductive bias, the double descent phenomenon, minimum norm, and Sobolev norm.
- ML-based methods effectively approximate solutions in macroeconomic models with multiple steady states, bubbles, and explosive trajectories.

Connection to ML literature

Inductive Bias

= A preference for a simpler model when fitting a general model with limited observations.

- In essence, inductive bias guides how an algorithm selects among the many possible models or solutions that are consistent with the observed data.
- Since learning from data alone does not provide enough information to choose a unique model, the algorithm relies on certain built-in biases or preferences. These biases help the model favor one solution over another, ideally leading to better generalization on new, unseen data.
- Inductive bias helps prevent overfitting by pushing the model toward simpler or more general solutions that are more likely to generalize well to unseen data.
- Types of Inductive Bias:
    - *Occam's Razor, Smoothness Bias, Bias Toward Flat Solutions*

Connection to ML literature

Double descent phenomena

= An unexpected behavior in ML models, particularly in overparameterized models such as deep neural networks.

- In **classical ML**, with a few parameters, the model *underfits* the data, and with too many parameters, it *overfits* by capturing noise in the training data. ⇒ a U-shaped curve in terms of generalization error
- In the **double descent phenomenon**, this degradation in performance does not continue indefinitely.
  – After an initial increase in error (overfitting), as the number of parameters grows further, the model's performance improves again.
  – This results in a second descent in the error curve ⇒ "double descent".
- This phenomenon is especially pronounced in neural networks, where models with billions of parameters can outperform simpler models, as they avoid the traditional overfitting issues.

Connection to ML literature

Sobolev norm

= A mathematical tool used to measure both the value of a function and the size of its derivatives.

- It helps assess the smoothness of a function by taking into account not just the function's values but also its gradients.
- The authors consider this norm in the context of penalizing "non-smooth" or "explosive" solutions to functional equations.
- In dynamic economic models, the objective is often to find solutions that not only satisfy the functional equations but also respect boundary conditions like transversality (i.e., solutions that do not "explode" over time).
- The Sobolev norm becomes useful here because it penalizes functions with large derivatives, effectively discouraging steep or explosive trajectories that violate economic assumptions about long-run stability.

Comments

- More complex applications are needed
- Not importance of checking transversality condition (TVC)
- Non-smooth problems
- Relation to Turnpike Theorem
- A salt of grain about neural networks

Comment 1: More complex models are needed I

- **Applications Studied: Two Canonical Models**
  – A linear asset pricing model with bubbles.
  – A neoclassical growth model with multiple steady states (including a steady state where TVC is violated and two steady states, each with its own domain of attraction).
- These applications are too simple to draw broad conclusions about the usefulness of double descent and inductive bias in DL.
- In particular, for high-dimensional models, it is unclear whether an overfitted DL model would even be computationally feasible.
- Additionally, would overfitting affect the predictions of high-dimensional models in the same way as it affects low-dimensional models?

Comment 1: More complex models are needed II

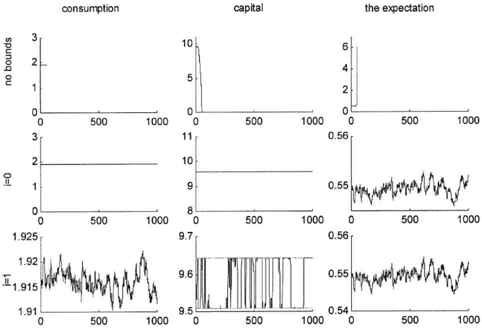
- Adding more examples and clarifying certain technical aspects could strengthen the contribution.
- More complex (and more interesting) macroeconomic examples to consider would be large-scale New Keynesian models, Krusell-Smith models (1998) with distributional dynamics or climate change models.
- The paper does not discuss many computational details. However, they are important for the overall performance of the DL solution method.
- For example, in large-scale new Keynesian models with distributions, one neural network is not enough.

Comment 2: How important is it to check TVC? I

- Solution algorithms differ in their degrees of instability / stability.
- Stability depends on many choices: type of regression, integration method, approximation function, ...
- For example, early versions of PEA (Marcet, 1998) were highly unstable.
- But next generation stochastic simulation algorithms, SSA (e.g., generalized SSA of Maliar et al. 2011) is highly stable.
- *Question:* Do we really need to check TVC if a solution algorithm is stable?

Comment 2: How important is it to check the TVC? II

Maliar and Maliar (2003): moving bounds in PEA



...

Comment 3: Inductive bias and non-smooth problems I

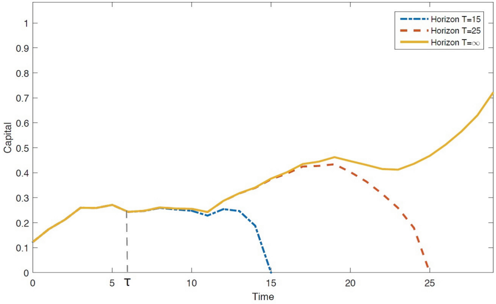
- Inductive bias in neural networks seems to work in smooth problems.
- How smooth do these problems need to be?
- Is it possible to characterize technical conditions for solution smoothness?
- For example, many orthogonal polynomials (like Chebyshev polynomials) are applied to smooth problems, and approximation theory provides bounds on the derivatives of approximation functions. Can we establish similar bounds in this context?
- Will it be possible to require smoothness of the Sobolev norm in models with large shocks or with uncertainty in general?

Comment 3: Inductive bias and non-smooth problems II

- Many interesting economic problems are not smooth.
- These include default models, models with occasionally binding constraints, discrete choice models, etc.
- Will encouraging the selection of solutions with small functional norms still be effective in these cases?
- Further research is required to explore this question.

Comment 4: Relation to Turnpike Theorem I

Turnpike Theorem: If we are interested in the behavior of infinite-horizon non-stationary economy during some initial number of periods τ, we can accurately approximate the infinite-horizon solution by solving the finite-horizon model.



Comment 4: Relation to Turnpike Theorem II

Argument in the paper: "inductive bias favors transition dynamics that tend toward the turnpike without explicitly characterizing it, given that the turnpike trajectory is the unique path that does not diverge."

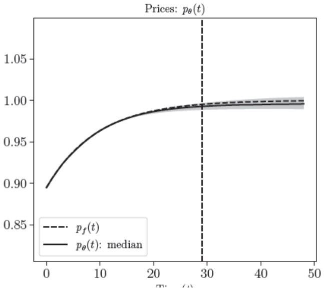


Figure 1

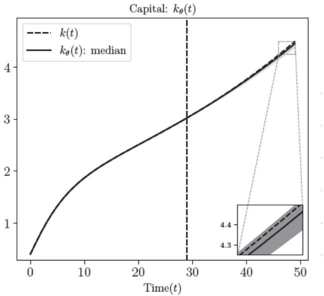
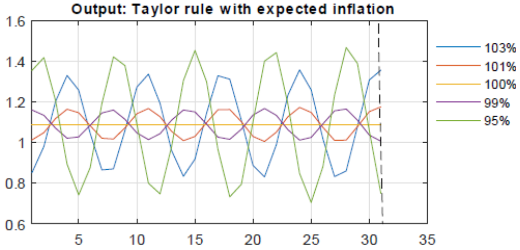


Figure 3

Comment 4: Relation to Turnpike Theorem III

- Turnpike Theorem holds under certain assumptions (see Maliar, Maliar, Taylor and Tsener, 2020). It allows to analyze non-stationary models.
- My guess is that the DL method based on inductive bias would not work for non-stationary models or for the models with complex roots and backward stability:



see Maliar and Taylor (2021).

Comment 5: Grain of salt about neural networks

- 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.
- Own experience points to the fact that neural networks are very difficult to work with:
  - Maliar, Maliar and Winant (2021) show how to use DL for solving Krusell-Smith (1998) model.
  - Gorodnichenko, Maliar, Maliar and Naubert (2021) use it to solve a new Keynesian model with distributions.

Other comments

- The Sobolev norm is just one of many ways to measure function smoothness. A numerical comparison of different norms would be valuable.
- There is limited discussion on potential risks of using inductive bias in practice. For instance, under what conditions might the ML model fail to meet long-run boundary conditions or produce inaccurate results?
- The paper notes that the solutions generated by ML methods are "almost stable," implying that periodic retraining may be necessary to prevent divergence in the long run. How often should we retrain?
- Some of the more technical aspects (the precise role of functional norms in ML models or the implications of the double descent phenomenon) could benefit from further elaboration.

Thank you!