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

Mahdi Ebrahimi Kahou¹ Jesús Fernández-Villaverde² Sebastián Gómez-Cardona³ Jesse Perla⁴ Jan Rosa⁴

Conference on Frontiers in Machine Learning and Economics, Federal Reserve Bank of Philadelphia

¹Bowdoin College

²University of Pennsylvania

³Morningstar, Inc.

⁴University of British Columbia

Motivation, Question, and

Contribution

Motivation

In the long run, we are all dead—J.M. Keynes, A Tract on Monetary Reform (1923)

- Numerical solutions to dynamical systems are central to many quantitative fields in economics.
- Dynamical systems in economics are **boundary value** problems:
 - 1. The boundary is at **infinity**.
 - 2. The values at the boundary are potentially unknown.
- Resulting from forward looking behavior of agents.
- Examples include the *transversality* and the *no-bubble* condition.
- Without them, the problems are ill-posed and have infinitely many solutions:
 - The problems are ill-posed in the Hadamard sense, meaning the solutions are not unique.
 - These forward-looking boundary conditions are a key limitation on increasing dimensionality.

1

Question

Question:

Can we (economists and agents) **ignore** these long-run boundary conditions and still have accurate short/medium-run dynamics disciplined by these long-run conditions?

Contribution

- 1. **Yes**, it is possible to meet long-run boundary conditions **without** strictly enforcing them as a constraint on the model's dynamics.
 - We show how using Machine Learning (ML) methods achieve this method.
 - This is due to the **inductive bias** of ML methods.
 - In this paper focusing on deep neural networks
- 2. We argue how inductive bias can serve as a micro-foundation for modeling forward-looking behavioral agents.
 - Easy to compute.
 - Provides short-run accuracy.
 - Satisfies the necessary long-run constraints.

Background: Economic Models, Deep learning and inductive bias

Economic Models: functional equations

Many theoretical models can be written as functional equations:

- Economic object of interest: f where $f: \mathcal{X} \to \mathcal{R} \subseteq \mathbb{R}^N$
 - e.g., asset price, investment choice, best-response, etc.
- Domain of $f: \mathcal{X}$
 - e.g. space of dividends, capital, opponents state or time in sequential models.
- The "model" error: $\ell(x, f) = \mathbf{0}$, for all $x \in \mathcal{X}$
 - e.g., Euler and Bellman residuals, equilibrium FOCs.

Then a **solution** is an $f^* \in \mathcal{F}$ where $\ell(x, f^*) = \mathbf{0}$ for all $x \in \mathcal{X}$.

Approximate solution: deep neural networks

- 1. Sample \mathcal{X} : $\mathcal{D} = \{x_1, \dots, x_N\}$
- 2. Pick a deep neural network $f_{\theta}(\cdot) \in \mathcal{H}(\theta)$:
 - θ : parameters for optimization (i.e., weights and biases).
- 3. To find an approximation for f solve:

$$\min_{\theta} \frac{1}{N} \sum_{x \in \mathcal{D}} \|\ell(x, f_{\theta})\|_2^2$$

- Deep neural networks are highly over-parameterized.
- Formally, $|\theta| \gg N$

Over-parameterized interpolation

- Over-parameterized ($|\theta| \gg N$), the optimization problem can have many solutions.
- ullet Since individual heta are irrelevant it is helpful to think of optimization directly within ${\cal H}$

$$\left(\min_{f_{\theta} \in \mathcal{H}} \sum_{x \in \mathcal{D}} \|\ell(x, f_{\theta})\|_{2}^{2} \right)$$

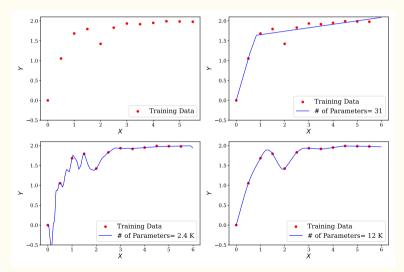
- But which f_{θ} ?
- ullet Mental model: chooses min-norm interpolating solution for a (usually) unknown functional norm ψ

```
egin{aligned} \min_{\hat{f} \in \mathcal{H}} & \|f_{	heta}\|_{\psi} \ & 	ext{s.t.} \ \ell(x, f_{	heta}) = 0, \quad 	ext{ for all } x \in \mathcal{D} \end{aligned}
```

- That is what we mean by **inductive bias** (see Belkin, 2021 and Ma and Yang, 2021).
- Characterizing *S* (e.g., sobolev norms or semi-norms?) is an active research area in ML.

Smooth interpolation

• Intuition: biased toward solutions which are flattest and have smallest derivatives



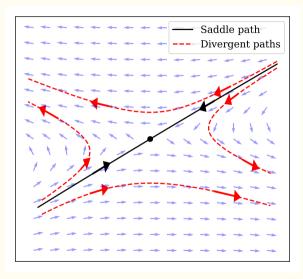
Intuition of the paper

Minimum-norm implicit bias:

- Over-parameterized models (e.g., large neural networks) interpolate the train data.
- They are biased towards interpolating functions with smaller norms.
- So they dont like explosive functions.

Violation of economic boundary conditions:

- Sub-optimal solutions diverge (explode) over time.
- They have large or explosive norms.
- This is due to the saddle-path nature of econ problems.



Outline

Outline of the Talk

To explore how we can ignore events after "we are all dead", we show deep learning solutions to

- 1. Classic linear-asset pricing model.
- 2. Sequential formulation of the neoclassical growth model.
- 3. Sequential formulation of the neoclassical growth model with non-concave production function.
- 4. Equivalent for a recursive formulation of the neoclassical growth model.

Linear asset pricing and he no-bubble condition

Linear asset pricing: setup

• The risk-neutral price, p(t), of a claim to a stream of dividends, y(t), is given by the recursive equation:

$$p(t) = y(t) + \beta p(t+1)$$
, for $t = 0, 1, \cdots$

- $\beta < 1$, and y(t) is exogenous, y(0) given.
- This is a two dimensional dynamical system with unknown initial condition p(0). This problem is ill-posed.
- A family solutions

$$p(t) = \underbrace{p_f(t)}_{\text{fundamentals}} + \underbrace{\zeta(\frac{1}{\beta})^t}_{\text{explosive bubble}}$$

• $p_f(t) \equiv \sum_{\tau=0}^{\infty} \beta^{\tau} u(t+\tau)$. Each solution corresponds to a different $\zeta > 0$.

Linear asset pricing: the long-run boundary condition

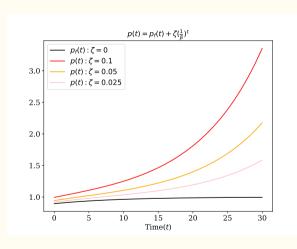
• Long-run boundary condition that rule out the explosive bubbles and chooses $\zeta = 0$

$$\lim_{t\to\infty}\beta^t p(t)=0$$

 Any norm with positivity preservation in norms (think of L_p or Sobolev (semi-)norms)

$$\min_{\zeta\geq 0}\|p(t)\|_{\psi}=\|p_f\|_{\psi}$$

 Therefore, using a deep neural network and ignoring the no-bubble condition should give us p_f(t)



Linear asset pricing: numerical method

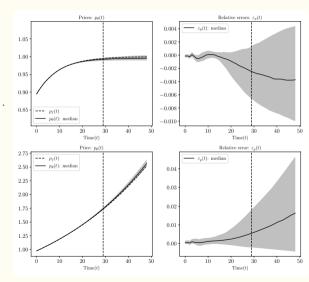
- Sample for time: $\mathcal{D} = \{t_1, \dots, t_N\}.$
- Dividend process: y(t+1) = c + (1+g)y(t), given y(0).
- A over-parameterized neural network $p_{\theta}(t)$, ignore the long run boundary condition and solve

$$\min_{\theta} \frac{1}{N} \sum_{t \in \mathcal{D}} \left[p_{\theta}(t) - y(t) - \beta p_{\theta}(t+1) \right]^2 \tag{1}$$

• This minimization should provide an accurate short- and medium-run approximation for price based on the fundamentals $p_f(t)$.

Linear asset pricing: results

- Two cases: g < 0 and g > 0.
- Relative errors: $\varepsilon_p(t) \equiv \frac{p_{\theta}(t) p_f(t)}{p_f(t)}$.
- for g > 0: $p_{\theta}(t) = e^{\phi t} NN_{\theta}(t)$, ϕ is "learnable".
- Results for 100 different seeds (initialization of the parameters):
 - important for non-convex optimizations.
- Very accurate short- and medium-run approximation.



Sequential neoclassical growth model and transversality condition

neoclassical growth model: setup

• Total factor productivity z(t)

$$p(t) = y(t) + \beta p(t+1)$$
, for $t = 0, 1, \cdots$

- $\beta < 1$, and y(t) is exogenous, y(0) given.
- This is a two dimensional dynamical system with unknown initial condition p(0). This problem is ill-posed.
- A family solutions

$$p(t) = \underbrace{p_f(t)}_{\text{fundamentals}} + \underbrace{\zeta(\frac{1}{\beta})^t}_{\text{explosive bubble}}$$

• $p_f(t) \equiv \sum_{\tau=0}^{\infty} \beta^{\tau} u(t+\tau)$. Each solution corresponds to a different $\zeta > 0$.

Conclusion

- Long-run (global) conditions can be replaced with appropriate regularization (local) to achieve the optimal solutions.
- The minimum-norm implicit bias of large ML models aligns with optimality in economic dynamic models.
- Both kernel and neural network approximations accurately learn the right steady state(s).
- Proceeding with caution: can regularization be thought of as an equilibrium selection device?

Appendix