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

Mahdi Ebrahimi Kahou[1] Jesús Fernández-Villaverde[2] Sebastián Gómez-Cardona[3]
Jesse Perla[4] Jan Rosa[4]

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

[1]Bowdoin College

[2]University of Pennsylvania

[3]Morningstar, Inc.

[4]University 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.

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 <u>deep neural networks</u>
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 $f : \mathcal{X} \rightarrow \mathcal{R} \subseteq \mathbb{R}^N$
    - 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 $\mathcal{X}$: $\mathcal{D} = \{x_1, \cdots, x_N\}$
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.
- Since individual θ are irrelevant it is helpful to think of optimization directly within $\mathcal{H}$

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

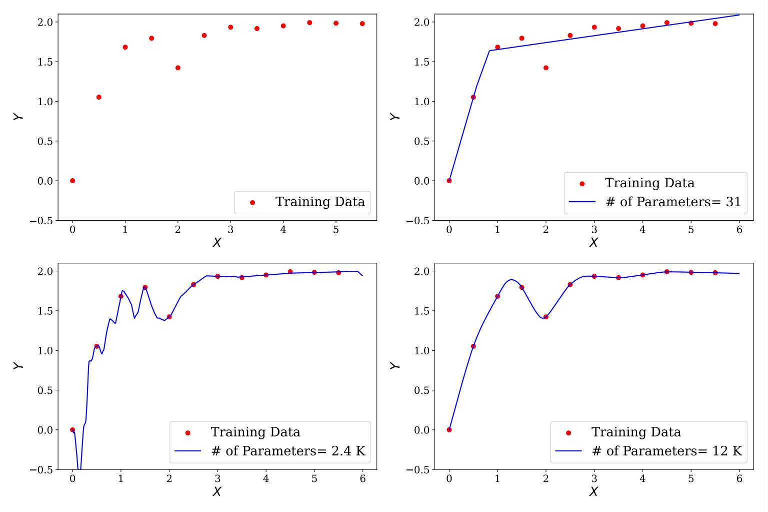
- But which f_θ?
- **Mental model:** chooses min-norm interpolating solution for a (usually) unknown functional norm ψ

$$\begin{aligned} &\min_{\hat{f} \in \mathcal{H}} ||f_\theta||_\psi \\ &\text{s.t.}\, \ell(x, f_\theta) = 0, \quad \text{for all } x \in \mathcal{D} \end{aligned}$$

- 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.

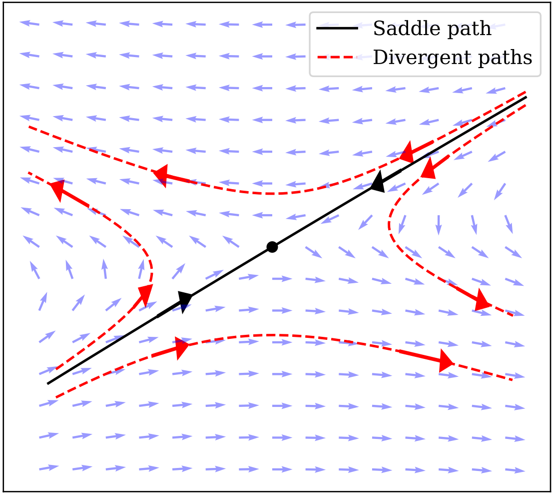
Smooth interpolation

- 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 the 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:

$$p(t) = y(t) + \beta p(t+1), \quad \text{for } t = 0, 1, \cdots$$

- $\beta < 1$, 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

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

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

Linear asset pricing: the long-run boundary condition

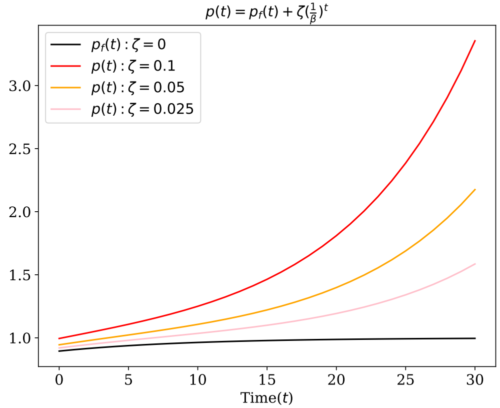
- Long-run boundary condition that rule out the explosive bubbles and chooses $\zeta = 0$

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

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

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

- Ignoring the no-bubble condition and using a deep neural network provides an accurate approximation for $p_f(t)$.



Linear asset pricing: numerical method

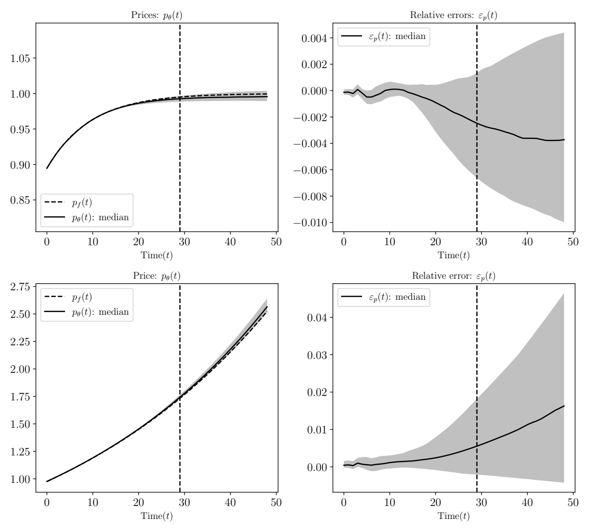
- Sample for time: $\mathcal{D} = \{t_1, \cdots, t_N\}$.
- Dividend process: $y(t+1) = c + (1+g)y(t)$, given $y(0)$.
- A over-parameterized neural network $p_\theta(t)$, **ignore** the non-bubble condition and solve

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

- 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 $z(t)$ exogenously given, capital $k(t)$ with given $k(0)$, consumption $c(t)$, production function $f(\cdot)$, depreciation rate $\delta < 1$, discount factor β :

$$\underbrace{k(t+1) = z(t)^{1-\alpha} f\left(k(t)\right) + (1-\delta)k(t) - c(t)}_{\text{feasibility constraint}},$$

$$\underbrace{c(t+1) = \beta c(t)\left[z(t+1)^{1-\alpha} f'\left(k(t+1)\right) + 1 - \delta\right]}_{\text{Euler equation}}.$$

- This is a three dimensional dynamical system with unknown initial condition $c(0)$. This problem is **ill-posed**.
- A family of solutions, each solution corresponds to a different $c(0)$.

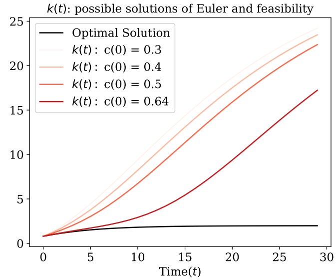
Linear asset pricing: the long-run boundary condition

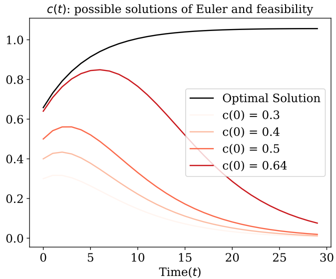
- To rule out sub-optimal solutions, transversality condition

$$\lim_{t\to\infty} \beta^t \frac{k(t+1)}{c(t)} = 0$$

- Any norm with positivity preservation in norms (think of L_p or Sobolev (semi-)norms), the optimal capital path has the lowest norm.

- using a deep neural network and ignoring the transversality condition provides a an accurate approximation for the optimal capital path.





Neoclassical growth model: numerical method

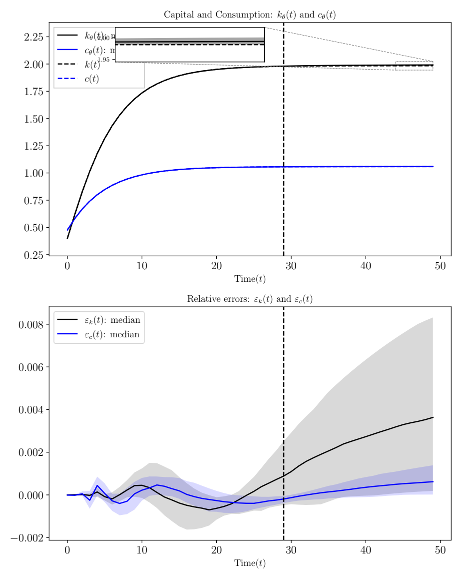
- Sample for time: $\mathcal{D} = \{t_1, \cdots, t_N\}$.
- TFP process: $z(t+1) = (1+g)z(t)$, given $z(0)$.
- A over-parameterized neural network $k_\theta(t)$,
- Given $k_\theta(t)$, define the consumption function $c(t; k_\theta) = z(t)^{1-\alpha} f(k_\theta(t)) + (1-\delta)k_\theta(t) - k_\theta(t+1)$
- **Ignore** the transversality condition and solve

$$\boxed{\min_{\theta \in \Theta} \left[\frac{1}{N} \sum_{t \in \mathcal{D}} \left(\underbrace{\frac{c(t+1; k_\theta)}{c(t; k_\theta)} - \beta \left[z(t+1)^{1-\alpha} f'\left(k_\theta(t+1)\right) + (1-\delta) \right]}_{\text{Euler residuals}} \right)^2 + \left(\underbrace{k_\theta(0) - k_0}_{\text{Initial condition residual}} \right)^2 \right]}$$

- This minimization should provide an accurate short- and medium-run approximation for the optimal capital and consumption path.

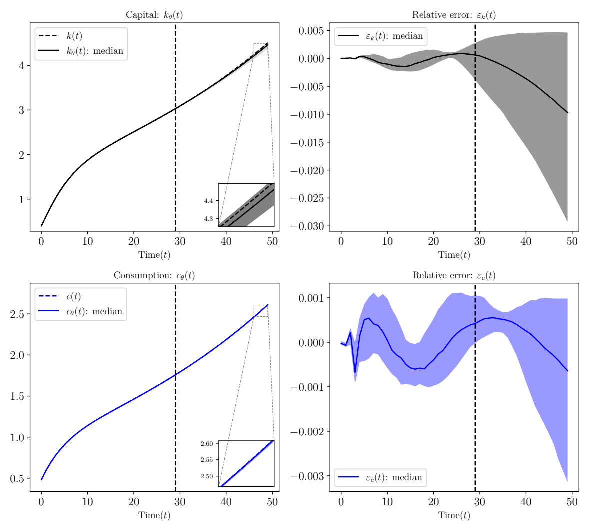
Neoclassical growth model, no TFP growth: results

- $g = 0$, $z(0) = 1$.
- $\varepsilon_k(t) \equiv \frac{k_\theta(t) - k(t)}{k(t)}$, and $\varepsilon_c(t) \equiv \frac{c(t;k_\theta) - c(t)}{c(t)}$
- Benchmark solution: value function iteration.
- Results for 100 different seeds (initialization of the parameters):
  - important for non-convex optimizations.
- Very accurate short- and medium-run approximation.



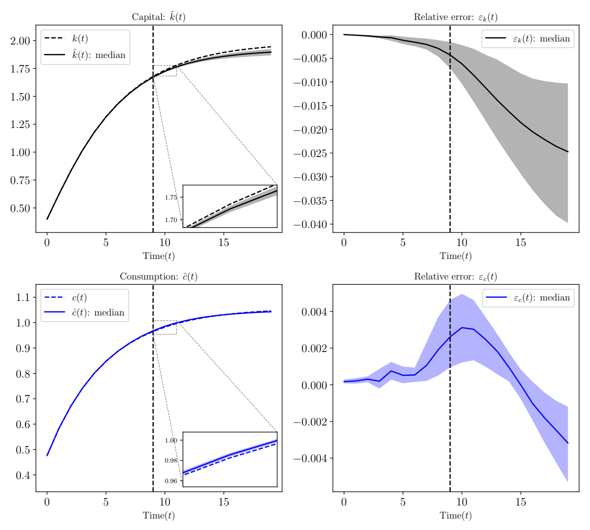
Neoclassical growth model with TFP growth: results

- $g > 0$ and $z(0) = 1$.
- $k_\theta(t) = e^{\phi t} NN_\theta(t)$, ϕ is "learnable".
- Results for 100 different seeds (initialization of the parameters):
  - important for non-convex optimizations.
- Very accurate short- and medium-run approximation.



But seriously, “in the long run, we are all dead”

- So far, we have used long time-horizon $\mathcal{D} = \{0, 1, \cdots, 29\}$.
- In other methods, choosing the time-horizon T is a challenge:
    - Too large $\rightarrow$ accumulation of errors, and numerical instability. We don't have that problem.
    - Too small $\rightarrow$ convergence to the steady state too quickly.
- An accurate short-run solution, even for a medium-sized T.



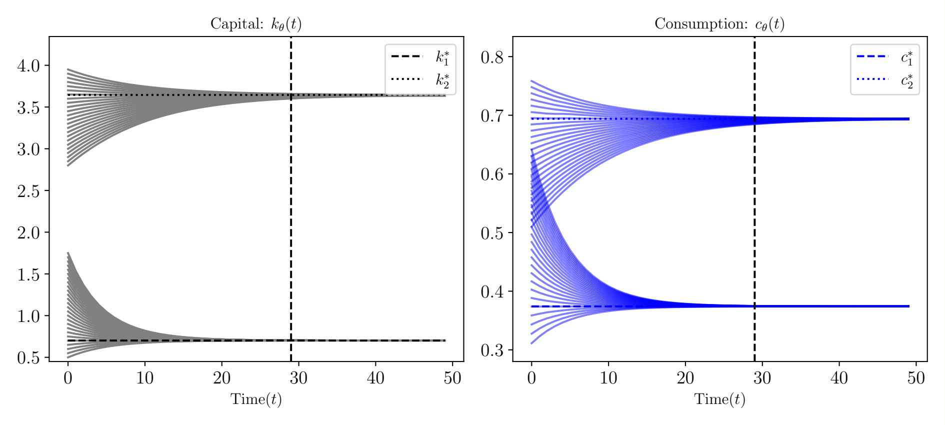
Neoclassical growth model: multiple steady-states and hysteresis

- When there are multiple (saddle-path) steady states, each with its domain of attraction:
  - Can the inductive bias detect there are multiple basins of attraction?
  - How does the inductive bias move us toward the correct steady state for a given initial condition?
- Consider a non-concave production function:

$$f(k) \equiv a \max\{k^{\alpha}, b_1 k^{\alpha} - b_2\}$$

- Two steady-states k_1^* and k_2^*.
- The same numerical procedure, different production function.

Neoclassical growth model with non-concave production function: results



-

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