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]

–

[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.
- Many 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:
  - 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 the 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.
   - This is due to the **inductive bias** of ML methods.
   - In this paper focusing on deep neural networks
2. We argue the inductive bias provides a foundation for modeling forward-looking behavioral agents with self-consistent expectations.
   - Easy to compute.
   - Provides short-run accuracy.

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 "Economics model" error: $\ell(x, f)$
  - e.g., Euler and Bellman residuals, equilibrium FOCs.

Then a **solution** is $f^* \in \mathcal{F}$ 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:

$$\min_\theta \frac{1}{N} \sum_{x \in \mathcal{D}} \| \underbrace{\ell(x, f_\theta)}_{\text{Econ model error}} \|_2^2$$

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

Over-parameterized interpolation

- Being 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 $\mathcal{H}$

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

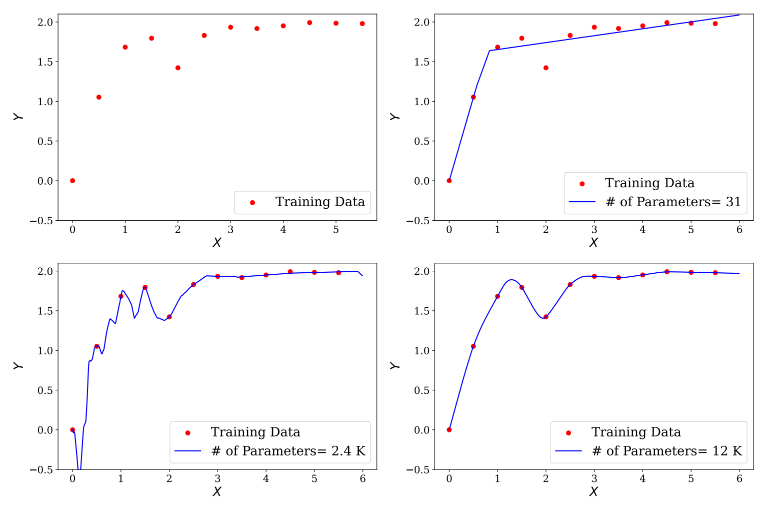
- But which f_θ?
- **Mental model:** chooses min-norm interpolating solution for a (usually) unknown functional norm ψ

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

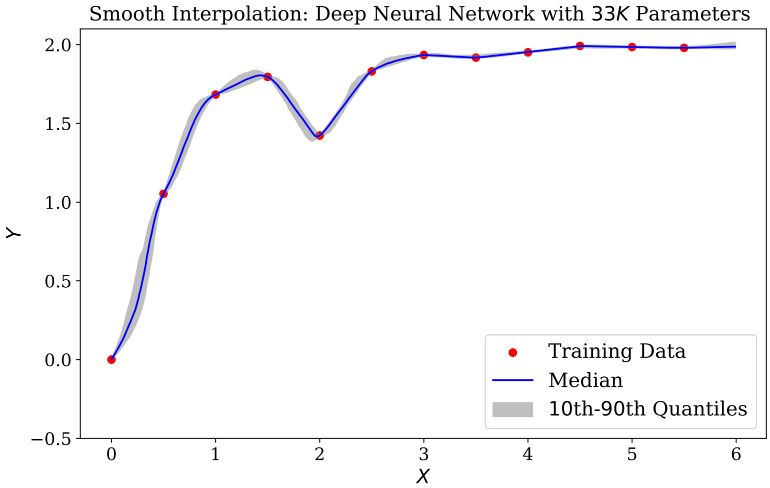
- That is what we mean by **inductive bias** (see Belkin, 2021 and Ma and Yang, 2021).
- Characterizing ψ (e.g., Sobolev norms or semi-norms?) is an active research area in ML.

Over-parameterization and smooth interpolation

- Intuition: biased toward solutions which are flatter and have smaller derivatives

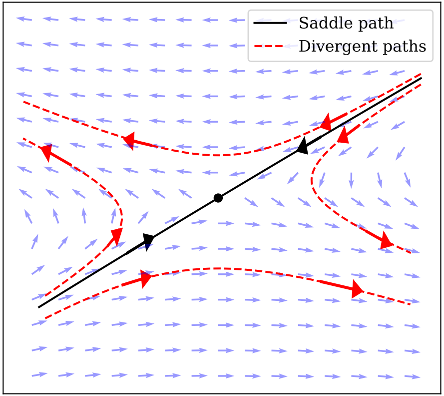


Deep Learning: Smooth Interpolation



Intuition of the paper

- **Minimum-norm inductive 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.
  - This is due to the **saddle-path** nature of econ problems.
  - The long-run boundary conditions rule out the explosive solutions.



Outline

Outline of the talk

To explore how we can ignore the long-run boundary conditions, 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 of solutions

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

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

Linear asset pricing: the long-run boundary condition

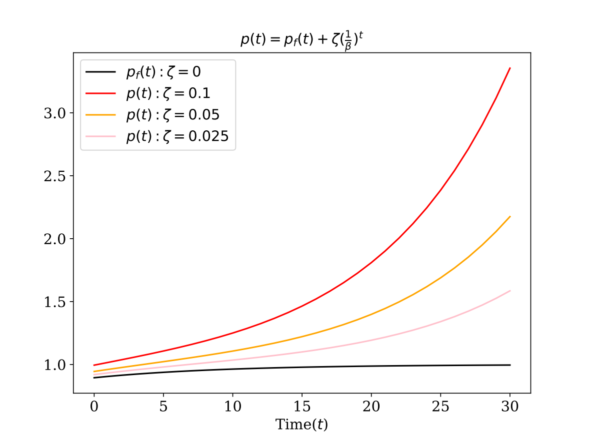
- Long-run boundary condition that rule out the explosive bubbles and chooses $\zeta = 0$

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

- Any norm that preserve monotonicity, like L_p and Sobolev (semi-)norms

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

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



Linear asset pricing: numerical method

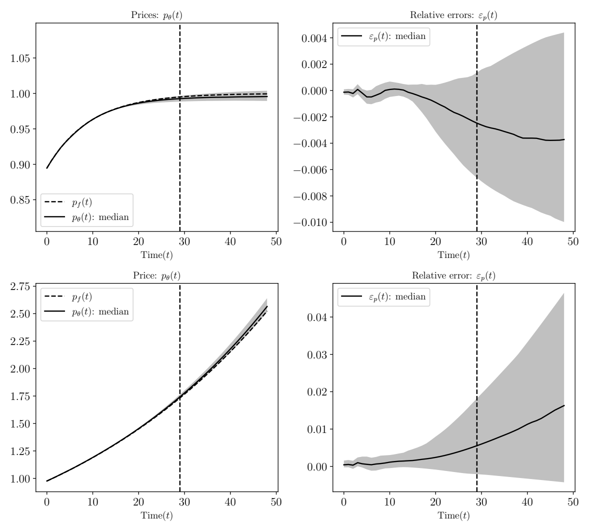
- Sample for time: $\mathcal{D} = \{t_1, \cdots, t_N\}$.
- Generating the dividend process: $y(t+1) = c + (1+g)y(t)$, given $y(0)$.
- An over-parameterized neural network $p_\theta(t)$, **ignore** the non-bubble condition and solve

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

- 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 the 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 :

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

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

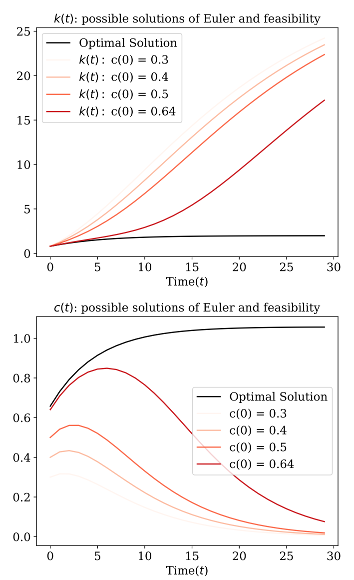
- This is a three dimensional dynamical system with unknown initial condition $c(0)$. This problem is **ill-posed**.
- A family of solutions, each solution corresponds to a different $c(0)$. Only one of them is the optimal solution.

Neoclassical growth model: the long-run boundary condition

- To rule out sub-optimal solutions, transversality condition

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

- 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

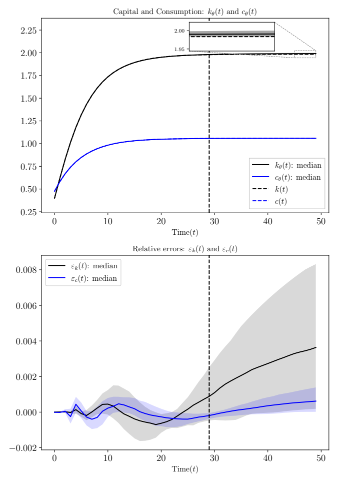
- Sample for time: $\mathcal{D} = \{t_1, \cdots, t_N\}$.
- Generating the TFP process: $z(t+1) = (1+g)z(t)$, given $z(0)$.
- A over-parameterized neural network $k_\theta(t)$,
- Given $k_\theta(t)$, define the consumption function $c(t;k_\theta) = z(t)^{1-\alpha} f(k_\theta(t)) + (1-\delta)k_\theta(t)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

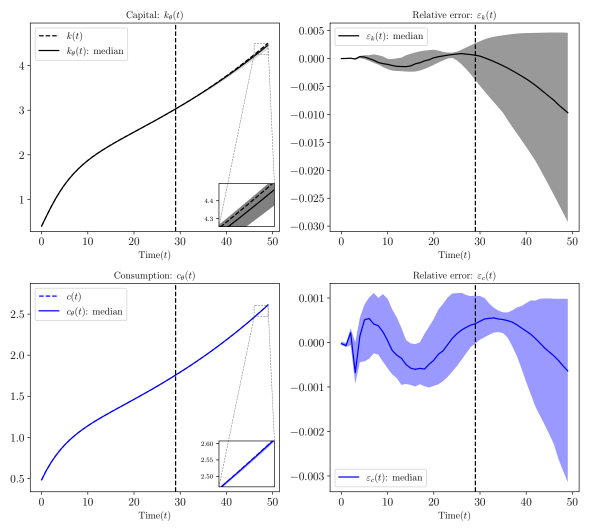
Neoclassical growth model, no TFP growth: results

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



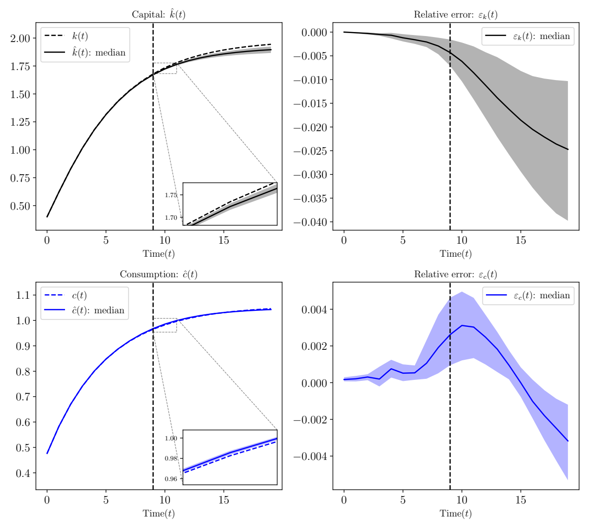
Neoclassical growth model with TFP growth: results

- $g > 0$ and $z(0) = 1$.
- $k_\theta(t) = e^{\phi t} NN_\theta(t)$, ϕ is "learnable".
- Results for 100 different seeds.
- 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.

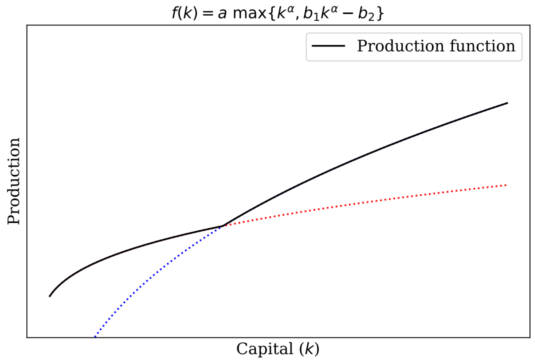


Neoclassical growth model: multiple steady-states and hysteresis

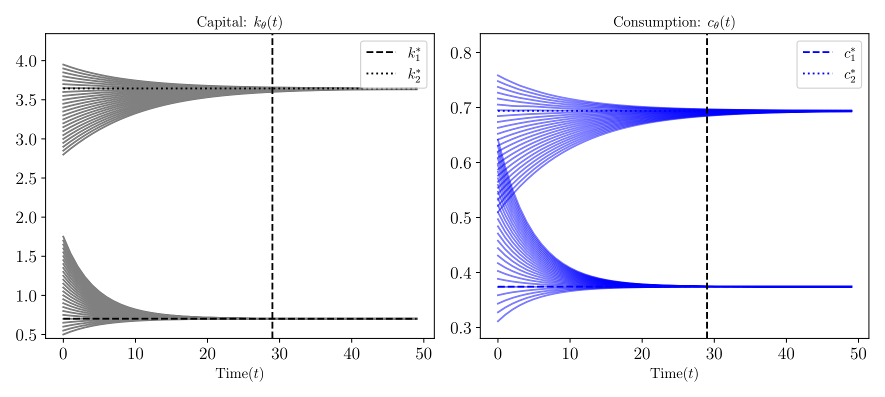
- When there are multiple steady states with saddle-path stability, each with its domain of attraction:
  - Can the inductive bias detect there are multiple basins of attraction?
  - How does the inductive bias move us toward the correct steady state for a given initial condition?
- Consider a non-concave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



Neoclassical growth model with non-concave production function: results



- Different initial conditions in $k_0 \in [0.5, 1.75] \cup [2.75, 4]$.
- In the vicinity of k_1^* and k_2^* the paths converge to the right steady-states.

Deep learning is not the only option: kernels

- Deep learning might be too "spooky".
- We can use kernels, $K(\cdot,\cdot)$, instead of neural networks and control the RKHS norms. For instance:

$$\frac{dk}{dt} = \sum_{i=1}^{N} \alpha_i^k K(t, t_i), \quad \frac{dc}{dt} = \sum_{i=1}^{N} \alpha_i^c K(t, t_i)$$

- The same results, theoretical guarantees, very fast and robust.

How Inductive Bias in Machine Learning Aligns with Optimality in Economic Dynamics

Mahdi Ebrahimi Kahou[1] James Yu[2] Jesse Perla[2] Geoff Pleiss[2,3]
[1]Bowdoin College [2]University of British Columbia [3]Vector Institute
m.ebrahimikahou@bowdoin.edu
yuyuming@student.ubc.ca
jesse.perla@ubc.ca
geoff.pleiss@stat.ubc.ca

Abstract

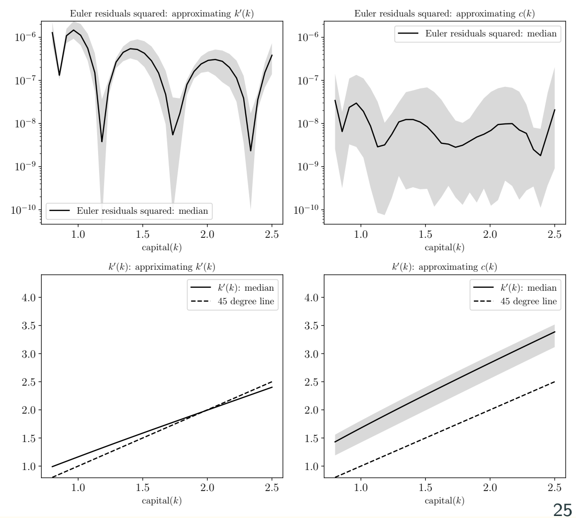
This paper examines the alignment of inductive biases in machine learning (ML) with structural models of economic dynamics. Unlike dynamical systems found in physical and life sciences, economics models are often specified by differential equations with a mixture of easy-to-enforce initial conditions and hard-to-enforce infinite horizon boundary conditions (e.g. transversality and no-ponzi-scheme conditions). Traditional methods for enforcing these constraints are computationally expensive and unstable. We investigate algorithms where those infinite horizon constraints are ignored, simply training unregularized kernel machines and neural networks to obey the differential equations. Despite the inherent underspecification of this approach, our findings reveal that the inductive biases of these ML models innately enforce the infinite-horizon conditions necessary for the well-posedness. We theoretically demonstrate that (approximate or exact) min-norm ML solutions to interpolation problems are sufficient conditions for these infinite-horizon boundary conditions in a wide class of problems. We then provide empirical evidence that deep learning and ridgeless kernel methods are not only theoretically sound with respect to economic assumptions, but may even dominate classic algorithms in low to medium dimensions. More importantly, these results give confidence that, despite solving seemingly ill-posed problems, there are reasons to trust the plethora of black-box ML algorithms used by economists to solve previously intractable, high-dimensional dynamical systems—paving the way for future work on estimation of inverse problems with embedded optimal control problems.

Extensions in the paper

- Sequential models:
    - Shorter time-horizons.
    - Misspecification of growth.
- Recursive neoclassical growth model
    - Accurate short- and medium-run dynamics.
    - Accurate solutions even with TFP growth.
    - Deep learning solutions can go very wrong
        - We should use the information in the transversality condition to know what to approximate.

Deep learning solutions can be misleading: approximating capital vs. consumption

- Capital k is the state variable.
- Two options: approximating capital policy $k'_\theta(k)$ or $c_\theta(k)$
- left panels: results for $k'_\theta(k)$ approximation.
- Right panels: results for $c_\theta(k)$ approximation.
- Only the left panel results are correct. $k'_\theta(k)$ has a fixed point at the right steady state.
- However, the wrong solution has lower Euler residuals.



Conclusion

- Short- and medium-run accurate solutions can be obtained **without** strictly enforcing the long-run boundary conditions on the model's dynamics.
- Long-run (**global**) conditions can be replaced with appropriate regularization (**local**) to achieve optimal solutions, hence the title of the paper.
- Inductive bias provides a foundation for modeling forward-looking behavioral agents with self-consistent expectations.

Discussion: where to go from here?

- Can inductive bias/regularization be thought of as an equilibrium selection device?
  - In this paper it is used to select solutions.
- This method (mostly the kernel method) can be used for sampling high-dimensional state spaces when there is stochasticity.
  - Solve the deterministic in short-run and use the points as sample of the state-space.
  - Then solve the stochastic problem.