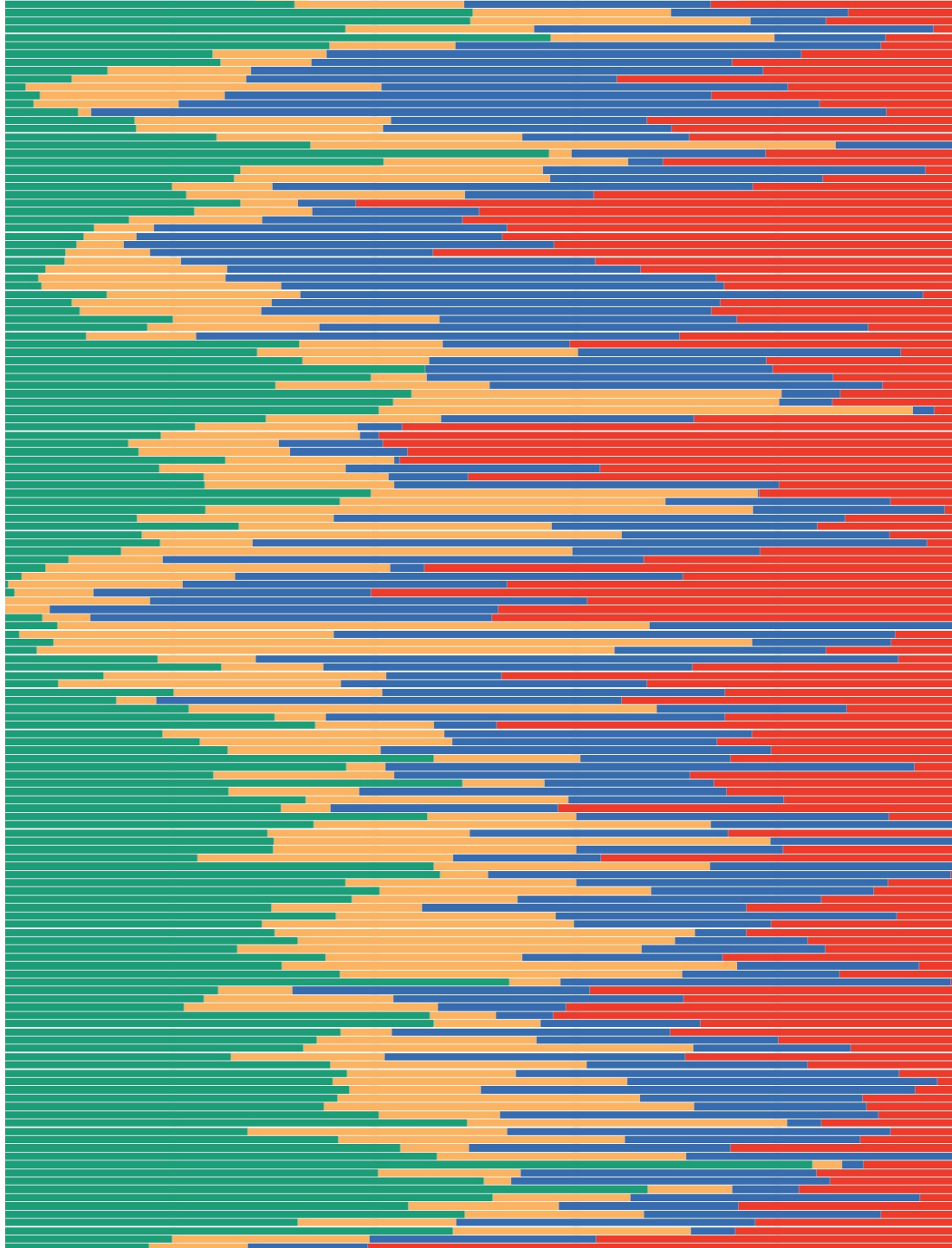
Deep Learning Vignettes for Economists

ECO 930j

By PGC



What’s bad about DNNs

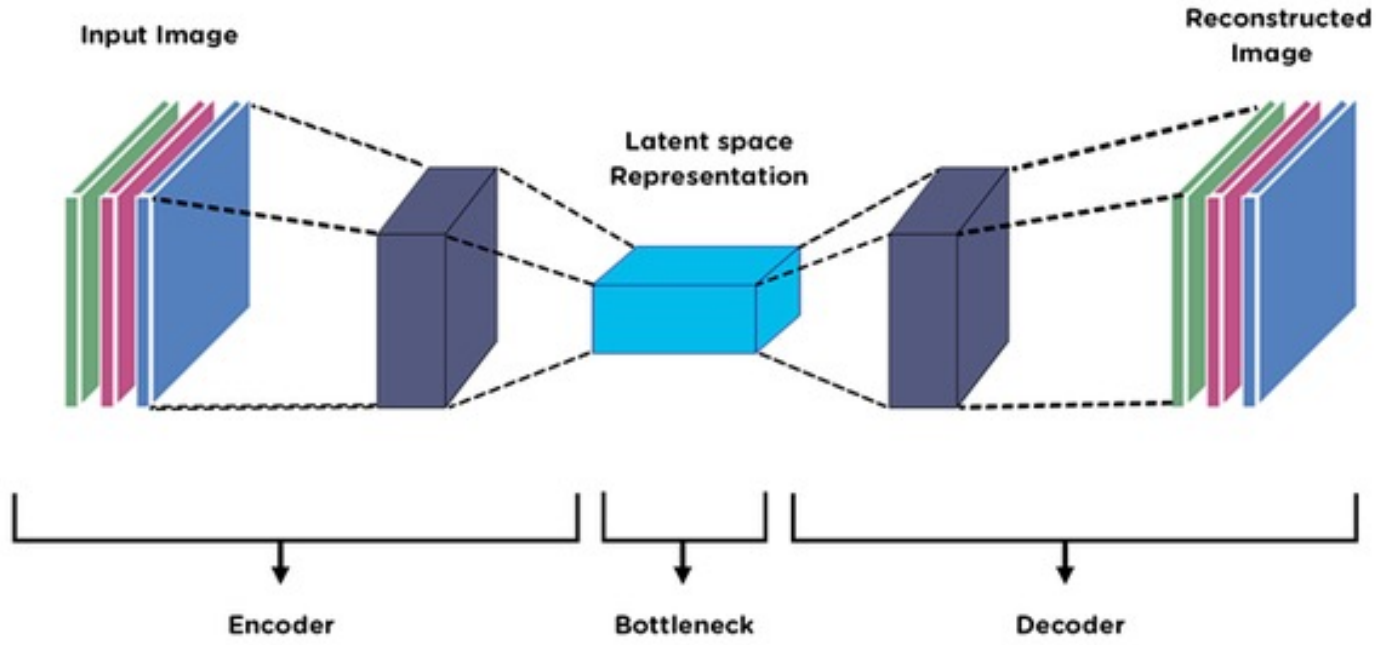
- sensitivity to HPs
- often worse than tree-based models for most tasks that econometricians do, but could be about to change
- black-box, but not so much worse than tree ensembles, and perhaps less
- Incomprehensible new language for stats (”epoch”=optimization steps, “learning”=estimation)

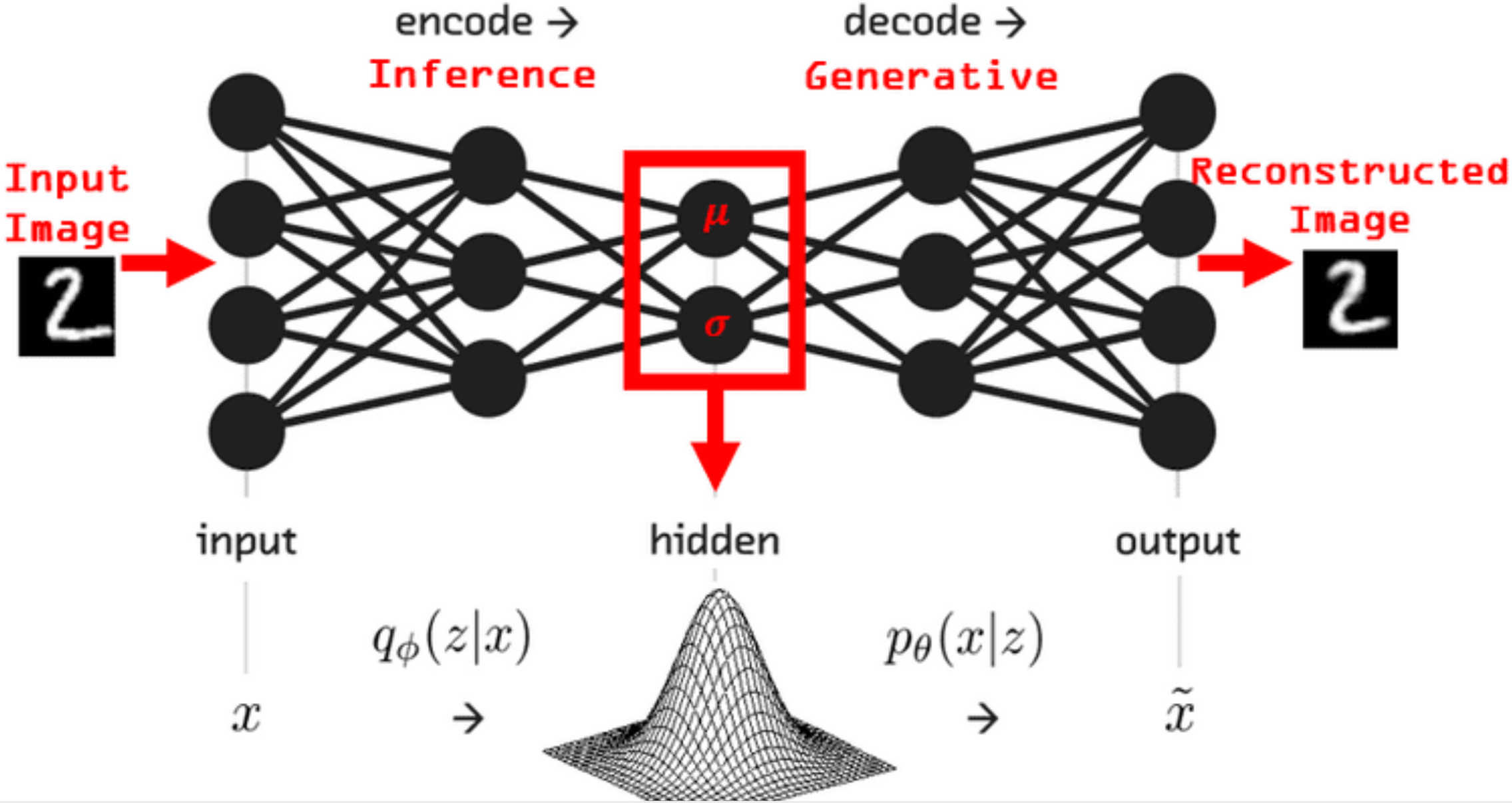
What’s good about DNNs

- Can work well, but rarely outperforms significantly the much easier RF and BT
- But the reason for DL prominence is not typical regression problem for typical tabular data
- Very personalizable, can easily design archi for specification problem other than just extracting E(y|x)

Some well known architectures

- CNN
- RNN
  - These are generalizations of state space models
- Autoencoders
- Transformers and Attention




encode →
Inference
decode →
Generative
Input Image
Reconstructed Image
μ
σ
input
hidden
output
$q_\phi(z|x)$
$p_\theta(x|z)$
x
$\tilde{x}$

Other time series models

- DeepAR
- Prophet (FB)
- Many others (MIT)

Some other cool words you should know

- Transfer Learning
- Multitask Learning
- Stacking
- Variational Inference
- Auto-ML

Avenues for Micro

- Heterogeneity of treatment or other regression parameters
- Farrel et al. (2021)

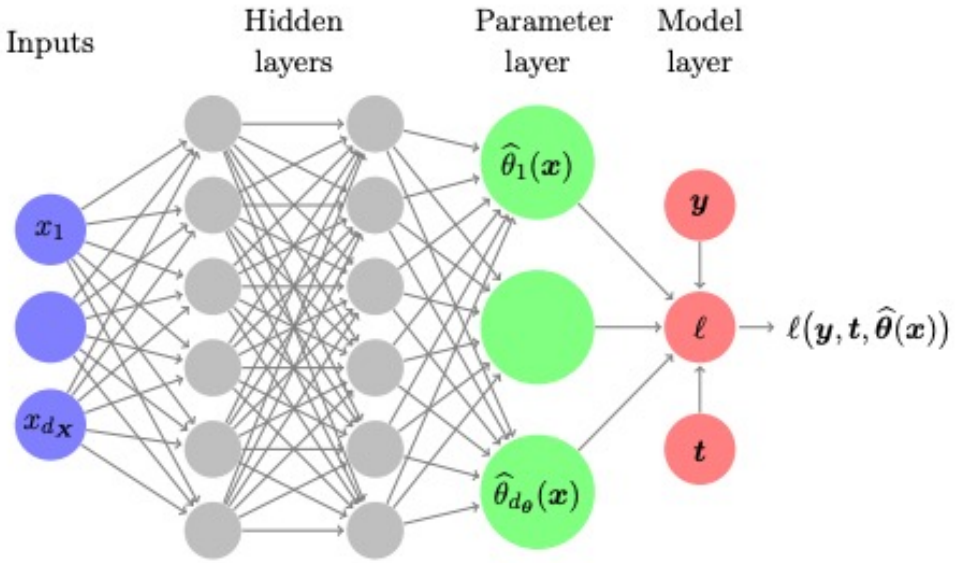


Figure 1: Illustration of the deep neural network estimation of the parameter functions $\boldsymbol{\theta}(\boldsymbol{x})$ for a generic structured model (2.1)

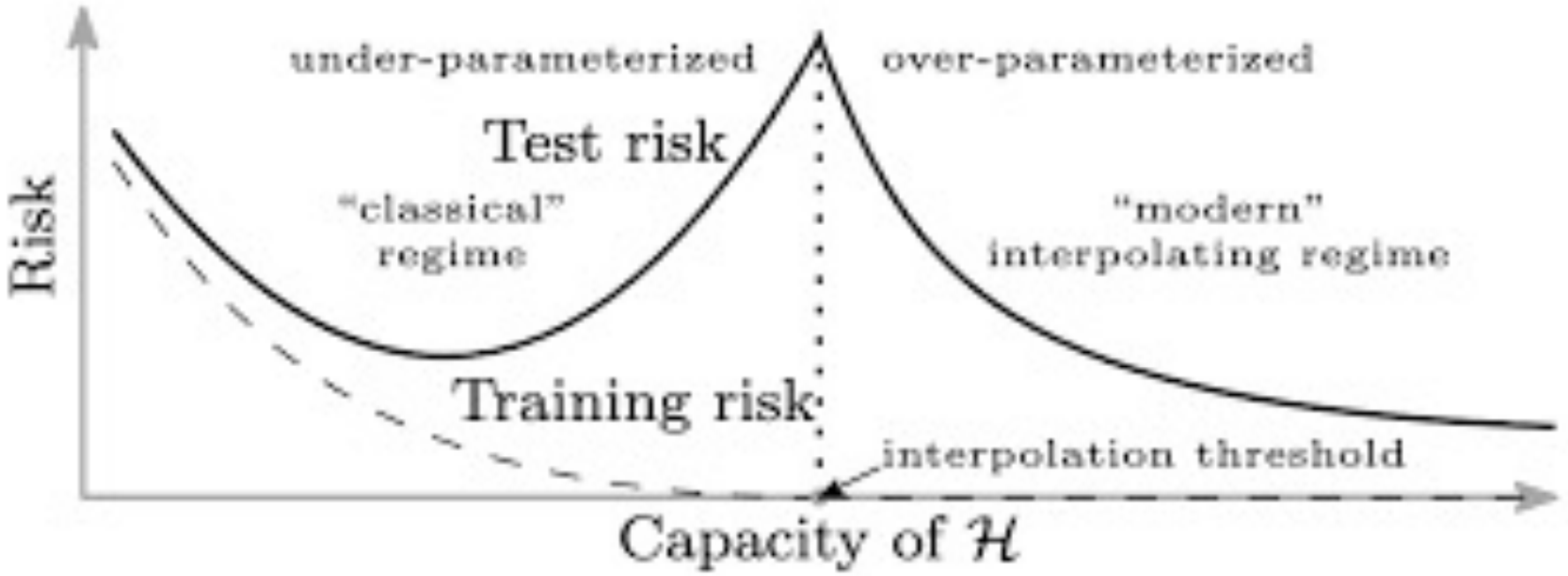
will have $\boldsymbol{\theta}(\boldsymbol{x})$, mapping $\mathbb{R}^{d_X} \mapsto \mathbb{R}^{d_\theta}$, and we assume that the true parameter functions $\boldsymbol{\theta}_0(\cdot)$ solve

$$\boldsymbol{\theta}_0(\cdot) = \underset{\boldsymbol{\theta} \in \mathcal{H}}{\arg\min}\, \mathbb{E}\left[\ell\big(\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{\theta}(\boldsymbol{X})\big)\right], \tag{2.1}$$

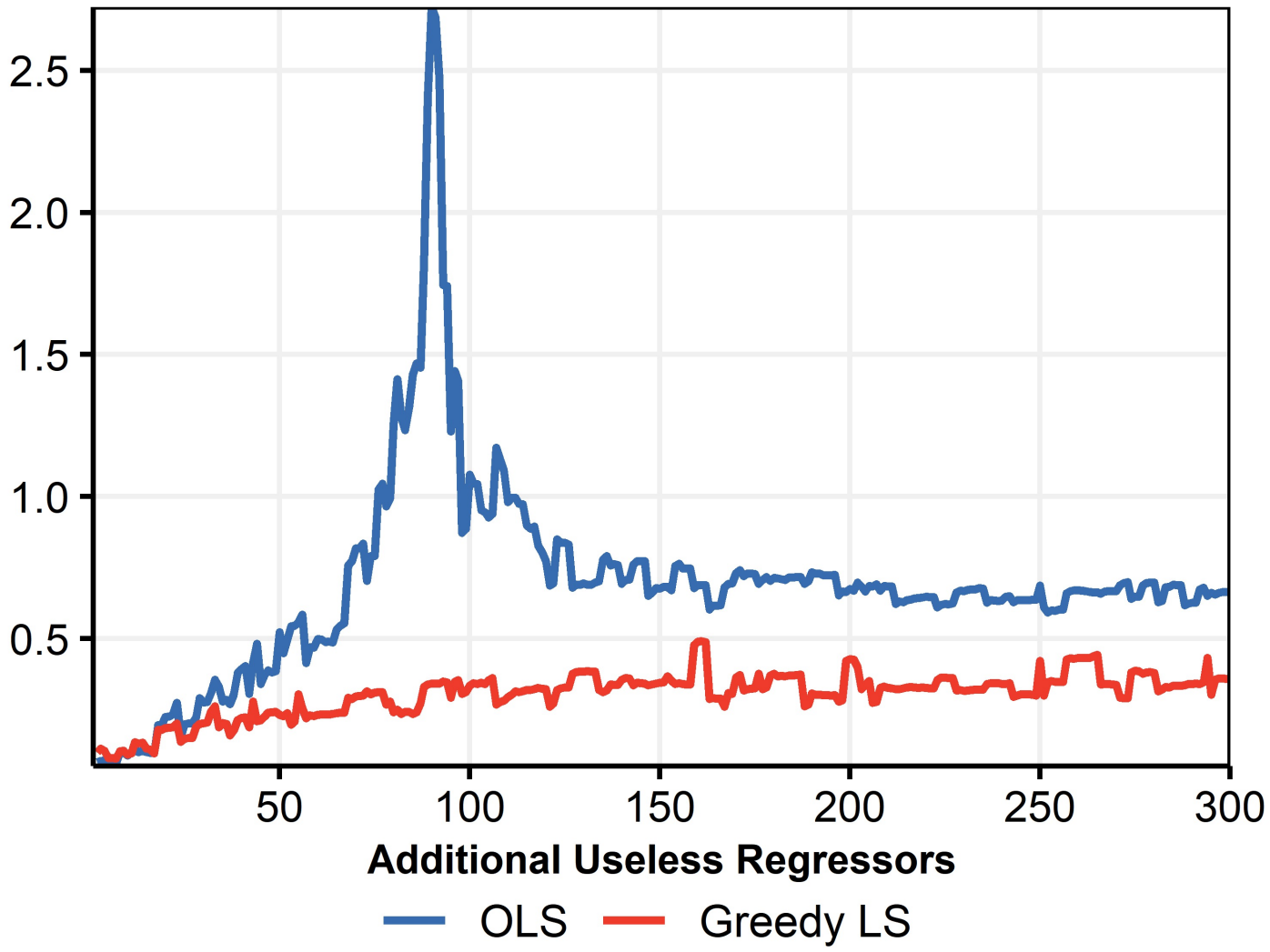
Avenues for Macro

- Sentiment analysis, LDA, NLP, anything with text data
  - Monetary policy shocks
- Time-varying Parameters
- Conditional Mean *and* Volatility, and skewness even (my slides)
- Latent states extraction, like “economic slack” for inflation (my slides)

Bending of the Statistical Mind: Double Descent

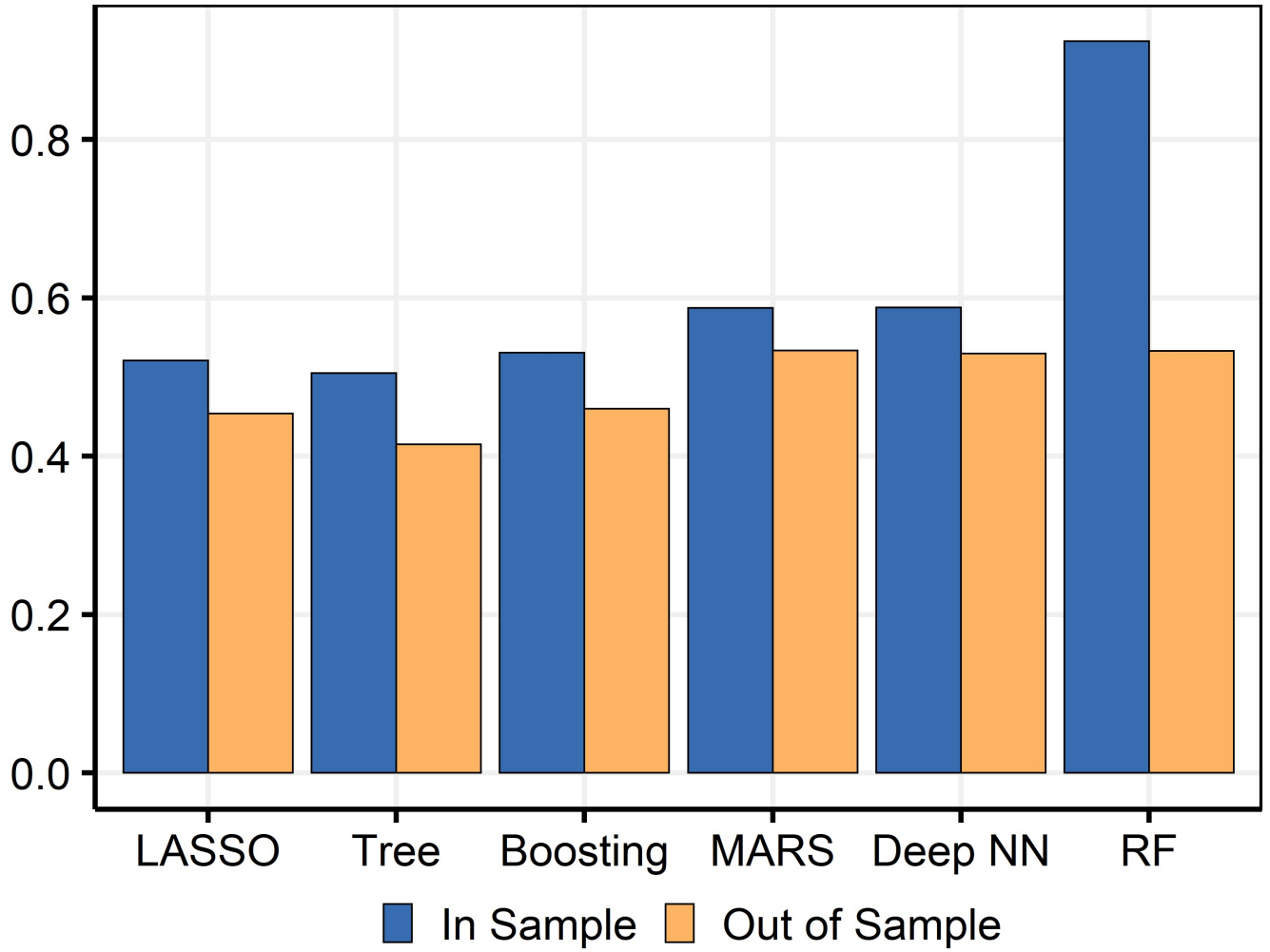


https://hippocampus-garden.com/double_descent/

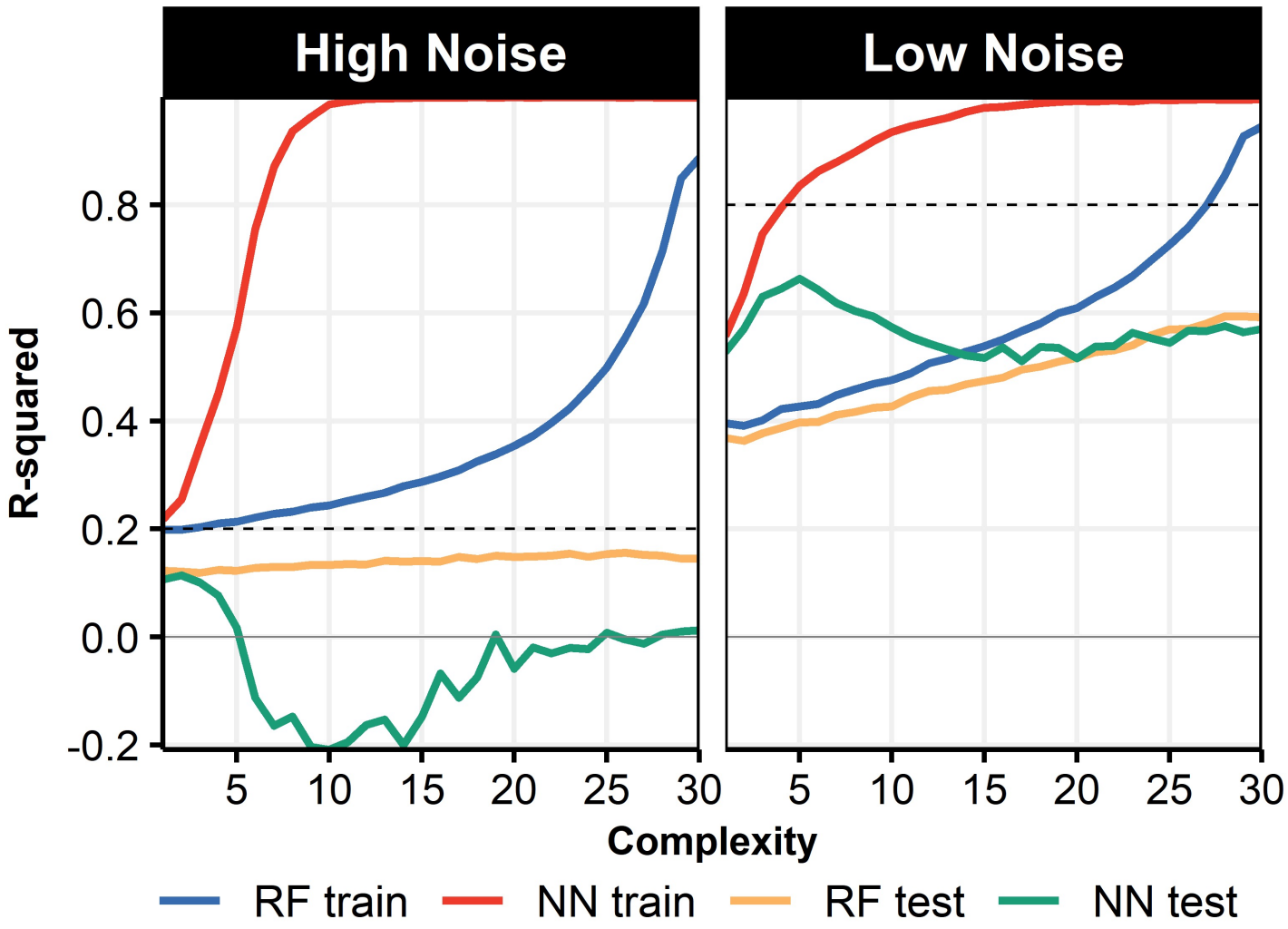


Happens even with... OLS.

RF is a weird animal.



Dominant vs non-dominant tuning strategies



Lessons of DD for economic analysis of ML models

- ***Never*** look at training results, except maybe for Lasso, Ridge and other simpler alternatives
- Look at out-of-sample if you can, or out-of-bag if necessary (historical analysis)

Interpretable ML

- Quite important, sometimes more than the prediction itself (inflation outlook, medicine, etc.)
- Variable importance
- Partial dependance plots
- Shapley values (quite popular now)
- Write a model that is flexible yet inherently interpretable (HNN, MRF)

Parting Words and Outlook for ML Econ Research

- Write specialized DNN architectures
- Macro: get more data (dis-aggregated)
- Macro: write architectures that learn what equations to pool together for increased statistical efficiency, i.e., some (novel?) form of multitask learning