Deep Surrogates for Finance: With an Application to Option Pricing

Yaqi Zeng

CUHK ECON

October 20, 2023

Outline

Outline

Literature

- Methods for constructing surrogate models in general and their application to high-dimensional models in economics and finance (Fernandez-Villaverde and Guerron-Quintana 2020; Kaji, Manresa, and Pouliot 2020)
- Applications of deep learning in finance and economics (Hutchinson, Lo, and Poggio 1994; Becker, Cheridito and Jentszen 2018)
- Empirical option pricing (Heston 1993; Bates 1996)

A Shining Contribution

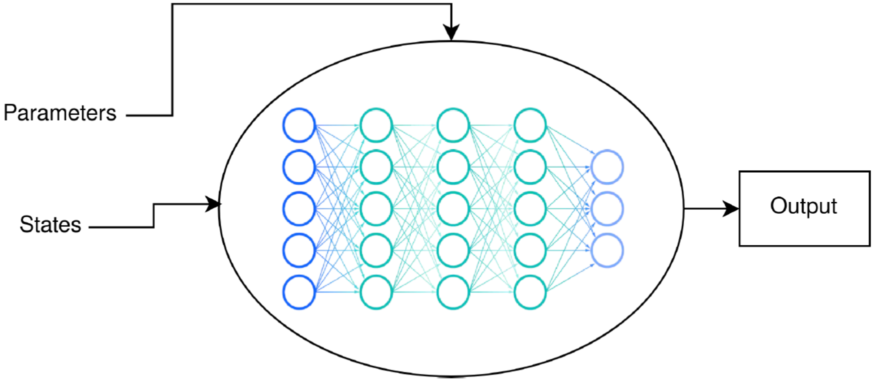
- The deep surrogate not only price options and produces the greeks accurately, but also helps identify parameters accurately in GMM estimations
- Helps reduce the average time for estimating the parameters on a cross-section of 1,000 options from over 40 minutes using traditional methods to less than one second in our numerical experiments

Outline

An economic model

- An economic model
$$\boldsymbol{y}_t = f(\boldsymbol{s}_t|\theta)$$
where $\boldsymbol{s_t} \in \mathbb{R}^n$ is a vector of (observable or hidden) states, $\theta \in \mathbb{R}^p$ is a vector of model parameters, and $\boldsymbol{y}_t \in \mathbb{R}^k$ is a vector of model outputs.
- By treating the model parameters θ as pseudo-states, we can focus on the augmented state $\boldsymbol{x}_t = [\boldsymbol{s}_t, \theta] \in \chi \subset \mathbb{R}^m$ with $m = n + p$, and
$$\boldsymbol{y}_t = f(\boldsymbol{x}_t)$$

Deep Surrogate



$$\phi(\mathbf{x}_t|\theta_{NN}) = y_t$$

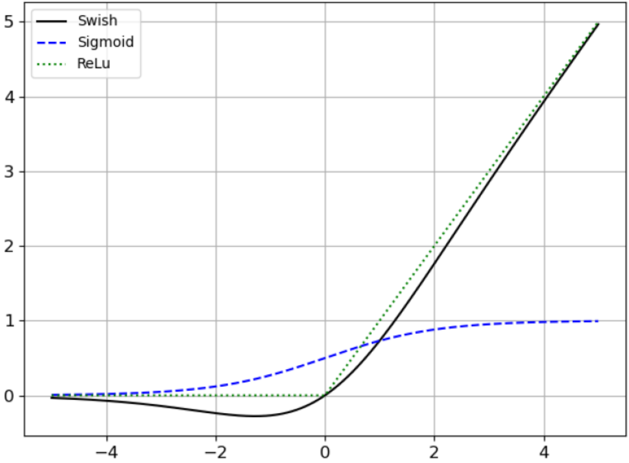
Deep Neural Network

- Neural networks consist of multiple stacked-up layers of neurons
- A given layer l takes a vector I_l as input and produces a vector O_l as output:

$$O_l = \sigma(W_l I_l + b_l),$$

where W_l and b_l consist of unknown parameters for the network; $\sigma(\cdot)$ is a activation function (non-linear).

Activation Functions



Generating the Training Sample

For each element $x_t^{(j)}$ in $\boldsymbol{x}_t$, let $\underline{x}^{(j)}$ and $\overline{x}^{(j)}$ be its lower bound and upper bound, and

$$\underline{\mathbf{x}} = \left[\underline{x}^{(1)}, \underline{x}^{(2)}, \cdots, \underline{x}^{(m)}\right], \quad \bar{\mathbf{x}} = \left[\bar{x}^{(1)}, \bar{x}^{(2)}, \cdots, \bar{x}^{(m)}\right]$$

Draw $\tilde{\boldsymbol{x}}_i$ according to

$$\tilde{\mathrm{x}}_i = \underline{\mathrm{x}} + \tilde{R}_i(\overline{\mathrm{x}} - \underline{\mathrm{x}})$$

where $\tilde{R}_i = diag(\tilde{r}_i)$, with $\tilde{r}_i \in \mathbb{R}^m$ following an appropriate multivariate distribution.

Surrogate Training

Train the neural network $\phi(\boldsymbol{x}_t|\xi)$ by

$$\xi^* = \arg\max_\xi \; L(\boldsymbol{x}, \boldsymbol{y}; \xi),$$

where the loss function is given by

$$L(\boldsymbol{x}, \boldsymbol{y}; \xi) = \frac{1}{N} \sum_{t=1}^{N} ||\phi(\boldsymbol{x}_t|\xi) - \boldsymbol{y}_i||_1$$

Surrogate Validation

The acceptance of the surrogate:

$$\sup_{\boldsymbol{x}_t \in \chi} d(\phi(\boldsymbol{x}_t|\xi), f(\boldsymbol{x}_t)) \leq \epsilon,$$

where $d(\cdot)$ is a distance metric between the two functions, and ϵ is some small positive constant.

Surrogate validation

- Generate a new random sample of size N^o, $(\boldsymbol{x}_j^o, \boldsymbol{y}_j^o)_{j=1}^{N^o}$,
- Evaluate a finite-sample version under the l_1-norm:

$$\sup_j ||\phi(\boldsymbol{x}_i^o|\xi^*) - \boldsymbol{y}_i^o||_1 \leq \epsilon.$$

- If condition is not satisfied, add flexibility of $\phi(\cdot)$ by modifying, for instance, its architecture, and increase the size of the training sample to further reducing training error, return to step one to revalidate the model. Otherwise $\phi(\boldsymbol{x}_t|\xi^*)$ is accepted as a deep surrogate of the model $f(\cdot)$.

Benefits

- Act as a (high-dimensional) lookup table.
- Highly parallelizable in the evaluation of a deep surrogate.
- Get gradient of a surrogate model for little to no computational costs thanks to the backward propagation algorithm.

Deep surrogates for estimation

An example in GMM:

- Moment condition:

$$E[f(\boldsymbol{s}_t|\theta) - \boldsymbol{y}_t] = 0.$$

- Assume $k > p$ and the system of moment restrictions is over-identified. With a sample of T observations, define

$$g_T(\theta) \equiv \frac{1}{T}\sum_{t=1}^{T}(\phi(\boldsymbol{x}_t(\theta)) - \boldsymbol{y}_t),$$

where $x_t(\theta) = [\boldsymbol{s}_t, \theta]$.

Deep surrogates for estimation

An example in GMM:

- GMM estimator with a $k \times k$ weighting matrix W:

$$\hat{\theta} = \arg\max_{\theta} g_T(\theta)^T W g_T(\theta).$$

- The first-order condition is given by

$$d_T(\theta)^T W d_T(\theta) = 0,$$

where

$$d_T(\theta) \equiv \frac{\partial g_T(\theta)}{\partial \theta^T} = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \phi(\boldsymbol{x}_t(\theta)}{\partial \theta^T}$$

Outline

The Bates Model

Under the risk-neutral measure $\mathbb{Q}$, the stock price S_t and the conditional variance ν_t are assumed to follow the process:

$$\frac{dS_t}{S_{t^-}} = (r - d - \lambda\bar{\nu})dt + \sqrt{v_t}dW_1 + dZ_t$$
$$dv_t = \kappa(\theta - v_t)dt + \sigma\sqrt{v}dW_2$$

- r: (constant) instantaneous risk-free rate
- d: dividend yield
- v_t: conditional variance which follow a Feller process under $\mathbb{Q}$, with the speed of mean reversion κ, long-run mean θ, and volatility parameter σ.
- $W_{1,t}$ and $W_{2,t}$: two standard Brownian motions under $\mathbb{Q}$, correlated with $E[dW_{1,t}dW_{2,t}] = \rho dt$,
- Z: a pure jump process with constant arrival intensity λ.

Price and BSIV

- Price of a European put option with maturity T and strike price K at time t:

$$P(S_t, t, v_t) \equiv E^{\mathbb{Q}}\left[e^{-r(T-t)}(K - S_T)^+\right]$$

- Map option prices to the Black-Scholes implied volatility (BSIV) to facilitate a comparison of the pricing errors for options with different moneyness and maturities.
- A normalized moneyness measure

$$m_t = \frac{\ln\left(\frac{K}{S_t F_{t,T}}\right)}{\sqrt{T - t}\sigma_{atm}}$$

where $F_{t,T}$ is the forward price of the stock for the same maturity as that of the option; σ_{atm} is a constant volatility measure, which we set to be the average BSIV of at-the money options from the entire sample.

Mapping to Deep Surrogate

- Three state variables $\boldsymbol{s}_t = [m_t, T - t, v_t]$, where v_t is latent
- Ten parameters $\theta = [r, d, \kappa, \theta, \sigma, \rho, \lambda, \nu_u, \nu_d, p]$.

Training Sample

Table A.1: This table presents the ranges for the surrogate model of the HM and Bates's training sample.

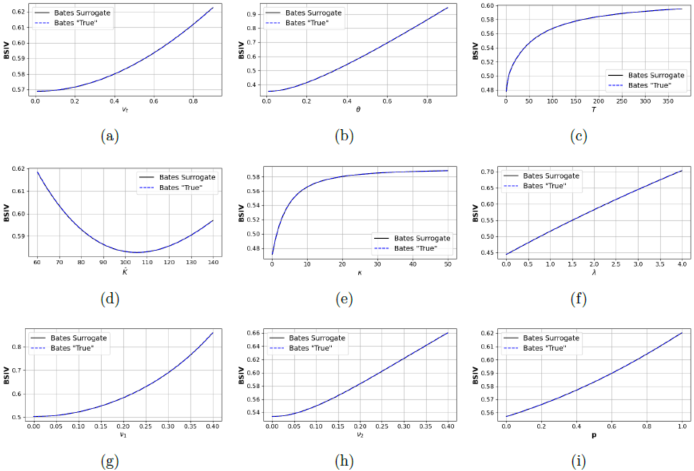
j	HM: $\underline{x}^{(j)}$	HM: $\bar{x}^{(j)}$	Bates: $\underline{x}^{(j)}$	Bates: $\bar{x}^{(j)}$
m	-9.00	5.000	-9.00	5.000
rf	0.00	0.075	0.00	0.075
dividend	0.00	0.050	0.00	0.050
v_t	0.01	0.900	0.01	0.900
T	1.00	365.000	1.00	365.000
κ	0.10	50.000	0.10	50.000
θ	0.01	0.900	0.01	0.900
σ	0.10	5.000	0.10	5.000
ρ	-	-	-1.00	-0.000
λ	-	-	0.00	4.000
ν_1	-	-	0.00	0.400
ν_2	-	-	0.00	0.400
p	-	-	0.00	1.000

Deep Surrogates for Option Pricing

- Training sample: increasing the training sample size and depth of the deep surrogate gradually until the validation criterion based on mean absolute pricing error is met.
- Network architecture design: 7 hidden layers, 400 neurons each (a total 967201 trainable parameters).
- a mini-batch stochastic gradient descent algorithm to determine the neural network's parameters

Accuracy of the Surrogates

How accurately the surrogate model captures the sensitivity of option value to various states and parameters.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Accuracy of the Surrogates

The ability of the surrogate to recover parameter values in structural estimations.

- Randomly draw volatility v_t and parameters θ
- Compute the option prices implied by the Bates model on 1000 options, with the moneyness and maturity drawn from a uniform distribution. Convert these option prices into BSIVs.
- Use deep surrogate function $\phi(\cdot)$ to run GMM estimation from the cross-section o f option data

$$\hat{v}_t, \hat{\theta} = \arg\min_{v,\theta} \left(\phi(\mathbf{x}(v,\theta)) - \mathbf{y}_t\right)^T \left(\phi(\mathbf{x}(v,\theta)) - \mathbf{y}_t\right).$$

Accuracy of the Surrogates

- The estimation errors for each element of θ:

$$e_i = \frac{|\theta_i - \hat{\theta}_i|}{\overline{\theta_i} - \underline{\theta_i}}$$

where $\overline{\theta_i}$ and $\underline{\theta_i}$ represent the maximum and minimum values in the training sample of the respective surrogate model.

- Same error measure applies to the estimate of the conditional variance, v_t.

Accuracy of the Surrogates

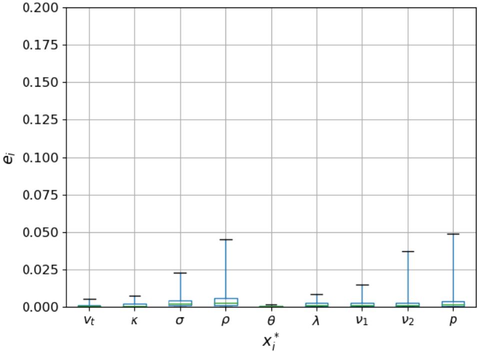


Figure 1: The figures above show that the Bates surrogate can be used to estimate the parameters of the models on a small data sample. We simulated market days with the original pricing models and used our surrogates to estimate the parameters and states simultaneously. Above, we show the resulting prediction errors (see Eq. (20)) for each hidden state and parameter of both models. The green vertical line represents the median prediction error computed across 1000 simulations. The box shows the interquartile range, while the whiskers show the 1st and 99th percentile errors across the simulations.