

Bridging Structured Knowledge and Data: A Unified Framework with Finance Applications

Yi Cao^{*} Zexun Chen[†] Lin William Cong[‡] Heqing Shi[†]

^{*}Xi'an Jiaotong-Liverpool University, Dept. Fin Maths

[†]University of Edinburgh Business School

[‡]Nanyang Technological University NBS & CCDS, ABFER, CEPR, and NBER

18 May, 2026

Motivation: Numbers Speak for Themselves


The success of data-driven modeling and a popular narrative (Anderson, 2008):

WIRED MAGAZINE: 16.07

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson 06.23.08



 Illustration: Marian Bantjes

— “This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. [...] With enough data, the numbers speak for themselves.”

Motivation: Structural vs. Data-Driven Models

STRUCTURAL MODELS

Principled but Rigid



Robust
to noisy observations.



Interpretable "white box"
by construction; popular
among econometricians.

BUT...



Rely on **idealized assumptions**
(e.g., market efficiency,
frictionless physics).



Possess **limited expressive capacity**
to capture complex,
real-world phenomena.

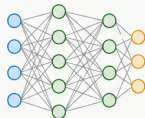


Prone to **parameter instability**
and model **misspecification**
in non-stationary data.

VS.

DATA-DRIVEN MODELS

Flexible but Fragile



No model-by-model
assumptions.
"Data speaks for itself."



Strong **expressive capacity**
(infinite-dimensional
functional space).

BUT...



Overfit to noise, spurious correlations
when in low signal-to-noise ratio
settings.

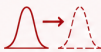


Exhibit **poor generalization**
when the domain shifts
significantly.



Function as an **uninterpretable**
"black box", hampering trust and
scientific insight.

Literature

Domain-specific theories/knowledge still useful? How to reconcile such a dichotomy?

- Raissi et al. (2017, 2019): PINNs which train/penalize NN with PDE residual.
↪ Most work solves PDE; little work fuses data- and PDE-knowledge.
- Aboussalah et al. (2024); Hainaut and Casas (2024); Wang et al. (2023): further empirical works on option pricing PINNs.
- Ackerer et al. (2020); Chataigner et al. (2020); Zhang et al. (2023); Zheng et al. (2021): model-free shape constraints (boundedness, monotonicity, convexity, etc. of the target function, e.g., static no-arbitrage). → Limited economic insights.
- Bai (2024); Chen et al. (2023, 2026): deep surrogate modeling; transfer learning.
↪ Catastrophic forgetting; heavy (unmanageable) pre-training.
- Almeida et al. (2023): correct structural models via learning residuals.
↪ Two-stage design; hard to learn from residuals.

Example: PINNs

PINNs as a trending AI4Sci method, still not practical or versatile enough, particularly for Finance and Econometrics.

- Mainly used for solving PDEs (with known PDE parameters), rather than out-of-sample reasoning.
- Wang et al. (2021, 2022): pathological gradient dynamics.

For example, the Heston's PDE:

$$\partial_T V + \frac{1}{2} v S^2 \partial_{SS} V + \rho \sigma v S \partial_{Sv} V + \frac{1}{2} \sigma^2 v \partial_{vv} V + r S \partial_S V + [\kappa (\theta - v)] \partial_v V - rV = 0$$

The differentiation feature v (instantaneous variance) is not observable!

- PDEs are not always there. Often, theories/models are prescribed by symbolic (parametric) equations, SDEs, and distributions.

Contribution: Conceptual and Methodological Innovations

- **Structured-Knowledge-Informed Neural Networks (SKINNs)**— a novel framework to better reconcile the dichotomy.
- SKINNs allow diverse formats of structured knowledge to represent “theories”, embedding, e.g., theoretical, simulated, pre-trained, or cross-domain insights.
- SKINNs jointly learn the structural parameters in these SK representations, at a rate of $\mathcal{O}(1/\sqrt{N})$, with the nuisance NN parameters.
- SKINNs are scalable to complex, (ultra-) high-dimensional SK presentations.
- Companion paper: "SKINN: Joint Data- and Structural-Parameter Learning with Structured Knowledge"

Contribution: Empirical Application in Finance

- Empirical study on panels of S&P 500 index options, spanning over two decades.
- SKINNs achieve SOTA performance:
 - Option pricing → (10%-15% better) OOS pricing & hedging performance
 - Asset pricing → (29%-42% better) OOS Sharpe ratio, implementable portfolios
 - Realized vol forecasting → (~20% better) OOS forecasting performance
- SKINNs-learned SK parameters carry significant economic information.

Model: The SKINNs Framework

Consider a data-driven function approximator (NN component):

$$\begin{aligned} f(\mathbf{X} \subseteq \mathbb{R}^d; \theta) &: \mathbb{R}^d \mapsto \mathbb{R}, \\ f^{(l)} &= h\left(\mathbf{b}^{(l-1)} + \mathbf{W}^{(l-1)} f^{(l-1)}\right), \forall l \in \mathbb{N}^+, 1 \leq l \leq L, \\ f^{(0)} &= \mathbf{X} \end{aligned}$$

One typically learns the NN parameters $\theta = \{\mathbf{b}^{(l)}, \mathbf{W}^{(l)}\}_{l=1}^L$ by minimizing a data loss:

$$\mathcal{L}_{\text{Data}}(\theta; \mathcal{D}_{\text{Obs}} \subseteq \{\mathbf{X}_{\text{Obs}}, \mathbf{y}_{\text{Obs}}\}) = \frac{1}{N_{\text{Obs}}} \|\mathbf{f}_{\theta}(\mathbf{X}_{\text{Obs}}) - \mathbf{y}_{\text{Obs}}\|_2^2$$

- **No** awareness of principled scientific structures (no structured knowledge)!
- Result in a local optimal with spurious patterns \rightarrow does not generalize well.

Model: The SKINNs Framework

SKINNs embed structured knowledge into the above learning process (SK component):

Definition (Structured Knowledge Representation)

Structured Knowledge is a representation function

$$g(\mathbf{X}^{SK} \subseteq \mathbb{R}^{d_{SK}}; \phi \subseteq \mathbb{R}^{d_\phi}) : \mathbb{R}^{d_{SK}} \mapsto \mathbb{R},$$

which is C^1 -smooth w.r.t. \mathbf{X}^{SK} , ϕ , or Lipschitz continuous at least, such that ϕ is Gradient Descent learnable. Moreover, $\mathbf{X}^{SK} \subseteq \mathbf{X}$, $d_{SK} \ll d$, $d_\phi \ll d_\theta$.

- Function space of $g_\phi(\mathbf{X}^{SK})$ given by theories \rightarrow known SK representation
- Structural parameters $\phi \in \Phi$ unknown \rightarrow undetermined SK parameters

Model: The SKINNs Framework

Force the data-driven function approximator $f_\theta(\mathbf{X})$ to be SK aware (via a SK loss):

$$\mathcal{L}_{\text{SK}}(\theta, \phi; \mathbf{X}_{\text{Colloc}}) \equiv \frac{1}{N_{\text{Colloc}}} \|f_\theta(\mathbf{X}_{\text{Colloc}}) - \mathbf{g}_\phi(\mathbf{X}_{\text{Colloc}}^{\text{SK}})\|_2^2$$

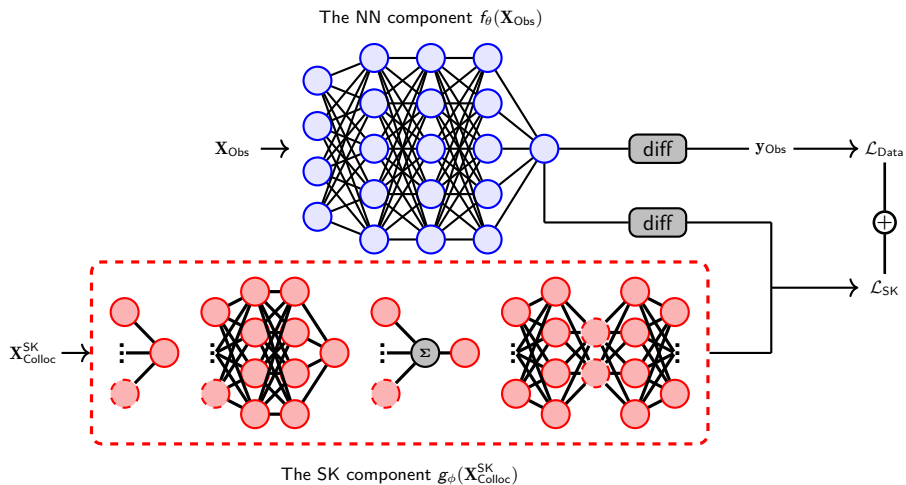
SKINNs inform the gradient dynamics of training $f_\theta(\mathbf{X})$ through a composite loss:

$$\mathcal{L}_{\text{Total}}(\theta, \phi; \mathbf{X}_{\text{Obs}} \cup \mathbf{X}_{\text{Colloc}}^{\text{SK}}, \mathbf{y}_{\text{Obs}}) = \lambda_{\text{Data}} \underbrace{\mathcal{L}_{\text{Data}}(\theta; \mathbf{X}_{\text{Obs}}, \mathbf{y}_{\text{Obs}})}_{\text{Data fidelity}} + \lambda_{\text{SK}} \underbrace{\mathcal{L}_{\text{SK}}(\theta, \phi; \mathbf{X}_{\text{Colloc}}^{\text{SK}})}_{\text{Theory consistency}}$$

Find the optimal θ, ϕ simultaneously by solving the optimization with (S)GD:

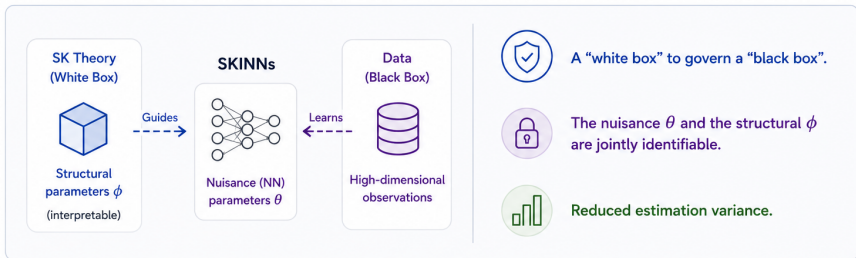
$$(\theta^*, \phi^*) =: \arg \min_{\theta \in \Theta, \phi \in \Phi} \mathcal{L}_{\text{Total}}(\theta, \phi; \mathbf{X}_{\text{Obs}} \cup \mathbf{X}_{\text{Colloc}}^{\text{SK}}, \mathbf{y}_{\text{Obs}})$$

Model: The SKINNs Framework



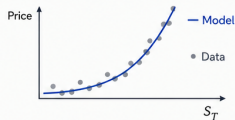
Model: The SKINNs Framework

SKINNs work as a generalizability-improved **predictor** and an econometric **estimator** for SK parameters.

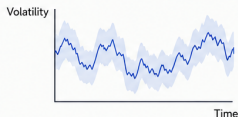


Our SK representation allows various forms of "theories"

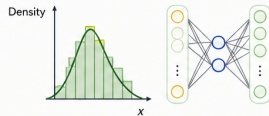
- 1 Parametric SK rep
e.g., Black-Scholes-Merton



- 2 Semi-parametric SK rep
e.g., non-affine stochastic volatility



- 3 Non-parametric SK rep
e.g., probability distribution, (variational) autoencoder



SK Representations: Parametric

Main Idea

The structural map is specified by a known theory:

$$g_{\phi}(\mathbf{X}^{\text{SK}})$$

where the functional form is fixed and only a low-dimensional parameter vector ϕ is learned.

- Strongest theoretical discipline and highest interpretability.
- Limited flexibility when the market departs from the assumed structure.
- In the benchmark tables, this corresponds to:

+**BSM**(Black-Scholes-Merton Stochastic Volatility) +**ABSM**(Ad-hoc BSM) +**HSV**(Heston's Stochastic Volatility) +**HSVJ**(Heston's Stochastic Volatility with Jump)

SK Representations: Semi-parametric

Main Idea

The theory defines the economic mechanism, but the pricing map is not available in closed form:

$$g_{\phi}(\mathbf{X}^{\text{SK}}) \approx f_{\theta}^{\text{SR}}(\mathbf{X}^{\text{SK}}, \phi)$$

where a deep surrogate learns the solution operator.

- Keeps low-dimensional structural parameters.
- Allows richer dynamics, such as non-affine stochastic volatility.
- In the benchmark tables, this corresponds to:

+**DSNN-HSV**(Deep Surrogate NN- HSV)

+**DSNN-NASV**(Deep Surrogate NN- Non-Affine Stochastic Volatility)

SK Representations: Non-parametric

Idea

The structural object itself is learned, such as a risk-neutral distribution or a latent representation:

$$g_{\phi}(\mathbf{X}^{\text{SK}}) = D(t, T) \times \mathbf{V} \times \text{flatten}\{\mathbb{Q}(S_T)\} \quad \text{or} \quad f_{\theta}^{\text{Dec}}(\mathbf{X}^{\text{SK}}, \phi) \quad \text{or} \quad f_{\theta}^{\text{AE}}(\mathbf{X}^{\text{SK}}, \phi).$$

- Most flexible representation of structural knowledge.
- Can learn high-dimensional probabilities or latent factors directly from data.
- In the benchmark tables, this corresponds to:

+**MOPA**(Martingale Option Pricing Approach)

+**AE-BSM**(Autoencoder BSM)

Financial Application: Dataset

We use comprehensive S&P 500 index option data:

- Daily S&P 500 index call options (European style).
- Covering the period from January 4, 1996, to December 31, 2022.
- Options with maturities between 7 and 365 calendar days.

Model training schedule:

- Forward rolling-window approach.
- Train on three-month panels.
- Test over the next two consecutive months.

Empirical Study: Benchmark Models

Model	Structured Knowledge	$\dim(\phi)$ (learnable/total)	Activation function
Benchmark NNs			
NN	No	0/0	ReLU
NN+Bnd	Bnd	0/0	ReLU
NN+Shape	Shape	0/0	SiLU
PINN+BSM	BSM	1/1	SiLU
TLNN+HSV	HSV	0/5	ReLU
(PSKR) SKINNs			
SKINN+BSM	BSM	1/1	ReLU
SKINN+HSV	HSV	5/5	ReLU
SKINN+ABSM	ABSM	6/6	ReLU
SKINN+HSVJ	HSVJ	9/9	ReLU
(SPSKR) SKINNs			
SKINN+DSNN-HSV	DSNN-HSV	5/5	ReLU
SKINN+DSNN-NASV	DSNN-NASV	6/6	ReLU
(NPSKR) SKINNs			
SKINN+AE-BSM	AE-BSM	2/2	ReLU
SKINN+MOPA	MOPA	2,000/2,000	ReLU

Empirical Study: OOS Pricing Error (Diebold-Mariano Test)

Longer prediction horizons (RMSE)								
Model	Structural models				Benchmark NNs			
	BSM	ABSM	HSV	HSVJ	NN	NN +Bnd	NN +Shape	PINN +BSM
BSM	-	2.49***	0.62	-5.23***	2.67***	1.36*	2.51***	2.54***
ABSM	-	-	-0.57	-5.83***	2.56***	1.26	2.40***	2.42***
HSV	-	-	-	-12.95***	2.63***	1.33*	2.45***	2.47***
HSVJ	-	-	-	-	3.09***	1.74**	2.94***	2.97***
NN	-	-	-	-	-	-2.51***	-0.44	-0.38
NN+Bnd	-	-	-	-	-	-	1.30*	1.14
NN+Shape	-	-	-	-	-	-	-	0.01
PINN+BSM	-	-	-	-	-	-	-	-

Model	PSKR				SPSKR		NPSKR	
	SKINN +BSM	SKINN +ABSM	SKINN +HSV	SKINN +HSVJ	SKINN +DSNN-HSV	SKINN +DSNN-NASV	SKINN +MOPA	SKINN +AE-BSM
BSM	-3.61***	-3.61***	-3.06***	-3.56***	-2.09**	-2.78***	-2.53***	0.04
ABSM	-3.85***	-3.87***	-3.30***	-3.80***	-2.31**	-3.01***	-2.74***	-0.10
HSV	-3.89***	-3.91***	-3.37***	-3.89***	-2.33**	-3.08***	-2.74***	-0.03
HSVJ	-2.69***	-2.63***	-2.23**	-2.69***	-1.26	-1.88**	-1.76**	0.56
NN	-3.98***	-3.95***	-3.83***	-4.02***	-3.51***	-3.70***	-3.83***	-3.61***
NN+Bnd	-2.72***	-2.69***	-2.63***	-2.77***	-2.28**	-2.44***	-2.61***	-2.54***
NN+Shape	-3.55***	-3.57***	-3.44***	-3.66***	-3.06***	-3.27***	-3.43***	-3.05***
PINN+BSM	-3.50***	-3.53***	-3.42***	-3.62***	-3.02***	-3.23***	-3.41***	-2.82***

Empirical Study: OOS Hedging Error (Diebold-Mariano Test)

Longer prediction horizons (MHE)								
Model	Structural models				Benchmark NNs			
	BSM	ABSM	HSV	HSVJ	NN	NN +Bnd	NN +Shape	PINN +BSM
BSM	-	-4.30***	7.69***	7.67***	3.44***	3.48***	3.52***	3.20***
ABSM	-	-	7.83***	7.81***	3.68***	3.83***	3.72***	3.43***
HSV	-	-	-	-1.42*	0.25	-1.15	1.11	0.20
HSVJ	-	-	-	-	0.39	-0.99	1.21	0.32
NN	-	-	-	-	-	-2.12**	1.40*	-0.05
NN+Bnd	-	-	-	-	-	-	2.76***	1.47*
NN+Shape	-	-	-	-	-	-	-	-1.94**
PINN+BSM	-	-	-	-	-	-	-	-

Model	PSKR				SPSKR		NPSKR	
	SKINN +BSM	SKINN +ABSM	SKINN +HSV	SKINN +HSVJ	SKINN +DSNN-HSV	SKINN +DSNN-NASV	SKINN +MOPA	SKINN +AE-BSM
BSM	-5.07***	-6.33***	0.84	0.35	0.72	1.45*	0.26	0.68
ABSM	-3.66***	-4.96***	1.78**	1.55*	1.45*	2.01**	1.25	1.35*
HSV	-8.32***	-8.12***	-8.42***	-8.11***	-6.37***	-5.63***	-6.69***	-5.64***
HSVJ	-8.26***	-8.03***	-8.34***	-8.07***	-6.18***	-5.58***	-6.47***	-5.57***
NN	-4.71***	-4.59***	-3.44***	-3.63***	-3.41***	-2.89***	-3.58***	-3.83***
NN+Bnd	-5.07***	-5.12***	-3.40***	-3.71***	-3.14***	-2.56***	-3.50***	-4.07***
NN+Shape	-4.41***	-4.46***	-3.49***	-3.67***	-3.31***	-3.03***	-3.60***	-3.88***
PINN+BSM	-4.32***	-4.29***	-3.01***	-3.23***	-2.94***	-2.45***	-3.29***	-3.34***

Empirical Study: SKINNs Outperformance & Market Vol

How is the model's outperformance relative to NN associated with market volatility?

- Bad model: performs worse when market vol is increasing (pro-cyclical).
- Good model: performs better when market vol is increasing (counter-cyclical).

Check this with: $\Delta\text{RMSE}^{\text{Model}} = \beta_0 + \beta_1 \text{AvgVIX} + \varepsilon$

- $\Delta\text{RMSE}^{\text{Model}}(t_j) := \text{RMSE}^{\text{Model}}(t_j) - \text{RMSE}^{\text{NN}}(t_j)$, $\text{AvgVIX}(t_j) := \frac{1}{N_{t_j}} \sum_{i=1}^{N_{t_j}} \frac{\text{VIX}_i}{100}$
- t_j denotes the OOS evaluation period j , $1 \leq j \leq 317$

Empirical Study: SKINNs Outperformance & Market Vol

Longer prediction horizons (64 low-vol periods)								
Model Panel (A)	BSM	ABSM	HSV	HSVJ	NN +Bnd	-	-	-
β_1	-0.1979 (0.1657)	-0.1172 (0.1644)	-0.1354 (0.1646)	-0.1288 (0.1671)	-0.0346 (0.1192)	-	-	-
β_0	0.0230 (0.0205)	0.0125 (0.0204)	0.0167 (0.0204)	0.0119 (0.0207)	0.0024 (0.0148)	-	-	-
Obs	64	64	64	64	64	-	-	-
Adj. R ²	0.0067	-0.0079	-0.0052	-0.0065	-0.0148	-	-	-
Model Panel (B)	SKINN +BSM	SKINN +ABSM	SKINN +HSV	SKINN +HSVJ	SKINN +DSNN-HSV	SKINN +DSNN-NASV	SKINN +MOPA	SKINN +AE-BSM
β_1	-0.2303 (0.1504)	-0.1690 (0.1528)	-0.0628 (0.1614)	0.0021 (0.1836)	-0.6827*** (0.1828)	-0.1875 (0.1524)	-0.1741 (0.1270)	-0.1343 (0.1275)
β_0	0.0252 (0.0186)	0.0162 (0.0189)	0.0037 (0.0200)	-0.0035 (0.0227)	0.0912*** (0.0226)	0.0221 (0.0189)	0.0175 (0.0157)	0.0167 (0.0158)
Obs	64	64	64	64	64	64	64	64
Adj. R ²	0.0209	0.0035	-0.0137	-0.0161	0.1705	0.0081	0.0138	0.0018

Empirical Study: SKINNs Outperformance & Market Vol

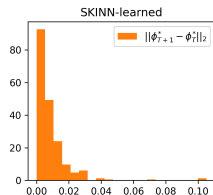
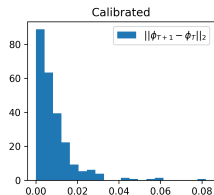
Longer prediction horizons (64 high-vol periods)								
Model Panel (A)	BSM	ABSM	HSV	HSVJ	NN +Bnd	-	-	-
β_1	-0.0364 (0.0783)	-0.0416 (0.0787)	-0.0345 (0.0788)	-0.0272 (0.0791)	-0.0466 (0.0287)	-	-	-
β_0	0.0051 (0.0262)	0.0077 (0.0264)	0.0025 (0.0264)	-0.0013 (0.0265)	0.0094 (0.0096)	-	-	-
Obs	64	64	64	64	64	-	-	-
Adj. R ²	-0.0126	-0.0116	-0.0130	-0.0142	0.0255	-	-	-
Model Panel (B)	SKINN +BSM	SKINN +ABSM	SKINN +HSV	SKINN +HSVJ	SKINN +DSNN-HSV	SKINN +DSNN-NASV	SKINN +MOPA	SKINN +AE-BSM
β_1	-0.1464* (0.0780)	-0.1372* (0.0761)	-0.1516** (0.0739)	-0.1474** (0.0724)	-0.1555** (0.0720)	-0.1373* (0.0767)	-0.1526** (0.0660)	-0.1825*** (0.0453)
β_0	0.0282 (0.0261)	0.0270 (0.0255)	0.0308 (0.0248)	0.0295 (0.0243)	0.0307 (0.0241)	0.0263 (0.0257)	0.0315 (0.0221)	0.0478*** (0.0152)
Obs	64	64	64	64	64	64	64	64
Adj. R ²	0.0385	0.0344	0.0484	0.0475	0.0549	0.0337	0.0645	0.1946

Empirical Study: The Estimator Side of SKINNs

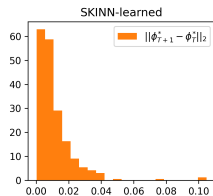
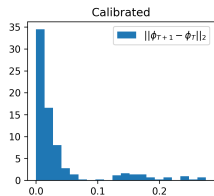
Beyond functioning as predictive devices, SKINNs also work as econometric estimators:

- Nest functional Generalized Method of Moments (GMM).
- Close to semi-parametric models, but more flexible.
- (Ultra-) high-dimensional SK parameters are identifiable, with reduced variance.
- SKINNs-learned SK parameters carry significant economic information.

Empirical Study: The Estimator Side of SKINNs

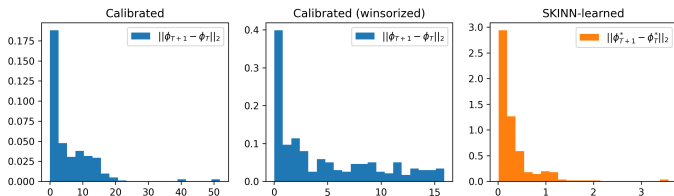


(a) SKINN+BSM

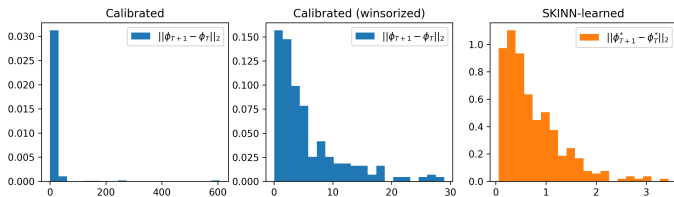


(b) SKINN+ABSM

Empirical Study: The Estimator Side of SKINNs



(a) SKINN+HSV



(b) SKINN+HSVJ

Empirical Study: SKINNs Asset Pricing

- An asset manager optimizes his/her portfolio by solving a Mean-Variance Optimization (MVO):

$$\begin{aligned} \mathbf{w}^* &:= \arg \min_{\mathbf{w}} - \mathbf{w}^\top \hat{\mathbf{r}} + \eta \mathbf{w}^\top \Sigma \mathbf{w} \\ \text{s.t. } & \mathbf{w}^\top \mathbf{1} = 1 \\ & \mathbf{l} \leq \mathbf{w} \leq \mathbf{u} \end{aligned}$$

The MVO is typically solved given a learned asset-pricing model $\hat{\mathbf{r}} = f_\theta(\mathbf{X})$
 \hookrightarrow Two-stage optimization; no interaction with each other

Turn such a two-stage optimization into an SK component:

$$\begin{aligned} g^{\text{MVO}}(f_\theta, \phi \in \{\mathbf{w}\}) &= \mathbf{w}^\top f_\theta + \eta \mathbf{w}^\top \Sigma \mathbf{w} \\ \text{s.t. } & \left. \begin{aligned} \mathbf{w}^\top \mathbf{1} &= 1 \\ \mathbf{l} &\leq \mathbf{w} \leq \mathbf{u} \end{aligned} \right\} \text{Pass to clamping functions} \end{aligned}$$

Similar concept used in Jensen et al. (2026); Wang et al. (2026)

Empirical Study: SKINNs Asset Pricing

- We trade 50 stocks from S&P 500.
- Forward rolling-window test: train one-year daily OHLC, trade daily over the next month. Stock data spans 2010-2023.
- Set $\mathbf{l} = 0.0$, $\mathbf{u} = 0.2$. Decile-sorting on NN/SKINN predicted returns.

Model	Group	NN			SKINN ($\eta = 0$)			SKINN ($\eta = 10$)		
		Avg %	SD %	SR	Avg %	SD %	SR	Avg %	SD %	SR
1-32	H	60.73	130.07	0.47	68.22↑	109.78↓	0.62↑	70.83↑	102.91↓	0.69↑
	L	59.29	136.54	0.43	54.78	133.41	0.41	57.77	134.48	0.43
2-32	H	57.01	127.73	0.45	71.86↑	119.78↓	0.60↑	65.32↑	99.71↓	0.66↑
	L	72.12	145.66	0.50	70.98	141.31	0.50	57.19	140.62	0.41
3-32	H	56.57	127.98	0.44	68.94↑	122.60↓	0.56↑	59.45↑	100.10↓	0.59↑
	L	76.98	132.41	0.58	71.67	132.81	0.54	60.34	134.04	0.45
4-32	H	57.71	124.46	0.46	67.11↑	118.09↓	0.57↑	61.96↑	97.56↓	0.64↑
	L	65.51	136.01	0.48	68.98	133.11	0.52	54.35	134.98	0.40
5-32	H	56.54	125.43	0.45	70.35↑	122.43↓	0.57↑	65.48↑	100.00↓	0.65↑
	L	77.84	134.57	0.58	73.08	132.93	0.55	70.90	135.16	0.52

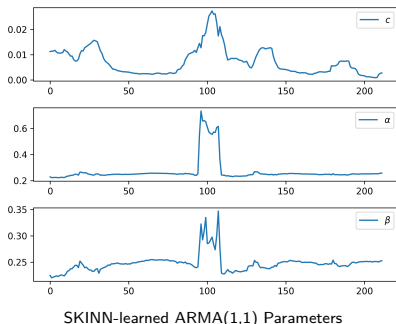
Empirical Study: Realized Vol Forecasting

- Market (S&P 500 index) 5-min realized vol from January 3, 2000 to 26 June, 2018
- Train on one-year realized vols, test over the next month (212 training & test subsamples in total)
- Features: lagged FF5 factors (daily)

Average OOS RMSE		
Model	NN	SKINN+ARMA(1,1)
	0.0150	0.0121

Diebold-Mariano Test		
Model	NN	SKINN+ARMA(1,1)
NN	-	-2.13**
SKINN+ARMA(1,1)	-	-

Δ RMSE and Market Vol		
	NN	SKINN+ARMA(1,1)
β_1	-	-1.8469**
β_0	-	-0.0001
Obs	-	212
Adj. R^2	-	0.0186



Conclusions

- All models/theories are wrong? Feed them to DNN!
- SKINNs provide a novel unified framework to embed theoretical structures into machines (much more versatile than PINNs, Transfer Learning, etc.).
- Data-driven is not almighty, still need theories for interpretability & generalizability.



Check out our preprint in arXiv and SSRN

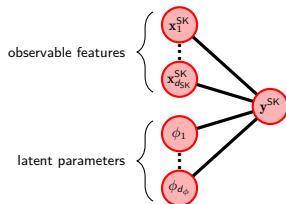
References I

- Aboussalah, A.M., Li, X., Chi, C., Patel, R., 2024. The AI Black-Scholes: Finance-informed neural network. arXiv preprint arXiv:2412.12213 .
- Ackerer, D., Tagasovska, N., Vatter, T., 2020. Deep smoothing of the implied volatility surface. *Advances in Neural Information Processing Systems* 33, 11552–11563.
- Almeida, C., Fan, J., Freire, G., Tang, F., 2023. Can a machine correct option pricing models? *Journal of Business & Economic Statistics* 41, 995–1009.
- Anderson, C., 2008. The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine* 16, 16–07.
- Bai, C., 2024. Teach machine asset pricing. Available at SSRN 4598745 .
- Chataigner, M., Crépey, S., Dixon, M., 2020. Deep local volatility. *Risks* 8, 82.
- Chen, H., Cheng, Y., Liu, Y., Tang, K., 2023. Teaching economics to the machines. Available at SSRN 4642167.
- Chen, H., Didisheim, A., Scheidegger, S., 2026. Deep surrogates for finance: With an application to option pricing. *Journal of Financial Economics* 177, 104222.
- Hainaut, D., Casas, A., 2024. Option pricing in the heston model with physics inspired neural networks. *Annals of Finance* 20, 353–376.
- Jensen, T.I., Kelly, B., Malamud, S., Pedersen, L.H., 2026. Machine learning and the implementable efficient frontier. *The Review of Financial Studies* , hhag022.
- Raissi, M., Perdikaris, P., Karniadakis, G.E., 2017. Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations. doi:10.48550/arXiv.1711.10566, arXiv:1711.10566.

References II

- Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics* 378, 686–707.
- Wang, S., Teng, Y., Perdikaris, P., 2021. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing* 43, A3055–A3081.
- Wang, S., Yu, X., Perdikaris, P., 2022. When and why pinns fail to train: A neural tangent kernel perspective. *Journal of Computational Physics* 449, 110768.
- Wang, X., Li, J., Li, J., 2023. A deep learning based numerical pde method for option pricing. *Computational economics* 62, 149–164.
- Wang, Y., Gao, H., Harvey, C.R., Liu, Y., Tao, X., 2026. Machine learning meets Markowitz. Technical Report. National Bureau of Economic Research.
- Zhang, W., Li, L., Zhang, G., 2023. A two-step framework for arbitrage-free prediction of the implied volatility surface. *Quantitative Finance* 23, 21–34.
- Zheng, Y., Yang, Y., Chen, B., 2021. Incorporating prior financial domain knowledge into neural networks for implied volatility surface prediction, in: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3968–3975.

SK Representations: Parametric



- The spec of $g_\phi(\mathbf{X}^{SK})$ is fully determined by theories (closed/semi-closed form).
- Only the latent structural parameters ϕ are unknown but learnable.

SK Representations: Parametric

Black-Scholes-Merton case (one-dimensional):

$$g^{\text{BSM}}(\mathbf{X}^{\text{SK}}; \phi \in \{\sigma\}) = S_t \Phi(d_1(\sigma)) - Ke^{-r(T-t)} \Phi(d_2(\sigma)),$$
$$d_1(\phi \in \{\sigma\}) = \frac{1}{\sigma \sqrt{T-t}} \left(\log(S_t/K) + \left(r + \frac{\sigma^2}{2}\right)(T-t) \right),$$
$$d_2(\phi \in \{\sigma\}) = d_1(\sigma) - \sigma \sqrt{(T-t)}$$

Ad-hoc BSM case (six-dimensional):

$$g^{\text{ABSM}}(\mathbf{X}^{\text{SK}}; \phi \in \{a_0, a_1, a_2, a_3, a_4, a_5\}) = S_t \Phi(d_1(a_0, \dots, a_5)) - Ke^{-rT} \Phi(d_2(a_0, \dots, a_5)),$$
$$d_1(\phi \in \{a_0, \dots, a_5\}) = \frac{1}{(a_0 + a_1 m + a_2 m^2 + a_3 T + a_4 T^2 + a_5 mT) \sqrt{T}}$$
$$\times \left(\log(S_t/K) + \left(r + \frac{(a_0 + a_1 m + a_2 m^2 + a_3 T + a_4 T^2 + a_5 mT)^2}{2}\right) T \right),$$
$$d_2(\phi \in \{a_0, \dots, a_5\}) = d_1(a_0, \dots, a_5) - (a_0 + a_1 m + a_2 m^2 + a_3 T + a_4 T^2 + a_5 mT) \sqrt{T}$$

SK Representations: Parametric

Heston's Stochastic Volatility (five-dimensional, semi-closed form):

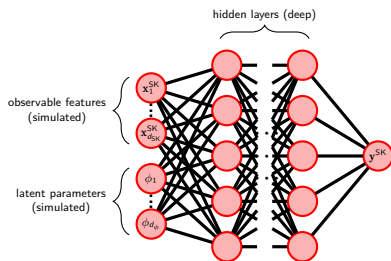
$$g^{\text{HSV}}(\mathbf{X}^{\text{SK}}; \phi \in \{\mathbf{v}_\theta, \mathbf{v}_0, \sigma_v, \rho, \kappa\}) = \frac{1}{2} e^{-r(T-t)} \Re \left\{ \psi_{T|t}(0; S, r, \mathbf{v}_\theta, \mathbf{v}_0, \sigma_v, \rho, \kappa) \right\} V_0(K) \\ + e^{-r(T-t)} \sum_{w=1}^{N-1} \Re \left\{ \psi_{T|t} \left(\frac{w\pi}{b-a}; S, r, \mathbf{v}_\theta, \mathbf{v}_0, \sigma_v, \rho, \kappa \right) e^{-iw\pi \frac{a}{b-a}} \right\} V_w(K)$$

Heston's Stochastic Volatility with Jump (nine-dimensional, semi-closed form):

$$g^{\text{HSVJ}}(\mathbf{X}^{\text{SK}}; \phi \in \{\mathbf{v}_\theta, \mathbf{v}_0, \sigma_v, \rho, \kappa, \rho, \eta_1, \eta_2, \lambda\}) = \frac{1}{2} e^{-r(T-t)} \\ \times \Re \left\{ \psi_{T|t}(0; S, r, \mathbf{v}_\theta, \mathbf{v}_0, \sigma_v, \rho, \kappa) \times \psi_{T|t}^J(0; \rho, \eta_1, \eta_2, \lambda) \right\} V_0(K) \\ + e^{-r(T-t)} \sum_{w=1}^{N-1} \Re \left\{ \psi_{T|t} \left(\frac{w\pi}{b-a}; S, r, \mathbf{v}_\theta, \mathbf{v}_0, \sigma_v, \rho, \kappa \right) \psi_{T|t}^J \left(\frac{w\pi}{b-a}; \rho, \eta_1, \eta_2, \lambda \right) e^{-iw\pi \frac{a}{b-a}} \right\} V_w(K),$$

$$\text{where, } \psi_{T|t}^J(u; \phi \in \{\rho, \eta_1, \eta_2, \lambda\}) = \exp \left[\lambda (T-t) \left(\frac{\rho \eta_1}{\eta_1 - iu} + \frac{(1-\rho) \eta_2}{\eta_2 + iu} - 1 \right) \right]$$

SK Representations: Semi-parametric



- The SK offers no (semi-) closed-form solutions, but admits intractable systems.
- E.g., SDEs (non-affine jump-diffusions). Only solvable with numerical methods.
- Unknown function space of $g_{\phi}(\mathbf{X}^{SK})$; known input and parameter space (low dim).
- Chen et al. (2026) deep surrogate: parameterize $g_{\phi}(\mathbf{X}^{SK})$ using a DNN.

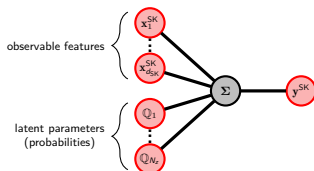
SK Representations: Semi-parametric

Non-affine Stochastic Volatility (six-dimensional):

$$\begin{aligned}dS_t &= rS_t dt + \sqrt{v_t} S_t dW_t^1, \\dv_t &= \kappa (v_\theta - v_t) dt + \sigma_v v_t^{\gamma_v} dW_t^2, \\ \langle dW^1, dW^2 \rangle_t &= \rho dt, \text{ initial instantaneous variance} = v_0\end{aligned}$$

- Draw $\{\mathbf{x}_1^{\text{SK}}, \mathbf{x}_2^{\text{SK}}, \dots, \mathbf{x}_{d_{\text{SK}}}^{\text{SK}}, v_\theta, v_0, \sigma_v, \rho, \kappa, \gamma_v\}$ large number of times (50 million) from a multivariate uniform distribution. All solved with a preferred numerical method, such as MC.
- Pre-train a deep surrogate $f_\theta^{\text{SR}}(\mathbf{X}^{\text{SK}}, \phi = \{v_\theta, v_0, \sigma_v, \rho, \kappa, \gamma_v\})$. The SK rep is now a DSNN $g^{\text{NASV}}(\mathbf{X}^{\text{SK}}, \theta^{\text{SR}}; \phi \in \{v_\theta, v_0, \sigma_v, \rho, \kappa, \gamma_v\}) = f_\theta^{\text{SR}}(\mathbf{X}^{\text{SK}}, \phi)$.
- Plug in the DSNN $g^{\text{NASV}}(\mathbf{X}^{\text{SK}}, \phi)$ to SKINN. Switch $\phi \in \{v_\theta, v_0, \sigma_v, \rho, \kappa, \gamma_v\}$ in the DSNN input layer as **learnable** latent structural parameters.

SK Representations: Non-parametric



- SK is often derived from distribution-based principles, e.g., the law of no-arbitrage:

$$\begin{aligned} g_{\phi}(\mathbf{X}^{\text{SK}}) &= e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}[\underbrace{u(z, \mathbf{X}^{\text{SK}})}_{\text{Payoff function}} | \mathcal{F}_t] \\ &= e^{-r(T-t)} \sum_{i=1}^{\infty} u(z^{[i]}, \mathbf{X}^{[i], \text{SK}}) \mathbb{Q}(z^{[i]}) \end{aligned}$$

- SKINNs embed such SK by directly learning the probabilities $\mathbb{Q}(z^{[i]})$, $1 \leq i \leq N_z$.
- The learnable probabilities (parameters) can be ultra-high-dimensional.

SK Representations: Non-parametric (MOPA)

Martingale Option Pricing Approach (ultra-high-dimensional):

Consider a panel of European call options, with maturities $T_1 \leq T_2 \leq \dots \leq T_s$, and strikes $K_1 \leq K_2 \leq \dots \leq K_{N_h}$ for each maturity $T_h, 1 \leq h \leq s$. The payoff $u(z, \mathbf{X}^{\text{SK}})$ can be written analytically, for T_h ,

$$\mathbf{V}_{T_h} = \begin{bmatrix} (S_{T_h}^{[1]} - K_1)^+ & (S_{T_h}^{[2]} - K_1)^+ & \dots & (S_{T_h}^{[q]} - K_1)^+ \\ (S_{T_h}^{[1]} - K_2)^+ & (S_{T_h}^{[2]} - K_2)^+ & \dots & (S_{T_h}^{[q]} - K_2)^+ \\ \vdots & \vdots & \ddots & \vdots \\ (S_{T_h}^{[1]} - K_{N_h})^+ & (S_{T_h}^{[2]} - K_{N_h})^+ & \dots & (S_{T_h}^{[q]} - K_{N_h})^+ \end{bmatrix}$$

Assuming q states of nature $S^{[1]}, \dots, S^{[q]}$.

The payoff for the panel (block-diagonal),

$$\mathbf{V}_{N \times sq} = \begin{bmatrix} \mathbf{V}_{T_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{T_2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{V}_{T_s} \end{bmatrix}$$

SK Representations: Non-parametric (MOPA)

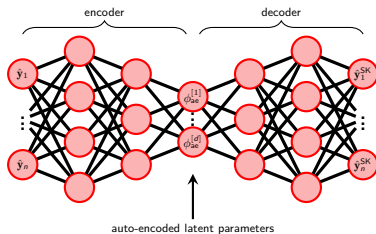
Pack up a learnable probability matrix,

$$\mathbb{Q}(\mathcal{S}_T)_{s \times q} = \begin{bmatrix} \mathbb{Q}(\mathcal{S}_{T_1}^{[1]}) & \mathbb{Q}(\mathcal{S}_{T_1}^{[2]}) & \cdots & \mathbb{Q}(\mathcal{S}_{T_1}^{[q]}) \\ \mathbb{Q}(\mathcal{S}_{T_2}^{[1]}) & \mathbb{Q}(\mathcal{S}_{T_2}^{[2]}) & \cdots & \mathbb{Q}(\mathcal{S}_{T_2}^{[q]}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{Q}(\mathcal{S}_{T_s}^{[1]}) & \mathbb{Q}(\mathcal{S}_{T_s}^{[2]}) & \cdots & \mathbb{Q}(\mathcal{S}_{T_s}^{[q]}) \end{bmatrix},$$

The MOPA SK representation (non-parametric, unknown probabilistic distribution):

$$g^{\text{MOPA}}(\mathbf{X}^{\text{SK}}; \phi \in \{\mathbb{Q}(\mathcal{S}_{T_h}^{[j]})\}_{1 \leq h \leq s, 1 \leq j \leq q}) = D(t, T) \times \mathbf{V} \times \text{flatten}(\mathbb{Q}(\mathcal{S}_T))$$

SK Representations: Non-parametric



- The SK representation can be machine-learned (a very general way).
- Unknown function space, and unknown structural parameter space.
- But have observations \hat{y} which should contain theoretical structures.

SK Representations: Non-parametric

(V)AE-derived Representation:

Given cross-sections of option prices $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$, we ask (V)AE to directly learn the latent structural parameter space Φ^{AE} , $\dim(\Phi^{\text{AE}}) \ll n$.

Observations $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$ can be constructed according to Mixture of Models, Expert Domain Knowledge, Historical Data, ...

The decoder of the pre-trained (V)AE serves as the differentiable SK rep:

$$g^{\text{AE}}(\mathbf{X}^{\text{SK}}; \phi \in \{\phi_{\text{AE}}^{[1]}, \phi_{\text{AE}}^{[2]}, \dots, \phi_{\text{AE}}^{[\phi_d]}\}) = f_{\theta}^{\text{Dec}}(\mathbf{X}^{\text{SK}}, \phi_{\text{AE}}^{[1]}, \phi_{\text{AE}}^{[2]}, \dots, \phi_{\text{AE}}^{[\phi_d]})$$