Incorporating Prior Financial Domain Knowledge into Neural Networks for Implied Volatility Surface Prediction

Yu Zheng

1. School of Finance
Southwestern University of Finance and Economics
2. Inboc Technologies Ltd.
zhengyu@swufe.edu.cn

Yongxin Yang

1. School of Informatics
University of Edinburgh
2. ArrayStream Technologies Ltd.
yongxin.yang@ed.ac.uk

Bowei Chen*

Adam Smith Business School
University of Glasgow
bowei.chen@glasgow.ac.uk

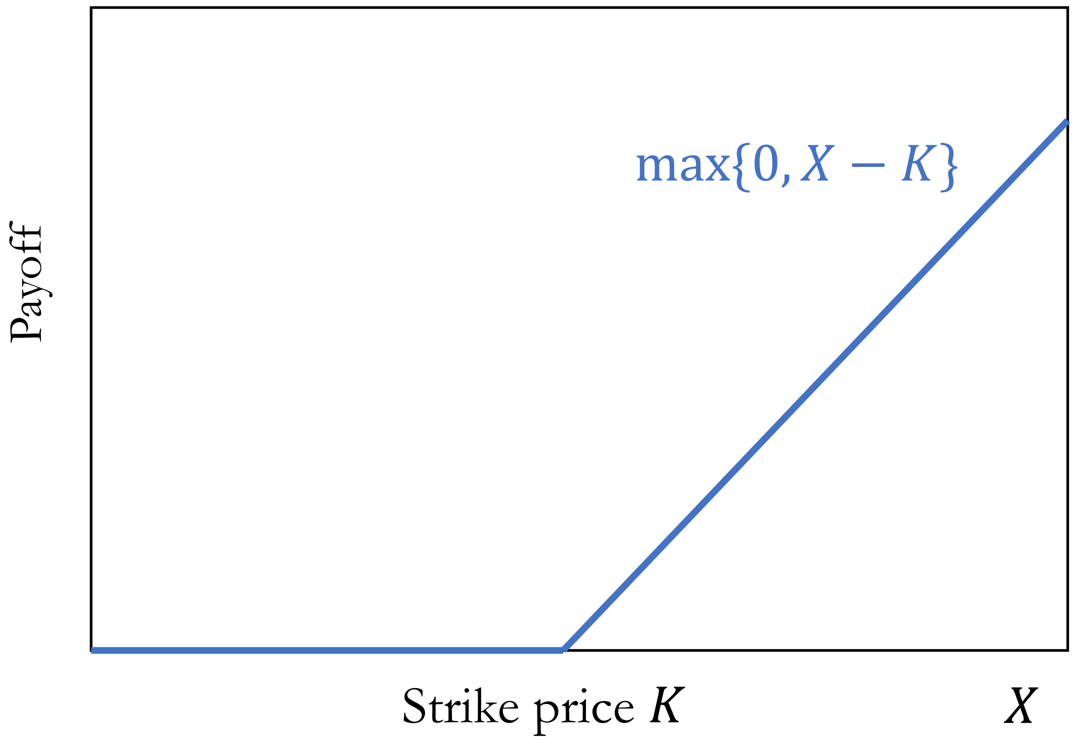
KDD 2021

* Corresponding author. | DOI: 10.1145/3447548.3467115

Option on S&P 500 index

S&P 500 index

Call option payoff



Stochastic underlying models

- Brownian motion (or Wiener process, or continuous-time random walk)
- Geometric Brownian motion
- Jump diffusion process
- Levy model

- Bachelier, L. (1900) Theorie de la speculation. *Annales Scientifiques de l'Ecole Normale Superieure*, 3(17):21–86.
- Samuelson, P. (1965) Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 41– 49.
- Kou, S. (2002) A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101.
- Hobson, D. (2004) A survey of mathematical finance. *Proceedings of the Royal Society: Mathematical, Physical and Engineering Sciences*. 460:2052, 3369-3401

Black-Scholes model

Given a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, X_t can be modeled by the following stochastic differential equation

$$dX_t = \mu X_t dt + \sigma X_t dW_t,$$

where μ is a constant drift, σ is a constant volatility, W_t is a Brownian motion under $\mathbb{P}$.

The value of an European call option contract is then given by

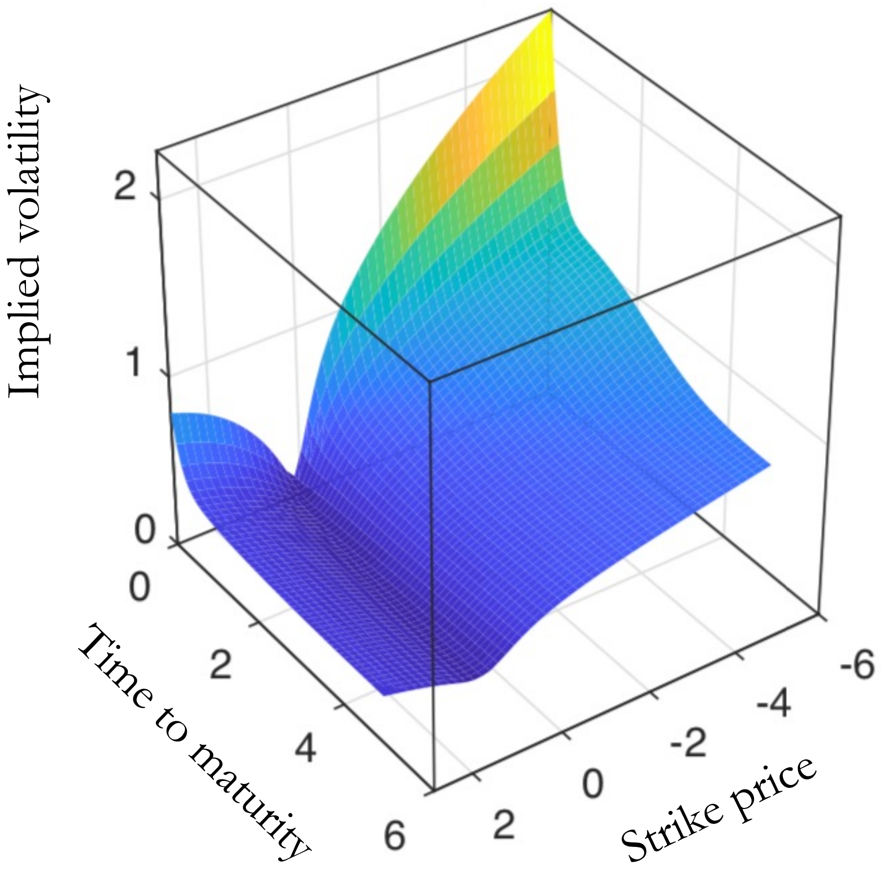
$$\begin{aligned}
C(X_t, t) &= N(d_1)X_t - N(d_2)Ke^{-r(T-t)}, \\
d_1 &= \frac{1}{\sigma\sqrt{T-t}}\left[\ln\left(\frac{X_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right], \\
d_2 &= d_1 - \sigma\sqrt{T-t},
\end{aligned}$$

where r is the risk-less interest rate, T is the expiration date of the option, and t is the current time.

Black, F., & Scholes, M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, 81, 637–654.

Implied volatility

The *implied volatility* of an option is defined as the inverse problem of option pricing, mapping from the option price in the current market to a single value. When it is plotted against the option price strike price and the time to maturity, it is referred to as the *implied volatility surface*.



To avoid dealing with interest rates and dividends, the forward measure is used

The implied volatility $v(m, \tau)$ can be written as a function of m and τ, where m is the log forward moneyness and τ is the annualized time to maturity. The value of v can be obtained by inverting the Black–Scholes formula. The log forward moneyness can be computed by $\log\{\frac{K}{F_{t,T}}\}$, where K is the strike price, $(F_{t,T})_{t \geq 0}$ is the forward price of the asset with maturity date T so $F_{t,T} := \frac{X_t}{B(t,T)}$ where $B(t, T)$ is the price at time t of a zero-coupon bond paying one unit at time T.

Cont, R., & Da Fonseca, J. (2002) Dynamics of implied volatility surfaces. *Quantitative Finance*, 2, 45–60.

Prior financial domain knowledge

1. financial conditions studied in the existing financial mathematics studies
2. empirical evidence volatility smile

Financial conditions

Absence of arbitrage conditions

Gulisashvili, A. (2012) Analytically Tractable Stochastic Stock Price Models. *Springer*.

Boundaries conditions

Carr, P., & Wu, L. (2007) Stochastic skew in currency options. *Journal of Financial Economics*, 86, 213–247.

Asymptotic slope condition

Lee, R. (2004) The moment formula for implied volatility at extreme strikes. *Mathematical Finance*. 14(3), 469–480.

THEOREM 1. *Let* $d_{\pm}(m,\tau) = -\frac{m}{\sqrt{\tau}v(m,\tau)} \pm \frac{\sqrt{\tau}v(m,\tau)}{2}$, $n(\cdot)$ *and* $N(\cdot)$ *be the probability density and cumulative functions of a standard normal distribution, respectively. The following conditions should hold for an implied volatility surface* $v(m,\tau)$*:*

1. ***(Positivity)*** $v(m,\tau) > 0$ *for all* $(m,\tau) \in \mathbb{R} \times \mathbb{R}^{+}$.
2. ***(Twice Differentiability)*** *For every* $\tau > 0$, $m \to v(m,\tau)$ *is twice differentiable on* $\mathbb{R}$.
3. ***(Monotonicity)*** *For every* $m \in \mathbb{R}$, $\tau \to \sqrt{\tau}v(m,\tau)$ *is increasing on* $\mathbb{R}^{+}$, *therefore*

$$v(m,\tau) + 2\tau\partial_{\tau}v(m,\tau) \geqslant 0.$$

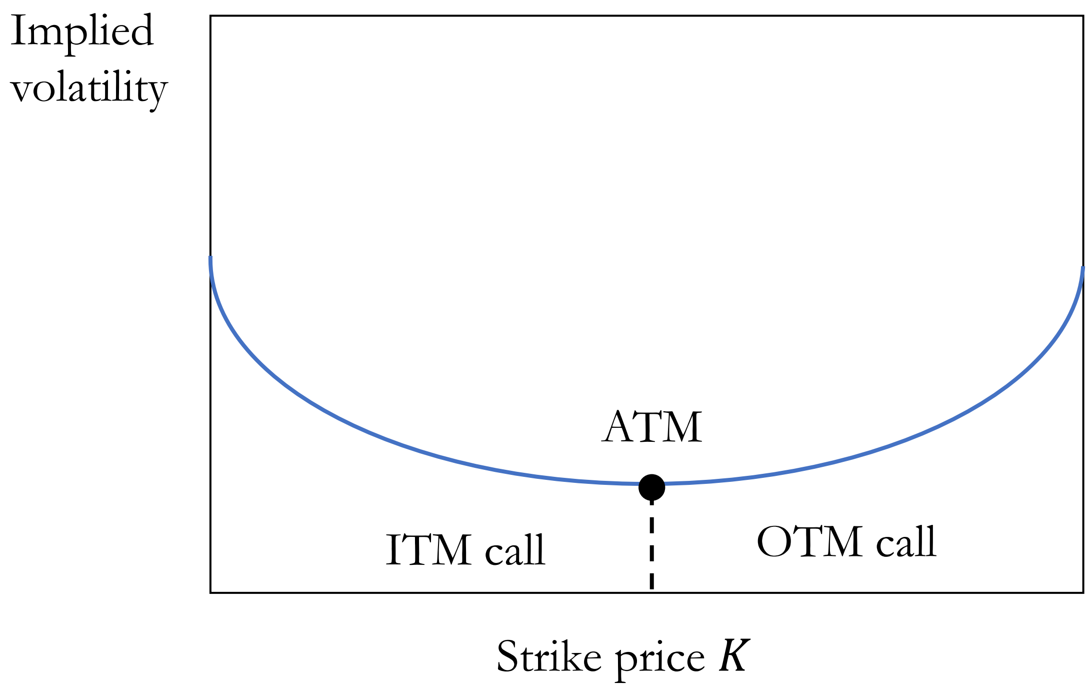
4. ***(Absence of Butterfly Arbitrage)*** *For all* $(m,\tau) \in \mathbb{R} \times \mathbb{R}^{+}$,

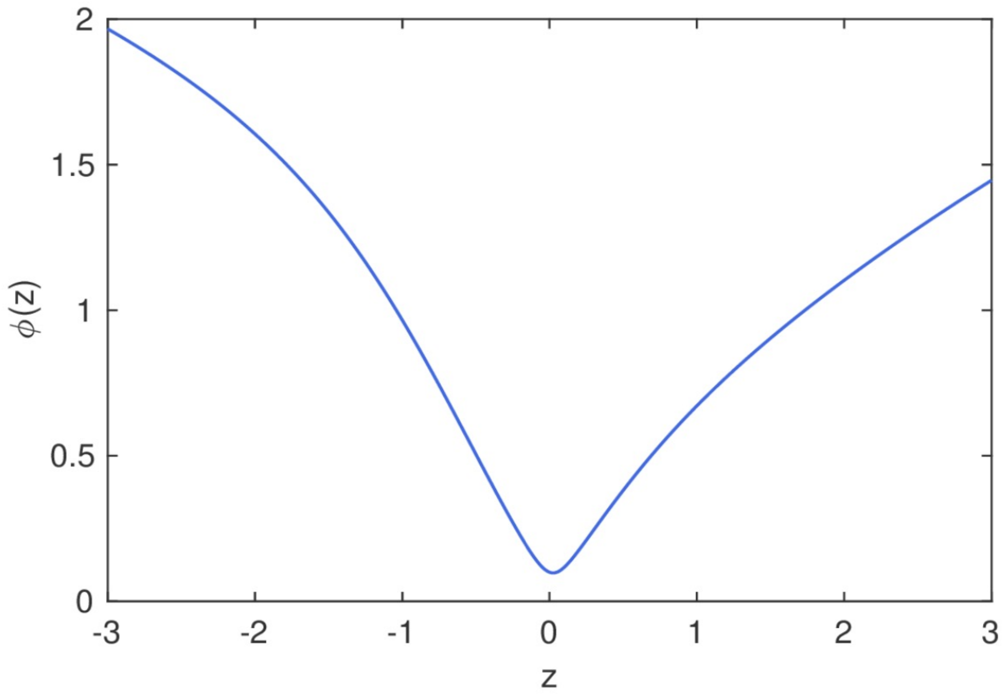
$$\left(1 - \frac{m\partial_m v(m,\tau)}{v(m,\tau)}\right)^2 - \frac{(v(m,\tau)\tau\partial_m v(m,\tau))^2}{4} + \tau v(m,\tau)\partial_{mm}v(m,\tau) \geqslant 0.$$

5. ***(Limit Condition)*** *For every* $\tau > 0$, $\lim_{m\to+\infty} d_{+}(m,\tau) = -\infty$.
6. ***(Right Boundary)*** $N(d_{-}(m,\tau)) - \sqrt{\tau}\partial_m v(m,\tau) n(d_{-}(m,\tau)) \geqslant 0$ *if* $m \geq 0$.
7. ***(Left Boundary)*** $N(-d_{-}(m,\tau)) + \sqrt{\tau}\partial_m v(m,\tau) n(d_{-}(m,\tau)) \geqslant 0$ *if* $m < 0$.
8. ***(Asymptotic Slope)*** *For every* $\tau > 0$, $2|m| - v^2(m,\tau)\tau > 0$.

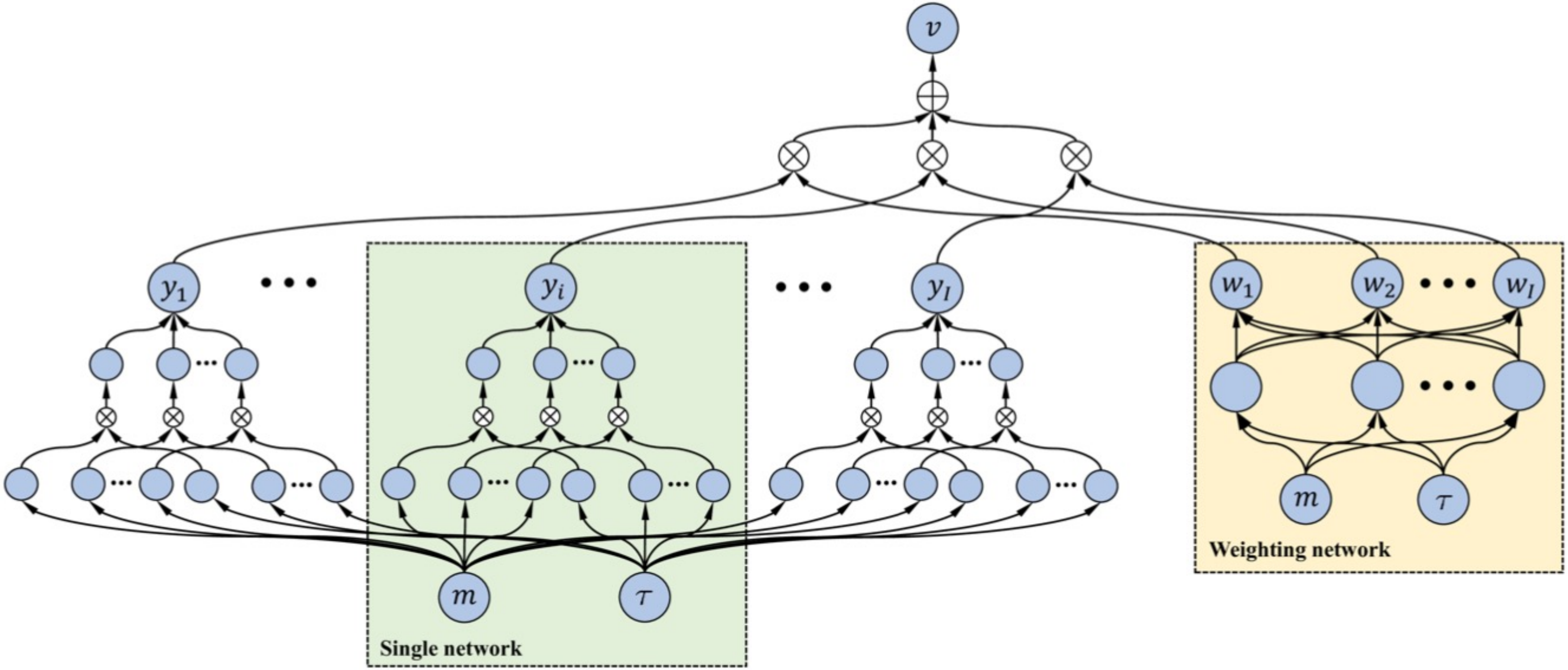
Volatility smile and smile function

$$\phi(z) = \sqrt{z\tanh(z+\frac{1}{2}) + \tanh(-\frac{1}{2}z+\epsilon)}, \quad z \in \mathbb{R},$$





Neural network architecture



Yang, Y, Zheng, Y, & Hospedales, T. (2017). Gated neural networks for option pricing: rationality by design. *AAAI*.

Embedding constraints into optimisation

$$\min \ell = \ell_0 + \gamma\ell_1 + \delta\ell_2 + \eta\ell_3 + \rho\ell_4 + \omega\ell_5,$$

$$\ell_0 = \alpha\underbrace{\left(\frac{1}{N}\sum_{n=1}^{N}(\log(v_n) - \log(\hat{v}_n))^2\right)}_{\text{MSLE}} + \beta\underbrace{\left(\frac{1}{N}\sum_{n=1}^{N}(\frac{v_n - \hat{v}_n}{v_n})^2\right)}_{\text{MSPE}}, \quad \longleftarrow \text{Joint data loss}$$

$$\ell_1 = \sum_{p=1}^{P}\sum_{q=1}^{Q}\max(0, -a(m_p, \tau_q)), \longleftarrow \text{Monotonicity condition} \quad \triangleright\, a(m,\tau) := v(m,\tau) + 2\tau\partial_\tau v(m,\tau)$$

$$\ell_2 = \sum_{p=1}^{P}\sum_{q=1}^{Q}\max(0, -b(m_p, \tau_q)), \longleftarrow \text{Absence of the butterfly arbitrage condition}$$

$$\triangleright\, b(m,\tau) := (1 - \tfrac{m\partial_m v(m,\tau)}{v(m,\tau)})^2 - \tfrac{(v(m,\tau)\tau\partial_m v(m,\tau))^2}{4} + \tau v(m,\tau)\partial_{mm}v(m,\tau)$$

$$\ell_3 = \sum_{p_1=1}^{P_1}\sum_{q=1}^{Q}\max(0, -c_1(m_{p_1}, \tau_q)) + \sum_{p_2=1}^{P_2}\sum_{q=1}^{Q}\max(0, -c_2(m_{p_2}, \tau_q)), \longleftarrow \text{Boundary condition}$$

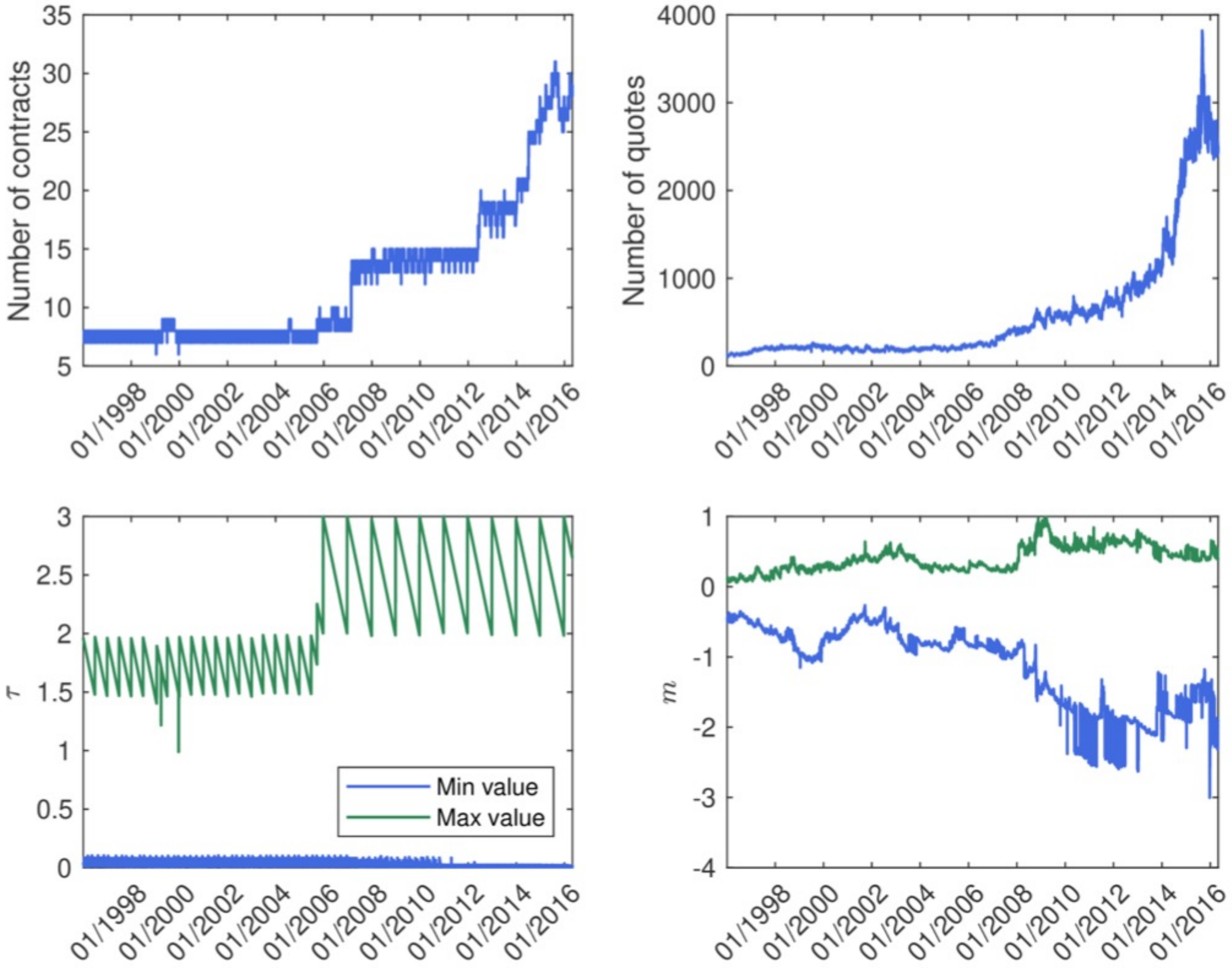
$$\triangleright\, c_1(m,\tau) := N(d_-(m,\tau)) - \sqrt{\tau}\partial_m v(m,\tau)n(d_-(m,\tau))$$

$$\triangleright\, c_2(m,\tau) := N(-d_-(m,\tau)) + \sqrt{\tau}\partial_m v(m,\tau)n(d_-(m,\tau))$$

$$\ell_4 = \sum_{p=1}^{P}\sum_{q=1}^{Q}\max(0, -(g(m_p, \tau_q) - \epsilon)), \longleftarrow \text{Asymptotic condition} \quad \triangleright\, g(m,\tau) := 2|m| - v^2(m,\tau)\tau$$

$$\ell_5 = \begin{cases} ||\bar{W}||_F^2 + ||\tilde{W}||_F^2 + ||\hat{W}||_F^2, \quad \longleftarrow \text{Regularization} & \text{for the single model,} \\ \sum_{i=1}^{I}||\bar{W}^{(i)}||_F^2 + \sum_{i=1}^{I}||\tilde{W}^{(i)}||_F^2 + \sum_{i=1}^{I}||\hat{W}^{(i)}||_F^2 + ||\dot{W}||_F^2 + ||\ddot{W}||_F^2, & \text{for the multimodel.} \end{cases}$$

Data


Number of contracts
Number of quotes
τ
m
Min value
Max value
01/1998
01/2000
01/2002
01/2004
01/2006
01/2008
01/2010
01/2012
01/2014
01/2016

Examined models & hyperparameters' setting

Model	Description
SSVI	[17]
Multi	The proposed model specified in Eqs. (2)-(11).
Multi†	The Multi model trained without embedding $\ell_1, \ell_2, \ell_3, \ell_4$.
Single	The single network model so there is no weighting network, and $\|\|\dot{w}\|\|_F^2$ and $\|\|\ddot{w}\|\|_F^2$ are not included in the regularization term ℓ_5 for the model training
Single†	The Single model trained without embedding $\ell_1, \ell_2, \ell_3, \ell_4$.
Vanilla	The neural network model with the simplest architecture – it has a single hidden layer which only uses the sigmoid activation function and the model's output is censored to be non-negative.
Vanilla†	The vanilla model trained without embedding $\ell_1, \ell_2, \ell_3, \ell_4$.

Model	Hyperparameter									
	I	J	K	α	β	γ	δ	η	ρ	ω
Multi	4	8	5	1	1	10	1	10	1	5e-5
Multi†	4	8	5	1	1	0	0	0	0	5e-5
Single	1	32	-	1	1	10	1	10	1	5e-5
Single†	1	32	-	1	1	0	0	0	0	5e-5
Vanilla	1	32	-	1	1	10	1	10	1	5e-5
Vanilla†	1	32	-	1	1	0	0	0	0	5e-5

- Gatheral, J., & Jacquier, A. (2014). Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14, 59–71.
- Kingma, D, & Ba, J. (2015). Adam: a method for stochastic optimization. *ICLR*, pp. 1–13.
- Yang, Y, Zheng, Y, & Hospedales, T. (2017). Gated neural networks for option pricing: rationality by design. *AAAI*.

Overall performance – mean absolute percentage error (MAPE)

Implied volatility

Model	Training		Test	
	Mean	STD	Mean	STD
Multi	1.74	0.50	3.34	2.18
Multi†	1.76	0.50	3.35	2.17
Single	2.15	0.67	3.60	2.12
Single†	1.82	0.52	3.38	2.16
Vanilla	3.21	0.98	4.46	2.07
Vanilla†	2.87	0.80	4.18	2.04
SSVI	2.59	0.85	3.73	2.18

Option price

Model	Training		Test	
	Mean	STD	Mean	STD
Multi	5.97	1.86	10.64	6.72
Multi†	6.03	1.86	10.67	6.70
Single	7.38	2.57	11.64	6.68
Single†	6.20	1.91	10.77	6.67
Vanilla	11.31	3.57	14.61	6.42
Vanilla†	10.53	3.34	14.17	6.60
SSVI	8.71	2.72	12.74	6.74

Conditions check

Model	Monotonicity	Absence of butterfly arbitrage	Left boundary	Right boundary	Asymptotic slope
Multi	0.00%	7.02e-6%	0.00%	0.00%	0.00%
Multi†	1.28%	4.87%	0.00%	14.06%	0.00%
Single	0.00%	5.56e-3 %	0.00%	0.05%	0.00%
Single†	0.00%	14.88%	0.95%	5.16%	0.00%
Vanilla	3.75e-3%	1.53e-2%	4.07e-3%	0.00%	0.00%
Vanilla†	5.32%	5.72%	14.63%	0.00%	0.54%

Conclusion

- Technology wise, we propose a framework of incorporating prior financial domain knowledge into neural network design and training. This is an important step for interpretable machine learning, and we hope the framework can motivate many other investigations of machine learning applications in finance.
- From the application perspective, we develop a best-performing prediction model, and the conventional financial conditions and empirical evidence are met empirically. To the best of our knowledge, this is one of the very first neural networks tailored for implied volatility surface.

Thank you!