Deep Smoothing of the Implied Volatility Surface

Damien Ackerer, Natasa Tagasovska, Thibault Vatter

NeurIPS 2020

Background
What is implied volatility?

- Implied volatility surface

  derived from BS model

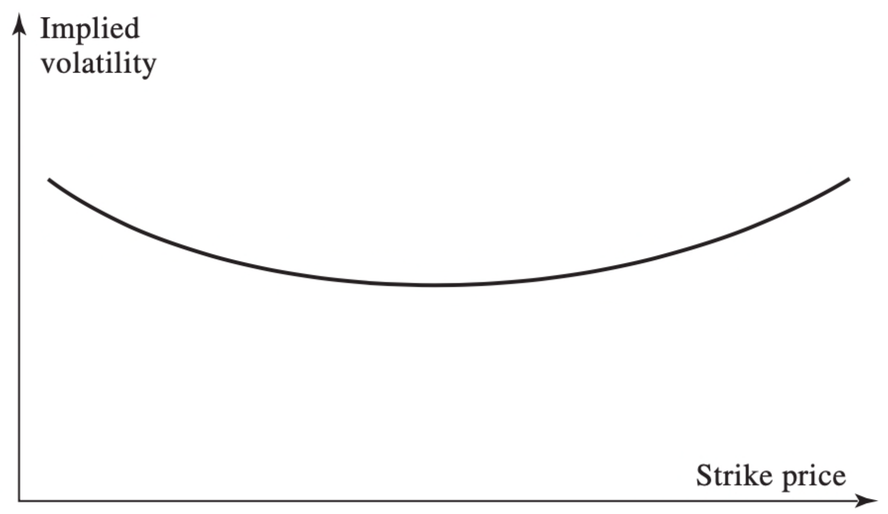
  volatility smile & volatility skew

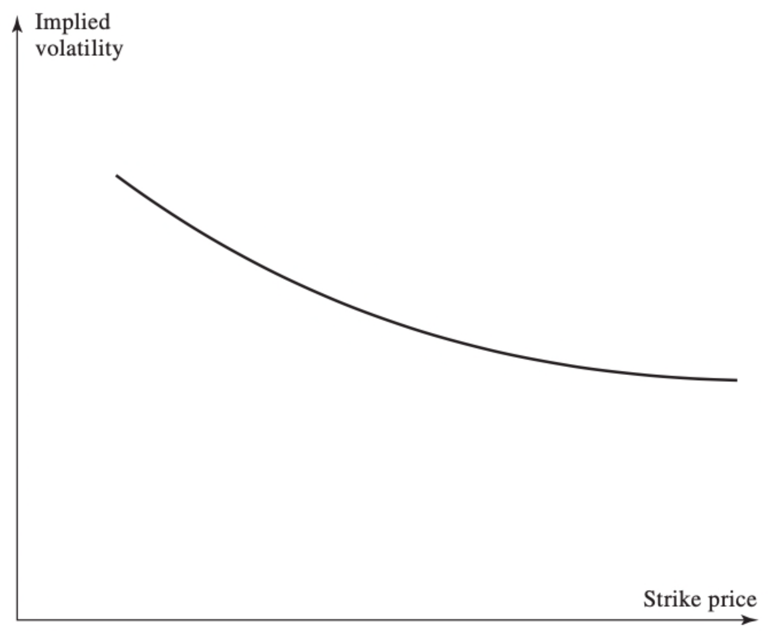
$$k = \log(K/Se^{-\delta\tau})$$

$$\pi(K,\tau) = C(S, \sigma(k,\tau), r, \delta, K, \tau)$$
$$= Se^{-\delta\tau}\Phi(d_+) - e^{-r\tau}K\Phi(d_-),$$

where $d_\pm = (\log(S/K) + (r-\delta)\tau)/(\sigma\sqrt{\tau}) \pm (1/2)\sigma\sqrt{\tau}$

$\Phi(\cdot)$ is a Gaussian CDF





	K/S_0				
	0.90	*0.95*	*1.00*	*1.05*	*1.10*
1 month	14.2	13.0	12.0	13.1	14.5
3 month	14.0	13.0	12.0	13.1	14.2
6 month	14.1	13.3	12.5	13.4	14.3
1 year	14.7	14.0	13.5	14.0	14.8
2 year	15.0	14.4	14.0	14.5	15.1
5 year	14.8	14.6	14.4	14.7	15.0

Introduction

Why we need the method?

- Black-Scholes → constant volatility
- Stochastic volatility model → hard to do the calibration
- Neural network → arbitrage opportunity & shallow network
- Contribution: Prior model + NN model

→free of arbitrage + generalization + deep network

Dugas Charles, Bengio Yoshua, Bélisle François, Nadeau Claude, Garcia René. Incorporating second-order functional knowledge for better option pricing // Advances in Neural Information Processing Systems. 2001. 472–478.
Márquez-Neila Pablo, Salzmann Mathieu, Fua Pascal. Imposing hard constraints on deep networks: Promises and limitations // Unpublished manuscript. 2017.

Conditions of free of arbitrage call option

Roper Michael. Arbitrage free implied volatility surfaces // Unpublished manuscript. 2010.

$\omega(k,\tau) = \sigma^2(k,\tau)\tau$

$C1)$(positivity) for every $k \in R$ and $\tau > 0$, $\omega(k,\tau) > 0$

$C2)$(Value at maturity) for every $k \in R$, $\omega(k,0) = 0$

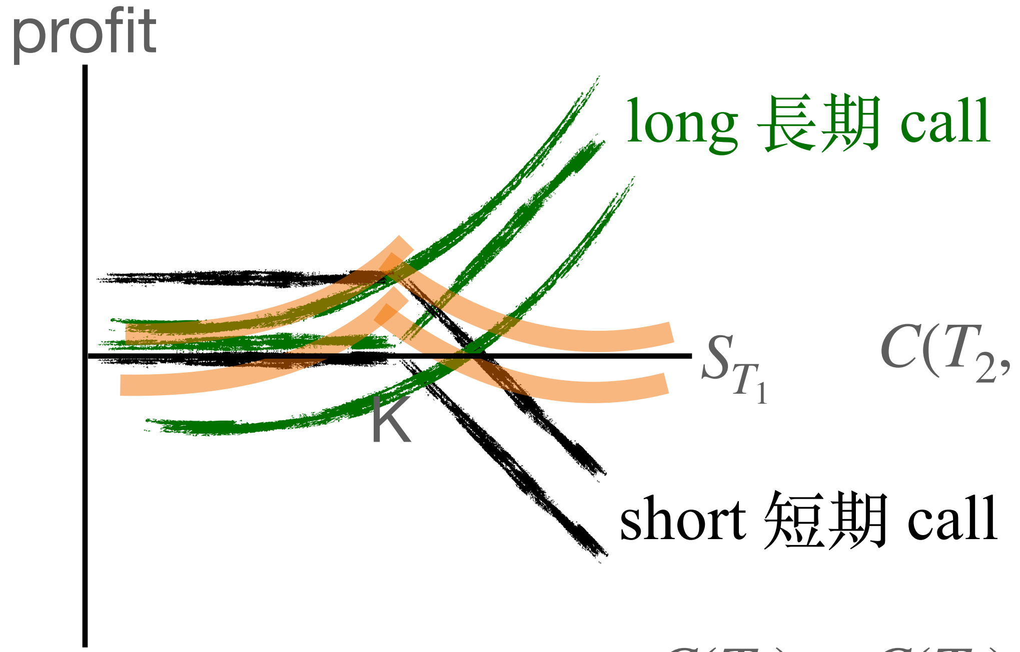
$C3)$(Smoothness) for every $\tau > 0$, $\omega(\cdot,\tau)$ is twice differentiable

$C4)$(Monotonicity in τ) to prevent calendar arbitrage, total variance $\omega(k,\cdot)$ should be non-decreasing along the time axis for any given forward-moneyness level (for every $k \in R$), $l_{cal}(k,\tau) = \partial_\tau \omega(k,\tau) \geq 0$

$C5)$(Durrleman's Condition)for every $\tau > 0$ and $k \in R$,

$$l_{but}(k,\tau) = \left(1 - \frac{k\partial_k \omega(k,\tau)}{2\omega(k,\tau)}\right)^2 - \frac{\partial_k \omega(k,\tau)}{4}\left(\frac{1}{\omega(k,\tau)} + \frac{1}{4}\right) + \frac{\partial^2_{kk}\omega(k,\tau)}{2} \geq 0$$

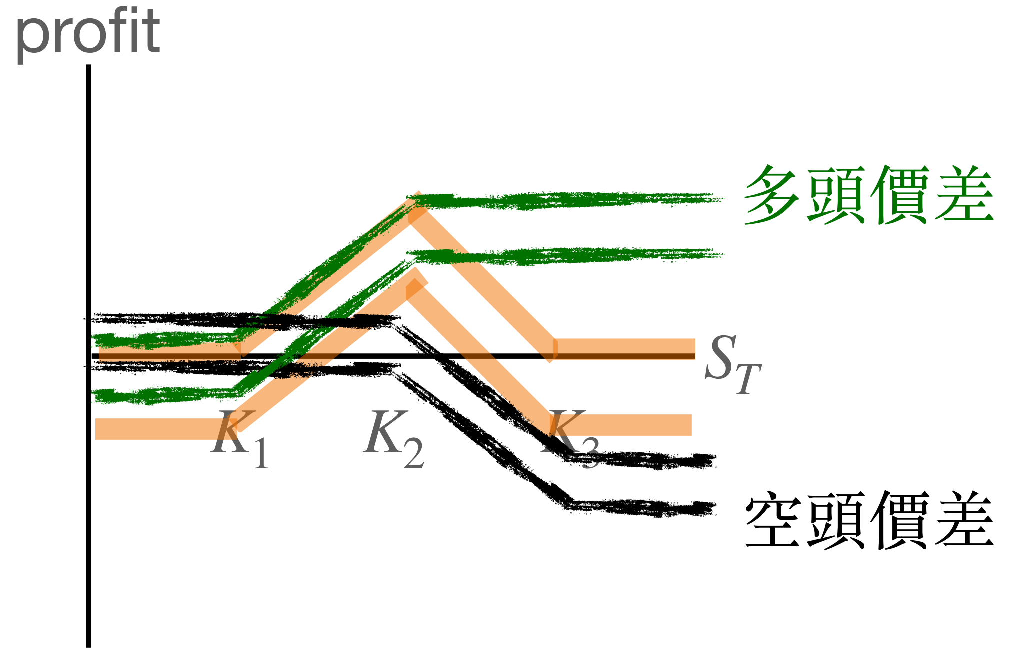
$C6)$(Large moneyness behavior) for every $\tau > 0$, $\sigma^2(k,\tau)$ is linear for $k \to \pm\infty$



S_{T_1} $\quad C(T_2,K) - \max\{S_{T_1} - K,0\} \geq 0$

$C(T_1) - C(T_2)$

$C(T_1) - C(T_2) + C(T_2,K) - \max\{S_{T_1} - K,0\}$



$\max(S_T - K_1,0) - \max(S_T - K_2,0)$
$-C(K_1) + C(K_2)$

≥ 0

$-\max(S_T - K_2,0) + \max(S_T - K_3,0)$
$C(K_2) - C(K_3)$

Methodology

Model

$$\sigma_\theta(k,\tau) = \sqrt{\omega_\theta(k,\tau)/\tau}$$

correction

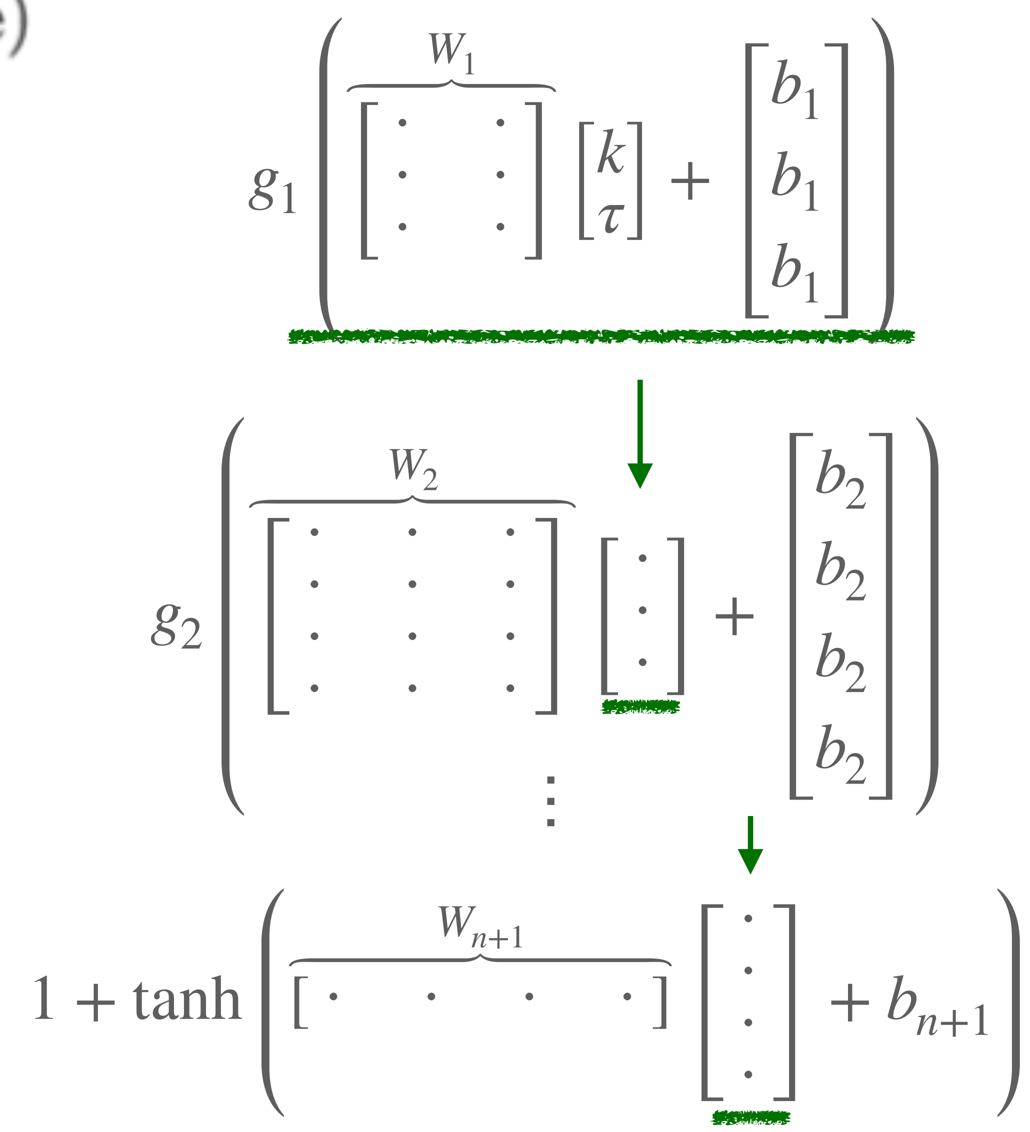
$$\omega_\theta(k,\tau) = \omega_{\text{nn}}(k,\tau;\theta_1) \times \omega_{\text{prior}}(k,\tau;\theta_2), \text{ where the parameter } \theta = \{\theta_1, \theta_2\}$$

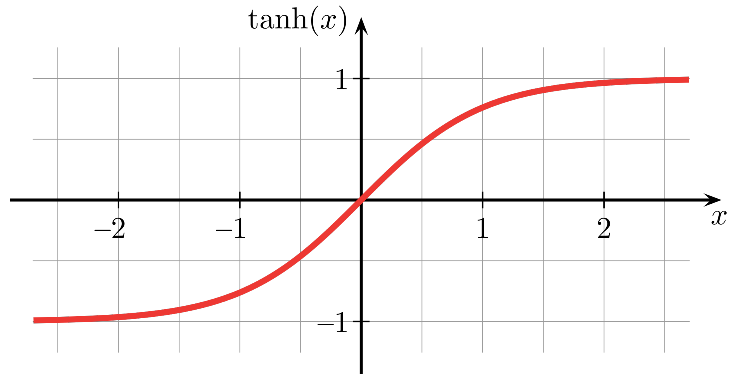
$$\omega_{\text{prior}}(k,\tau;\theta_2) = \begin{cases} \omega^{\text{bs}}_{\text{prior}}(k,\tau) = \sigma^2\tau \to \omega^{\text{bs}}_{\text{prior}}(k,\tau) = \omega_{\text{atm}}(\tau) \\ \omega^{\text{ssvi}}_{\text{prior}}(k,\tau) = \frac{\omega_{\text{atm}}(\tau)}{2}\left(1 + \rho\phi(\omega_{\text{atm}}(\tau))k + \sqrt{(\phi(\omega_{\text{atm}}(\tau))k + \rho)^2 + 1 - \rho^2}\right), \text{ where } \phi(x) = \frac{\eta}{x^\gamma(1+x)^{1-\gamma}} \\ \text{where } \omega_{\text{atm}} \text{ is extracted from the market data by collecting the total variance } \sigma^2\tau \text{ values} \\ \quad \text{corresponding to the contract closest to } k = 0 \text{ for each maturity } \tau \\ \quad \text{p.s. } \sigma^2 \text{ here is historical volatility (hence it's the "theoretical" option price)} \end{cases}$$

composition operator

$$\omega_{\text{nn}}(k,\tau;\theta_1) = \circ_{i=1}^{n+1} f_i^{W_i,b_i}(k,\tau) \text{ with } f_i(x) = \begin{cases} g_i(W_i x + b_i) & i < n+1 \\ \alpha(1 + \tanh(W_{n+1}x + b_{n+1})) & i = n+1 \end{cases},$$

where $\theta_1 = \{W_1, b_1, W_2, b_2, \ldots, \alpha\}$, g_i is the activation function, α is the scaling parameter letting ω_{nn} take values in $[0, 2\alpha]$





Methodology

Loss function

$C1$)(positivity) for every $k \in R$ and $\tau > 0$, $\omega(k, \tau) > 0$

$C2$)(Value at maturity) for every $k \in R$, $\omega(k,0) = 0$

$C3$)(Smoothness) for every $\tau > 0$, $\omega(\cdot, \tau)$ is twice differentiable

$C4$)(Monotonicity in τ) to prevent calendar arbitrage, total variance $\omega(k, \cdot)$ should be non-decreasing along the time axis for any given forward-moneyness level (for every $k \in R$), $l_{cal}(k, \tau) = \partial_\tau \omega(k, \tau) \geq 0$

$C5$)(Durrleman's Condition)for every $\tau > 0$ and $k \in R$,

$$l_{but}(k, \tau) = \left(1 - \frac{k\partial_k \omega(k, \tau)}{2\omega(k, \tau)}\right)^2 - \frac{\partial_k \omega(k, \tau)}{4}\left(\frac{1}{\omega(k, \tau)} + \frac{1}{4}\right) + \frac{\partial^2_{kk}\omega(k, \tau)}{2} \geq 0$$

$C6$)(Large moneyness behavior) for every $\tau > 0$, $\sigma^2(k, \tau)$ is linear for $k \to \pm\infty$

prediction error

arbitrage penalty

$$L(\theta) = L_0(\theta) + \sum_{j=1}^{6} \lambda_j L_{Cj}(\theta)$$

$$\text{where } L_0(\theta) = \sqrt{1/|I_0| \sum_{\sigma_i, k_i, \tau_i \in I_0} (\sigma_i - \sigma_\theta(k_i, \tau_i))^2} + 1/|I_0| \sum_{\sigma_i, k_i, \tau_i \in I_0} |\sigma_i - \sigma_\theta(k_i, \tau_i)|/\sigma_i$$

$$L_{C1} \equiv L_{C2} \equiv L_{C3} \equiv 0$$

$$L_{C4}(\theta) = 1/|I_{C45}| \sum_{(k_i, \tau_i) \in I_{C45}} \max(0, -l_{\text{cal}}(k_i, \tau_i))$$

$$L_{C5}(\theta) = 1/|I_{C45}| \sum_{(k_i, \tau_i) \in I_{C45}} \max(0, -l_{\text{but}}(k_i, \tau_i))$$

$$L_{C6}(\theta) = 1/|I_{C6}| \sum_{(k_i, \tau_i) \in I_{C6}} |\partial^2 \omega_\theta(k_i, \tau_i)/\partial k \partial k|$$

$$\text{where } I_{C45} = \left\{(k, \tau) : k \in \{x^3 : x \in [-(-2k_{\min})^{1/3}, (2k_{\max})^{1/3}]_{100}\}, \tau \in \mathrm{T}\right\}$$

$$I_{C6} = \{(k, \tau) : k \in \{6k_{\min}, 4k_{\min}, 4k_{\max}, 6k_{\max}\}, \tau \in \mathrm{T}\}$$

$$\mathrm{T} = \{\exp(x) : x \in [\log(1/365), \max(\log(\tau_{\max} + 1))]_{100}\}$$

$$\text{where } k_{\min} = \min(I_0^k) < 0, k_{\max} = \max(I_0^k) > 0, \tau_{\max} = \max(I_0^\tau),$$

$[a, b]_x$ indicates an equidistant set of x points between a and b

Results

Setting

- Underlying Asset: S&P 500
- Hyper-parameters: $\lambda_4 = \lambda_5 = \lambda_6 = \lambda = 10$
- Prior model: SSVI
- NN model: 4 layers and 40 neurons per layer
- Synthetic data: from Bates model
- Market data: from market

Results

Synthetic Data

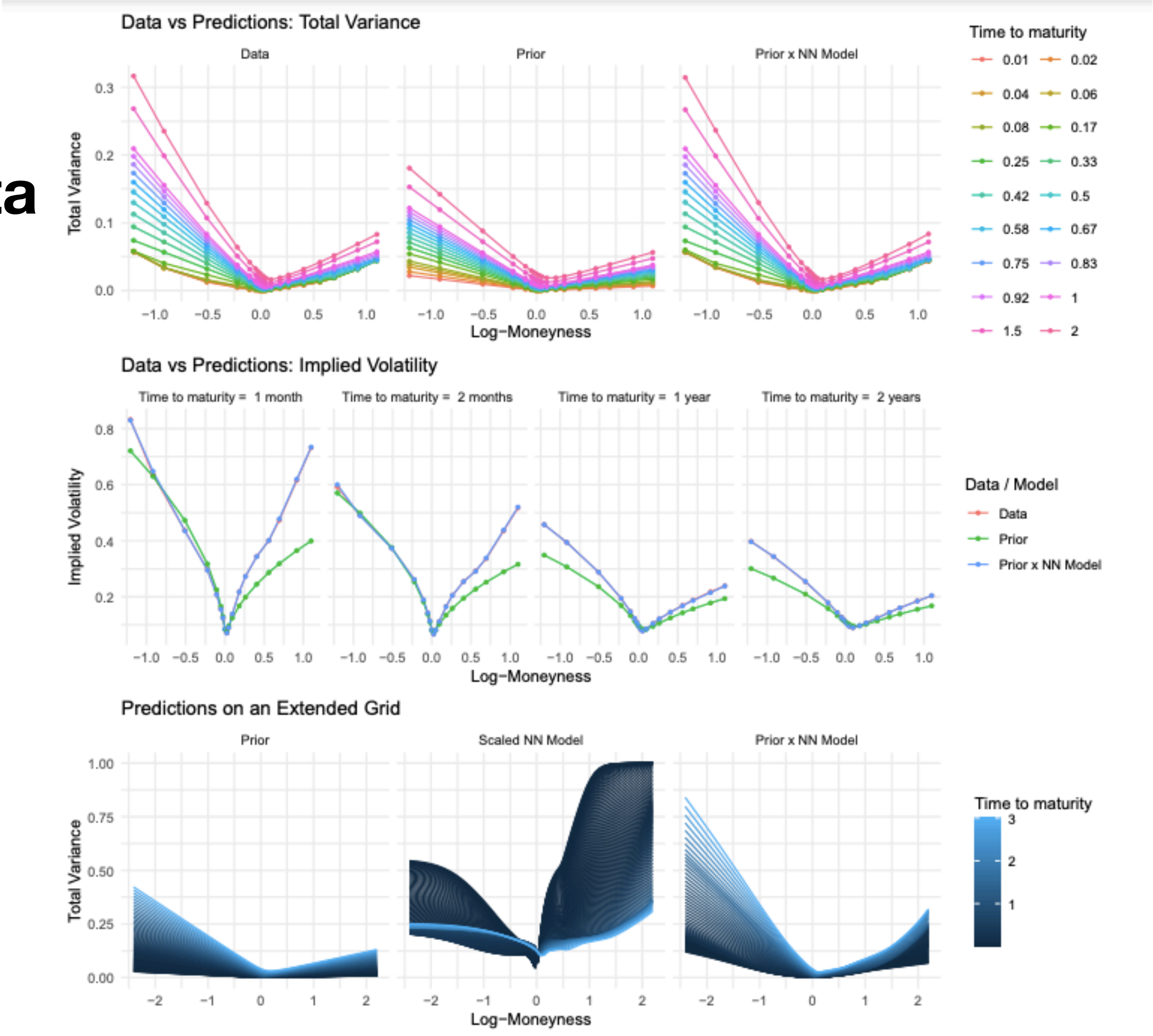


Figure 1: Synthetic data and trained model predictions for a specific configuration (scenario 12).

Results

Comparing different configurations

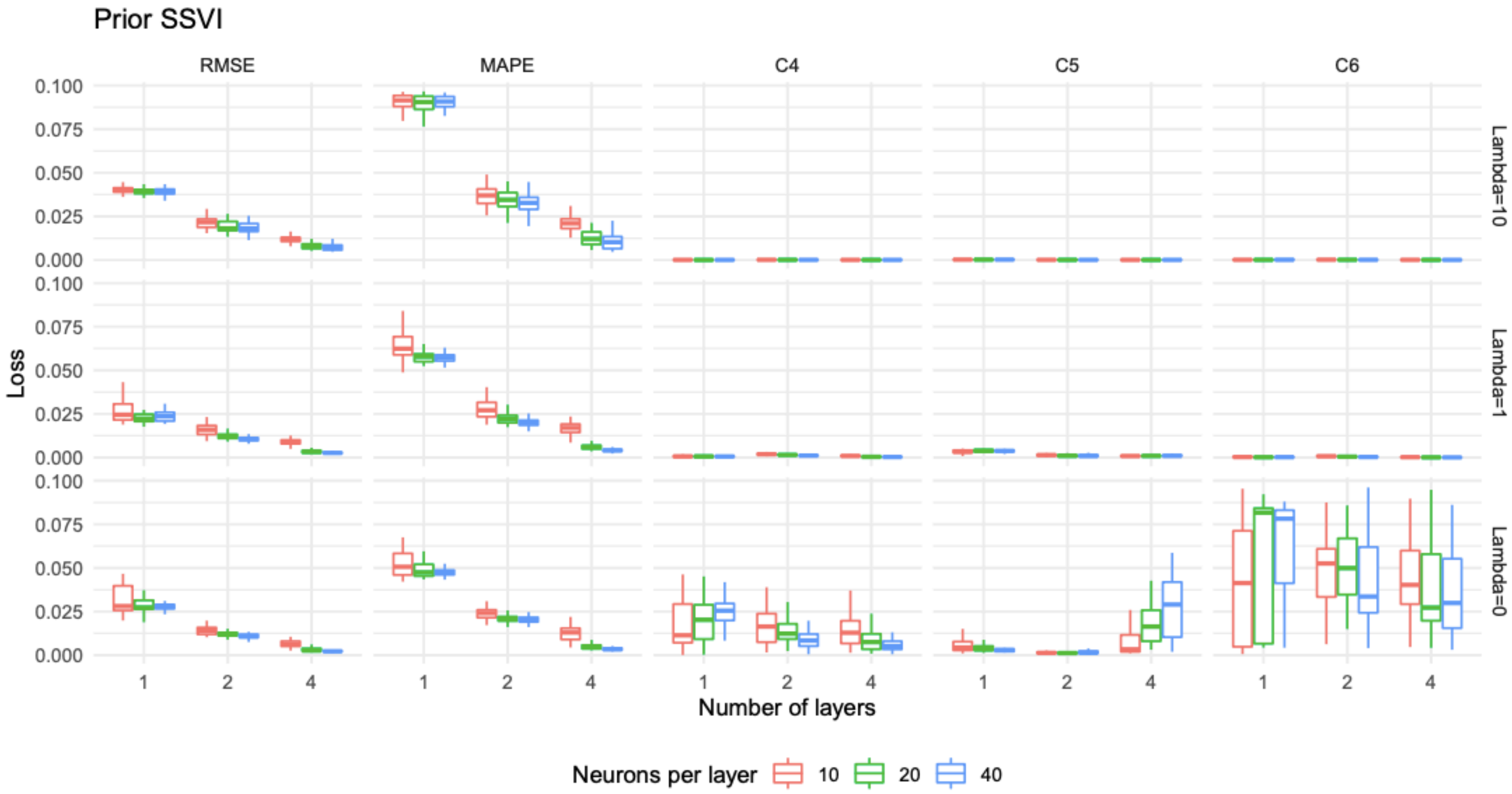


Figure 2: Losses for different number of layers, neurons per layer, and penalty value.

Results

Market data

$$C(S, \sigma, r, \delta, K, \tau) = S_0 e^{-\delta\tau}\Phi(d_+) - Ke^{-r\tau}\Phi(d_-)$$

where $d_+ = \frac{\ln(S_0/K) + ((r-\delta) + 0.5 * \sigma^2)\tau}{\sigma\sqrt{\tau}}$

$$d_- = \frac{\ln(S_0/K) + ((r-\delta) - 0.5 * \sigma^2)\tau}{\sigma\sqrt{\tau}} = d_+ - \sigma\sqrt{\tau}$$

$\Phi(d)$ is the cumulative probability distribution for a variable that has a standard normal distribution with mean of zero and standard deviation of one.

- Dataset: Option Metric IvyDB US, S&P500 index option price (European)
- Date: 2008/9 ~ 2008/12、2018/1 ~ 2018/4 (Daily)
- Steps:
  - Pick the mid price $\frac{\text{bid price + ask price}}{2}$
  - Calculate risk free rate and dividend rate using option around the money (± 7.5%) and put-call parity $C_t - P_t = S_t e^{\delta\tau} + Ke^{-r\tau}$ $\quad C_t - P_t = S_t\beta^S_{t,T} + K\beta^K_{t,T}$

    $e^{-r\tau} = \beta^K_{t,T}$ $\quad e^{-\delta\tau} = \beta^S_{t,T}$

    $\Rightarrow \log e^{-r\tau} = \log \beta^K_{t,T}$ $\quad \Rightarrow \log e^{-\delta\tau} = \log \beta^S_{t,T}$

    $\Rightarrow -r\tau = \log \beta^K_{t,T}$ $\quad \Rightarrow -\delta\tau = \log \beta^S_{t,T}$

    $\Rightarrow r = -\log \beta^K_{t,T}/\tau$ $\quad \Rightarrow \delta = -\log \beta^S_{t,T}/\tau$
  - Calculate the moneyness $k = \log(K/Se^{e-\delta}\tau) = \ln(K/S_t) - (r_{t,T} - \delta_{t,T})(T-t)$
  - Calculate the implied volatility using Black-Scholes model and Brent's method

Results

Backtesting

Table 1: Backtesting results for the SSVI prior (quantiles in %, Jan-Apr 2018 / Sep-Dec 2008)

		Interpolation						Extrapolation					
		Train			Test			Train			Test		
Loss	λ	q_{05}	q_{50}	q_{95}	q_{05}	q_{50}	q_{95}	q_{05}	q_{50}	q_{95}	q_{05}	q_{50}	q_{95}
RMSE	10	0.4 / 0.7	0.5 / 3.1	1.3 / 19.9	0.4 / 0.9	0.5 / 3.3	1.4 / 18.8	0.2 / 0.3	0.3 / 1.1	0.4 / 3.5	2.2 / 2.7	5.0 / 6.3	8.0 / 12.0
	1	0.3 / 3.0	0.4 / 6.8	1.0 / 13.9	0.4 / 2.9	0.5 / 7.1	1.3 / 13.5	0.2 / 1.9	0.2 / 4.4	0.4 / 10.6	3.7 / 5.0	6.6 / 10.6	11.7 / 20.5
	0	0.2 / 0.4	0.3 / 1.3	0.5 / 4.0	0.3 / 1.6	0.5 / 3.5	4.8 / 10.8	0.2 / 0.3	0.2 / 0.9	0.3 / 3.3	3.1 / 5.5	7.5 / 11.6	18.1 / 26.7
MAPE	10	0.5 / 0.9	0.7 / 2.1	1.2 / 17.4	0.5 / 1.1	0.8 / 2.4	1.2 / 18.5	0.4 / 0.5	0.6 / 0.9	0.9 / 2.1	1.2 / 2.4	1.7 / 3.3	2.4 / 6.5
	1	0.5 / 4.0	0.6 / 8.2	1.2 / 12.2	0.5 / 4.3	0.7 / 8.1	1.3 / 12.9	0.3 / 2.2	0.5 / 6.3	0.9 / 11.0	1.5 / 5.6	2.3 / 9.7	3.3 / 13.1
	0	0.4 / 0.5	0.6 / 1.0	0.9 / 1.8	0.5 / 1.2	0.7 / 1.9	0.9 / 3.2	0.3 / 0.3	0.5 / 0.7	0.8 / 1.6	1.5 / 4.3	2.2 / 7.4	4.8 / 14.3
C4	10	0.0 / 0.0	0.0 / 0.0	0.0 / 0.2	0.0 / 0.0	0.0 / 0.0	0.0 / 0.5	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 14.1
	1	0.0 / 0.0	0.0 / 0.0	0.2 / 0.1	0.0 / 0.0	0.0 / 0.0	0.2 / 0.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.1	0.0 / 0.0	0.0 / 0.0	0.1 / 0.1
	0	1.6 / 1.5	18.0 / 9.8	99+ / 99+	1.6 / 1.2	40.7 / 12.1	99+ / 99+	0.3 / 1.1	2.4 / 3.7	44.2 / 83.7	0.4 / 1.0	2.8 / 4.3	46.7 / 30.4
C5	10	0.0 / 0.0	0.0 / 0.0	0.0 / 0.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 1.2
	1	0.0 / 0.0	0.0 / 0.0	0.1 / 0.0	0.0 / 0.0	0.0 / 0.0	0.5 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.5 / 0.0
	0	2.6 / 99+	72.1 / 99+	99+ / 99+	2.6 / 99+	63.8 / 99+	99+ / 99+	2.8 / 99+	24.2 / 99+	99+ / 99+	1.7 / 99+	19.0 / 99+	99+ / 99+
C6	10	0.0 / 0.0	0.0 / 0.0	0.0 / 0.2	0.0 / 0.0	0.0 / 0.0	0.0 / 0.6	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 44.3
	1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.6 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
	0	0.8 / 0.0	48.2 / 1.7	99+ / 99+	0.8 / 0.0	78.4 / 0.4	99+ / 99+	0.3 / 0.1	12.0 / 7.5	99+ / 99+	0.0 / 0.0	2.4 / 0.0	99+ / 99+

Results

Market Data

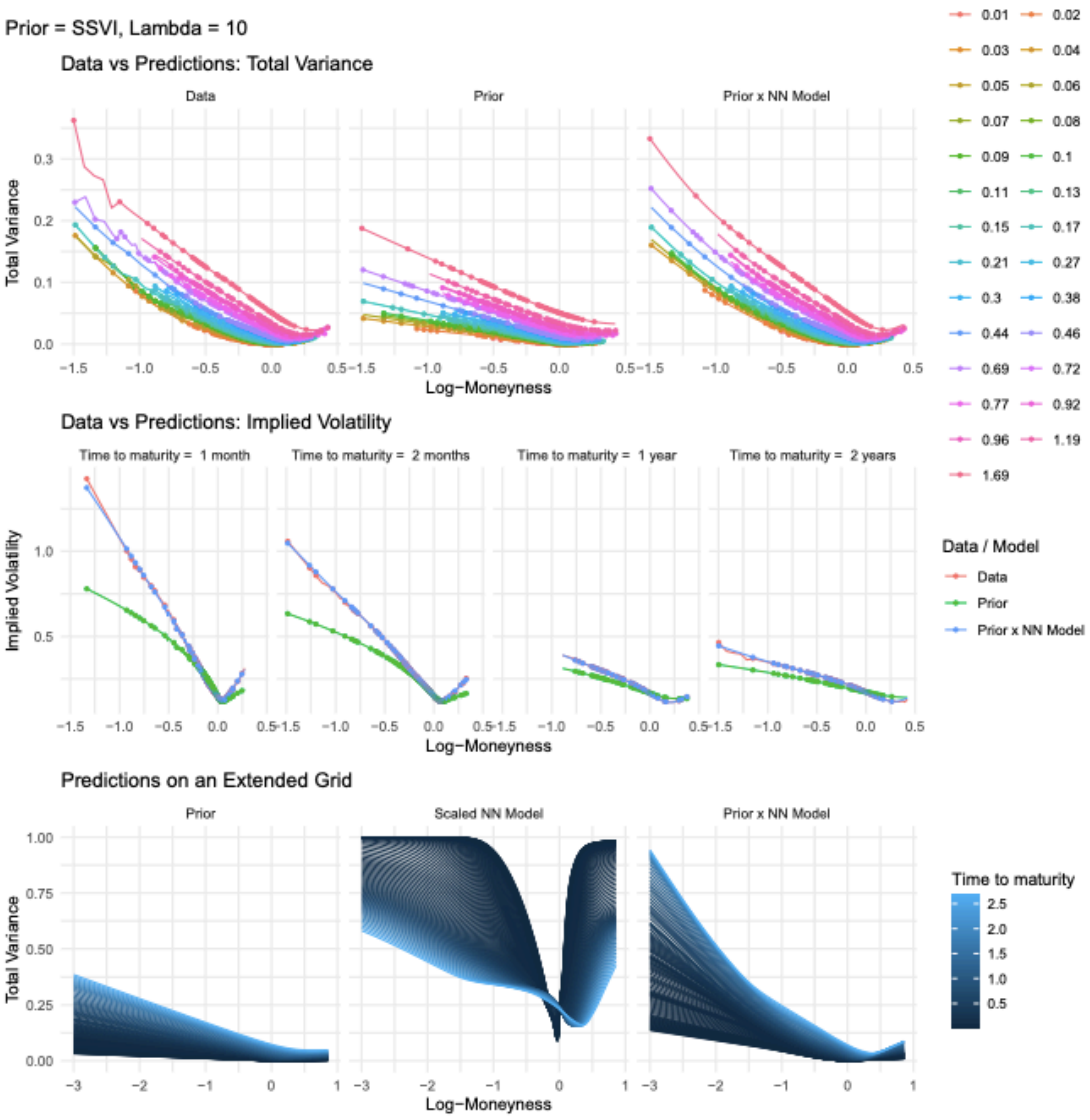


Figure 3: Market data and trained model predictions for a specific configuration (scenario 12).

Conclusion

- Construct the implied volatility surface based on standard volatility model.
- Design an economically sensible model by penalizing the arbitrage free opportunity.
- Use the deep neural network framework successfully.