A Bayesian nonparametric approach to option pricing

Zhang Qin[†] Caio Almeida[‡]

Abstract Accurately modeling the implied volatility surface is of great importance to option pricing, trading and hedging. In this paper, we investigate the use of a Bayesian nonparametric approach to fit and forecast the implied volatility surface with observed market data. More specifically, we explore Gaussian Processes with different kernel functions characterizing general covariance functions. We also obtain posterior distributions of the implied volatility and build confidence intervals for the predictions to assess potential model uncertainty. We apply our approach to market data on the S&P 500 index option market in 2018, analyzing 322,983 options. Our results suggest that the Bayesian approach is a powerful alternative to existing parametric pricing models.

Keywords: Machine Learning, Gaussian processes, Bayesian nonparametrics, Options, Kernels.

JEL Code: C13, C11.

1. Introduction

Implied volatility is a measure of the future expected risk of the underlying price. Accurately modeling and forecasting implied volatility is of great importance in both finance theory and practical decision making. In this project, we propose a method to approach the option pricing problem in a Bayesian framework.

Implied volatility tends to increase for in-the-money (ITM) and out-ofthe-money (OTM) options. This pattern is usually described as the "volatility smile" as the implied volatility plotted against strike or moneyness looks like the shape of a smile. Dupire (1994) extends the Black-Scholes model by allowing the volatility to be a deterministic function of the spot price and time. Heston (1993) develops a mean-reverting stochastic volatility model, with closed-form solution for the implied volatility surface up to solving an inverse Fourier transform. Other stochastic volatility models include Hull-White (Hull and White; 1987), SABR (Hagan et al.; 2003), etc. Despite their popularity in the industry, these models are very complicated and sometimes without closed-form solution. Among others, Dumas et al. (1998) show that

Submitted on July 18, 2020. Revised on August 9, 2020. Accepted on August 10, 2020. Published online in December 2020. Editor in charge: Marcelo Fernandes.

[†]Bendheim Center for Finance, Princeton University, USA: zqin@princeton.edu

[‡]Department of Economics, Princeton University, USA: calmeida@princeton.edu

most local volatility models are not stable over time and that the model overfits the current data. In addition, the one-week-ahead pricing error is usually very large.

With recent developments in machine learning, data-driven methods of fitting the implied volatility surface have become a heated area of research, with promising results. Most machine learning attempts involve the use of deep learning in option pricing. For example, Buehler et al. (2019) use neural networks and reinforcement learning to hedge options with a model-independent algorithm. Fan and Mancini (2009) use a model-guided nonparametric method to learn and correct systematic bias of pricing errors. Although these models are able to incorporate multiple sources of information available in the market and achieve good accuracy, they are usually very computationally intensive. Also, since most work is done with synthetic data, sometimes the advantage of adopting such methods in real markets is unclear.

In addition to getting the point estimates as accurate as possible, some information on the variance and distribution of the predicted price would also be useful. Model-implied prices always differ from the true price, so it matters to investors to have accurate price confidence intervals. Usually, incorporating model uncertainty significantly increases hedging performance (Branger and Schlag; 2004; Gupta and Reisinger; 2011). That said, it is important to emphasize that Bayesian models are well-equipped for this task. In fact, Bayesian probabilistic modeling allows us to obtain posterior distributions and confidence intervals for the predicted implied volatility, and consequently for the predicted prices as well.

In this paper, building on the idea of integrating machine learning and finance theory, we propose a Bayesian nonparametric approach to model implied volatility surfaces with the goal of pricing options. We aim at improving accuracy and stability of pricing with less demand for data. To that end, we adopt Gaussian Processes (GP) to model the implied volatility as a function of moneyness and time to maturity.

GP models assume a flexible infinite-dimensional Gaussian distribution over the functional space determined by the features selected to compose the model. GP models are general, yet simple to understand, since they are completely characterized by a mean function and a covariance kernel function. By varying the kernel function, one can obtain a rich class of stochastic processes that can fit the dynamics of financial variables well.

Based on a fixed subset of options, we estimate the mean and covariance kernel functions of different GP models to test their ability to price out-of-sample options. Using market data on S&P 500 index options, we compare the performance of four GP models with different kernel functions to the suc-

cessful parametric model proposed by Dumas et al. (1998) that provides very stable implied volatility surfaces over time. We show that the GP models not only outperform the parametric model in terms of out-of-sample pricing errors, but they also provide posterior distributions for option prices that, when translated into confidence intervals, are informative about the accuracy of the estimated prices.

Our method may be seen as a good alternative to traditional parametric option pricing models, offering a novel probabilistic and data-driven approach when the underlying price dynamics are unknown but can be approximated by a GP model. Moreover, the available inference results, which appear as a byproduct of our method, can help decision-making in realistic tasks such as portfolio hedging and risk management, since they allow us to evaluate the uncertainty embedded in all option price predictions.

The paper is organized as follows. In Section 2 we review the most relevant background and references on option pricing and Bayesian modeling literature. Section 3 discusses the formulation of our model. Section 4 reports the result from the empirical analysis we conduct with the S&P 500 index option prices. Finally, we discuss our results and conclusions, as well as potential ideas for further development in Section 5.

2. Background and literature review

Options are contracts between the buyer and seller that give the purchaser the right to buy (call options) or sell (put options) an underlying asset at a predetermined time and price. Option contracts are important financial instruments for risk hedging and are actively traded in the market. In this section, we review widely used models and some previous work on option pricing with machine learning methods.

2.1 Black-Scholes model and implied volatility

The Black-Scholes (BS) model (Black and Scholes; 1973) is the fundamental framework in option pricing. The BS model assumes the underlying stock price *S* follows a geometric Brownian motion. Formally,

$$dS_t = \mu S_t dt + \sigma S_t dB_t,$$

where B_t is a standard Brownian motion. The payoff of a European call option is $(S_T - K)^+$, whereas the payoff of a put option is $(K - S_T)^+$, with S_T denoting the underlying price at maturity T and K the strike price. Solving the corresponding partial differential equation with appropriate boundary condi-

tions, we can obtain a closed-form solution of the price of a non-dividendpaying European call option under the Black-Scholes model:

$$C(S_0, K, r, \sigma, T) = N(d_1)S_0 - N(d_2)Ke^{-rT}$$

$$d_1 = \frac{1}{\sigma\sqrt{T}} \left[\ln\left(\frac{S}{K}\right) + T\left(r + \frac{\sigma^2}{2}\right) \right]$$

$$d_2 = d_1 - \sigma\sqrt{T},$$

where

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}z^2} dz$$

is the cumulative distribution of the standard normal distribution, and r is the interest rate in the market. A similar closed-form formula is available for put options.

In this case, the parametric assumption on the dynamic of the underlying asset price is such that there is a closed-form solution. The parameters S_0, K, r , and T are either available in the market or specified in the option contract, with the exception of the volatility σ . The option price C is a monotonically increasing function in σ . Given the market value of the option, we can invert the function to get the corresponding volatility, which we refer to as "implied volatility." Although the Black-Scholes model assumes constant volatility, the implied volatility is not constant with respect to strike price and maturity in the market. Most studies either interpolate or forecast the implied volatility as a function of strike price, maturity, or other available information in the market.

2.2 Local volatility and stochastic volatility models

There have been substantial improvements to the Black-Scholes model since their original work, for example local volatility and stochastic volatility models.

Dupire (1994) proposed the local volatility models, where σ is assumed to be a deterministic function of the underlying price. Specifically,

$$dS_t = \mu(t)S_t dt + \sigma(t, S_t) dt.$$

Although simple closed-form solutions are not always available, it is possible to calibrate the model with binomial-tree-based approaches or through Monte Carlo simulations. Dumas et al. (1998) assume a polynomial function $\sigma(t, S_t)$ on S_t as a way to avoid the potential overfitting. In Dumas' work, the authors identify an important phenomenon: that the local volatility models

are not always better than the original Black-Scholes in terms of hedging and predictive performance.

Heston (1993) introduces a diffusion process for the volatility itself:

$$dS_t = \mu S_t dt + \sigma_t S_t dB_t^1$$

$$d\sigma_t = a(b - \sigma_t) dt + \xi \sqrt{\sigma_t} dB_t^2$$

where B_t^1 and B_t^2 are independent Brownian motions and *a*, *b* and ξ are parameters to be estimated. Numerical integration methods and complex-number computation must be employed to get the solution. For some other stochastic volatility models such as the SABR model (Hagan et al.; 2003), we can only approximate the solution.

The local and stochastic volatility models are the most widely used in the industry because of their good performance. However, they are sometimes unstable and rely on complicated calibration methods.

2.3 Machine Learning and data-driven methods

In addition to the parametric models based on solving SDEs, there are a number of data-driven machine learning studies on option pricing.

Hutchinson et al. (1994) propose learning-network-based pricing and hedging methods. They compare the pricing performance of radial-basis-function networks, multilayer perceptron networks, and projection pursuit,¹ with ordinary least squares as a baseline comparison. They find that the learning networks could produce less hedging error compared to the Black-Scholes model when applied to S&P 500 data. Buehler et al. (2019) present a reinforcement learning approach to price and hedge a portfolio of derivatives under market frictions such as trading costs and liquidity. They train Feedforward Neural Network (FFN) and Long Short-Term Memory (LSTM)² models to find the optimal hedge ratio in simulated market data.

There are also previous attempts to approach the option pricing problem in a Bayesian framework. Tegnér and Roberts (2019) use a Gaussian Process prior with a squared exponential kernel on the implied volatility and then a Markov Chain Monte Carlo algorithm to fit market prices. The prediction of the VIX index is also studied employing the same set-up.

²FFN is the type of artificial neural network where the connections between the nodes do not form a cycle, whereas LSTM is a recurrent neural network with feedback connections which can process an entire sequence of data.



¹A multilayer perceptron network is one of the most popular and simplest types of neural networks, composed of multiple layers of perceptrons. It has activation functions in each neuron mapping weighted inputs to outputs. Radial basis function networks are neural networks with radial basis functions as activation functions. The projection pursuit is used to reveal structure in the original multi-dimensional data by offering orthogonal projections in lower dimension.

Deep learning methods only give point price estimates and are very data intensive. We hope to build on the previous Bayesian approach to option pricing and explore different model and kernel specifications to better exploit the underlying structure of the implied volatility functions. We hope to add to the rich literature on option pricing by proposing a relatively simple and intuitive method to nonparametrically calibrate the implied volatility function, while also accounting for model uncertainty.

3. A Bayesian nonparametric approach

3.1 Model

We use a Gaussian Process (GP) model, assuming an infinite-dimensional Gaussian distribution over the function space. The process we would like to model can be viewed as a collection of random variables, and any finite number of elements have a joint Gaussian distribution (Rasmussen and Williams; 2006). Denote the process by $f(\mathbf{x})$, the GP prior is fully characterized by the mean function $m(\mathbf{x})$ and covariance (kernel) function $k(\mathbf{x}, \mathbf{x}')$:

$$m(\mathbf{x}) = \mathbb{E}(f(\mathbf{x}))$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))].$$

And then we can write the Gaussian Process as

$$f(\mathbf{x}) \sim \mathscr{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')).$$

In our specific task of modeling the implied volatility function, the input is defined as $\mathbf{x}_i := [M_i \quad T_i] \in \mathcal{X}$, $i = 1 \dots n$. T_i is the time to maturity (in days) for option *i* in the dataset. M_i is the forward moneyness F_i/K_i , where F_i is the forward price of the underlying at maturity, and K_i is the strike price specified in the option contract. The volatility is a degree of variation and should be a positive function. Therefore, we model the logarithm of the implied volatility, instead of the implied volatility itself. Under the assumption that $\sigma(x) : \mathcal{X} \to \mathbb{R}$ is a continuous positive function, we model the log transform of implied volatility, $\ln \sigma$ with a Gaussian process:

$$\ln \sigma(\mathbf{x}) = f(\mathbf{x}) \qquad f(\mathbf{x}) \sim \mathscr{GP}(0, k(\mathbf{x}, \mathbf{x}')).$$

The mean function is assumed to be zero for simplicity, since we standardize the data.

We further assume that market prices for the *i*th option \mathbf{y}_i are noisy observations of the true underlying value:

$$\mathbf{y}_i = f(\mathbf{x}_i) + \boldsymbol{\varepsilon}_i \qquad \boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2),$$



where ε is assumed to be independent of $f(\mathbf{x})$ and each ε_i is independent and identically distributed.

3.2 Kernel functions

In GP models, the covariance is characterized by a kernel function $k(\mathbf{x}, \mathbf{x}')$. The kernel function is a way to evaluate the similarity across data points and is usually a symmetric function based on a distance metric. We compare the performance of four formulations of the kernel functions in order to explore which best capture the structure in implied volatility functions. The kernel functions are: Squared Exponential (SE), Rational Quadratic (RQ), Matérn 3/2, and Matérn 5/2 (Rasmussen and Williams; 2006). Let

$$r = \sqrt{\frac{(M - M')^2}{2l_M^2} + \frac{(T - T')^2}{2l_T^2}}.$$

The formulations of the kernels are

Squared Exponential Kernel:

$$k_{\rm SE}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{r^2}{2}\right)$$

Rational Quadratic Kernel:

$$k_{\rm RQ}(\mathbf{x}, \mathbf{x}') = \left(1 + \frac{r^2}{2\alpha}\right)^{-\alpha}$$

Matérn 3/2 Kernel:

$$k_{\text{Matern32}}(\mathbf{x}, \mathbf{x}') = \left(1 + \sqrt{3}r\right) \exp\left(-\sqrt{3}r\right)$$

Matérn 5/2 Kernel:

$$k_{\text{Matern52}}(\mathbf{x}, \mathbf{x}') = \left(1 + \sqrt{5}r + \frac{5}{3}r^2\right) \exp\left(-\sqrt{5}r\right)$$

Here we use the automatic relevance determination (ARD) version of all the kernels. The ARD version allows different length-scale *l*'s for the two dimensions. ARD kernels adjust the magnitude of l_T and l_M to automatically determine the relative importance of the two features. If a feature is less important, its length-scale will be adjusted to be larger to make the data points seems "farther" in the dimension and contribute less to the posterior.



3.2.1 Properties of the kernel functions

The SE kernel is one of the most widely used kernels, which is continuous and infinitely differentiable. The GP function process $f(\mathbf{x})$ with this kernel has mean square derivatives of all orders and is very smooth. The RQ kernel can be seen as an infinite sum of the squared exponential kernels with different length-scales. The parameter α is assumed to be greater than 0. The RQ kernel with parameters α and *l* converges to an SE kernel with length-scale *l* when $\alpha \rightarrow \infty$. The process characterized by the rational quadratic kernels is also infinitely mean square differentiable and very smooth, but it offers more flexibility than the SE kernel.

However, the strong smoothness assumption is sometimes unrealistic in the applications. Therefore, we also consider the Matérn kernels. The Matérn family of kernels is also widely used to model less smooth functions, with v generally equal to 3/2 or 5/2. The process will be *k* times differentiable if and only if v > k. When $v \to \infty$, the kernel again converges to the squared exponential kernel.

With the comparison among different kernel functions, we hope to exploit the structure in the implied volatility functions and gain accuracy.

3.2.2 Hyperparameters

Each kernel function above is parametrized by a few hyperparameter θ s. For example, in the RQ kernel, $\theta = \begin{bmatrix} \alpha & l_M & l_T \end{bmatrix}$. These hyperparameters are chosen by maximizing the marginal likelihood function:

$$\begin{split} \boldsymbol{\theta}^* &= \operatorname{argmax}_{\boldsymbol{\theta}} \log p(\mathbf{y} | \mathbf{X}, \boldsymbol{\theta}) \\ &= -\frac{1}{2} \mathbf{y}^\top (\mathbf{K}_{\mathbf{X}\mathbf{X}}(\boldsymbol{\theta}) + \sigma_{\varepsilon}^2 \mathbf{I}_n)^{-1} \mathbf{y} - \frac{1}{2} \log \left| \mathbf{K}_{\mathbf{X}\mathbf{X}}(\boldsymbol{\theta}) + \sigma_{\varepsilon}^2 \mathbf{I}_n \right| - \frac{n}{2} \log 2\pi, \end{split}$$

where **y** is the vector of training output, with the *i*th element being **y**_{*i*}, the training output for the *i*th option in the training set; **K**_{XX} is the covariance matrix with $(\mathbf{K}_{\mathbf{XX}})_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, and *n* is the number of training data. Due to the matrix inversion operation, the training process is of $\mathcal{O}(n^3)$. This cost dominates the difference caused by the number of hyperparameters for different kernel specifications, so the four kernels have similar efficiency in the estimation. If *n* is large, the model can take some time to run. Therefore, the GP models might not be suitable candidates in high-frequency settings.

3.3 Posterior distribution and inference

Given the GP prior, we can use the Bayes rule to get the posterior distribution.

Let \mathbf{X}, \mathbf{y} be the input and output in the training set. The *i*th row of \mathbf{X} is $\mathbf{x}_i = [M_i \quad T_i]$, the input for the *i*th option. $\widetilde{\mathbf{X}}$ is the matrix of testing input, with the *i*th row of $\widetilde{\mathbf{X}}$ equal to $\widetilde{\mathbf{x}}_i$. $\widetilde{\mathbf{f}}$ represents the function values at the testing location. In order to make predictions at the testing locations, we are concerned with the posterior distribution of the testing outputs given the data and testing input, $\widetilde{\mathbf{f}}|\widetilde{\mathbf{X}}, \mathbf{X}, \mathbf{y}$. According to Bayes' rule,

$$p(\widetilde{\mathbf{f}}|\widetilde{\mathbf{X}},\mathbf{X},\mathbf{y}) = \frac{p(\widetilde{\mathbf{f}},\mathbf{y}|\widetilde{\mathbf{X}},\mathbf{X})}{p(\mathbf{y}|\widetilde{\mathbf{X}},\mathbf{X})}.$$

The marginal distribution $\mathbf{y}|\mathbf{\tilde{X}}, \mathbf{X}$ follows from the prior assumption. The joint distribution of the training output values and the function values at the test locations $\mathbf{\tilde{f}}, \mathbf{y}$ under the prior is

$$\left[\begin{array}{c} y\\ \widetilde{f}\end{array}\right]\sim\mathcal{N}\left(0,\left[\begin{array}{cc} K_{XX}+\sigma_{\mathcal{E}}^{2}I & K_{X\widetilde{X}}\\ K_{\widetilde{X}X} & K_{\widetilde{X}\widetilde{X}}\end{array}\right]\right).$$

Using conditional probability on multivariate Gaussian distributions, we can arrive at the following posterior distribution (Rasmussen and Williams; 2006):

$$\begin{split} \widetilde{\mathbf{f}} | \widetilde{\mathbf{X}}, \mathbf{X}, \mathbf{y} &\sim \mathscr{N} \left(\widetilde{\mathbf{u}}, \operatorname{cov}(\widetilde{\mathbf{f}}) \right) \\ \widetilde{\mathbf{u}} &= \mathbf{K}_{\widetilde{\mathbf{X}}\mathbf{X}} \left(\mathbf{K}_{\mathbf{X}\mathbf{X}} + \sigma_{\varepsilon}^{2} \mathbf{I} \right)^{-1} \mathbf{y} \\ \operatorname{cov}(\widetilde{\mathbf{f}}) &= \mathbf{K}_{\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}} - \mathbf{K}_{\widetilde{\mathbf{X}}\mathbf{X}} \left(\mathbf{K}_{\mathbf{X}\mathbf{X}} + \sigma_{\varepsilon}^{2} \mathbf{I}_{N} \right)^{-1} \mathbf{K}_{\mathbf{X}\widetilde{\mathbf{X}}}. \end{split}$$

We will use the \tilde{u}_i , the *i*th element in the vector $\tilde{\mathbf{u}}$ and the expected value of the function value, as the prediction for the test input $\tilde{\mathbf{x}}_i$. Variance at the testing locations is given by the *i*th diagonal element of $\operatorname{cov}(\tilde{\mathbf{f}})$, \tilde{v}_i^2 .

Under the formulation that the log implied volatility follows a normal distribution, the implied volatility $\tilde{\sigma}$ follows a lognormal distribution. We then use the formula for the mean and variance of the lognormal distribution to recover the model prediction for the implied volatility:

$$\widetilde{\sigma}(\widetilde{\mathbf{x}}_i) = \exp\left(\widetilde{u}_i + \frac{\widetilde{v}_i^2}{2}\right)$$
$$\operatorname{Var}(\widetilde{\sigma}(\widetilde{\mathbf{x}}_i)) = [\exp(v_i^2) - 1] \exp(2\widetilde{u}_i + \widetilde{v}_i^2).$$

We can then give the prediction and a confidence region with the model. Both $\tilde{\mathbf{u}}$ and $\text{cov}(\tilde{\mathbf{f}})$ will change with kernel specifications, and we will compare the performance in the following section.



4. The implied volatility surface of the S&P 500 index options

4.1 Data

The S&P 500 index (SPX) option is one of the most liquid and most studied option markets. We apply our Bayesian nonparametric models to the S&P 500 index European options in 2018 to assess whether it produces accurate results with real market data.

We obtain the data from the WRDS OptionMetrics database,³ which contains the date of the contract, the strike price, maturity date, implied volatility, best ask and best bid of each option, as well as the forward price of the SPX index. We also retrieve the SPX spot prices and expected dividends. According to the OptionMetric reference manual, the forward price is calculated based on last closing security price, plus interest, less projected dividends.

The bid price is the maximum price a buyer is willing to pay for the option, and the best bid is the highest among all the bids. Similarly, the ask is the lowest the seller is willing to accept for the same option, and the best ask is the lowest among all. The mid point between the best bid and best ask, or the mid price, is a widely used reference price in the market. The option price and the implied volatility calculations are both based on this mid price.

We select the options with trading volume greater than 0 to eliminate options that were not actually traded. We also eliminate options with implied volatility smaller than or equal to 0 or greater than 1, to avoid the outliers distorting the results. For the purpose of this study, we define at-the-money (ATM) options as those with moneyness between 0.95 and 1.05. In-the-money (ITM) options have moneyness between 1.05 and 1.3 for call options, and between 0.7 and 0.95 for put options. Conversely, out-of-themoney (OTM) is defined as moneyness between 0.7 and 0.95 for calls, and 1.05 and 1.3 for puts. We limit our study to options with moneyness between 0.7 and 1.3. We do not concern ourselves with deep in-the-money or out-of-the-money options, as they are relatively illiquid and do not contain much information about the overall structure of the implied volatility function.

The study is mainly concerned with options between 20 and 365 days to expiration, although short-term options with maturity of 1-20 days, and longer-term options with maturity of 365-1095 days, were also included for comparison. There are both A.M. settled and P.M. settled options in the dataset. The difference is that the A.M. settled ones expire at the market open of the expiring trading day, and P.M. settled ones at the market close. The maturity is calculated as the number of days between the date of the contract

³https://wrds-www.wharton.upenn.edu/

and the option maturity date for P.M. settled options. If the option is A.M. settled, the maturity date is one less than the P.M. settled options.

Table 1 contains a summary of the data. The numbers are an aggregation of the 251 recorded trading days in 2018. The volatility smile phenomenon is apparent, as the implied volatility is higher for ITM and OTM options than for ATM ones. OTM options are on average cheaper in dollar terms, while the options with longer maturity are more expensive because of their time value. There are more puts in the dataset than calls, which might also influence the accuracy of the models.

Tabla 1

Data summary									
			total	ATM	ITM	OTM			
		total	322983	223954	19355	79674			
	count	1-20 days	102761	79985	4090	18686			
	count	20-365 days	206457	137011	12895	56551			
		365-1095 days	9204	3298	2223	3683			
		total	14.46%	13.32%	23.50%	15.46%			
	man implied velotility	1-20 days	16.63%	14.47%	36.11%	21.58%			
call	mean implied volatility	20-365 days	13.38%	12.62%	20.51%	13.59%			
		365-1095 days	14.79%	15.19%	17.68%	12.69%			
		total	45.22	39.56	270.97	6.28			
		1-20 days	29.24	25.84	227.07	0.49			
	mean price	20-365 days	48.01	44.67	268.54	5.81			
		365-1095 days	169.13	173.42	370.26	43.89			
		total	501699	222914	12590	266195			
	count	1-20 days	170789	83490	3540	83759			
		20-365 days	322493	137036	8182	177275			
		365-1095 days	8417	2388	868	19355 79674 4090 18686 12895 56551 2223 3683 23.50% 15.46% 36.11% 21.58% 20.51% 13.59% 17.68% 12.69% 270.97 6.28 227.07 0.49 268.54 5.81 370.26 43.89 12590 266195 3540 83759 8182 177275 868 5161 21.11% 27.08% 32.40% 34.83% 16.98% 23.64% 14.13% 19.32% 262.29 13.28 212.01 1.93 261.53 15.21 474.44 131.34			
		total	21.73%	15.37%	21.11%	27.08%			
	man implied velotility	1-20 days	25.90%	16.66%	32.40%	34.83%			
put	mean implied volatility	20-365 days	19.62%	14.58%	16.98%	23.64%			
1		365-1095 days	17.83%	15.94%	14.13%	19.32%			
		total	34.53	47.05	262.29	13.28			
	maan price	1-20 days	18.85	27.63	212.01	1.93			
	mean price	20-365 days	38.63	55.61	261.53	15.21			
		365-1095 days	195.94	234.35	474.44	131.34			

4.2 Setup

For each trading day in 2018, we repeat the model estimation. The data are split into a training set and testing sets for each trading day. The training set is composed of the option with strike prices divisible by 10, and the rest is reserved for testing. With this method, about 60% of all the options on a particular day are used for training. This choice of training data made the



training points more evenly spread out across the curve and made most predictions take the form of interpolation. We think that this setting resembles the situation faced in applications where we always have the data for some very liquid options spread across maturities and we want to use the information to price the other options.

The performance is evaluated by three metrics: root mean squared error (RMSE), mean valuation error (VE) and mean outside error (MOE). The RMSE is calculated by the square root of the squared distance between model-predicted implied volatility $\tilde{\sigma}$ and the real implied volatility σ in the dataset squared. The RMSE computes the error after the model fit is exponentiated back to get the implied volatility. We then calculate the option price with the predicted implied volatility using the Black-Scholes model and compute the absolute value of the difference between this model-predicted price and the mid-price given by the market data. To put the VE in relative terms, we also calculate the error to price ratio VE/P = $\frac{VE}{\text{Mid Price}}$. The mean outside error is the average valuation error outside the bid-ask spread. As defined by Dumas et al. (1998), the MOE is the predicted price minus the ask price if below. If the theoretical value is between bid and ask prices, the value is set to 0. This measure is used to detect if there is bias in the models in one direction.

4.3 Baseline

We also compare our Bayesian models with a relatively simple baseline, the best performing "ad-hoc" (AH) model used in Dumas' paper in 1998 (Dumas et al.; 1998). The model is a truncated quadratic function. The specific formulation is

 $\widetilde{\sigma}_{\text{Baseline}} = \max(0.01, \min(a_0 + a_1M + a_2T + a_3M^2 + a_4T^2 + a_5MT, 0.5)).$

Although simple, this model achieves quite accurate results. For each trading day in the dataset, we fit this baseline model together with our Bayesian models with the same training set and record the same metrics for the same testing set.

4.4 Cross-sectional interpolation

As mentioned above, we fit the GP model with the four kernels for each trading day in 2018, and compare the result together with the baseline model. Call options and put options are fitted separately. Table 2 reports the average of the performance metrics on the test set across the 251 days. The models are fitted with all the options with moneyness between 0.7 and 1.3 and maturity between 20-365 days.

4.4.1 Performance comparison of different kernel functions

GP models outperform the baseline uniformly, across kernel choices and performance metrics. The RMSE is at least 3 times smaller (SE kernel), and the best performing Matérn 3/2 kernel is almost 10 times better than the baseline. The valuation error is quite small for the GP models, with errors around 10 cents and about 1% of the option price. The models have better performance for puts than calls, probably because the put options are more liquid and there is more put data in the dataset. The best performing GP models have MOE around or less than 1 cent, which indicates that the models produce fairly accurate results with very small deviation outside the bid-ask spread. The GP models tend to undervalue calls and overvalue puts, while the baseline overvalues both.

In Table 2, the smallest errors are marked with bold font. Among all the GP models, the Matérn 3/2 kernel has the best performance. As described in Section 3.2.1, the Matérn 3/2 kernel is the least-smooth kernel among all the kernels tested. Although they generally look very smooth, there must be some non-smooth regions in the volatility surface that prevents us from getting the most accurate results with the SE kernel. The RQ kernel has a better result than the SE kernel, and better than the Matérn 5/2 kernel as well for calls. Although the RQ kernel is also very smooth, it has more flexibility than the SE kernel. The failure of the smoothness assumption might also be the reason why the baseline model performs worse than the GP models, because the baseline is a very smooth quadratic function.

			Table	2					
	Performance comparison for different kernels								
		SE	RQ	Matérn32	Matérn52	Baseline			
	RMSE	0.0037	0.0016	0.0013	0.0021	0.0124			
	VE	\$ 0.27	\$ 0.13	\$ 0.08	\$ 0.14	\$ 2.61			
aa11	VE/P	1.57%	1.04%	0.74%	0.99%	20.58%			
call	MOE	\$(0.0746)	\$(0.0162)	\$(0.0120)	\$ (0.0283)	\$ 0.2692			
	SD	0.0014	0.0011	0.0007	0.0010	-			
	RMSE	0.0020	0.0012	0.0008	0.0011	0.0116			
	VE	\$ 0.20	\$ 0.15	\$ 0.06	\$ 0.09	\$ 2.09			
	VE/P	1.04%	0.92%	0.33%	0.55%	11.16%			
put	MOE	\$(0.0055)	\$ 0.0013	\$ 0.0026	\$ 0.0018	\$ 0.2781			
	SD	0.0018	0.0022	0.0009	0.0012	-			

Therefore, the implied volatility functions may not be very smooth, and we need to use less smooth kernels to try to model them.

*Models fitted with options with moneyness 0.7-1.3 and maturity 20-365 days.

4.4.2 Performance comparison across moneyness

To discover how the models perform for options with different moneyness, we compute the same set of performance evaluation metrics for ATM, ITM and OTM options separately. Again, we use options with maturities from 20-365 days as examples. The results are presented in Table 3.

The estimation errors evaluated by the RMSE are lowest for ATM options across kernel specifications and models. This might be a result of more data points concentrated in the ATM region, as also demonstrated in Figure 1, panels (a) (calls) and (b) (puts). The OTM and ITM regions both have more spread-out training data.

Furthermore, we also need to take into account the fact that implied volatility is lowest for ATM options. ITM options have the highest RMSE and VE, but the lowest VE/P ratio because the ITM options are inherently more expensive, due to their intrinsic value. Evaluated by the VE/P ratio, the OTM options performed the worst for both calls and puts, and for both kernels. The Matérn 3/2 kernel is still the best-performing kernel across moneyness.

In addition, the GP models with all the kernels tend to undervalue ATM and ITM calls and overvalue OTM calls, with the reverse true for puts. This phenomenon might be caused by the assumption that the mean function in the GP prior is 0 for standardized data. In an implied volatility function, the ends of the curve are usually higher than the mean implied volatility. The prior has more weights on the prediction at the ends of the curve because less information is available from the data. As demonstrated in the visualization Figure 1, the ITM region is usually higher than the OTM region, but the model prediction might be pulled towards the center by the prior mean function. This suggests a potential improvement of the results by adjusting the prior mean function that can be explored in future studies.

In general, the GP models have relatively accurate performances across moneyness, with best performance in ATM and ITM options.

4.4.3 Results with shorter or longer maturities

Many studies find unstable behavior in their models when adjusting the maturities of the options, especially on the short end (Tegnér and Roberts; 2019). From previous results in this paper, the GP processes with the Matérn 3/2 kernel have the best results. Therefore, we test whether our GP models with this kernel are more stable for shorter (1-20 days) and longer (365-1095 days) maturity options. We also compare the results to the baseline model using the same training and testing set. Tables 4 and 5 display the results for short and long term options, respectively.

	Perfo	rmance	compariso	n across m	oneyness	
			total	ATM	OTM	ITM
		RMSE	0.0037	0.0011	0.0011	0.0161
		VE	\$ 0.27	\$ 0.22	\$ 0.05	\$ 2.45
	call	VE/P	1.57%	1.14%	3.44%	0.81%
		MOE	\$(0.0746)	\$(0.0065)	\$ 0.0018	\$(1.7998)
SE		SD	0.0014	0.0011	0.0011	0.0074
		RMSE	0.0020	0.0014	0.0018	0.0061
		VE	\$ 0.20	\$ 0.27	\$ 0.13	\$ 1.09
	put	VE/P	1.04%	0.76%	1.30%	0.44%
		MOE	\$(0.0055)	\$ 0.0062	\$(0.0194)	\$ 0.4626
		SD	0.0018	0.0013	0.0023	0.0033
		RMSE	0.0016	0.0007	0.0009	0.0058
		VE	\$ 0.13	\$ 0.12	\$ 0.04	\$ 0.79
	call	VE/P	1.04%	0.68%	2.56%	0.26%
		MOE	\$(0.0162)	\$(0.0035)	\$ 0.0058	\$(0.4131)
RO		SD	0.0011	0.0009	0.0009	0.0057
		RMSE	0.0012	0.0011	0.0009	0.0042
		VE	\$ 0.15	\$ 0.22	\$ 0.08	\$ 0.53
	put	VE/P	0.92%	0.75%	1.08%	0.23%
		MOE	\$ 0.0013	\$ 0.0045	\$(0.0026)	\$ 0.0933
		SD	0.0022	0.0015	0.0028	0.0025
		RMSE	0.0013	0.0004	0.0009	0.0048
		VE	\$ 0.08	\$ 0.07	\$ 0.03	\$ 0.61
	call	VE/P	0.74%	0.34%	2.48%	0.20%
		MOE	\$(0.0120)	\$(0.0045)	\$ 0.0028	\$(0.2622)
Matérn32		SD	0.0007	0.0005	0.0006	0.0054
		RMSE	0.0008	0.0006	0.0005	0.0042
		VE	\$ 0.06	\$ 0.07	\$ 0.04	\$ 0.47
	put	VE/P	0.33%	0.15%	0.49%	0.21%
		MOE	\$ 0.0026	\$ 0.0058	\$(0.0011)	\$ 0.1019
		SD	0.0009	0.0005	0.0012	0.0017
		RMSE	0.0021	0.0006	0.0009	0.0088
		VE	\$ 0.14	\$ 0.11	\$ 0.04	\$ 1.17
	call	VE/P	0.99%	0.60%	2.66%	0.37%
		MOE	\$(0.0283)	\$(0.0047)	\$ 0.0033	\$(0.7053)
Matérn52		SD	0.0010	0.0007	0.0008	0.0061
		RMSE	0.0011	0.0008	0.0009	0.0044
		VE	\$ 0.09	\$ 0.12	\$ 0.06	\$ 0.56
	put	VE/P	0.55%	0.36%	0.72%	0.24%
		MOE	\$ 0.0018	\$ 0.0104	\$(0.0077)	\$0.1435
		SD	0.0012	0.0007	0.0016	0.0022

 Table 3

 Performance comparison across moneyness

For short term options, the GP models still significantly outperform the baseline. The RMSE increases slightly for options across moneyness. Although valuation error is still relatively low and no more than a few cents, the VE/P is larger because short term options are cheaper. The GP model with the Matérn kernel performs well in terms of the MOE metrics. The small MOE number means that the valuation is generally not far from the bid-ask spread. For the ITM put options, all the predictions are within the bid-ask spread, although the predicted option price is not exactly the midprice. Since it is common for shorter term options to exhibit some anomalies, many researchers eliminate the shortest maturity. We think the results from our GP model can be a candidate model in estimating option prices with short maturities.

Performance for short term (1-20 days maturity) options							
			total	ATM	OTM	ITM	
Matérn32	call	RMSE VE VE/P MOE SD RMSE	0.0041 \$ 0.04 3.54% \$ 0.0001 0.0021 0.0041	0.0018 \$ 0.04 1.93% \$ 0.0000 0.0016	0.0052 \$ 0.01 17.87% \$ 0.0003 0.0038 0.0048	0.0166 \$ 0.20 0.10% \$(0.0014) 0.0174	
	put	VE VE/P MOE SD	\$ 0.0041 \$ 0.04 2.30% \$ 0.0000 0.0031	\$ 0.05 0.54% \$(0.0000) 0.0016	\$ 0.02 4.37% \$ 0.0001 0.0049	\$ 0.25 0.14% \$ - 0.0053	
Baseline	call	RMSE VE VE/P MOE	0.0237 \$ 1.21 41.81% \$ 0.2252	0.0187 \$ 1.42 38.52% \$ 0.2016	0.0269 \$ 0.14 82.20% \$ 0.0672	0.0761 \$ 2.03 1.03% \$ 0.2729	
	put	RMSE VE VE/P MOE	0.0372 \$ 1.30 28.33% \$ 0.3523	0.0243 \$ 2.06 25.02% \$ 0.7947	0.0446 \$ 0.44 33.38% \$(0.1807)	0.1101 \$ 2.23 1.20% \$(0.2067)	

 Table 4

 Performance for short term (1-20 days maturity) options

Similarly, the RMSE increases for long term options. While GP models still significantly outperform the baseline, both the RMSE and VE increase. However, due to the fact that options are inherently more expensive in dollar terms, the VE/P ratio is still low, generally less than 1%, with the exception of the OTM calls.

Thus, we can conclude that the GP models are relatively stable when we extend them to fit both shorter and longer term options.

Figure 1 Cross-sectional interpolation result with the Matérn 3/2 kernel.

The plots show the fitting result for implied volatility function on Jan 2, 2018 for four different maturities. The upper panels are for calls and the lower for puts. The models are fit on options with moneyness 0.7-1.3 and maturity 20-365 days. The red error bar shows the 95% confidence interval for the test output at the given point. Confidence intervals are wider where less training data are available or beyond the last training point.



4.4.4 Results with less training data

Moreover, we also test our GP models in the situation where less training data is available. In this case, instead of divisible by 10, we select options with strike prices divisible by 100 in the training set. The training set is still



		0		ť		<u> </u>
			total	ATM	OTM	ITM
Matérn32	call	RMSE VE VE/P MOE SD	0.0018 \$ 1.16 0.65% \$(0.2811) 0.0017	0.0009 \$ 0.82 0.46% \$(0.1337) 0.0011	0.0007 \$ 0.58 1.49% \$ 0.1120 0.0007	0.0027 \$ 2.28 0.61% \$(0.7657) 0.0034
	put	RMSE VE VE/P MOE SD	0.0014 \$ 0.98 0.72% \$ 0.0460 0.0013	0.0013 \$ 1.26 0.61% \$ 0.3284 0.0011	0.0012 \$ 0.86 0.83% \$(0.0863) 0.0014	0.0012 \$ 0.99 0.26% \$ 0.0484 0.0010
Baseline	call	RMSE VE VE/P MOE	0.0038 \$ 3.24 2.22% \$ 0.1353	0.0023 \$ 2.54 1.59% \$(0.6358)	0.0032 \$ 2.76 6.39% \$ 1.2736	0.0063 \$ 5.24 1.42% \$ 0.9713
	put	RMSE VE VE/P MOE	0.0016 \$ 1.38 0.91% \$ 0.0189	0.0014 \$ 1.51 0.68% \$ 0.3011	0.0014 \$ 1.21 1.03% \$(0.1853)	0.0025 \$ 2.44 0.68% \$ 0.4724

 Table 5

 Performance for long term (365-1095 day maturity) options

evenly spread out across the input domain, but now composed of about only 15% of the entire dataset. The results are presented in Table 6.

Again, the RMSE and VE both increase, and the VE/P ratio also increases. The GP models still outperform the baseline model with the same training set. The VE/P ratio is relatively low (around 2%) for the best-performing RQ and Matérn 3/2 kernel. The GP models generally undervalue both call and put options, which differs from the previous results with more training data. We think that this is possibly because the prior mean function has greater influence on the model outputs.

The RQ kernel has better performance on average than the Matérn 3/2 kernel for pricing call options. The performance of RQ kernel is also very close to the Matérn kernels for put options. It is possible that a smoother function would be better for interpolation in regions where less training data are available.

In contrast to some data-intensive machine learning methods, the GP models are able to get a relatively accurate result even with significantly less training data. This means that one could even think of employing GP models with illiquid markets or situations where market data are more difficult to obtain.

A Bayesian nonparametric approach to option pricing

	Performance with less training data								
		SE	RQ	Matérn32	Matérn52	Baseline			
call	RMSE VE VE/P MOE SD	0.0048 \$ 0.41 3.17% \$(0.0518) 0.0022	0.0029 \$ 0.27 2.45% \$(0.0070) 0.0017	0.0027 \$ 0.28 2.63% \$(0.0002) 0.0023	0.0032 \$ 0.31 2.75% \$(0.0143) 0.0021	0.0166 \$ 3.56 37.48% \$ 1.3371			
put	RMSE VE VE/P MOE SD	0.0036 \$ 0.39 2.62% \$(0.0225) 0.0030	0.0025 \$ 0.31 2.00% \$(0.0116) 0.0032	0.0023 \$ 0.28 1.70% \$(0.0116) 0.0033	0.0024 \$ 0.31 1.83% \$(0.0159) 0.0029	0.0154 \$ 2.85 15.51% \$ 1.0524			

Table 6Performance with less training data

*Models fitted with options with moneyness 0.7-1.3 and maturity 20-365 days.

4.4.5 Inference and confidence interval

The advantage of the Bayesian framework is the ability to provide statistical inference in addition to point estimates with the posterior distribution. The inference result could help with practical applications to quantify the uncertainty for a given testing input. The SD rows of Tables 2 - 6 present the standard deviations from the model fit.

In general, less accurate fitting results in larger RMSEs and have larger average SDs. The posterior uncertainty evaluation is built into the GP models because covariance kernels are based on the Euclidean distance metric. If there are fewer points adjacent to a given testing input, the model will naturally use less training points and a larger confidence interval ensues as a result. This phenomenon is also shown in Figure 1. Confidence intervals are very tight in the ATM region where training points are concentrated, and become wider in the ITM and OTM regions where the training data are more spaced out. When the testing point is outside the grid covered by training inputs and the prediction takes the form of extrapolation, the estimated variance also increases. For short-term options where the option price fluctuates more with the implied volatility, the confidence band differs more with the change in moneyness, compared to the longer term options, where the misspecification of the implied volatility has smaller impact on the price (Tables 4 and 5).

The inference results from the Bayesian framework provide a method to quantify uncertainty when calibrating implied volatility models, which is a rather unique advantage. The posterior uncertainty should be considered when pricing options with less known quotes. This can be implemented by taking into account information from more liquid options similar to them in moneyness and maturity, and with shorter maturities where the price changes more with the implied volatility.

To summarize, GP models entail good performance for implied volatility interpolation. Across kernel specifications, the Matérn 3/2 kernel performs best. The models also have relatively accurate performance for shorter- and longer- term options, as well as with less training data. The inference results can also help quantify uncertainty in the estimation.

4.5 One-Day Forecast

In addition to interpolating the implied volatility surface on a given day, we also test if the model can be stable to forecast the implied volatility of the next day. We calibrate a GP model to predict implied volatility with a one-day lag.

4.5.1 Setup

The model setup is similar to the cross-sectional interpolation task, except that the training set is composed of all the options on a particular date t, and the testing set is all the options on date t + 1. The samples are still limited to have maturity from 20-365 days, and moneyness 0.7-1.3. The GP model is calibrated with the Matérn 3/2 kernel, as it is the best-performing kernel in the cross-sectional interpolation. The results are also compared to the baseline truncated quadratic model on the same training and testing set. We repeat the calibration and testing procedure in a rolling fashion for all the trading days in 2018. Table 7 presents the average results across the 250 rolling estimations.

4.5.2 Discussion

The testing error increases significantly compared to the cross-sectional interpolation. The RMSEs are about 10 times larger, whereas the VEs are about 2 dollars for both call and put options. The VEs are more than 10% of the mid-price of the options. The VE is still less than 1.5% of the price for ITM options, which is much better than ATM and OTM options. The GP model performs better for calls than for puts, contrary to the cross-sectional case, even though we still have more put quotes in the dataset. The posterior standard deviation also increases, especially for puts, indicating a high level of uncertainty in the forecast. The GP models are still better than baseline models, but not to a significant degree.

The results indicate that the structure of implied volatility functions is not stable even through a short time horizon. As prediction errors are larger for put options, it is possible that the structure for puts fluctuates more. The instability and the difficulty it poses on local volatility models have been discussed in several previous works, such as Dumas et al. (1998). However, there are

A Bayesian nonparametric approach to option pricing

One-day-anead prediction results							
			total	ATM	OTM	ITM	
Matérn32	call	RMSE VE VE/P MOE SD	0.0108 \$ 2.21 13.15% \$ 0.0056 0.0015	0.0099 \$ 2.93 10.33% \$ (0.1200) 0.0010	0.0064 \$ 0.72 21.13% \$ 0.0488 0.0017	0.0224 \$ 3.29 1.36% \$ (0.2308) 0.0079	
	put	RMSE VE VE/P MOE SD	0.0126 \$ 2.44 10.62% \$(0.1581) 0.1951	0.0113 \$ 3.36 7.67% \$ (0.1274) 0.0758	0.0103 \$ 1.70 13.48% \$ (0.1748) 0.2765	0.0345 \$ 3.01 1.23% \$ 0.4241 0.0122	
Baseline	call	RMSE VE VE/P MOE	0.0199 \$ 3.58 30.23% \$ 0.1732	0.0147 \$ 4.32 20.40% \$ (0.3328)	0.0214 \$ 1.42 62.65% \$ 0.7009	0.0420 \$ 8.28 2.87% \$ 0.8081	
	put	RMSE VE VE/P MOE	0.0196 \$ 3.52 16.59% \$ 0.1670	0.0157 \$ 4.82 10.54% \$ 0.9333	0.0173 \$ 2.48 21.63% \$ (0.4510)	0.0682 \$ 7.64 2.82% \$ 3.8143	

 Table 7

 One-day-ahead prediction results

*Models fitted with options with moneyness 0.7-1.3 and maturity 20-365 days.

some potential ways to improve the prediction with GP models. For example, it is possible to include time as another input dimension to the GP model and use several past days as training data, in order to learn how the implied volatility functions evolve with time.

5. Conclusion

Bayesian Gaussian Process models are powerful alternatives to existing models for interpolating implied volatility surface and pricing options. In a cross-sectional interpolation setting, GP models yield rather accurate results. By changing the covariance kernel, we are able to exploit the structure of the implied volatility function and find the best calibration. From our results, the Matérn 3/2 kernel best fits the implied volatility function for the S&P 500 option in 2018. The models are also good candidates in fitting options with very short or relatively long time to maturity, or in a situation with less training data. However, the forecasting models still need some improvement. As a unique advantage of Bayesian models, we get posterior variance for the testing inputs, which would serve as a valuable reference for evaluating the uncertainty in model prediction and real-world decision making.

There are multiple interesting directions to conduct more in-depth studies for option pricing in a Bayesian framework. First of all, we could explore some time-series formulations to get a better forecast result, as forecasting could also be important in application. We can also explore some other kernel specifications in an attempt to better exploit the smoothness and structure of the implied volatility surface. In addition, there are other more complicated variants of the GP model, such as GP latent variable models. It would be interesting to check if the more complicated models produce a more accurate fit. We could also apply Bayesian models in portfolio hedging with the inference results characterizing the risk measures, and study their effects on the portfolio. We leave these for future work.

References

- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* **81**(3): 637–654.
- Branger, N. and Schlag, C. (2004). Model risk: A conceptual framework for risk measurement and hedging. Working Paper, EFMA Basel Meetings.
- Buehler, H., Gonon L., Teichmann J. and Wood, B. (2019). Deep hedging. Quantitative Finance 19(8): 1271–1291.
- Cohen, S. N. and Tegnér, M. (2018). European option pricing with stochastic volatility models under parameter uncertainty. Preprint. URL: https://arxiv.org/abs/1807.03882.
- Dumas, B., Fleming, J. and Whaley, R. E. (1998). Implied volatility functions: Empirical tests. *Journal of Finance* 53: 2059-2106, 1998.
- Dupire., B. (1994). Pricing with a smile. Risk 7(1): 18-20.
- Fan, J. and Mancini, L. (2009). Option pricing with model-guided nonparametric methods. *Journal of the American Statistical Association*, 104(488): 1351–1372.
- Gupta, A. and Reisinger, C. (2011). Optimal Bayesian hedging strategies in the context of model uncertainty. Preprint. URL: http://citeseer x.ist.psu.edu/viewdoc/summary?doi=10.1.1.258.8863.
- Hagan, P. Kumar, D., Lesniewski A. S. and Woodward, D. E. (2003). Managing smile risk. *WILMOTT magazine*.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**(2): 327–343.

- Hutchinson, J. M., Lo, A. W. and Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance* **49**(3): 851–889.
- Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance* **42**(2): 281–300.
- Jung, K. H., Kim, H. C., Lee, J. (2006). A novel learning network for option pricing with confidence interval information. In: Wang J., Yi Z., Zurada J.M., Lu BL., Yin H. (eds) *Advances in Neural Networks*: 491–497.
- Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(1): 141–183.
- Rasmussen, C. E. and Williams, C. K. (2006). Gaussian Processes for Machine Learning. MIT Press: Cambridge(2006).
- Tegnér, M. and Roberts, S. (2019). A probabilistic approach to nonparametric local volatility. Preprint. URL: https://arxiv.org/pdf/1901.0 6021.pdf.