A Black-Scholes-integrated Gaussian Process Model for American Option Pricing

Chiwan Kim

April 16, 2020

Honors Undergraduate Thesis

Faculty Supervisor: Prof. Simon Mak (Duke University)

Abstract

Acknowledging the lack of option pricing models that simultaneously have high prediction power, high computational efficiency, and interpretations that abide by financial principles, we suggest a Black-Scholes-integrated Gaussian process (BSGP) learning model that is capable of making accurate predictions backed with fundamental financial principles. Most data-driven models boast strong computational power at the expense of inferential results that can be explained with financial principles. Vice versa, most closed-form stochastic models (principle-driven) exhibit inferential results at the cost of computational efficiency. By integrating the Black-Scholes computed price for an equivalent European option into the mean function of the Gaussian process, we can design a learning model that emphasizes the strengths of both data-driven and principle-driven approaches. Using American (SPY) call and put option price data from 2019 May to June, we condition the Black-Scholes mean Gaussian Process prior with observed data to derive the posterior distribution that is used to predict American option prices. Not only does the proposed BSGP model provide accurate predictions, high computational efficiency, and interpretable results, but it also captures the discrepancy between a theoretical option price approximation derived by the Black-Scholes and predicted price from the BSGP model.

Keywords: Gaussian Process, American Options, Black-Scholes, Machine Learning, Option Pricing

1 Introduction

It is widely known that, with certain assumptions held, the theoretical price for a European option can be represented as the solution to a partial differential equation (PDE) system, known as the Black-Scholes equation. This PDE system can be solved in closed-form via the Black-Scholes formula (Black and Scholes, 1973), allowing us to theoretically map out the price behavior for European options when the risk-free rate and implied volatility is held constant. American options, however, are drastically harder to price than European options, because the former can be exercised at *any* time before its expiration date, whereas the latter must be exercised at its expiration date. Because of this variable exercise time, the corresponding PDE system for the theoretical price of an American option requires numerical approximation, which can be quite time-consuming, and is more complex. With options traded so often in financial markets today, solving the price of an American option remains a fundamentally challenging problem in finance (Mitchell et al., 2014). In this paper, we present a new Gaussian process learning model, called the Black-Scholes-integrated Gaussian Process (BSGP) model, for efficient pricing of American options. The proposed BSGP model integrates financial market data with financial principles from the Black-Scholes model to yield more accurate and interpretable predictions of American options compared to existing pricing models.

Existing American option pricing models can be roughly categorized into two categories: principle-driven and data-driven. Principle-driven models make use of an assumed PDE system to model the price of American options. This includes several notable extensions of the Black-Scholes equations for American options (Zhu, 2005; Wang et al., 2007; do Rosario Grossinho et al., 2018). While such model extensions provide valuable insights, these methods are limited to very specific option assumptions (e.g., non-dividend-paying American options), and ultimately cannot provide a general solution to price American options. One way to compensate for this limitation is via the Binomial Options Pricing Model (BOPR) proposed by Cox, Ross, Rubenstein (Cox et al., 2003), which performs a time-discretized approximation of the underlying PDE system for theoretical American option prices. Although BOPR is widely used in the industry, one key disadvantage of binomial options pricing is that it can be computationally expensive for precise approximations, and hence may not be appropriate when quick pricing decisions are needed. Furthermore, such models do not account for discrepancies between the theoretical pricing model and market prices, which is known to be present in practice (MacKenzie, 2006).

With exciting developments in machine learning in the past decade, there has been renewed interest in so-called "data-driven" models for American option pricing. Such models aim to use historical data on financial markets to train a predictive statistical model, then use the trained model to predict option prices. Notable works in this category include Johnson (Johnson, 1983), who applied an analytical approach to approximate non-dividend paying American put option prices by using weighted average prices of two European puts that are at the boundary prices, and Broadie and Detemple (Broadie and Detemple, 1996), who suggested the LUBA approximation that takes into account both the lower and the upper bounds of the American option price and reduces the computation speed comparable to a 50-step binomial tree while sustaining accuracy of a 1000-step binomial tree. Recently, Liu, Oosterlee, and Bohte (Liu et al., 2019) applied artificial neural networks (Cybenko, 1989) to train a data-driven model for American option pricing.

Data-driven models have several advantages over principle-driven models: the former addresses the need for efficient pricing in practical problems, and provides a data-driven approach to learn market price discrepancies. Despite this, data-driven models also have key limitations: the fitted models (e.g., from a neural network) are oftentimes complex and black-box, and fail to capture key financial principles which underlie options pricing. By neglecting such information, these data-driven models often violate fundamental pricing principles and lack interpretability, and are therefore difficult to implement in trading strategies where actual risks must be taken. Furthermore, by ignoring underlying financial principles, the data-trained pricing model may also yield poor predictive power especially in the presence of limited or unbalanced market data.

Clearly, on the spectrum of principle-driven models on one end and data-driven models on the other, pricing models which lie on either end of this spectrum may have serious limitations that hamper its applicability for practical American option pricing. Principle-driven models, while theoretically grounded and interpretable, can be computationally expensive and have notable discrepancies with market data. Data-driven models, while computationally efficient and trained using market data, can easily violate fundamental financial principles and may have poor predictive power given unbalanced data. To address this, we propose a new statistical learning model which trades-off between the two ends of this spectrum, by integrating financial principles within a Gaussian process (GP) model – a flexible Bayesian non-parametric model widely used in machine learning (Williams and Rasmussen, 2006) and engineering (Mak et al., 2018).

The proposed model, called the Black-Scholes-integrated Gaussian Process, integrates guiding principles from the Black-Scholes model within a data-driven GP learning model. The key idea is to embed the Black-Scholes equation (which solves the Black-Scholes PDE system for European options) as the mean function for the underlying stochastic process in the BSGP model. This "hybrid" model inherits the key advantages on both ends of the aforementioned principle-driven / data-driven spectrum. It contains much of the interpretability and insights from principle-driven pricing models, while also enjoying the computational efficiency and market adaptivity of data-driven models. Furthermore, the proposed BSGP model can also parse out the discrepancy between theoretically-approximated option prices and real market prices via the underlying data-driven GP model. We illustrate the effectiveness of the BSGP model in several pricing experiments on the Standard & Poor's 500 Exchange Traded Fund, showing that the proposed model outperforms standard principles-driven and data-driven pricing models in the literature.

This thesis is organized as follows. Section 2 provides an overview of financial options, the options market, and existing approaches for option pricing. Section 3 presents the proposed BSGP model, and discusses implementation details. Section 4 compares the BSGP pricing model with existing models for pricing American options on the SPY. Section 5 concludes with direction for future work.

2 Options and options pricing

We first provide a brief overview of European and American options, then present the Black-Scholes model for pricing European options, as well as existing pricing models for American options.

2.1 Call and Put Options

An option contract is a financial instrument whose value is derived from the value of another asset, hence called a financial derivative. The asset on which an option contract is based on is commonly referred to as the underlying asset. There are two common types of options: a call and a put. Call options give the buyer the right, but not the obligation, to buy a given quantity of the underlying asset at a given price (called the strike price K) on a particular date. Conversely, put options give the buyer the right, but not the obligation, to sell a given quantity of the underlying asset at a given price on a particular date. As options represent the opportunity but not the obligation to purchase/sell the underlying asset, one must pay a premium (or a price) to own such options, just like any other financial instrument such as stocks or gold. If the buyer of the option contract exercises the right to buy or sell as agreed upon, the seller is then obligated to fulfill the transaction.

To illustrate this visually, Figure 1 shows the payoff diagrams of both the buyer and the seller of a put and a call option. The diagrams show the payoff one receives across a range of possible underlying stock prices at option expiration for a fixed strike price K. Here, a buyer of the option would adopt a *long* position, and a seller would adopt a *short* position. For a long call, a buyer would have a positive payoff from the option only when the stock price exceeds the strike price (the buyer would exercise the option, and sell the stock in the market at a higher price); for a short call, a seller would have the opposite

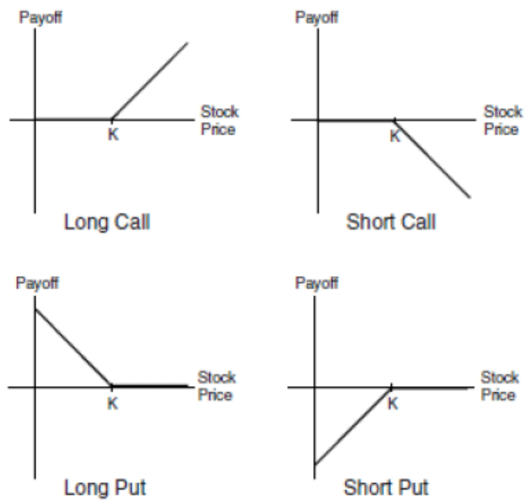


Figure 1: *Payoff diagrams of different positions for calls and puts*

payoff. Similarly, for a long put, the buyer would have a positive payoff from the option only when the strike price exceeds the stock price (the buyer would exercise the option, and buy the stock cheaper in the market); for a short put, vice versa. Note that the payoff diagrams in the figure disregard the price of the option; it shows only the payoff of the option at expiration, hence there are no negative payoffs for long positions and no positive payoffs for short positions.

2.2 European and American Options

We review two common types of options, European and American options. There are of course many more styles of options, e.g., Asian or Bermudan, but this paper will focus on European and American options, which are the most popular options used in financial markets. *European options* are options that may be exercised *only* at the expiration of the option contract, which are predetermined during the issuance of the contract. *American options*, on the other hand, are options that may be exercised at *any* trading day or before the expiration of the option contract. Note that an American option is always more valuable than its European option counterpart since the former perfectly entails the rights of a

European option and provides additional rights for potential early exercising.

For an option holder, there are two important questions to answer (a) "When is it optimal to exercise?" and (b) "Is it optimal to exercise?". Both questions are simple to answer for European options. For (a), a European option can only be exercised at expiration. For (b), a European option should be exercised only when it results in a positive payout. For call options, it will only be rational to exercise when the strike price is lower than the market value of the underlying asset, since one may buy the underlying asset at a sub-market level. Similarly, for put options, it will only be rational to exercise when the strike price is above the market value of the underlying, since one may sell the asset at a higher price than the market.

For American options, however, both questions become much harder to answer and it involves a much more complicated procedure. This is because they depend on the detail of the path that the underlying takes on its way to the expiry date, unlike European's which simply depends on the terminal value. The challenge of pricing American options can easily be understood when thinking in the perspective of the seller of the option who wants to put a price tag on the American option. It becomes much more difficult to hedge risks and anticipate future profit because you have no way to ensure when the buyer is going to exercise the option because the buyer can exercise the option at any time one sees fit. The proposed model addresses this inevitable challenge in pricing American options and takes a hybrid approach of the principle-driven and data-driven models, which can leverage market data within a learning model for efficient forecasting of American option prices.

2.3 The Importance of Pricing Options

Pricing options are extremely important for trading, investing, hedging risk and even structuring derivative products. For example, an investor who might be interested in buying a share of stock due to a predicted appreciation of the price, can easily put on the same view by buying a call option on the stock. This can be favorable as the call option price is generally much cheaper than the cost of actually buying the stock.

Another example can be companies that need to hedge unwanted exposure to price risk. Oil companies that sell oil to their clients are always exposed to price fluctuations of the market price of oil. If the market price for oil were to tank in the future, oil companies will have to take huge losses. One way to mitigate this risk is to buy put options on oil which will gain profit that can cancel out the potential loss of price depreciation of oil.

Options are used in many different ways and are utilized not only by sophisticated investors but also by retail investors as well. Just like any other financial instruments like stocks or bonds, it is extremely important to understand how to appropriately price these options as, at times, the market might be irrationally mispricing.

2.4 Pricing Models for Options

We now review the well-known Black-Scholes model (Black and Scholes, 1973), which is a widely-used theoretical model for pricing European options under certain assumptions. This model assumes the option price V follows the differential equation:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0 \tag{1}$$

The Black-Scholes Model is the closed-form solution to the above equation and can be written as:

$$Call = Se^{-qt}N(d_1) - Ke^{-rt}N(d_2)$$

$$Put = Ke^{-rt}N(-d_2) - Se^{-qt}N(-d_1)$$

where,

$$d_1 = \frac{\ln(\frac{S}{X}) + (r - q + \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}$$

$$d_2 = \frac{\ln(\frac{S}{X}) + (r - q - \frac{\sigma^2}{2})t}{\sigma\sqrt{t}} = d_1 - \sigma\sqrt{t}$$

Given the six Black-Scholes parameters, one can simply substitute the values in the equation to derive the price of a European option under conditions that implied volatility and the risk-free rate stay constant throughout the life of the option. For American options, however, the partial differential equation of the Black-Scholes Formula changes to inequality due to the variational optimal stopping problem of finding the time to execute the option as visualized in Figure 2; therefore, it becomes a much harder problem to solve. The inequality is shown in Equation (2).

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV \leq 0 \tag{2}$$

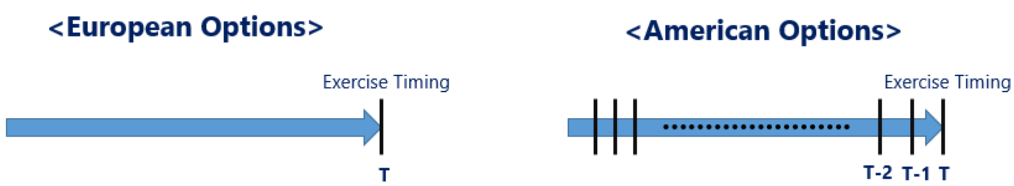


Figure 2: *European options vs American options*

2.5 Principle-Driven Approach vs Data-Driven Approach

Then how are American options currently priced? The first group of approaches can be summarized under the umbrella of "principle-driven" approaches as they are motivated by financial intuition and principle or simply stochastic calculus. Approaches that serve as extensions of the Black-Scholes model that attempt to provide a closed-form solution to the inequality above have all only done so by limiting the scope of American options to specific conditions such as assuming non-dividend paying stocks. Consequently, there is no generalized analytical model for American options. Works that fall under this category are Zhu (2005), Wang et al. (2007), and do Rosario Grossinho et al. (2018).

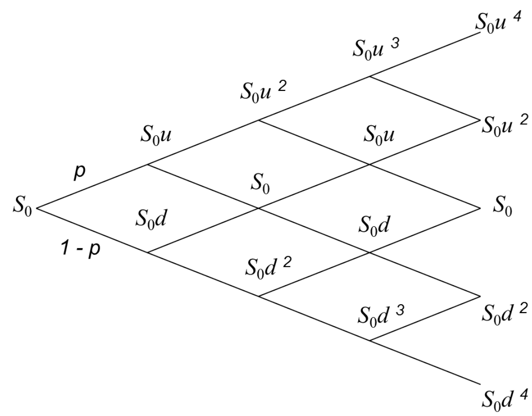


Figure 3: *Binomial Options Pricing Model*

A widely used principle-driven model is the Binomial Options Pricing Model (Cox et al., 2003) which was developed to help price options that the Black-Scholes Equation could not. This model, often referred to as CRR trees, makes use of binomial trees and follows a simple 3 step algorithm. Create the binomial price tree, find the option value at each final node, then find the option value at earlier nodes.

Using a binomial distribution to map out the potential prices of the option at each point in time, the CRR tree as shown in Figure 3 provides a discrete-time approximation to the solution of the partial differential equation of the American option price. However, for one to compute American option prices using this model, one must expand binomial trees which require the extensive calculation of $O(n^2)$, which can easily be computationally inefficient when using large values of n for accurate predictions. In this notation, n is denoted for the number of partitions for dividing the time horizon of our desired predictions. The more accurate results we want, the more partitions we need to construct the CRR tree, which will have computation inefficiency growing in a quadratic function. In practice, a value of $n > 5000$ can be common but also easily computationally heavy.

Not only do such principle-driven approaches fail to provide a generalizable closed-form, but are also computationally inefficient. In a world where time equals money, those who need to price options need a model that can efficiently compute prices, ideally generalizable to all American options.

As an alternative approach to the principle-driven approach and for purposes of improving computational efficiency, data-driven approaches are used in the industry as well. Recent work under this umbrella includes work done by Liu et al. (2019), who applied Artificial Neural Networks, a commonly used machine learning technique, to the American option pricing challenge. While this was successful in improving computational efficiency, it failed to provide intuition or interpretable results that can prove that the model was learning within a rational boundary of prices that abide fundamental option theory. In fact, such data-driven approaches are also called black-box models due to their opaque learning style. However, in the financial world, it is also important to keep a level of transparency in the model as it leads directly to credibility.

What if we wanted a good trade-off in the middle of these two approaches that somehow combines financial principles with market data while keeping computational efficiency, prediction accuracy, and interpretability high? Motivated by this question, we propose a new model to the American option pricing challenge and name it the ***Black-Scholes-integrated Gaussian Process***, or ***BSGP*** in short.

3 Black-Scholes-integrated Gaussian Process model

We use this general sampling model in order to predict the prices of American options:

$$\mathbf{Y_i} = f(\mathbf{x_i}) + \epsilon_\mathbf{i} \tag{3}$$

for: $\mathbf{Y_i}$ = the true price of American options

$\mathbf{x_i}$ = predictors for the American option (S, K, r, σ, t, d)

$f(\mathbf{x_i})$ = the expected price of the American option at predicted $\mathbf{x_i}$ values

$\epsilon_\mathbf{i}$ = the noise term, $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$

Our goal is to predict the unknown function $\mathbf{f}$ from option data and quantify the uncertainty of our predictions, where function **f** predicts the true price for an American option. As we do not know the true form of the function **f**, we are going to put a prior then condition it on the data to implement a Bayesian approach in estimating the posterior distribution.

The idea is to first assign a prior stochastic process to $f(\cdot)$, which captures our prior beliefs on the unknown function **f**. Then, we condition on observed values in the training data to obtain the posterior distribution Equation (11).

In selecting the prior stochastic process, we propose using a Gaussian process prior. A Gaussian process prior is a flexible Bayesian non-parametric model used in many applications of astrophysics (Kaufman et al., 2011), healthcare (Chen et al., 2019), and engineering (Mak et al., 2018). The BSGP model extends the applications to the domain of financial engineering.

3.1 Gaussian Process Modeling

Let us assume that $f(\cdot)$ follows a Gaussian process (GP) prior (Santner et al. 2003; Rasmussen & Williams 2006):

$$f(\cdot) \sim GP(\mu(\cdot), k(\cdot, \cdot)) \tag{4}$$

Mean Function:

$$\mu(x) = E[f(x)] \tag{5}$$

Covariance Function:

$$k(x_1, x_2) = Cov(f(x_1), f(x_2)) \tag{6}$$

A Gaussian process is completely specified by its mean function and covariance function, just as how a normal distribution is completely specified by its mean and variance. The mean function serves as the center for the function and the covariance function serves as a similarity metric that determines the shape of each function drawn in the learning process.

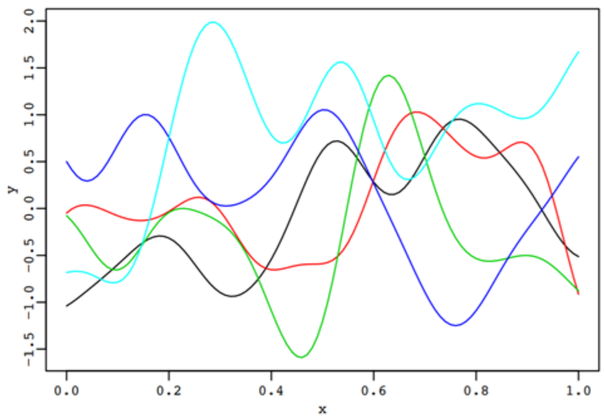


Figure 4: *Gaussian process prior of the unknown function* $\boldsymbol{f}$

Consequently, the variance will simply be $k(x, x)$ and can be computed easily for points that are equal. Essentially the prior will try to learn the unknown function **f** based on the two mean and covariance functions it is fed with as we can easily visualize with a 1 dimensional case as in Figure 4.

Once we set some prior, we are able to condition the prior with observed data. The observed data serves as points that the model has to pass with zero variance and will have higher variance at parts of the model where the input data points are further from the observed data points. The model is, therefore, conditioned on these data points. Since the Gaussian process models the distribution of x_i together with Y_i as a multivariate normal distribution, we are able to compute the conditional distribution as already well defined for multivariate normal distributions.

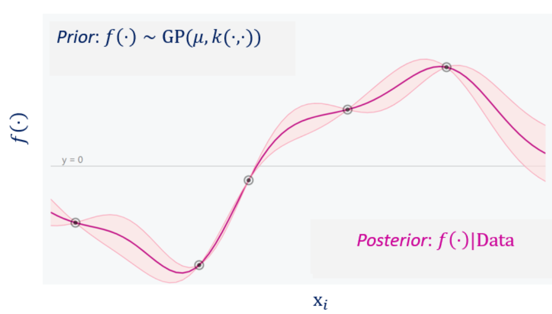


Figure 5: *Visualizing the Gaussian process model conditioned on observed data.*

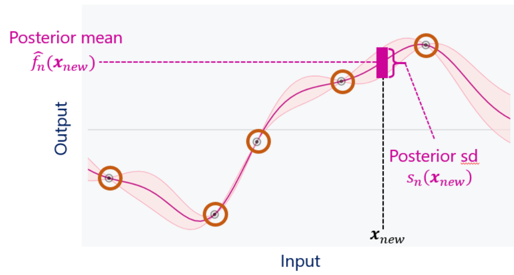


Figure 6: *Posterior distribution* $f(x_{new}|Data) \sim N(\hat{f}_n(x_{new}), {s_n}^2(x_{new}))$

Consequently, the posterior distribution will have a form that follows Equation (7).

$$f(x_{new}|Data) \sim N(\hat{f}_n(x_{new}), {s_n}^2(x_{new})) \tag{7}$$

with a mean and variance that follow Equations (8) & (9)

$$\hat{f}_n(x_{new}) = \mu_{x_{new}} + k(x_{new}, x_{tr})^T (K_{x_{tr},x_{tr}} + \sigma^2 I)^{-1}(y - \mu) \tag{8}$$

$${s_n}^2(x_{new}) = k(x_{new}, x_{new}) + \sigma^2 - k(x_{new}, x_{tr})^T (K_{x_{tr},x_{tr}} + \sigma^2 I)^{-1} k(x_{new}, x_{tr}) \tag{9}$$

Thus, allowing us to sample from the posterior distribution to make predictions as can be visualized in Figure 6.

In practice, generic choices of $\mu(\cdot)$ and $k(\cdot,\cdot)$ are chosen for standard Gaussian process models where the mean function $\mu(\cdot)$ is often a constant μ or simply even 0 unless there is strong evidence for a trend. A popular correlation function for $k(\cdot,\cdot)$ is the Gaussian correlation function:

$$k(x_1, x_2) = exp(-\sum_{l=1}^{d} \theta_l (x_{1,l} - x_{2,l})^2) \tag{10}$$

We will refer to this as the Standard Gaussian Process or simply SGP.

3.2 Model Specification of BSGP

If we were to use the SGP model to learn from the data and predict, we will be utilizing another purely data-driven approach. In order to avoid the limitations of purely data-driven approaches, the main purpose of the BSGP model is to integrate financial principles with the Gaussian process prior. This can be viewed as setting an informative prior which entails financial domain knowledge from the partial differential equation of the Black-Scholes. Thus, the BSGP is a new model that combines the data-driven model with the Black-Scholes.

Mathematically, we do this integration by modifying the mean function $\mu(\cdot)$ from the SGP. Instead of learning the mean function purely from the data, we want to reflect a prior domain knowledge through μ; therefore, our prior belief will be set to center on the Black-Scholes model.

Using the BSGP, we can fit historical American option data to predict future American option data using the posterior distribution:

$$[f_A(\cdot)|Data_A] \tag{11}$$

Our BSGP model will be specified in the following way:

$$Y_i(\cdot) = f(x_i) + \epsilon_i, \epsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2) \tag{12}$$

$$f(x_i) = BS(x_i) + \delta(x_i) \tag{13}$$

with a mean function equivalent to the Black-Scholes computed price of the option with same inputs parameters, i.e., $BS(x_i)$. The delta function is the discrepancy component of the model that follows a Gaussian Process:

$$\delta(x_i) \sim GP(0, K_A(x_1, x_2)) \tag{14}$$

Finally, the covariance function follows the Gaussian correlation with an extra γ^2 parameter that controls the effect of the covariance function on the shape of the overall

Gaussian process as shown in equation (15).

$$k_A(x_1, x_2) = \gamma^2 exp(-\sum_{l=1}^{d} \theta_l (x_{1,l} - x_{2,l})^2) \tag{15}$$

Notice that the unknown function has two components as we add the Black-Scholes computed price to the delta function that follows a standard Gaussian process. The model is structured this way so that we are able to isolate the delta function, which we will now be referring to the discrepancy function. With how the model is formulated, the Gaussian process essentially has a mean function of the Black-Scholes price only that we add the Black-Scholes price to the zero mean. This discrepancy term will effectively capture the difference between the theoretical European option price computed through the Black-Scholes and empirically observed price data. We later observe this discrepancy function separately to extract further insight.

3.3 Posterior Inference for BSGP

With hyperparameters, γ^2, σ^2, θ in our specification, we perform MCMC sampling of $[\gamma^2, \sigma^2, \theta_l | Data]$ using RStan (Team et al., 2016). By this sampling, we are able to use the posterior means of the samples to generate values $\hat{\gamma^2}$, $\hat{\sigma^2}$, $\hat{\theta}_l$. These average values are used for plug-ins for equations (12) & (15) into our Gaussian process predictions.

Priors for the hyper-parameters are set to weak priors in the form of inverse gamma distributions:

$$\theta_l, \gamma^2, \sigma^2 \sim InverseGamma(0.1, 0.1) \tag{16}$$

With only limited prior information on the options, the hyper-parameter values will differ drastically for different data sets; therefore, we chose to use weak priors instead of using strong priors for these hyper-parameters.

Computational efficiency in general for the BSGP model is $\mathcal{O}(N^3)$, which is the same as a standard Gaussian process (Belyaev et al., 2014). N denotes the number of input and output pairs in the data set. This is the case because we have only changed the mean function from the standard Gaussian process. Comparing this to $\mathcal{O}(n^2)$, the computational efficiency for CRR trees where n is the number of time partitions, we can easily see that the BSGP model has better computational efficiency as n increases and N decreases. A realistic value for n can easily exceed multiple thousands when N, the number of data points used in the regression, can be as small as a week's amount of data which allows the BSGP model to be more computationally efficient than the CRR tree.

4 Predictive Performance of the SPY

In order to assess the performance of the proposed BSGP model, we constantly compare the outputs with two other models, each representative of a purely principle-driven model and a purely data-driven model. With approaches on the opposite ends of the spectrum, we will see how the BSGP provides a middle ground trade off that carries the strengths of each approach while mitigating the limitations of each as well. We use the Black-Scholes model to represent a purely principle-driven model and the standard Gaussian process with a mean equivalent to the posterior mean that was sampled and learned along with the hyper-parameters for the Gaussian process to represent a purely data-driven model.

To evaluate the performance, we will look at prediction accuracy, interpretations, uncertainty of predictions, and insights from the discrepancy function. For visual examples, we use the outputs generated when using 1st week of 2019 June put option data as training to predict 2nd week of 2019 June put option prices. Other combinations of training and testing have consistent results.

4.1 SPY Options

The data used for analysis are price data on both call and put options on the Standard & Poor's 500 Exchange Traded Fund, or simply SPY, sourced from Wharton Research Data Services (WRDS, 2019). The SPY is simply an exchange traded fund that replicates the S&P 500 and options on the SPY are heavily traded with enormous trading volumes in a daily basis. The S&P 500 tracks the largest 500 companies in the United States and weight them by market capitalization. Options on the SPY are used to fulfill various purposes as they can serve as tools to express a view on the overall market, hedge risk or structure derivative products. SPY options are also well-suited for the analysis as the options are all American. We will be using 2019 May and June price data of SPY options for our analysis.

When combined, the data set has 859,658 rows and 31 columns with each row representing a particular option and each column representing a condition that characterizes the option contract. The columns that are used for analysis are limited to ***date*** (the date of the option when transaction was made), ***exdate*** (the expiration date of the option), ***time to exp*** (the time left until expiration), ***strike price*** (the strike price), ***mid price*** (the price of the option calculted by averaging the best bid and offer), ***volume*** (the trading volume of the option), ***forward price*** (the underlying price), ***interest rate*** (the risk-free rate), and ***dividend yield*** (the dividend earned as a percentage to the underlying price).

The data was scaled and normalized accordingly to match convention. For example, ***time to exp*** was transformed to units of years by dividing days by 250 (it is common practice to divide by the actual number of trading days in a year). Other transformations include scaling values of ***forward price***, ***strike price*** and ***time to exp*** so that the values are contained between 0 and 1 for regression conditions to meet later on in the analysis.

Options that had ***volume*** less than 10 were excluded from the analysis as we suspect such transactions to have special purposes or to be insignificant indicators that can potentially mislead the experiment.

4.2 Prediction Accuracy

In order to assess prediction accuracy of the BSGP model, we look at the Mean Squared Error (MSE) of our predictions across all three models throughout the testing period. Results are shown in Figure 7.

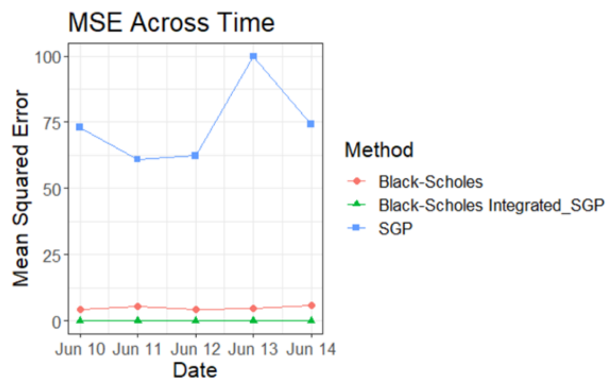
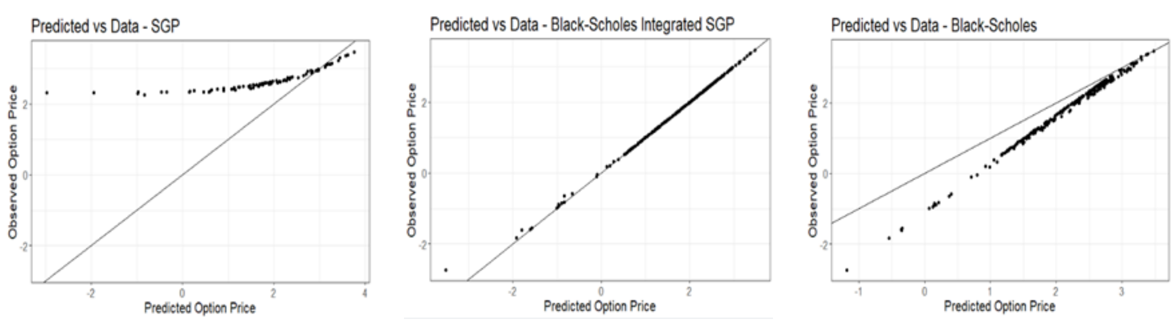


Figure 7: *MSE across testing dates 06/10/2019 to 06/14/2019*

The average MSE is computed and plotted for each testing day. We can easily see that the purely data-driven SGP model performs the worst among the three models with very high MSE values all across the testing period. More importantly, we can see that the BSGP model outperforms all other models consistently throughout the testing period. Taking a closer look, we can see how the BSGP model outperforms both models with noticeable MSE reductions in Figure 8.



(a) *SGP model* **(b)** *BSGP model* **(c)** *BS model*

Figure 8: *Plots of the predicted option price vs. the actual option price for the three pricing models.*

The three graphs above in Figure 8 have the observed option prices on the y-axis and predicted option prices on the x-axis, with the line $y = x$ plotted to show perfect prediction. Starting from the first plot, which is the output for the SGP model, we can see that there is a mixture of overpredictions and underpredictions without much of the dots plotted on the line. Ideally we would want the plotted dots to be on the $y = x$ slope. The MSE for the SGP model is 85.07085. On the opposite side, the purely principle-driven Black-Scholes model displays a consistently biased result as well with a MSE of 4.92335. The BSGP model, however, has a very low MSE value of 0.00262, and yields much better predictive performance (in terms of MSE) compared to both the SGP and the Black-Scholes model. Graphically, the superior prediction accuracy is shown by the plotted points well aligning with the $y = x$ slope.

4.3 Interpretations

As a limitation of data-driven approaches, we have pointed out earlier that a model has to be interpretable in order to stay transparent with its results and, more importantly, show that the model is learned under financial principles that govern. In order to check for such interpretability, we plot out contour plots of 2-dimensional slices of the predicted option price $\hat{f}$ for different pairs of the 6 input parameters. Below is an example of a 2-dimensional contour plot of strike price on the y-axis and the underlying price on the x-axis.

(a) *SGP model* **(b)** *BSGP model* **(c)** *BS model*

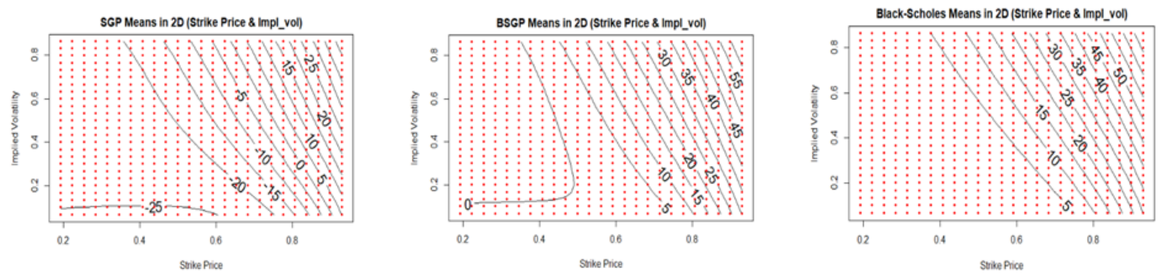
Figure 9: *Strike vs Underlying contours of the predicted option price $\hat{f}$ across all models*

The contour plots in Figure 9 are effective to see the influence each pair of parameters have on price when keeping all other 4 parameters constant. When plotting the contour plots, we keep the other 4 parameters that are not on the axes at its average value from the data set. Numbers on the line indicate the predicted option price. Note that strike price and the underlying price are scaled to fit between 0 and 1.

Starting from the purely data-driven model, we already notice negative predicted prices across the contour plots with the SGP model on the left. Clearly this is an indication that data-driven models cannot be used without financial intuitions as option prices cannot have negative values. In Figure 9c, we can see that the Black-Scholes contour plot shows a simple relationship between strike price and the underlying price. For put options, as used in this example, the value of the option increases as strike price increases. This is because the option gives the holder the right to sell the underlying at a higher price. Consequently, we are able to observe this relationship in the Black-Scholes contour plot where we see the lines increase from 5 to a max of 40 as strike price increases.

The BSGP model generates a contour plot that is quite similar to the Black-Scholes plot as it also depicts an increase in option price as strike price increases. This proves that the BSGP model, although learning from data, is abiding by the fundamental financial principles. Note that the plot is slightly different than the Black-Scholes plot with a max price of 35 instead of 40 and slightly different contour lines. This is reassuring as it shows the BSGP model is learning not only from the Black-Scholes mean but also from the observed data.

Another pair of 2-dimensional contour plots with implied volatility on the y-axis and strike price on the x-axis are shown in Figure 10.



(a) *SGP model* **(b)** *BSGP model* **(c)** *BS model*

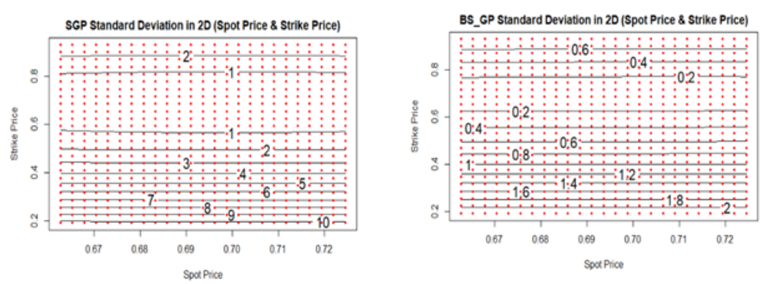
Figure 10: *Implied Volatility vs Strike contours for the predicted option price $\hat{f}$ across all models*

The contour plots show similar results in terms of proving that the BSGP model is abiding by financial principles. Once again, we see the SGP model in Figure 10a, showing negative predicted values for option prices, and the principle-driven model in Figure 10c, indicating that option prices rise as strike price and implied volatility increase, when all other parameters are kept constant. Just like how strike price has a positive relationship with option price, higher implied volatility is correlated with higher option prices as well. This is because there is more chance that the option will become "In-the-Money", a term used for option contracts that have the underlying asset price within the range where executing the option leads to positive profit. For example, a put option that has a strike price higher than the underlying price is "In-the-Money". Once again, the BSGP model in Figure 10b not only shows the interpretations present in the Black-Scholes model but also provides an extra layer of information with the contour that indicates a price of zero for puts. This extra contour layer provides information that none of the other approaches were able to produce. Specifically, in this case, we have learned from the BSGP model that the price of a put does not vary much within strike price ranges of [0.2,0.5] when implied volatility is very low. This arguably matches financial intuition as an increase of strike price from 0.2 to 0.5 will not matter for puts with implied volatility around 0.1 because it is unlikely that the price of the underlying asset to change enough for the option to becom "In-the-Money". As seen in this single example, the BSGP model is able to tell us more about how option prices move by providing more insight than other approaches.

4.4 Predictive Uncertainty

In any rational pricing problem, there are two components one must take into account: the predictions, or expected value, and uncertainty of the predictions, typically quantified in variance or standard deviation. Uncertainty plays a fundamental role in pricing (Klugman et al., 2012; Mak et al., 2016) as one cannot invest or trade with a strategy solely based on the expected value. Every return is associated with some risk component, which is why it is extremely important to be able quantify the uncertainty when making predictions.

To observe the uncertainty of our predictions, we plotted out the standard deviation across the same contour plots above in Figure 11. Therefore, the numbers on the contour lines are not the predicted prices, but the standard deviation of the predictions. Similar to the contour plots in Figure 9 & 10, the 4 parameters other than the strike price and underlying price are kept constant at their mean values.

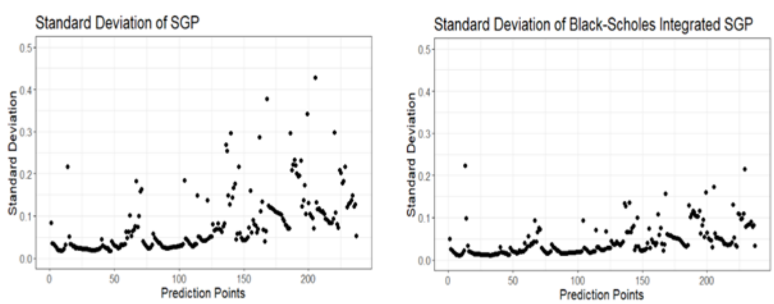


(a) *SGP model* **(b)** *BSGP model*

Figure 11: *Strike vs Underlying contours for the standard deviation of predictions $\hat{f}$ across all models*

Starting from the principle-driven approach, the Black-Scholes fail to plot any contour plots as it is a deterministic model that does not involve sampling. Having a standard deviation of zero is not a desired quality in predicting prices because we want a method that can quantify the inevitable risk associated with the predictions. The fact that we cannot quantify the associated risk is actually a critical disadvantage of the principle-driven approach. In Figure 11a, we have the data-driven approach with contour plots indicating that the uncertainty of the predictions increase as the options are further "In-the-Money" and "Out-the-Money", where "Out-the-Money" is the opposite concept of "In-the-Money" referring to option contracts with underlying prices at ranges where the option is completely worthless. The increase in prediction uncertainty as strike price deviates from the underlying price aligns well with intuition as one will expect that the further the strike price is, the more fluctuation of the option price there will be for a unit change in the underlying price. The increase in the value of the put with a strike price much further "In-the-Money" would be larger than the increase in the value of a put with a strike price much closer to the current price of the underlying. This results in more volatile predictions for such options which we can see even through our standard deviation contour plots. The BSGP model in Figure 11b shows the same relationship observed in the SGP model but with much lower uncertainty. In fact, we can see that the max standard deviation captured in the BSGP model is 2 whereas the max standard deviation for the SGP model is 10. Essentially we see the BSGP model providing much more certain predictions compared to the purely data-driven model. This is because the integration of both principle and data allows the model to learn with efficient constraints. As of result, the BSGP model serves better than both approaches for quantifying uncertainty.

A more explicit comparison of the variance in each model is shown in Figure 12



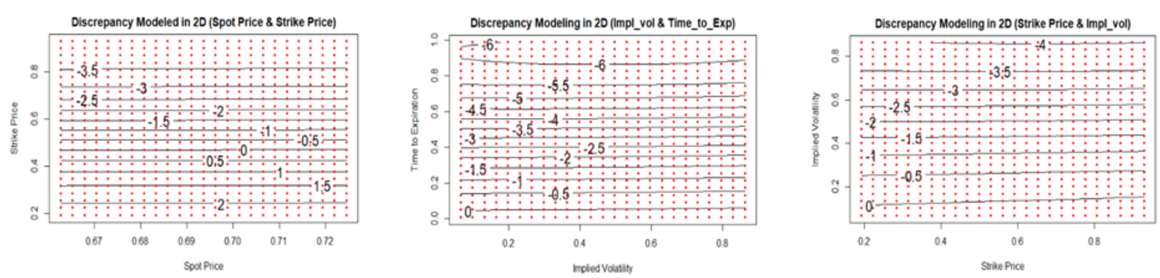
(a) *SGP model* **(b)** *BSGP model*

Figure 12: *Standard deviation of predicted option prices* $\hat{f}$

The average standard deviation for predictions for the data-driven model, the BSGP model, and the principle-driven model are 0.08607, 0.04442, and 0 accordingly. As these predictions do not fix certain parameters at their mean value as done in the contour plots in figures 9, 10, and 11, we see that the overall standard deviation in Figure 12 is significantly lower than values seen in the contour plots in Figure 11 for both the SGP and BSGP model. Regardless, the visualization in Figure 12 helps us compare the low standard deviation of the BSGP model ans how it lies in the middle of the spectrum between the two other approaches by providing a quantifiable yet decreased uncertainty for predictions.

4.5 Learning Market Discrepancies

Referring back to the BSGP model framework, equations (12) and (13), we are able to isolate the discrepancy between the Black-Scholes computed price $BS(\cdot)$ and the fitted option price $\hat{f}(\cdot)$. The discrepancy can be simply calculated by subtracting the Black-Scholes price from the predictions of the BSGP model. We then plot this estimated discrepancy $\hat{\delta}(\cdot)$ on the contour plots of different pairs of the 6 parameters in Figure 13. The contour plots in Figure 13 show the values of the expected discrepancy on the contour lines. These plots can point us to where the theoretical Black-Scholes price is under-predicting or over-predicting.



(a) *Strike vs Underlying* **(b)** *Time to Exp vs Impl Vol* **(c)** *Impl Vol vs Strike*

Figure 13: *Discrepancy function mapped on 2-dimensional contour plots*

Figure 13a shows contour plots of strike price on the y-axis and the underlying price on the x-axis where we can see that the Black-Scholes tend to overprice for puts that are further "In-the-Money" while underprcing generally for options that are "Out-the-Money". This can be an indication that the market is not realizing the value of being "In-the-Money" as much as theory suggests. Figure 13b has the scaled values of time to expiration on the y-axis and implied volatility on the x-axis. Here we see that the Black-Scholes predictions tend to overprice for options that have longer lives. This can be the case because the Black-Scholes price assumes a constant volatility and interest rate for its computation. With the two parameters assumed to be constant, it would make sense that the discrepancy would increase as predictions are made on options that have much higher time to expiration as the theory will deviate further from reality with such an aggressive assumption. Finally, Figure 13c has implied volatility on the y-axis and strike price on the x-axis where we see over predicted values of the Black-Scholes for options that have higher implied volatility. This is most likely because of the same reasons for the results showing in Figure 13b and the fact that the theoretical price is realizing volatility to be more influential on the option price than how the market actually does.

Many more insights and interpretations can be generated by plotting different pairs of parameters other than the ones used in Figure 13 across contour plots. Another key advantage of the BSGP model is clearly its ability to capture the discrepancy function and map out the conditions where theoretical price and predicted price differs the most.

Discrepancy learning becomes extremely useful for traders and portfolio managers who are facing the market at all times looking for mispriced opportunities. Once identifying these discrepancies, one can potentially put on an arbitrage strategy that can lead to profit with theoretically zero risk. In addition, learning the discrepancy is extremely useful for pinpointing the weaknesses in the current principle-driven models as we have already done briefly above. Understanding where the principle-driven models fail to explain option prices can guide us to how such models can be improved.

5 Conclusion

The BSGP model is a newly proposed model for pricing American options. American options are heavily used on a daily basis, yet we have not been able to come up with an effective model that can accurately predict American options while maintaining computational efficiency and interepretability. Consequently, the BSGP model serves to be a model that is capable of doing so, showing its superior performance in many metrics over both purely data-driven models and purely principle-driven models. The main advantages of using the BSGP model is summarized below.

- High prediction accuracy
- Quantifiable and low prediction uncertainty
- Interpretable results that show the model is learning within reasonable boundaries governed by financial principles
- Insights on price behavior that weren't provided by other models
- Explanation of the discrepancy of theoretical price and predicted price

With the strengths above, it can be said with high confidence that the BSGP model provides a practical and satisfactory solution for pricing American options.

Potential extensions of the model can include research in maximizing the discrepancy function, incorporating boundary information, utilizing data fusion and more. As mentioned earlier in section 4.5, one may be able to maximize the learned discrepancy to locate arbitrage opportunities. One might also be able to integrate more financial principles such as the boundary prices of American options (having a minimum price equal to a European option) using different correlation functions for the Gaussian process. In fact, if such boundary information is successfully integrated, the BSGP model will no longer have to limit its conditioning to American options as the posterior distribution can be generated with the prior conditioning on both American and European options (data fusion).

References

Belyaev, M., Burnaev, E., and Kapushev, Y. (2014). Exact inference for gaussian process regression in case of big data with the cartesian product structure. *arXiv preprint arXiv:1403.6573*.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Broadie, M. and Detemple, J. (1996). American option valuation: new bounds, approximations, and a comparison of existing methods. *The Review of Financial Studies*, 9(4):1211–1250.

Chen, J., Mak, S., Joseph, V. R., and Zhang, C. (2019). Function-on-function kriging, with applications to 3D printing of aortic tissues. *arXiv preprint arXiv:1910.01754*.

Cox, J., Ross, S., and Rubinstein, M. (2003). The cox—ross—rubinstein model. In *Mathematical Finance and Probability*, pages 201–219. Springer.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314.

do Rosario Grossinho, M., Kord, Y. F., Sevcovic, D., et al. (2018). Pricing american call options by the black-scholes equation with a nonlinear volatility function. Technical report.

Johnson, H. E. (1983). An analytic approximation for the american put price. *Journal of Financial and Quantitative Analysis*, 18(1):141–148.

Kaufman, C. G., Bingham, D., Habib, S., Heitmann, K., Frieman, J. A., et al. (2011). Efficient emulators of computer experiments using compactly supported correlation functions, with an application to cosmology. *The Annals of Applied Statistics*, 5(4):2470–2492.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: From Data to Decisions*, volume 715. John Wiley & Sons.

Liu, S., Oosterlee, C. W., and Bohte, S. M. (2019). Pricing options and computing implied volatilities using neural networks. *Risks*, 7(1):16.

MacKenzie, D. (2006). Is economics performative? Option theory and the construction of derivatives markets. *Journal of the History of Economic Thought*, 28(1):29–55.

Mak, S., Bingham, D., and Lu, Y. (2016). A regional compound Poisson process for hurricane and tropical storm damage. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(5):677–703.

Mak, S., Sung, C.-L., Wang, X., Yeh, S.-T., Chang, Y.-H., Joseph, V. R., Yang, V., and Wu, C. J. (2018). An efficient surrogate model for emulation and physics extraction of large eddy simulations. *Journal of the American Statistical Association*, 113(524):1443–1456.

Mitchell, D., Goodman, J., and Muthuraman, K. (2014). Boundary evolution equations for american options. *Mathematical Finance*, 24(3):505–532.

Team, S. D. et al. (2016). RStan: the R interface to Stan. *R package version*, 2(1).

Wang, I. R., Wan, J. W., and Forsyth, P. A. (2007). Robust numerical valuation of european and american options under the cgmy process. *Journal of Computational Finance*, 10(4):31.

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

WRDS (2019). Wharton research data services.

Zhu, S.-P. (2005). A closed-form exact solution for the value of american put and its optimal exercise boundary. In *Noise and Fluctuations in Econophysics and Finance*, volume 5848, pages 186–199. International Society for Optics and Photonics.