A STUDY ON RISK NEUTRAL DENSITY ESTIMATION USING SUPPORT VECTOR REGRESSION

Project report submitted for the degree of

B.Sc.(Hons.) Statistics and Data Analytics

by

Sahazadi Khatun

AU/2021/0006523
UG/05/BSTDA/2021/006

Under the supervision of
Dr. Arindam Kundu
Assistant Professor

Department of Mathematics
School of Basic and Applied Sciences
Adamas University
Kolkata-700126

June,2024

A STUDY ON RISK NEUTRAL DENSITY ESTIMATION USING SUPPORT VECTOR REGRESSION

Project report submitted for the degree of

B.Sc.(Hons.) Statistics and Data Analytics

by

Sahazadi Khatun

AU/2021/0006523
UG/05/BSTDA/2021/006

Under the supervision of
Dr. Arindam Kundu
Assistant Professor

Department of Mathematics
School of Basic and Applied Sciences
Adamas University,Kolkata-700126

June,2024
IS APPROVED FOR THE DEGREE OF
B.Sc.(Hons.) Statistics and Data Analytics

Supervisor

HOD, Mathematics

Dean, SOBAS

Dedicated To....

Certificate of Approval

Acknowledgement

I would like to express my deepest gratitude to my project supervisor, Dr. Arindam Kundu, Assistant Professor, Adamas University for his invaluable guidance, support, and encouragement throughout this project.

I also extend my thanks to my professors and mentors at Department of Mathematics for their insights and knowledge.

My heartfelt appreciation goes to my family and friends for their strong support and patience.

Finally, I would like to acknowledge the resources and facilities provided by Adamas University which were crucial in the completion of this project. Without the contributions of these individuals and institutions, this project would not have been possible.

Abstract

The accurate estimation of the risk-neutral density function (RNDF) is crucial for pricing financial derivatives and managing risk in financial markets. In this final-year project, we propose a novel approach utilizing Support Vector Regression (SVR) along with Quadratic Programming (QP) to estimate the RNDF from option prices. SVR offers a robust framework for modeling complex relationships between option prices and the underlying asset's risk-neutral distribution. By incorporating QP, we enhance the estimation process by optimizing the SVR model parameters to better fit the observed option prices while ensuring smoothness and adherence to no-arbitrage constraints [3]. Our methodology begins by collecting high-quality option price data and pre-processing it to remove noise and outliers. Next, we formulate the SVR model with appropriate kernel functions to capture the nonlinear relationships inherent in option pricing. The QP optimization step fine-tunes the SVR model parameters, balancing the trade-off between model complexity and fit to observed prices. We evaluate the performance of our approach using historical option data and compare the estimated RNDF with traditional methods. The results demonstrate the effectiveness of our proposed approach in accurately estimating the RNDF from option prices, providing valuable insights for pricing and risk management in financial markets. This research contributes to the advancement of quantitative finance by offering a reliable method for extracting implicit market expectations from option data.

Contents

Chapter 1

Introduction

Risk-neutral density (RND) is crucial in option pricing as it represents the probability distribution of an asset's future price under a risk-neutral measure, essential for valuing derivatives. By extracting RND from market prices of options, analysts can infer market expectations and implied volatilities. This aids in pricing and hedging options, assessing market sentiment, and managing risk. The RND can be recovered from cross sections of traded European option prices as it is proportional to the second derivative of the option prices for the strike.
Under the no-arbitrage assumption, asset pricing theory guarantees the existence of the RND. However, estimating the RND is often an underdetermined problem because the number of possible future prices of a security usually exceeds the number of observed prices. Therefore, additional constraints are needed to obtain a unique solution.

The estimation of the risk-neutral density (RND) from option prices has been a significant area of research due to the increasing availability of option databases and advancements in computational power. One of the foundational works in this area is by Sen Zhou(2018) [10], who provided a comprehensive method for estimating the RND function from option prices, highlighting the importance of accurate RND estimation for pricing and hedging derivatives (Zhou). Another pivotal study by Ian I-En Choo from Stanford University explored practical option pricing using Support Vector Regression (SVR) and Multiple Additive Regression Trees

(MART) [4], demonstrating the effectiveness of machine learning techniques in improving option pricing models (Choo).

Further research by Neetu Verma, Namita Srivastava, and Sujoy Das at Maulana Azad National Institute of Technology applied SVR for forecasting call option prices, showcasing the model's robustness in capturing market dynamics (Verma et al.) [7]. Negin Nayeri's study integrated ARIMA models with SVR to develop hybrid strategies for option pricing, addressing the limitations of traditional models by combining time series analysis with machine learning (Nayeri)[6]. J. Weston and colleagues investigated density estimation using Support Vector Machines, emphasizing its potential in accurately estimating the RND by leveraging statistical learning theories (Weston et al.) [8]. Pengbo Feng and Chuangyin Dang (2015) from City University of Hong Kong advanced this field by incorporating shape constraints in SVR for RND estimation, ensuring no-arbitrage conditions and improved accuracy (Feng and Dang)[3].

Additionally, Jianguo Zheng and co-authors utilized an SVR model optimized with a Bat Algorithm for stock index prediction, highlighting the versatility of SVR in different financial applications (Zheng et al.)[9]. In a notable study, Andreou, Charalambous, and Martzoukos employed SVR for European option pricing, presenting a novel application of SVR in the context of artificial neural networks (Andreou et al., 2009) [1]. Lastly, Du, Wang, and Du developed a general theory for inverting option prices to obtain implied risk-neutral probability density functions, applying their methodology to the natural gas market and demonstrating its broad applicability in quantitative finance (Du et al., 2012)[2].The incorporation of Bernstein polynomial basis for call option pricing with no-arbitrage inequality constraints further enhanced the precision and reliability of option pricing models (Arindam Kundu, Sumit Kumar, Nutan Kumar Tomar and Shiv Kumar Gupta)[5].

These studies collectively illustrate the evolution of RND estimation and option pricing methodologies, highlighting the increasing reliance on machine learning techniques like SVR to address the challenges of traditional parametric and non-parametric methods. Integrating these advanced techniques has led to more robust, accurate, and computationally efficient models, enhancing our understanding and capabilities in financial market analysis.

1.1 European Option Pricing and Related Terms

Understanding the fundamentals of option pricing requires familiarity with several key terms and concepts. Below is a detailed description of these terms and their significance, as outlined in prominent financial literature:

Asset An asset is any financial entity whose current price is known but is subject to fluctuation in the future. Examples include stocks, bonds, and currencies. These serve as the underlying elements in derivative contracts.

Derivative A derivative is a financial instrument whose value is contingent upon the price of an underlying asset at the time of expiration. Common examples include options and futures contracts. Derivatives are utilized for hedging risks or for speculative purposes in financial markets.

Writer The writer of an option is the entity that creates and sells the option contract. The writer is obligated to fulfill the contract's terms if the holder chooses to exercise the option.

Holder The holder is the purchaser of the option contract, possessing the right to exercise the option based on market conditions and personal preference. The holder is not obliged to exercise the option.

Strike Price (Exercise Price) The strike price is the predetermined price at which the underlying asset can be bought (in the case of a call option) or sold (in the case of a put option).

Expiration Date (Expiry Date) The expiration date is the specific date on which the option can be exercised. This date marks the end of the option's life.

Option An option is a contract between two parties, the writer and the holder, that grants the holder the right, but not the obligation, to buy or sell an underlying asset at a predetermined price (strike price) on a specified date (expiration date). Options can be classified into call options and put options.

Option Price (Premium) The option price, or premium, is the amount paid by the holder to the writer for the option contract at the inception of the agreement.

Call Option A call option grants the holder the right, but not the obligation, to purchase the underlying asset from the option writer at the strike price.

Put Option A put option grants the holder the right, but not the obligation, to sell the underlying asset to the option writer at the strike price.

European Options European options are a type of financial derivative that can only be exercised on the expiration date. This contrasts with American options, which can be exercised at any time before or on the expiration date. European option pricing involves determining the fair value of these options based on factors such as the spot price of the underlying asset, the strike price, the time to maturity, the risk-free interest rate, and the volatility of the asset's returns.

Types of European Options European Call Option: This option gives the holder the right to purchase the underlying asset at the strike price on the expiration date. European Put Option: This option gives the holder the right to sell the underlying asset at the strike price on the expiration date.

Fair Pricing of European Options One of the central questions in financial research is determining the fair price of European call and put options at the inception of the contract, prior to maturity. The Black-Scholes (BS) model, also known as the Black-Scholes-Merton (BSM) model, developed in 1973, provides a widely accepted method for calculating the fair price of an option.

These definitions provide a comprehensive foundation for understanding the complex mechanisms of option pricing and the factors influencing the valuation of European options. So, using the Black-Scholes formula we can determine the value of the European type call (C) option:

$$C = Se^{-\delta T} N(d_1) - Ke^{-rT} N(d_2) \tag{1.1}$$

where,

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + (r - \delta + 0.5\sigma^2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln\left(\frac{S}{K}\right) + (r - \delta - 0.5\sigma^2)T}{\sigma\sqrt{T}}$$

- S: price of underlying asset

- K: Strike Price
- σ: Volatility of the asset
- T: Time until expiration
- r: Interest rate
- δ: Dividend yield
- N: risk-adjusted probability

Similarly, the value of the European Put (P) option can be determined from the Black-Scholes formula:

$$P = Ke^{-rT}N(-d_2) - Se^{-\delta T}N(-d_1) \tag{1.2}$$

Thus, one can easily compute the price of the option using the BS model provided we know the market volatility of the underlying asset. But in general, the market volatility is unknown. Not only that, the BS model is derived with several assumptions. Among them, there are two main assumptions:

1. The asset price follows a geometric Brownian motion
2. Volatility is constant.

Volatility is the only parameter that cannot be observed directly, but we can extract the volatility implied from the market prices according to the BS model. The thesis is devoted to the pricing of European type call options only.

1.2 Risk-Neutral Density

The risk-neutral density function (RND) is a fundamental concept in mathematical finance and is heavily used in pricing financial derivatives. We give the definition as follows:

Definition: The risk-neutral density function for an underlying security is a probability density function for which the current price of the security is equal to the discounted expectation of its future prices. The price of a European call option of a stock is expressed as follows:

$$C(K) = e^{-rt} \int_{K}^{\infty} \max(0, S - K) f(S) \, dS \tag{1.3}$$

where f(·) is the RND, K is the strike price, S is the stock price at maturity, t is the time to maturity, r is the risk-free rate.

Differentiate the above equation with respect to the strike price K:

$$\frac{\partial C(K)}{\partial K} = (S - K) f(S) \mid_{\mathrm{S=K}} + e^{-rt} \int_{\infty}^{K} \frac{\partial (S - K) f(S)}{\partial K} \, dS \tag{1.4}$$

Differentiate it with respect to the strike price K again:

$$\frac{\partial^2 C(K)}{\partial K^2} = e^{-rt} f(S) \mid_{\mathrm{S=K}} = e^{-rt} f(K) \tag{1.5}$$

So the RND is :

$$f(K) = e^{rt} \frac{\partial^2 C(K)}{\partial K^2} \tag{1.6}$$

This relationship implies that if we have the option price function, differentiate it twice with respect to the strike price and multiply by the discounting factor, we will obtain the RND.

1.3 Constraints of the RND

1. The RND should be non-negative and integrate to 1, i.e;

$$f(K) \geq 0, \quad K \in [0, \infty) \tag{1.7}$$

$$\int_{0}^{\infty} f(S) \, dS = 1 \tag{1.8}$$

2. There should be no-arbitrage opportunities–

$$\max(0, S_0 e^{-\delta t} - K e^{-rt}) \leq C(K) \leq S_0 e^{-\delta t} \tag{1.9}$$

where K is the strike price, r is the risk-free rate, δ is the dividend yield rate, S_0 is the current stock price, t is the time to maturity.

1.4 Challenges in RND Estimation

Estimating the risk-neutral density (RND) from option prices is complex due to several key challenges:

1. **Data Limitations**
    - Option prices are available only at discrete and limited strike prices, often far from zero and infinity.
    - The tails of the RND, supported on $[0, +\infty)$, are difficult to estimate due to lack of data.
    - Curve fitting methods focus on interpolation between strikes, neglecting extrapolation beyond available data.
2. **Market Noise**
    - Option prices are noisy due to non-synchronous trading, discrete pricing, and bid-ask spreads.
    - Noise-sensitive methods, such as the maximum entropy method, become less attractive.
3. **No-Arbitrage Constraints**
    - The estimated RND must be positive and integrated into one to satisfy no-arbitrage conditions.
    - Ensuring these constraints can be difficult for methods like the expansion method.

4. **Curse of Differentiation**

   - Differentiation amplifies local irregularities, degrading the quality of the RND estimator compared to the option price estimator.
   - The estimation must balance accuracy and smoothness to mitigate this effect.

Chapter 2

Support Vector Regression

2.1 Introduction of SVR

Suppose we have a data set $\{(x_1, y_1), \ldots, (x_n, y_n)\} \subset X \times \mathbb{R}$, where X denotes the space of input patterns (e.g., $X = \mathbb{R}^d$). Our goal in ϵ-Support Vector Regression (SVR) is to find a function $f(x)$ that approximates the targets y_i within an ϵ deviation for all training data while maintaining minimal complexity. This is particularly relevant in contexts like predicting currency exchange rates using econometric indicators, where deviations beyond a certain threshold ϵ may be unacceptable. For simplicity, we begin with linear functions of the form:

$$f(x) = \langle w, x \rangle + b \quad \text{with} \quad w \in X,\, b \in \mathbb{R}$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product in X. To ensure flatness of f, we minimize the norm $\|w\|^2 = \langle w, w \rangle$. Thus, the optimization problem becomes:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|w\|^2 \\ \text{subject to} & y_i - \langle w, x_i \rangle - b \leq \epsilon \\ & \langle w, x_i \rangle + b - y_i \leq \epsilon \end{array}$$

This assumes the existence of a function f that approximates all (x_i, y_i) pairs within ϵ precision. However, this might not always be feasible. To address this, we introduce slack variables ξ_i and ξ_i^* to allow for some deviations, leading to the following formulation:

$$\begin{aligned} \text{minimize} \qquad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ \text{subject to} \qquad & y_i - \langle w, x_i\rangle - b \leq \epsilon + \xi_i \\ & \langle w, x_i\rangle + b - y_i \leq \epsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned} \tag{2.1}$$

Here, $C > 0$ determines the trade-off between flatness and the tolerance for deviations beyond ϵ. This leads to an ϵ-insensitive loss function defined as:

$$|\xi|_\epsilon = \begin{cases} 0 & \text{if } |\xi| \leq \epsilon \\ |\xi| - \epsilon & \text{otherwise} \end{cases} \tag{2.2}$$

This is the Support Vector (SV) expansion, where w is a linear combination of training patterns x_i. The complexity of f depends only on the number of support vectors, not the dimensionality of X. This property is crucial for extending SVR to nonlinear cases, as it allows the algorithm to be defined in terms of dot products between data points.

2.2 Kernel Tricks

Definition and Purpose

The kernel trick in SVM (Support Vector Machine) involves transforming non-linear data into a higher-dimensional feature space to make it linearly separable without explicitly computing the transformation.

Suppose we have a dataset $\{(x_1, y_1), \ldots, (x_n, y_n)\} \subset X \times \{1, -1\}$, where $X = (z_1, z_2) = \mathbb{R}^2$. As depicted in Figure 2.4(a), the optimal classification curve in this space is an ellipse:

$$w_1 z_1^2 + w_2 z_2^2 + b = 0 \tag{2.3}$$

To simplify the problem, we can map the data to a different feature space using a projection

$\phi : (z_1, z_2) \to (q_1, q_2)$:

$$(q_1, q_2) = \phi(z_1, z_2) = (z_1^2, z_2^2) \tag{2.4}$$

In this new feature space, the classification curve transforms into a line:

$$w_1 q_1 + w_2 q_2 + b = 0 \tag{2.5}$$

Mapping Function

A mapping function, such as Φ, is used to transform data from a lower-dimensional space (e.g., 2D) to a higher-dimensional space (e.g., 3D). This transformation helps in making complex data linearly separable. Mapping data explicitly to higher dimensions can be computationally expensive. Instead, we use kernel tricks to perform these mappings implicitly, leveraging inner products in the higher-dimensional feature space.

Consider a projection $\phi : \mathbb{R}^3 \to \mathbb{R}^9$:

$$\phi(Z) = \phi(z_1, z_2, z_3) = \begin{bmatrix} z_1 z_1 \\ z_1 z_2 \\ z_1 z_3 \\ z_2 z_1 \\ z_2 z_2 \\ z_2 z_3 \\ z_3 z_1 \\ z_3 z_2 \\ z_3 z_3 \end{bmatrix} \tag{2.6}$$

The inner product in $\mathbb{R}^9$ can be expressed as:

$$\langle \phi(Z), \phi(Y) \rangle = \phi(Z)^T \phi(Y) = \sum_{i,j=1}^{3} (z_i z_j)(y_i y_j) = \left(\sum_{i=1}^{3} z_i y_i \right)^2 = (Z^T Y)^2 \tag{2.7}$$

Thus, we define a kernel function:

$$K(Z, Y) := (Z^T Y)^2 = \phi(Z)^T \phi(Y) = \langle \phi(Z), \phi(Y) \rangle \tag{2.8}$$

The computation using the kernel function $K(Z, Y)$ takes $O(n)$ time, compared to the explicit computation of $\phi(Z)$ which takes $O(n^2)$, where n is the dimension of the input.

By focusing on the inner product in the feature space rather than the explicit mapping ϕ, we can simplify our computations using the kernel trick.

Characteristics of Kernel Functions

Kernel functions in machine learning, particularly in SVMs, possess several key characteristics:

- **Mercer's Condition**: Ensures the kernel is positive semi-definite.
- **Positive Definiteness**: Kernel values are positive except when inputs are equal.
- **Non-negativity**: Produces non-negative values for all inputs.
- **Symmetry**: Values are the same regardless of input order.
- **Reproducing Property**: Can reconstruct input data in the feature space.
- **Smoothness**: Produces smooth transformations of input data.
- **Complexity**: More complex kernels can lead to overfitting.

Selecting an appropriate kernel depends on the problem and data characteristics, significantly impacting algorithm performance.

Major Kernel Functions in SVMs

Linear Kernel

- **Definition**: $K(x, y) = x \cdot y$

- **Characteristics**: Simplest kernel; useful for linearly separable data.
- **Usage**: High-dimensional data where complex kernels may overfit.

Polynomial Kernel

- **Definition**: $K(x, y) = (x \cdot y + c)^d$
- **Characteristics**: Nonlinear; captures more complex relationships.
- **Usage**: Degree d determines nonlinearity; risk of overfitting with high d.

Gaussian (RBF) Kernel

- **Definition**: $K(x, y) = \exp\left(-\gamma \|x - y\|^2\right)$
- **Characteristics**: Nonlinear; captures complex relationships without explicit feature engineering.
- **Usage**: γ controls width and nonlinearity; challenging to select optimal γ.

Laplacian Kernel

- **Definition**: $K(x, y) = \exp\left(-\gamma \|x - y\|\right)$
- **Characteristics**: Nonlinear; robust to outliers; uses L1 norm.
- **Usage**: γ controls nonlinearity; robust but parameter tuning is crucial.

Selecting the right kernel function is critical for effective SVM performance, depending on the data and problem specifics.

Chapter 3

Quadratic Optimization: The Positive Definite Case

In this chapter, we focus on two prominent classes of quadratic optimization problems that are frequently encountered in engineering and computer science, particularly in fields such as computer vision:

Minimizing the quadratic function

$$f(x) = \frac{1}{2}x^T A x + x^T b \tag{3.1}$$

over all $x \in \mathbb{R}^n$ or subject to linear or affine constraints.

Minimizing the same function over the unit sphere, i.e., subject to

$\|x\| = 1$.

In both scenarios, A is a symmetric matrix. We aim to identify the necessary and sufficient conditions for f to achieve a global minimum.

Many problems in physics and engineering can be formulated as the minimization of an energy function, with or without constraints. It is a fundamental principle in mechanics that nature tends to minimize energy. Additionally, if a physical system is in a stable equilibrium state, the energy in that state is minimal. The simplest form of an energy function is a quadratic function.

Quadratic functions are conveniently represented as:

$$P(x) = \frac{1}{2}x^T Ax - x^T b \tag{3.2}$$

Here, A is a symmetric $n \times n$ matrix, and x and b are vectors in $\mathbb{R}^n$, viewed as column vectors. The inclusion of the factor $\frac{1}{2}$ in front of the quadratic term is for mathematical convenience and consistency with physical principles.

The crucial question is: under what conditions on A does $P(x)$ have a global minimum, ideally unique? We provide a comprehensive answer to this question in two parts:

Symmetric Positive Definite Case: If A is symmetric and positive definite, then $P(x)$ has a unique global minimum precisely when $Ax = b$. This implies that for the quadratic function to have a unique global minimum, A must be positive definite, meaning all its eigenvalues are strictly positive.

General Case: For the general case, we offer necessary and sufficient conditions in terms of the pseudoinverse of A. This will be discussed in the subsequent section.

To begin, we revisit the definition of positive definite matrices:

Definition 12.1: A symmetric positive definite matrix is a matrix whose eigenvalues are strictly positive, while a symmetric positive semidefinite matrix has nonnegative eigenvalues.

By considering the matrix form, the function

$$P(x) = \frac{1}{2}x^T Ax - x^T b$$

achieves its minimum when the gradient of $P(x)$ with respect to x vanishes. The gradient $\nabla P(x)$ is given by:

$$\nabla P(x) = Ax - b$$

Setting the gradient to zero for minimization, we get the linear system:

$$Ax = b$$

If A is positive definite, this system has a unique solution, ensuring a unique global minimum. The positive definiteness of A guarantees that the quadratic function $\frac{1}{2}x^T Ax$ is strictly convex, and hence, the optimization problem has a single, well-defined solution. This condition is critical for ensuring stability and uniqueness in various applications.

In summary, quadratic optimization problems with positive definite matrices A have unique global minima when $Ax = b$. This principle is pivotal in applications across engineering and computer science, providing a robust framework for solving a wide range of minimization problems effectively.

Chapter 4

Estimation of the RND

4.1 Real Dataset

Data Collection

- **Option Prices:** Collect historical option prices for a given underlying asset. Ensure the data includes a range of strike prices and maturities.

- **Underlying Asset Prices:** Gather historical prices for the underlying asset to provide context and additional features.

- **Interest Rates:** Obtain historical risk-free interest rates corresponding to the option maturity dates.

Dataset Summary

SNo.	Variable Name	Description
1.	Strike	The strike price of the option, representing the price at which the option can be exercised.
2.	VOL	The volume of options traded.
3.	Tau	The time to expiration, typically measured in days or trading days.
4.	UNDL_PRC (S)	The underlying asset price, denoted as S, at the time of the option pricing.
5.	L_BID	The last bid price for the option.
6.	L_ASK	The last ask price for the option.
7.	OIT	The open interest, indicating the total number of outstanding option contracts.
8.	HIGH	The highest price reached by the option during the trading day.
9.	LOW	The lowest price reached by the option during the trading day.
10.	OPEN	The opening price of the option at the start of the trading day.
11.	LAST	The last traded price of the option.
12.	Trade date	The date on which the option was traded.
13.	expiry date	The expiration date of the option, indicating when the option contract expires.

Data Preprocessing

- **Filtering:** Remove any incomplete or erroneous data points. Ensure the dataset only includes options with sufficient liquidity to avoid pricing anomalies.

- **Standardization:** Normalize the prices of options and the underlying asset to have a mean of zero and a standard deviation of one, if necessary.

- **Moneyness Calculation:** Calculate the moneyness of each option, defined as the ratio of the strike price to the underlying asset price.

Feature Engineering

- **Input Variables:** Create input features such as moneyness, time to maturity, and implied volatility.

- **Transformation:** Transform these variables if needed to improve the performance of the SVR model.

4.2 RND Estimation

4.2.1 Quadratic Programming SVR

To formulate the estimation of the risk-neutral density (RND) as an optimization problem within the framework of Quadratic Programming (QP) based on Support Vector Regression (SVR), we propose a scheme called QPSVR.

Let $\{(x_1, c_1), \ldots, (x_n, c_n)\}$ be the strike prices ($x_i \geq 0$) and the corresponding call option prices ($c_i \geq 0$) in the market for $i = 1, \ldots, n$. The objective is to estimate the RND $f(x)$ that best approximates these data points while ensuring that $f(x)$ remains as flat as possible. Assume the RND can be expressed as:

$$w = \sum_{i=1}^{n} \alpha_i \phi(x_i) \tag{4.1}$$

$$f(x) = w\phi(x) + b = \sum_{i=1}^{n} \alpha_i \langle \phi(x_i), \phi(x) \rangle + b = \sum_{i=1}^{n} \alpha_i K(x_i, x) + b \tag{4.2}$$

where $x \in [0, \infty)$ and $K(x_i, x)$ is a kernel function. To ensure the flatness of the RND, we use the following objective function:

$$\frac{1}{2}\|w\|^2 + \lambda \sum_{i=1}^{n} L(y_i, f(x_i)) \tag{4.3}$$

where $L(y_i, f(x_i))$ is the loss function describing how well $f(x)$ approximates the data points. Here, we use the ϵ-insensitive square loss function defined as:

$$|\xi_i|_\epsilon^2 := \begin{cases} 0, & \text{if } |\xi_i| \le \epsilon \\ (|\xi_i| - \epsilon)^2, & \text{otherwise} \end{cases}$$

Substituting this into the objective function gives:

$$\frac{1}{2}\|w\|^2 + \lambda \sum_{i=1}^{n} |y_i - f(x_i)|_\epsilon^2 \tag{4.4}$$

where y_i denotes the real RND. Since we do not have the real RND y_i directly, but instead have the market option prices c_i, we modify the objective function to:

$$\frac{1}{2}\|w\|^2 + \lambda \sum_{i=1}^{n} |c_i - C(x_i)|_\epsilon^2 \tag{4.5}$$

where

$$C(K) = e^{-rt} \int_K^\infty (S - K) f(S)\, dS \tag{4.6}$$

For a specified ϵ, incorporating no-arbitrage constraints and the RND constraints, the problem is formulated as:

$$\min_{b,\alpha_i,\xi_i,\xi_i^*} \frac{1}{2}\|w\|^2 + \lambda \sum_{i=1}^{n} (\xi_i^2 + (\xi_i^*)^2) \tag{4.7}$$

subject to:

$$
\begin{cases}
c_i - C(x_i) \leq \epsilon + \xi_i, & i = 1, \dots, n \\
C(x_i) - c_i \leq \epsilon + \xi_i^*, & i = 1, \dots, n \\
\xi_i, \xi_i^* \geq 0, & i = 1, \dots, n \\
f(K) \geq 0, & K \in [0, \infty) \\
\int_0^\infty f(S)\, dS = 1 \\
C(0) = S_0 e^{-\delta t}
\end{cases}
\tag{4.8}
$$

Given $c_i - C(x_i)$ is either non-positive or non-negative, only one of ξ_i, ξ_i^* will be non-zero. Thus, we minimize:

$$
\min_{b,\alpha_i,\xi_i} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) + \lambda \sum_{i=1}^{n} \xi_i^2
$$

subject to:

$$
\begin{cases}
-\epsilon - \xi_i \leq c_i - C(x_i) \leq \epsilon + \xi_i, & i = 1, \dots, n \\
f(K) \geq 0, & K \in [0, \infty) \\
\int_0^\infty f(S)\, dS = 1 \\
C(0) = S_0 e^{-\delta t}
\end{cases}
\tag{4.9}
$$

Using a quadratic objective function and a suitable kernel satisfying the Mercer Condition, we choose the Radial Basis Function (RBF) kernel:

$$
K(x_i, x) = \exp(-\gamma \|x - x_i\|^2) \tag{4.10}
$$

where $\gamma > 0$ is the scale parameter. The optimization problem has a global solution, and for the RBF kernel, we have:

$$
f(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x) = \sum_{i=1}^{n} \alpha_i \exp(-\gamma (x - x_i)^2) \tag{4.11}
$$

$$C(K) = e^{-rt}\sum_{i=1}^{n}\alpha_i\int_K^{\infty}(S-K)\exp(-\gamma(S-x_i)^2)\,dS \tag{4.12}$$

For the integration of the RBF kernel, we would need the Gauss Error Function which is defined as:

$$\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt, \quad x \geq 0$$

where $\operatorname{erf}(0) = 0$, $\operatorname{erf}(\infty) = 1$. Its integration is:

$$\begin{aligned}
\int_0^{\infty} K(x_i, S)\,dS &= \int_0^{\infty} e^{-\gamma(S-x_i)^2}\,dS \\
&= \int_{-x_i}^{\infty} e^{-\gamma t^2}\,dt \\
&= \int_{-x_i}^{0} e^{-\gamma t^2}\,dt + \int_0^{\infty} e^{-\gamma t^2}\,dt \\
&= \int_{-x_i}^{0} e^{-\gamma t^2}\,dt + \frac{1}{\sqrt{\gamma}}\int_0^{\infty} e^{-u^2}\,du \\
&= \frac{1}{\sqrt{\gamma}}\int_{-x_i\sqrt{\gamma}}^{0} e^{-u^2}\,du + \frac{1}{\sqrt{\gamma}}\sqrt{\frac{\pi}{2}} \\
&= \frac{1}{\sqrt{\gamma}}\sqrt{\frac{\pi}{2}}\operatorname{erf}(x_i\sqrt{\gamma}) + \frac{1}{\sqrt{\gamma}}\sqrt{\frac{\pi}{2}} \\
&= \frac{1}{\sqrt{\gamma}}\sqrt{\frac{\pi}{2}}(\operatorname{erf}(x_i\sqrt{\gamma}) + 1)
\end{aligned} \tag{4.13}$$

And its mean is:

$$\int_0^{\infty} SK(x_i, S)\,dS = \int_0^{\infty} Se^{-\gamma(S-x_i)^2}\,dS = \int_{-x_i}^{\infty}(t+x_i)e^{-\gamma t^2}\,dt = \int_{-x_i}^{\infty} te^{-\gamma t^2}\,dt + x_i\int_{-x_i}^{\infty} e^{-\gamma t^2}\,dt$$

$$= -\frac{1}{2\gamma}e^{-\gamma t^2}\Big|_{t=-x_i}^{\infty} + x_i\int_{-x_i}^{\infty} e^{-\gamma t^2}\,dt = \frac{1}{2\gamma}e^{-\gamma x_i^2} + x_i\sqrt{\frac{\pi}{\gamma}}\left(\frac{1}{2}\operatorname{erf}(x_i\sqrt{\gamma}) + 1\right) \tag{4.14}$$

$$C(0) = e^{-rt}\sum_{i=1}^{n}\alpha_i\int_0^{\infty} SK(x_i, S)\,dS = e^{-rt}\sum_{i=1}^{n}\alpha_i\left(\frac{1}{2\gamma}e^{-\gamma x_i^2} + \frac{x_i\sqrt{\pi}}{\sqrt{\gamma}}\frac{1}{2}(\operatorname{erf}(x_i\sqrt{\gamma}) + 1)\right)$$

The constraint on the RND f becomes:

$$f(K) \geq 0 \iff \sum_{i=1}^{n} \alpha_i e^{-\gamma(K-x_i)^2} \geq 0, \quad K \in [0,\infty)$$

$$\int_0^\infty f(S)\,dS = 1 \iff \int_0^\infty \sum_{i=1}^{n} \alpha_i K(x_i,S)\,dS = 1 \iff \sum_{i=1}^{n} \alpha_i \int_0^\infty K(x_i,S)\,dS = 1$$

$$\iff \sum_{i=1}^{n} \alpha_i \left(\frac{1}{\sqrt{\gamma}} \frac{\sqrt{\pi}}{2} (\operatorname{erf}(x_i\sqrt{\gamma}) + 1) \right) = 1 \tag{4.15}$$

and the constraint on $C(0)$ becomes:

$$C(0) = S_0 e^{-\delta t} \iff e^{-rt} \sum_{i=1}^{n} \alpha_i \left(\frac{1}{2\gamma} e^{-\gamma x_i^2} + \frac{x_i\sqrt{\pi}}{\sqrt{\gamma}} \frac{1}{2} (\operatorname{erf}(x_i\sqrt{\gamma}) + 1) \right) = S_0 e^{-\delta t} \tag{4.16}$$

The optimization problem becomes:

$$\min_{\alpha_i,\xi_i,i=1,\ldots,n} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j e^{-\gamma(x_i-x_j)^2} + \lambda \sum_{i=1}^{n} \xi_i^2$$

subject to

$$\begin{cases} -\epsilon - \xi_i \leq c_i - e^{-rt} \sum_{j=1}^{n} \alpha_i \int_{x_i}^{\infty} (S-x_i) e^{-\gamma(S-x_j)^2}\,dS, & i = 1,\ldots,n \\ c_i - e^{-rt} \sum_{j=1}^{n} \alpha_i \int_{x_i}^{\infty} (S-x_i) e^{-\gamma(S-x_j)^2}\,dS \leq \epsilon + \xi_i, & i = 1,\ldots,n \\ \sum_{i=1}^{n} \alpha_i e^{-\gamma(K-x_i)^2} \geq 0, & K \in [0,\infty) \\ \sum_{i=1}^{n} \alpha_i \left(\frac{1}{\sqrt{\gamma}} \frac{\sqrt{\pi}}{2} (\operatorname{erf}(x_i\sqrt{\gamma}) + 1) \right) = 1 \\ e^{-rt} \sum_{i=1}^{n} \alpha_i \left(\frac{1}{2\gamma} e^{-\gamma x_i^2} + \frac{x_i\sqrt{\pi}}{\sqrt{\gamma}} \frac{1}{2} (\operatorname{erf}(x_i\sqrt{\gamma}) + 1) \right) = S_0 e^{-\delta t} \end{cases} \tag{4.17}$$

The variables of the estimation problem $\alpha_i, \xi_i, i = 1,\ldots,n$. As we can see, the objective function is quadratic in terms of these variables, and the constraints are linear in terms of these variables.

Matrix formulation

The objective function can be written as:

$$u^T A u = u^T \begin{bmatrix} \frac{1}{2}M & 0 \\ 0 & \lambda I \end{bmatrix} u = \frac{1}{2} u_1^T M u_1 + \lambda u_2^T I u_2 = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j e^{-\gamma (x_i - x_j)^2} + \lambda \sum_{i=1}^{n} \xi_i^2$$

where $u \in \mathbb{R}^{2n}$ and $u^T = \{u_1^T, u_2^T\} = \{\alpha_1, \alpha_2, \dots, \alpha_n, \xi_1, \xi_2, \dots, \xi_n\}$, I is the $n \times n$ identity matrix, A is a $2n \times 2n$ matrix,

$$A = \begin{bmatrix} \frac{1}{2}M & 0 \\ 0 & \lambda I \end{bmatrix}$$

Recall $K(x_i, x) = e^{-\gamma (x - x_i)^2}, x \in [0, \infty)$. For distinct strike prices $\{x_1, \dots, x_n\}$, the Kernel Matrix M can be written as:

$$M = \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & K(x_1, x_3) & \dots & K(x_1, x_n) \\ K(x_2, x_1) & K(x_2, x_2) & K(x_2, x_3) & \dots & K(x_2, x_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K(x_n, x_1) & K(x_n, x_2) & K(x_n, x_3) & \dots & K(x_n, x_n) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & e^{-\gamma(x_1 - x_2)^2} & e^{-\gamma(x_1 - x_3)^2} & \dots & e^{-\gamma(x_1 - x_n)^2} \\ e^{-\gamma(x_2 - x_1)^2} & 1 & e^{-\gamma(x_2 - x_3)^2} & \dots & e^{-\gamma(x_2 - x_n)^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{-\gamma(x_n - x_1)^2} & e^{-\gamma(x_n - x_2)^2} & e^{-\gamma(x_n - x_3)^2} & \dots & 1 \end{bmatrix} \tag{4.18}$$

By theorem, M is strictly positive definite, i.e., for any vector $v_1 \in \mathbb{R}^n$,

$$v_1^T M v_1 \geq 0$$

The constraints can be written as:

1st Constraint

$$-\epsilon - \xi_i \leq c_i - e^{-rt} \sum_{j=1}^{n} \alpha_i \int_{x_i}^{\infty} (S - x_i) e^{-\gamma(S-x_j)^2} \, dS, i = 1, \ldots, n$$

implies,

$$-\epsilon - c_i \leq \xi_i - e^{-rt} \sum_{j=1}^{n} \alpha_i \int_{x_i}^{\infty} (S - x_i) e^{-\gamma(S-x_j)^2} \, dS, i = 1, \ldots, n$$

Now,

$$\int_{x_i}^{\infty} (S - x_i) e^{-\gamma(S-x_j)^2} \, dS$$

$$= \int_{x_i}^{\infty} S e^{-\gamma(S-x_j)^2} \, dS - x_i \int_{x_i}^{\infty} e^{-\gamma(S-x_j)^2} \, dS$$

Using equation (4.1 and 4.2):

$$= \frac{1}{2\gamma} e^{-\gamma(x_i-x_j)^2} + \frac{(x_i - x_j)\sqrt{\pi}}{\sqrt{\gamma}} \frac{1}{2} (\mathrm{erf}((x_i - x_j)\sqrt{\gamma}) + 1)$$

So the 1st constraint will be:

$$-\epsilon - c_i \leq \xi_i - e^{-rt} \sum_{j=1}^{n} \alpha_i (\frac{1}{2\gamma} e^{-\gamma(x_i-x_j)^2} + \frac{(x_i - x_j)\sqrt{\pi}}{\sqrt{\gamma}} \frac{1}{2} (\mathrm{erf}((x_i - x_j)\sqrt{\gamma}) + 1))$$

The constraint $GX \leq h$ can be written as follows, where G is an $n \times 2n$ matrix and X is a $2n \times 1$ matrix:

$$GX \leq h$$

Here,

$$G = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1,n} & -1 & 0 & \cdots & 0 \\ g_{21} & g_{22} & \cdots & g_{2,n} & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{n,n} & 0 & 0 & \cdots & -1 \end{pmatrix}, \quad X = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \\ \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}$$

and

$$g_{ij} = \frac{1}{2\gamma} e^{-\gamma(x_i - x_j)^2} + \frac{(x_i - x_j)\sqrt{\pi}}{\sqrt{\gamma}} \frac{1}{2} (\mathrm{erf}((x_i - x_j)\sqrt{\gamma}) + 1)$$

4.3 Result and Discussion

The results demonstrate the effectiveness of our proposed approach in accurately estimating the Risk-Neutral Density (RND) from option prices. This provides valuable insights for pricing and risk management in financial markets. Specifically, the application of Support Vector Regression (SVR) techniques yielded a high level of precision in capturing market expectations implied by option data. The graphical and numerical estimates results are computed using python programming and are compared with the observed method to validate the performance. Also, Figure-1 & Figure-2 show the relationship between Strike Price and Option Price for dataset-1 & dataset-2 respectively.

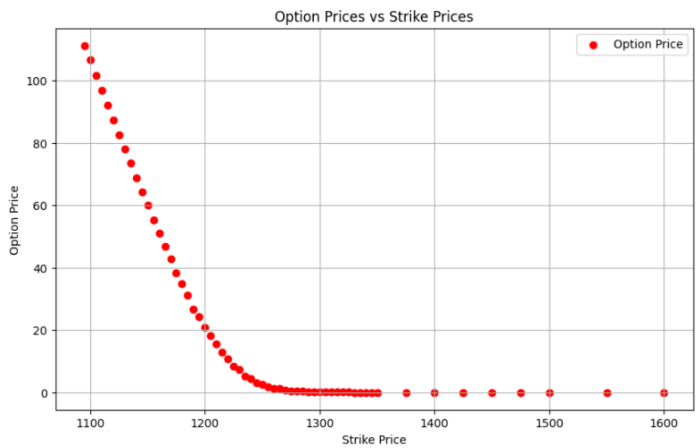

Figure 4.1: Scatter Plot of Observed Option Price from dataset-1

Figure 4.2: Scatter Plot of Observed Option Price from dataset-2

Figure-4.3 & Figure-4.4 depict the estimated arbitrage-free call price function for both of the datasets, which shows that our used estimator and QPSVR method match well with the mid-price of the observed bid-ask quotes.

Figure 4.3: Plot of observed option prices & estimated option prices obtained from support vectors regression for datase-1

Figure 4.4: Plot of observed option prices & estimated option prices obtained from support vectors regression for datase-2

After solving the optimization problem using the Python programming library "cvxpy" we get the numerical values of variables we estimate the RND function $f(x)$. Then plot of estimated RND w.r.t Strike Price for both datasets.

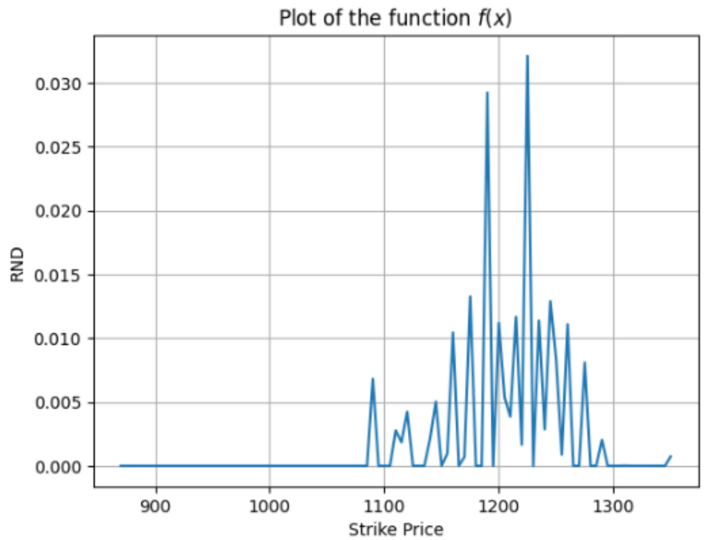

Figure 4.5: Plot of estimated RND w.r.t Strike Price for dataset-1

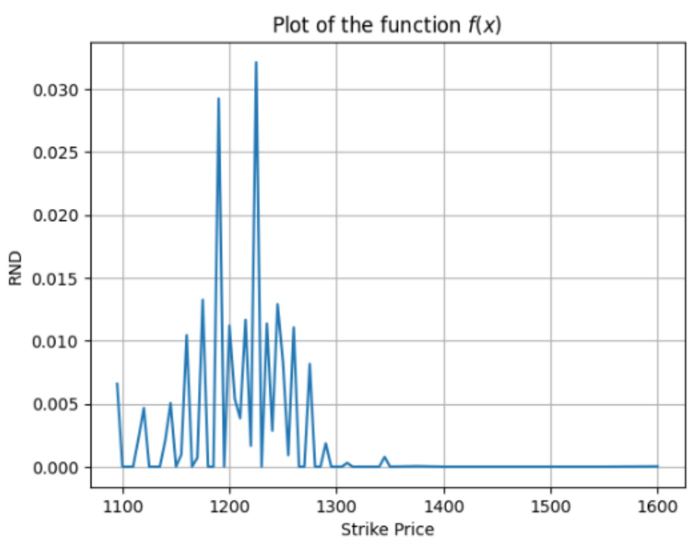

Figure 4.6: Plot of estimated RND w.r.t Strike Price for dataset-2

We also plotted the point-wise error b/w estimated option price and observed option price w.r.t strike price.

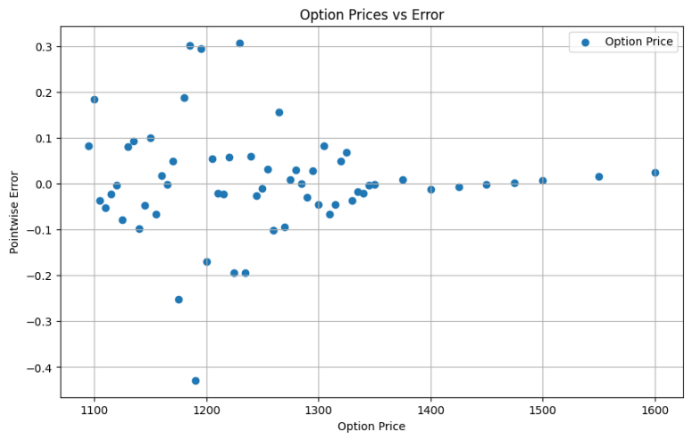

Figure 4.7: Plot of Point-wise Error w.r.t. Strike Prices for dataset-1

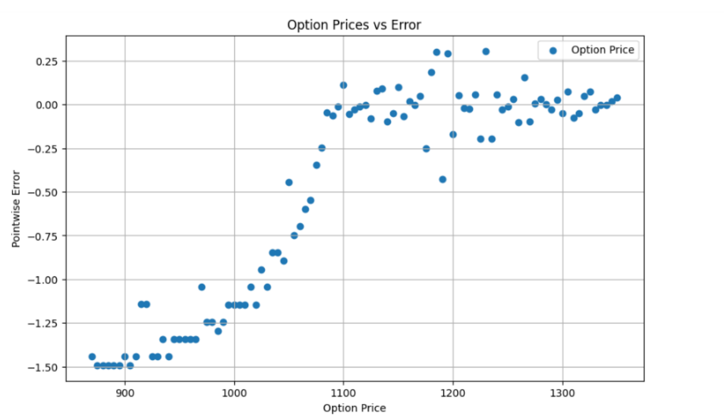

Figure 4.8: Plot of Point-wise Error w.r.t. Strike Prices for dataset-2

To evaluate the performance of the model, we can calculate the following metrics:

Model Evaluation

Regression

In regression tasks, the prediction/estimation is a continuous value. To measure the accuracy of the model, we can calculate the following metrics:

- **Mean Squared Error (MSE)** quantifies the average squared difference between the actual values (observed) and the estimated values produced by the model.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- **Mean Absolute Error (MAE)** is the average of the absolute differences between the actual and estimated values. It provides a linear measure of the average error.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- **Root Mean Squared Error (RMSE)** is the square root of the MSE. It provides an interpretable measure of the average error in the same units as the target variable.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

For those datasets the metrics values mention below:

	MSE	**MAE**	**RMSE**
Dataset-1 (Unequally Spaced)	0.0140	0.0765	0.1182
Dataset-2 (Equally Spaced)	0.6333	0.5514	0.7958

Chapter 5

Summary and Future Study

Summary

In this project, we delved into the complexities of estimating risk-neutral density (RND) from option prices and explored the critical role of kernel functions in Support Vector Machines (SVMs). From the result analysis, it is found that though the estimation prices have a good agreement with the observed prices but estimated RND is not smooth enough. However, our estimated RND satisfies all the no-arbitrage conditions. From this study, it is noticed that because of data restrictions and market noise, accurate RND estimation is important but difficult. The choice of kernel function has a major impact on model performance in SVR. Also, highlights the importance of careful selection and parameter adjustment.

Future Study

- **Hyperparameter Optimization:** Explore advanced hyperparameter optimization techniques to improve RND estimation accuracy.

- **RBF Kernel Exploration:** Conduct further investigations on using the Radial Basis Function (RBF) kernel for RND estimation.

- **Metaheuristic Approaches:** Utilize metaheuristic methods to identify optimal hyperparameters for Support Vector Regression (SVR).

- **Hybrid Models:** Develop and evaluate hybrid models combining SVR and other techniques for enhanced RND estimation.
- **Noise Reduction Techniques:** Implement and test advanced noise reduction methods to improve the quality of option price data used for RND estimation.
- **Real-time Data Integration:** Develop frameworks for integrating real-time market data to dynamically update RND estimations and adapt to changing market conditions.

Bibliography

[1] Panayiotis C Andreou, Chris Charalambous, and Spiros H Martzoukos. European option pricing by using the support vector regression approach. In *Artificial Neural Networks–ICANN 2009: 19th International Conference, Limassol, Cyprus, September 14-17, 2009, Proceedings, Part I 19*, pages 874–883. Springer, 2009.

[2] Yijun Du, Chen Wang, and Yibing Du. Inversion of option prices for implied risk-neutral probability density functions: general theory and its applications to the natural gas market. *Quantitative Finance*, 12(12):1877–1891, 2012.

[3] Pengbo Feng and Chuangyin Dang. Shape constrained risk-neutral density estimation by support vector regression. *Information Sciences*, 333:1–9, 2016.

[4] I Ian and En Choo. Practical option pricing with support vector regression and mart.

[5] Arindam Kundu, Sumit Kumar, Nutan Kumar Tomar, and Shiv Kumar Gupta. Call option price function in bernstein polynomial basis with no-arbitrage inequality constraints. *Journal of Inequalities and Applications*, 2016:1–16, 2016.

[6] Negin Nayeri. Option strategies using hybrid support vector regression-arima, 2020.

[7] Neetu Verma, Namita Srivastava, and Sujoy Das. Forecasting the price of call option using support vector regression. *IOSR Journal of Mathematics*, 10(6):38–43, 2014.

[8] J Weston, A Gammerman, MO Stitson, V Vapnik, V Vovk, and C Watkins. Density estimation using support vector machines. Technical report, Technical report, Royal Holloway College, Report number CSD-TR-97-23, 1997.

[9] Jianguo Zheng, Yilin Wang, Shihan Li, and Hancong Chen. The stock index prediction based on svr model with bat optimization algorithm. *Algorithms*, 14(10):299, 2021.

[10] Sen Zhou. *Estimation of the risk-neutral density function from option prices*. PhD thesis, Iowa State University, 2018.