GAUSSIAN PROCESS REGRESSION VS ARIMA IN THE ANALYSIS OF SHARE PRICES

Preprint · February 2025

DOI: 10.13140/RG.2.2.33795.41761

2 authors:

Robert Dalais
University of the Witwatersrand

Farai Mlambo
University of the Witwatersrand

GAUSSIAN PROCESS REGRESSION VS ARIMA IN THE ANALYSIS OF SHARE PRICES

Robert P. Dalais
School of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg, South Africa

Dr Farai F. Mlambo
School of Statistics and Actuarial Science & Wits Business School, University of the Witwatersrand, Johannesburg, South Africa

The analysis of financial time series, such as share price, has a wide range of practical applications in finance. This study compares the performance of Gaussian Process Regression (GPR) and Auto-Regressive Integrated Moving Average (ARIMA) models in forecasting the closing prices of Tesla stock. Five GPR models and six ARIMA models were evaluated using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The results indicate that while ARIMA models provided more accurate long-term forecasts, GPR models, particularly the Ornstein-Uhlenbeck and its combinations, performed better in short-term predictions. The Ornstein-Uhlenbeck X Rational Quadratic model achieved the highest short-term accuracy, while ARIMA(1,1,2) was the most accurate long-term model. The best overall model was the ARIMA(0,1,0) model. These findings suggest that GPR may be suitable for short-term financial predictions where the mean-reverting behaviour of the models did not compromise accuracy, though ARIMA remains the preferred method for long-term forecasting. The study highlights the trade-offs between model complexity and forecast accuracy and opens avenues for future research into non-mean-reverting kernels.

Key words: ARIMA, Gaussian process models, Financial times-series.

1. Introduction

Financial markets are complex and inherently dynamic systems influenced by a multitude of factors, including macroeconomic events, political developments, and investor sentiment. Forecasting stock prices plays a critical role in the decision-making processes of investors, portfolio managers, and financial institutions, as it provides valuable insights into future market movements and facilitates effective risk management. However, the volatile and stochastic nature of financial time series makes accurate forecasting an enduring challenge (Reddy, 2019a). Traditional time series models, such as the Auto-Regressive Integrated Moving Average (ARIMA), have been the preferred choice due to their simplicity, interpretability, and suitability for modelling linear trends. Yet, as financial environments evolve, these models struggle to capture non-linearity, market uncertainty, and abrupt shifts in patterns, necessitating more advanced forecasting approaches.

The challenge lies in achieving a balance between model complexity and forecasting accuracy. While ARIMA models offer reliable long-term predictions for stationary data, they perform poorly with non-stationary series and rapidly changing financial patterns (Jarrett and Kyper, 2011a). To address these limitations, recent research has explored the use of machine learning models, such as Gaussian Process Regression (GPR). GPR offers several advantages, including its non-parametric nature and the ability to generate probabilistic forecasts with uncertainty estimates. Despite its potential, the adoption of GPR in financial time series forecasting remains limited, and its robustness for long-term forecasts is still under investigation (Gonzálvez, Lezmi, Roncalli and Xu, 2019).

Existing literature highlights the strengths and limitations of both ARIMA and GPR models. ARIMA models are widely used for forecasting in global markets, such as the Chinese and Indian stock exchanges, owing to their reliability in long-term predictions (Jarrett and Kyper, 2011a; Reddy, 2019a). Meanwhile, GPR has demonstrated efficacy in other domains, such as spatial statistics and energy modelling, due to its flexibility in handling noisy and non-linear data (Ton, Flaxman, Sejdinovic and Bhatt, 2018a; Schulz, Speekenbrink and Krause, 2018a). However, in the context of financial forecasting, GPR's application remains underexplored, with most studies focusing narrowly on volatility estimation (Petelin, Šindelář, Přikryl and Kocijan, 2011a).

Several research gaps persist in the literature. First, there is limited empirical evidence comparing the performance of ARIMA and GPR models for financial time series forecasting. Second, while GPR's strength lies in uncertainty quantification, further exploration is needed to identify optimal kernel functions that improve short-term forecasting accuracy. Finally, the trade-offs between model complexity and forecasting performance are not sufficiently addressed in existing studies, despite their importance for practical financial applications.

The motivation for this study stems from the need to bridge these gaps. Financial time series forecasting is a crucial tool for calculating financial risk metrics, such as Value at Risk (VaR), which rely on accurate estimates and confidence intervals (Reddy, 2019b). Traditional models, including ARIMA, and machine learning approaches, such as artificial neural networks (ANNs), provide useful insights but often come with limitations. ANNs, for instance, lack explainability and struggle with uncertainty quantification, while ARIMA models are constrained by short-term forecasting horizons (Juan, Matutano and Valdecantos, 2023). GPR offers a promising alternative by providing flexible, probabilistic estimates across different time horizons, making it an ideal candidate for financial forecasting (Petelin et al., 2011a).

The primary aim of this study is to evaluate the applicability of GPR in forecasting the daily closing prices of Tesla Inc. (TSLA), listed on the NASDAQ, and to compare its performance against ARIMA models over a one-year period. The specific objectives are:

- **Determine the five most effective GPR models** and optimise their hyper-parameters by minimising the negative log-likelihood function.
- **Identify the six best ARIMA models** using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for optimal model selection.

- **Evaluate and compare the models** using performance metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).
- **Analyse short-term and long-term performance** by calculating rolling error measures, identifying the most suitable model for each forecasting horizon.
- **Provide practical recommendations for practitioners** and suggest future research directions, including the exploration of non-mean-reverting kernels and Bayesian approaches.

This study contributes to the financial forecasting literature by conducting a comprehensive comparison of GPR and ARIMA models, providing valuable insights into their relative strengths and limitations. The research offers practical guidance for selecting forecasting tools suited to various financial scenarios and explores the trade-offs between model complexity and forecast accuracy. Furthermore, it addresses the underutilisation of GPR in financial applications by identifying avenues for enhancing its predictive power through kernel selection and Bayesian frameworks.

The structure of this paper is as follows: Section 2 provides a detailed literature review, covering the theoretical foundations and previous research on ARIMA and GPR models. Section 3 presents the theoretical background of the methodologies used. Section 4 outlines the methodology employed for model development and evaluation. Section 5 discusses the results and insights from the comparative analysis. Finally, Section 6 concludes the study with recommendations for future research and practical applications.

Through this study, we aim to advance the understanding of financial time series forecasting by demonstrating the applicability of GPR models and providing new insights into the trade-offs involved in model selection. The findings will not only enrich the existing literature but also offer actionable insights for practitioners seeking to optimise their forecasting strategies in dynamic financial markets.

2. Literature review

Section 2.1 shows literature and theory relating to GPR. Section 2.2 displays the literature concerning ARIMA models. Section 2.3 covers literature that uses GPR for financial applications.

2.1 GPR

GPR is commonly used in the field of spatial statistics because it handles multivariate inputs easily which are used for 2 dimensional spatial statistics or 3 dimensional spatio-temporal statistics. GPR is known as "kriging" in spatial statistics and is one of the earliest methods used in spatial statistics (Ma and Kang, 2020). GPR also features heavily in machine learning research (Ton, Flaxman, Sejdinovic and Bhatt, 2018b).

Schulz, Speekenbrink and Krause (2018b) review the theory of GPR and the kernels and apply it to a few regression problems and time series. They also demonstrate the use of GPR for machine learning safe exploration tasks that require uncertainty quantification to learn functions while avoiding certain outcomes.

Ton et al. (2018b) examine the extension of GPR to a more general case that does not require the prior specification of a closed-form kernel. The paper also introduces a Fourier transform-based method for reducing the computational cost of GPR. The technique also allows for non-Gaussian kernels to be used - either via copulas or by training the kernel function directly from the data. This, however can induce heavy over-fitting which requires robust regularisation techniques to be applied.

2.2 ARIMA

Reddy (2019b) demonstrates the process of fitting and analyzing financial stock market prices using the Auto-Regressive Integrated Moving Average (ARIMA) method. This paper examines the rigorous use of the augmented Dickey-Fuller test to determine mean and trend stationarity under different lag conditions. Jarrett and Kyper (2011b) use ARIMA to model the share prices of Chinese stocks using an intervention model. The theory and methods used align with the theory shown by Reddy (2019b).

2.3 Use in financial time series

GPR is not widely used in the analysis of financial time series Petelin, Šindelář, Přikryl and Kocijan (2011b) compare Gaussian process regression (GPR) to Bayesian vector auto-regression to examine the application of GPR to financial time series. The paper concludes that GPR can be applied to the estimation of financial time series.

Chapados and Bengio (2007) use GPR to estimate financial time series data in the form of contract spreads. Their paper concludes that GPR offers no extra accuracy over other methods but is useful in applications requiring variance predictions.

Gonzalvez, Lezmi, Roncalli and Xu (2019) use GPR to estimate interest rates, fitting yield curves, and trend following strategies. The paper conclude that GPR is useful for yield curve estimation. is equivalent to other available methods for interest rate estimation. For trend following stratagies GPR allows for explainable hyper-parameters compared to other methods with no cost in accuracy .

Han, Zhang and Wang (2016) and Wu, Hernández-Lobato and Ghahramani (2014) use an adapted version of GPR called heteroskedastic GPR to estimate the volatility of returns. This requires the use of particle filters and Monte Carlo Markov Chains in order to work with the non-normal distribution of stock market returns. The GPR methods are compared to General Auto-Regressive Conditionally Heteroskedastic (GARCH) methods and conclude that GPR has more accurate long-term volatility predictions than the GARCH models.

3. Theoretical review

Sections 3.1,3.2 and 3.3 overviews the theory of GPR and optimization of hyper-parameters. Section 3.4 displays the formulae of the GPR kernels used in this paper. Section 3.5 explains the mean centring required for GPR models. Section 3.6 contains all the theory relating to ARIMA models. Section 3.7 provides the formulae for the evaluation metrics used in this paper.

3.1 Gaussian process regression overview

According to Schulz et al. (2018b) GPR is a non-parametric method that uses a Bayesian approach to regression. This is a multi-variate Gaussian process with theoretically infinite parameters with the exact number of parameters and Gaussian process density determined through a Bayesian process. Gaussian process regression can be applied to a large number of applications due to the mathematical simplicity of the conjugacy of the Gaussian distribution in the Bayesian context. Gaussian processes are also resistant to over-fitting and provide uncertainty levels automatically. Most versions of GPR rely on an assumption of stationary data but there are some that allow for non-stationary data.

One problem with the usage of GPR is that it has a complexity of O(n^3) (where n = the number of data points) due to a required Cholesky factorization of the data covariance matrix. To solve this problem sparse approximations, low-rank approximations, and spectral methods are used. Other approaches also include Fourier transforms and Nystrom approximation stochastic partial differential equation representations (Ton et al., 2018b).

3.2 Gaussian process regression theory

The theory of GPR covered below comes from the papers Ton et al. (2018b), Schulz et al. (2018b) and Gonzalvez et al. (2019). The assumption of the relationship between the response variable y and parameters X is given by the following equation:

$$y = f(x) + \varepsilon \quad , \varepsilon \sim Normal(0, \sigma_\varepsilon^2) \tag{1}$$

For Gaussian process regression f(x) is assumed to follow a Gaussian process with parameters m(x) (Expected value) and $k(x_i, x_i)$ (Variance). m(x)=0 is assumed for the simplicity that comes from the Gaussian process being described by $k(x_i, x_i)$ alone which is called the kernel of the distribution. The kernel can be defined as a kernel function with several different forms which represent different Gaussian distributions. The hyper-parameter θ contains values of constants for the kernel function set based on the prior assumptions of the behavior of the data.

Thus we have the Bayesian process:

$$\begin{aligned} &\theta \sim \pi(\theta) \\ &f|\theta \sim N(0, k(\theta)) \\ &y_i|f, x_i, \theta \sim N(f(x_i), \sigma_\varepsilon^2) \end{aligned} \tag{2}$$

Updating the prior with observed data $D_n = X_n, y_n$ that has observed covariance structure $K(X_n, X_n)$ And inputs X_*

$$\text{Where } K(X_n, X_n) = \begin{bmatrix} k(x_1,x_1) & k(x_2,x_1) & \dots & k(x_n,x_1) \\ k(x_1,x_2) & k(x_2,x_2) & \dots & k(x_n,x_2) \\ \dots & \dots & \dots & \dots \\ k(x_1,x_n) & k(x_2,x_n) & \dots & k(x_n,x_n) \end{bmatrix} \tag{3}$$

The posterior predictive distribution of output y_* based on X_* is $p(y_*|X_*, D_n, \theta) = N(y_*; \mu_\theta, \sigma_\theta^2)$.

$$\begin{aligned} \text{Where, } \mu_\theta &= K(X_*,X_n)[K(X_n,X_n)+\sigma_\varepsilon^2 I_n]^{-1} y_n \\ \sigma_\theta^2 &= K(X_*,X_*) - K(X_*,X_n)[K(X_n,X_n)+\sigma_\varepsilon^2 I_n]^{-1} K(X_n,X_*) \end{aligned} \tag{4}$$

Where the estimate $\widehat{y*}$ is μ_θ and confidence intervals are generated by σ_θ^2. Thus for a time-based input, estimates can be generated at any time horizon with an increasing σ_θ^2.

3.3 Hyper-parameter optimization

To optimize the hyper-parameters the log-likelihood function of the posterior distribution is obtained and maximized over θ. The marginal distribution is derived by the integral $p(y|\theta) = \int p(y|f,\theta)p(f|\theta)df$ and has the form:

$$log(p(y|\theta)) = -\frac{n}{2}log(2\pi) - \frac{1}{2}|K(X_n,X_n)+\sigma_\varepsilon^2 I_n| - \frac{1}{2}y^T[K(X_n,X_n)+\sigma_\varepsilon^2 I_n]^{-1}y \tag{5}$$

Where the dependence of y* on θ is induced via the kernel distribution, and θ can be maximized by any optimization methods such as gradient descent.

Maximizing this marginal likelihood automatically provides regularisation without the need for a specific regularisation method since maximizing the log likelihood function is efficient. The complexity of the model can be seen here since $[K(X_n,X_n)+\sigma_\varepsilon^2 I_n]^{-1}$ needs to be recalculated for each value of θ that is tested for maximization with a complexity of $O(n^3)$.

3.4 Common kernels and resulting Gaussian processes

The most common kernels used are the Matérn and Rational Quadratic kernels as they model such a wide variety of models. Schulz et al. (2018b) provides the formulas of several kernels displayed in this section.

3.4.1 Matérn kernel

Is the one of the most common types of kernels used.

$$\text{Matérn kernel } k_v(x_i,x_j) = \sigma_f^2 \frac{2^{1-v}}{\Gamma(v)}(\sqrt{2v}\frac{||x_i-x_j||}{p})B_v(\sqrt{2v}\frac{||x_i-x_j||}{p}) \tag{6}$$

p and v are non-negative covariance parameters, σ_f^2 is the signal variance and B_v is a modified Bessel function of the second kind. The kernel is v-1 times differentiable. The function often has hyper-parameter $\theta = (p, \sigma_f^2)$ where v is often set as 0.5+p, and p determines the decay of the covariance based on the distance between points.

3.4.2 Ornstein–Uhlenbeck kernel

A Matérn kernel where p = 0 which indicates the data contains very rough in the movements from one input to the next and follows an Ornstein–Uhlenbeck process.

$$\text{Ornstein–Uhlenbeck kernel } k(x_i,x_j) = \sigma_f^2 exp(-\frac{|x_i-x_j|}{\lambda}) \tag{7}$$

With hyper-parameter $\theta = (\lambda, \sigma_f^2)$

3.4.3 *Rational Quadratic kernel*

The Rational Quadratic kernel represents a combination of multiple wave functions.

$$\text{Rational Quadratic kernel } k(x_i, x_j) = \sigma_f^2 (1 + \frac{(x_i - x_j)^2}{2\alpha l^2})^{-\alpha} \tag{8}$$

With hyper-parameter $\theta = (\sigma_f^2, l, \alpha)$. Where l is a non-negative length scale parameter and α determines the weighting between large and small variations.

3.4.4 *Linear kernel*

This kernel returns simple Bayesian linear regression. Higher orders of Bayesian regression can be returned using the multiplicative combination of Linear kernels.

$$\text{Linear kernel } k_v(x_i, x_j) = \sigma_f^2 (x_i - c) * (x_j - c) \tag{9}$$

With hyper-parameter $\theta = (\sigma_f^2, c)$. Where is the x value at which the line representing the estimates will pass through.

3.4.5 *Combination of kernels*

All these kernels can be combined in a linear or multiplicative combination to form a kernel with a combination of the properties of the various kernel functions. E.g. a mean reverting kernel can be added to a kernel that can model trends to form a new process that reverts to a mean that linearly increases with time. A multiplicative model combination of a Matérn and linear kernel could be seen as a Matérn kernel with linearly increasing variance.

3.5 Mean centering of data for GPR

GPR requires mean-centered data as per Schulz et al. (2018b) to fit models since the mean function of the GPR is set to zero. Using the linear kernel, GPR can still be fitted to data with a linear or higher-order trend. One potential problem is that several kernels are mean-reverting kernels causing the mean to form an important part of the estimation. This can cause problems as the mean of the data may change due to choices like the number of observations included as well as outlier treatment.

3.6 ARIMA

Auto-regressive Integrated Moving Average(ARIMA) models as shown in Reddy (2019b) are often used to model time series data. ARIMA models are a combination of Moving Average [MA(q)] and Auto-regressive [AR(p)] and formed into an Auto-regressive Moving Average [ARMA(p,q)] model. The ARMA models can only be applied to stationary data. In order to extend ARMA to non-stationary data, data can be integrated (differenced) until stationarity is achieved which leads to ARIMA(p,d,q) where d is the order of differencing.

AR models rely on previous values of variables to estimate the current response variable. AR(p) models are defined by the equation:

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i} + \varepsilon_t \tag{10}$$

Where y_t is the value of the variable under consideration at time t. y_{t-i} for $i \in [1:p]$ is the value of the variable under consideration at time t-i. β_0 is an intercept term to be estimated. β_i for $i \in [1:p]$ are AR parameters to be estimated. ε_t is the error term associated with y_t and is assumed to follow a Normal$(0, \sigma_\varepsilon^2)$ distribution. The value of p is determined by the minimization of some criterion such as the Akaike Information Criterion (AIC) or some error term such as Mean Squared Error (MSE).

MA models rely only on error terms to estimate response data. MA(q) models are defined by the equation:

$$y_t = \beta_0 + \varepsilon_t + \sum_{i=1}^{q} \phi_i \varepsilon_{t-i} \tag{11}$$

Where y_t is the value of the variable under consideration at time t. β_0 is an intercept term to be estimated. ϕ_i for $i \in [1:q]$ are MA parameters to be estimated. ε_{t-i} for $i \in [0:q]$ is the error term associated with y_{t-i} for $i \in [0:q]$ and is assumed to follow a Normal$(0, \sigma_\varepsilon^2)$ distribution. The value of q is determined by the minimization of some criterion such as the Akaike Information Criterion (AIC) or some error term such as Mean Squared Error (MSE).

ARMA models are a combination of both AR and MA models where the value of a variable at time t is based on previous variable values and error terms. The ARMA(p,q) model is defined by the equation:

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \phi_j \varepsilon_{t-j} \tag{12}$$

Where y_t is the value of the variable under consideration at time t. β_0 is an intercept term to be estimated. y_{t-i} for $i \in [1:p]$ is the value of the variable under consideration at time t-i and ε_{t-j} for $j \in [0:q]$ is the error term associated with y_{t-j} for $j \in [0:q]$ and is assumed to follow a Normal$(0, \sigma_\varepsilon^2)$ distribution. β_i for $i \in [1:p]$ and ϕ_j for $j \in [1:q]$ are the AR and MA parameters to be estimated respectively. The values of p and q are determined by the minimization of some criterion such as the Akaike Information Criterion (AIC) or some error term such as Mean Squared Error (MSE).

ARIMA models are ARMA models where the data has been integrated[I(d)]. Integration is the $d^t h$ difference of the data. ARIMA(p,d,q) models are defined by the equation:

$$\Delta^d y_t = \sum_{i=1}^{p} \beta_i \Delta^d y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \phi_j \varepsilon_{t-j} \tag{13}$$

Where y_t is the value of the variable under consideration at time t. $\Delta^d y_{t-i}$ represents $y_{t-i} - y_{t-i-d-1}$ for $i \in [0:p]$. ε_{t-j} for $j \in [0:q]$ is the error term associated with the values of y_{t-j} for $j \in [0:q]$ and is assumed to follow a Normal$(0, \sigma_\varepsilon^2)$ distribution. β_i for $i \in [1:p]$ and ϕ_j for $j \in [1:q]$ are the AR and MA parameters to be estimated respectively. The values of p and q are determined by the minimization of some criterion such as the Akaike Information Criterion (AIC) or some error term such as Mean Squared Error (MSE). The order of integration d is defined as the minimum level of differencing required to make the data stationary. If the data is already stationary the d=0 and the model reverts to the same form as an ARMA(p,q) model and would include an intercept term β_0.

3.6.1 Akaike Information Criterion

The AIC is the negative log-likelihood function of the model with a penalty term for model complexity.

$$AIC = -2log(L) + 2 * k \tag{14}$$

Where L is the likelihood function of the model and k is the number of parameters used.

3.6.2 Bayesian Information Criterion

The BIC is very similar to the AIC but the penalty term for model complexity is much higher as log(n) is higher than 2.

$$BIC = -2log(L) + klog(n) \tag{15}$$

Where L is the likelihood function of the model, k is the number of parameters used and, n is the number of data points.

3.6.3 Normality of the data

The Shapiro-Wilk test was used to test the normality of the data as well as graphical methods such as the normal Quantile-Quantile (Q-Q) plot.

- H_0 The data is normally distributed
- H_1 The data is not normally distributed

(de Souza, Toebe, Mello and Bittencourt, 2023).

3.6.4 Power transformations of the data

The Box-Cox test is used to determine if the data required a power transform to induce normality in the case of non-normal data. The Box-Cox lambda is estimated by maximizing the log-likelihood with a 95% confidence interval constructed using χ_1^2 95 percentile (Proietti and Lütkepohl, 2013).

$$y_t(\lambda) = \begin{cases} \frac{y_t^{\lambda}-1}{\lambda} & \text{for } \lambda \neq 0 \\ ln(y_t) & \text{for } \lambda = 0 \end{cases} \tag{16}$$

The necessity of a transformation is determined by the following hypothesis test using the 95% confidence interval to determine if H_0 can be rejected.

- H_0 The data does not require a transformation ($\lambda = 1$)
- H_1 The data requires a transformation ($\lambda \neq 1$)

3.6.5 Stationarity of the data

ARIMA methods rely on the assumption of stationary data. To test this the Augmented Dickey-Fuller (ADF) unit root test is used to test stationarity of the time series. A time series is stationary if its mean and variance are constant with regard to time. The ADF test determines stationarity by testing for the presence of unit roots which indicate non-stationarity (Reddy, 2019b).

The regular Dickey-Fuller test is defined as 3 types for testing 3 types of data.

Type 0: No constant or trend $\Delta y_t = \beta_1 y_{t-1} + \varepsilon_t$

Type 1: Constant intercept but no trend $\Delta y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$

Type 2: Constant intercept and trend $\Delta y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 t + \varepsilon_t$

Where ε_t is the error term and Δy_t represents $y_t - y_{t-1}$. β_0 is a constant intercept term. $\beta_2 t$ is a linear trend term. β_1 is the unit root we are testing for the presence of.

The ADF test also has 3 different types for testing the presence of unit roots under different conditions as well as an augmented term which is used to test for stationarity in the presence of multiple lagged values.

Type 0: No constant or trend $\Delta y_t = \beta_1 y_{t-1} + \sum_{j=1}^{m} \alpha_j \Delta y_{t-j} + \varepsilon_t$

Type 1: Constant intercept but no trend $\Delta y_t = \beta_0 + \beta_1 y_{t-1} + \sum_{j=1}^{m} \alpha_j \Delta y_{t-j} + \varepsilon_t$

Type 2: Constant intercept and trend $\Delta y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 t + \sum_{j=1}^{m} \alpha_j \Delta y_{t-j} + \varepsilon_t$

Where ε_t is the error term and Δy_t represents $y_t - y_{t-1}$. The term $\sum_{j=1}^{m} \alpha_j \Delta y_{t-j}$ is the augmented term in the Dickey-Fuller test and allows the test to include multiple lags. The order of the additional lags m is determined via the AIC. β_0 is a constant intercept term. $\beta_2 t$ is a linear trend term. β_1 is the unit root we are testing for the presence of.
m can be found by representing the ADF test equations as AR(m) processes and then finding m that minimizes the AIC of the AR(m) processes. An alternate method can be used where m is increased until the last most significant lag.

The hypothesis test for the ADF test is as follows:

- H_0 The data contains a unit root and therefore is non-stationary.
- H_1 The data does not contain a unit root and therefore is stationary.

The ADF test is usually decided by test statistic defined as:

$$t = \frac{\hat{\delta} - \delta_{H_0}}{SE\ of\ \hat{\delta}} \tag{17}$$

We reject the null hypothesis H_0 if t is less than the critical values and conclude in favour of stationarity and vice versa.

3.7 Evaluation metrics

Several metrics will be used for the comparison and evaluation of the models applied namely the Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

3.7.1 *Mean Squared Error*

The MSE penalizes larger errors far more heavily than small errors as the errors are squared. The MSE is defined as:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{18}$$

(Chai and Draxler, 2014).

3.7.2 *Mean Absolute error*

The MAE does not induce extra penalties for large or small errors and reflects the average of the exact distance of estimates from real values. It is defined as:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{19}$$

(Chai and Draxler, 2014).

3.7.3 *Mean Absolute Percentage Error*

The MAPE represents the average percentage of the real value that estimated values vary by. It is defined as:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \tag{20}$$

(Reddy, 2019b).

4. Methodology

Section 4.1 provides a description of the data. Section 4.2 will cover normality, stationarity and power transformation tests. Section 4.3 contains the process for the selection of the ARIMA models. Section 4.4 discusses the selection of GPR kernels. Section 4.5 shows how the GPR models will be trained. Section 4.6 displays the evaluation metrics to be used for model comparison.

4.1 Data

This paper will look at the daily closing stock price of Tesla share prices over 12 months. The period of study is from 1 January 2023 to 31 December 2023. The data was obtained from Yahoo Finance which contains historical price data on many commodities, stocks, and foreign currency pairs, with all stocks dating back to their IPO. The data can be freely imported from the website https://finance.yahoo.com/quote/TSLA/.

There is no data for weekends and certain weekdays as the stock exchange does not operate on American religious and federal holidays. Both ARIMA and GPR can handle the missing data as the kernel for GPR can be calculated as long as there are more than 2 values and is capable of estimating the value of the intervening missing values. The ARIMA algorithm ignores the update step for the missing values but this does require the assumption that the missing values do not affect the stock returns.

Missing data was not imputed since to was not required since any imputation or estimation for day where the stock exchange is closed does not have a use and cannot be measured for accuracy as a value for those days will never exist. The data was split 80%/20% for training and testing purposes. Only training data was used for mean centering, normality testing, Box-Cox tests, and ADF tests to prevent information leakage.

4.2 Normality, stationarity and power transformation

The Shapiro-Wilk test will be used to test the normality of the data to determine if the data follows a white noise process. The shapiro.test R command from R stats will be used. If the data is not normal a Box-Cox test will then be applied in order to test if a power transform is appropriate and the transformed data will then be reassessed for normality. The boxcox R command from the MASS R package will be used. The Augmented Dickey-Fuller (ADF) test is used to determine if the data is stationary. If the data is non-stationary then it will be differenced and the ADF test will be applied again. The number of lags for the ADF test will be determined automatically using the AIC. The Box-Cox test will be applied to determine if a power transformation in required. The ur.df R command from the urca R package will be used.

4.3 Selection of ARIMA models

The top 6 ARIMA(p,d,q) will be chosen to be compared the the GPR models. As per Reddy (2019b) d will be determined by the ADF test to decide on the order of differencing that is required for stationarity or if differencing is needed at all. p and q will determined by using a grid search heuristic for a maximum AIC where AIC increases of 2 per parameter are considered significant. The selection will also consider the number of insignificant parameters present in each model. The Arima R command from the forecast R package will be used for training and the forecast R command from the same package will be used to retrieve estimates.

4.4 Selection of GPR models

Ton et al. (2018b) state that selection of kernel functions for GPR is done arbitrarily with no scientific way to evaluate the applicability of a kernel to the data until it has been evaluated. The kernels to be used are the Matérn and Rational Quadratic kernels as they are the most commonly used and most flexible kernels. An additive combination of a Matérn and linear kernel is used to test a kernel that is non-stationary and is not mean reverting. The Ornstein-Uhlenbeck kernel has also been selected as a candidate model as it represents data that is not smooth which matches the very rapid and sharp movements observed in the Tesla data. A multiplicative combination of the Ornstein-Uhlenbeck and Ration Quadratic kernels is also considered as it can be described as having the properties of an Ornstein-Uhlenbeck kernel with a variance that changes along a wave pattern which could capture the time-varying variance property of stock market data.

4.5 Training of GPR hyper-parameters

The data will be mean centred before training the GPR hyper-parameters and then the mean will be added back to the estimates to ensure it matches the data. The hyper-parameters of each GPR model will be calculated by using the optimise function in R to minimize the negative log-likelihood function of each model as suggested by Schulz et al. (2018b). The optimise function was rerun with the optimal values of the hyper-parameters and some random starting hyper-parameter values to test that a global minimum had been achieved. The constrOptim R command from the R stats package will be used for optimising the hyper-parameters. Original R code will be used for the rest of the analysis of the GPR models.

4.6 Evaluation metrics

The MSE, MAE, and MAPE defined in the Theoretical review will be calculated for the whole testing period for each model compared. The models will also be evaluated in 5-day rolling windows where the MSE, MAE, and MAPE are calculated for the next 5 days after which the model is refitted on the training data and the previously evaluated 5-day windows. These will be shown in a separate table for each metric and the average of the metrics of the 10 windows will be included in a table with the metrics for all the test observations. This will allow the long and short term predictive accuracy to be evaluated. The generalisation ability of each model can also be compared by looking at the accuracy of the later 5-day windows since the ARIMA models will not be re-specified each retrain and the hyper-parameters of the GPR models will not be re-optimised. The pape will use original R code supplied in the appendix for calculating these metrics.

5. Analysis and results

Section 5.1 covers the results of the normality, stationarity and power transformation tests. Section 5.2 displays the 6 best ARIMA models with comparisons metrics. Section 5.3 contains the MSE results for the rolling testing window. Section 5.4 contains the MAE results for the rolling testing window. Section 5.5 contains the MAPE results for the rolling testing window. Section 5.6 contains the overall comparison metrics as well as a discussion of the results.

5.1 Normality, stationarity, and power transform

The results of the Shapiro-Wilk test are a p-value of $\sim 0\%$ for the undifferenced data and a p-value of 2.68% for the differenced data. Therefore at a 95% confidence level we reject the null hypothesis that the data follows a normal distribution. This may cause some problems with the use of ARIMA and GPR but normal Q-Q plots also need to be considered.

In Figure 1 it can be seen that the distribution of the undifferenced data is very far from normally distributed. Using the undifferenced data will likely cause issues for estimation using ARIMA. The differenced data looks approximately normal albeit with very heavy tails. The data still looks symmetric for both the peak and heavy tails.

The log-likelihood method was used to find a Box-Cox lambda for the training data and was estimated at 1.515 with a 95% confidence interval of (0.91;2)[1]. This confidence includes $\lambda = 1$ therefore we fail to reject the null hypothesis H_0 that the data does not require a transformation ($\lambda = 1$) and ignore the alternate hypothesis H_1 that the data requires a transformation ($\lambda \neq 1$).

All three versions of the ADF test were used when testing stationarity. The results are shown in Table 1.

[1] The confidence interval extends beyond 2 but the R command used arbitrarily sets 2 as a maximum

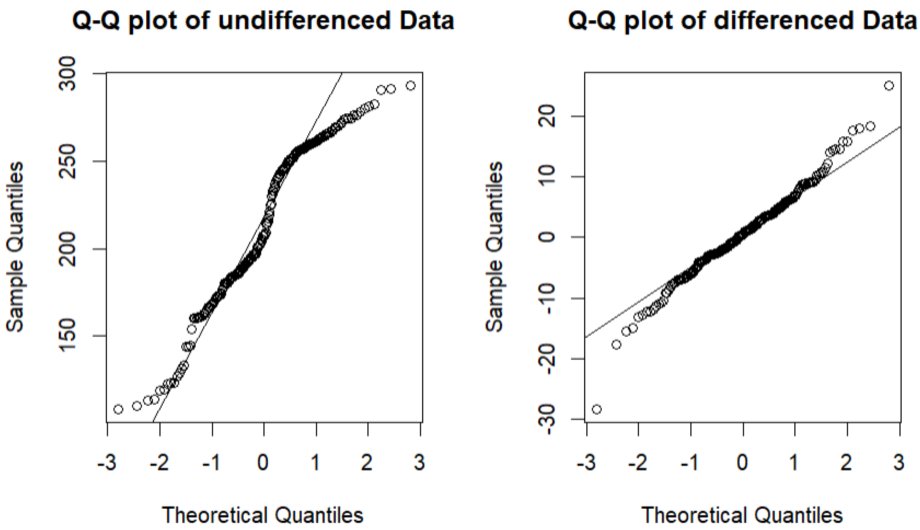

Figure 1. Q-Q plot of differenced and undifferenced data

Table 1. ADF results for TSLA stock price

	Type 1	Type 2	Type 3
90% confidence	-1.62	-2.57	-3.13
95% confidence	-1.95	-2.88	-3.43
99% confidence	-2.58	-3.46	-3.99
Calculated statistic	0.86	-2.17	-1.90
Number of lags (m)	1	1	1
β_0	-	Significant	Significant
β_2	-	-	Insignificant

Using the AIC method to determine the number of lags results in a lag order of m=1. β_0 is an intercept term and is found to be significant at a 95% confidence level implying the data should contain an intercept term. β_2 is found to be insignificant at a 95% confidence level implying that there is no linear trend in this data. From the table, it can be seen that the calculated test statistics are higher than the test statistics for all three types of the ADF test. Therefore we fail to reject the null hypothesis H_0 that the data contains a unit root and is therefore non-stationary and difference the series.

After differencing the series we retest the differenced series using the ADF test. The results for the differenced series are in 2.
Using the AIC method again to determine the number of lags still results in a lag order of m=1. β_0 and β_2 are found to be insignificant at a 95% confidence level and the data neither contains an intercept nor a linear trend. This follows the theory of differencing as differencing removes intercepts $[\Delta y_t = (y_t + \beta_0) - (y_{t-1} + \beta_0) = y_t - y_{t-1}]$. From the table, it can be seen that the calculated test

Table 2. ADF results for the differenced TSLA stock price

	Type 1	Type 2	Type 3
90% confidence	-1.62	-2.57	-3.13
95% confidence	-1.95	-2.88	-3.43
99% confidence	-2.58	-3.46	-3.99
Calculated statistic	-9.85	-9.96	-10.08
Number of lags (m)	1	1	1
β_0	-	Insignificant	Insignificant
β_2	-	-	Insignificant

statistics are much lower than the test statistics for all three types of the ADF test. Therefore we reject the null hypothesis H_0 that the data contains a unit root and is therefore non-stationary and accept the alternate hypothesis H_1 that the data does not contain a unit root and therefore is stationary. We conclude that d=1 and that no further differencing is required.

5.2 ARIMA

Using a grid search method and by maximising the likelihood value several ARIMA(p,1,q) models were selected as candidate models for further consideration. The models along with some comparison metrics are shown in table 3.

Table 3. Table showing the ARIMA results of the top-performing models for the Tesla share price

ARIMA	AIC	BIC	Insignificant parameters[2]
0,1,0	1349.50	1352.791	0/0
1,1,0	1351.30	1357.881	1/1
0,1,1	1351.30	1357.886	1/1
1,1,1	1350.59	1360.473	0/2
2,1,1	1352.41	1365.587	1/3
1,1,2	1352.38	1365.552	1/3

These ARIMA models along with a few GPR models are displayed in the following tables with detailed rolling metrics for MSE in Table 4, MAE in Table 5, and MAPE in Table 6. The metrics for the whole testing period as well as the average rolling window metrics are contained in Table 7. The GPR models picked, as per the Methodology 4, are the Matérn (Ma.) kernel, Matérn + Linear (Ma. + Li.) kernel, Ornstein-Uhlenbeck (OU), Rational Quadratic (RQ) kernel, and the multiplicative combination of the Rational Quadratic and Ornstein-Uhlenbeck (OU X RQ).

[2] Based on a 95% confidence level

Table 4. Table showing rolling MSE results for GPR and ARIMA models for the Tesla share price

	MSE for the following testing days:[3]									
Model	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50
ARIMA(0,1,0)	797.53	95.43	219.67	173.83	57.62	58.92	6.14	24.7	14.46	18.89
ARIMA(1,1,0)	819.04	94.62	218.36	173.84	60.1	61.05	5.96	24.77	13.48	18.91
ARIMA(0,1,1)	819.03	94.65	218.41	173.85	59.8	61.04	5.97	24.77	13.5	18.91
ARIMA(1,1,1)	801.24	102.47	217.81	167.84	94.39	61.01	5.98	24.76	13.49	18.92
ARIMA(2,1,1)	776.43	99.38	203.35	175.13	75.99	62.24	5.76	25.15	13.28	18.78
ARIMA(1,1,2)	772.24	98.51	200.79	175.57	76.05	63.44	5.83	25.29	12.96	18.85
Ma.	474.96	363.06	268.49	275.88	468.49	361.41	20.02	34.67	53.3	275.63
Ma. + Li.	979.05	902.88	94.78	118.01	993.61	54.68	84.15	80.99	43.05	40.29
OU	784.1	98.48	219.44	177.8	47.37	64.67	6.07	25.28	20.56	20.28
RQ	693.5	220.15	124.43	174.74	108.05	90.18	12.32	29.6	25.38	61.95
OU X RQ	692.19	38.43	341.7	188.06	68.13	42.62	8.5	24.91	18.66	19.16

5.3 MSE results

Analysis of the rolling window results of the MSE from Table 4 the ARIMA models all performed similarly with the (2,1,1) and (1,1,2) models outperforming the others. These two models contained statistically insignificant parameters and should not be significantly different from the (1,1,1) model which may indicate over-fitting however the metrics for the later windows should be worse that (1,1,1) if there was over-fitting. The GPR models' metrics vary heavily from one another and from period to period. The OU model behaves the most like the ARIMA models with only minor differences in each period. The Ma. process appears to be the least accurate over the periods except the first one where it is the most accurate. As can be seen in Figures 2, 3, 4, and 5, there is a sudden drop in Tesla share price at the beginning of the training period. The mean reverting nature of the Ma. and other GPR processes can also been seen in Figure 2. This estimation of which drops to the mean may have been why the Ma. process is so accurate for the window containing a drop in price but suffers in estimating the other windows. This may also explain the behavior of other mean reverting GPR models and how GPR is more accurate in some windows, namely the first, while being less accurate for the 4th window due to the behaviour of the data in those windows. Another model that supports this is the Ma. + Li. model as in Figure 4. The Ma. + Li. is not mean reverting and is actually increasing with time. It is the least accuracy in some windows like the first and the most accuracy in others like the 4^{th}.

The Ornstein-Uhlenbeck model appearing to be the most accurate GPR model. The Ornstein-Uhlenbeck model while mean reverting like the Matérn model has a much slower mean reversion seen in Figure 5. The Ornstein-Uhlenbeck also seems to be closer in accuracy to the ARIMA models then the other GPR models especially in the mirroring of accuracy behaviour of the ARIMA models for all windows.

[3] Where the model is refitted on training and previously tested days

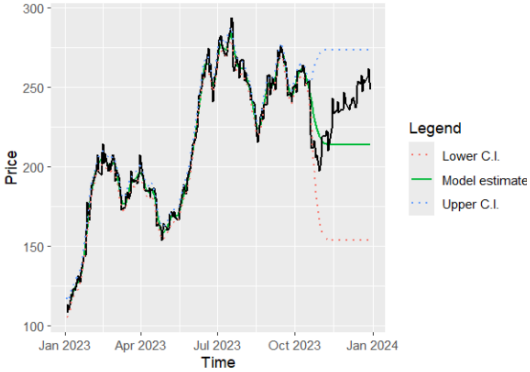

Figure 2. Matérn kernel model

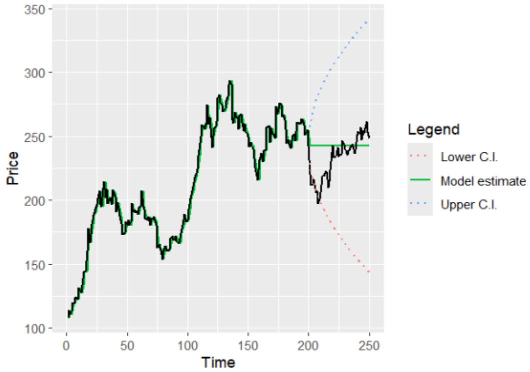

Figure 3. ARIMA (0,1,0) model

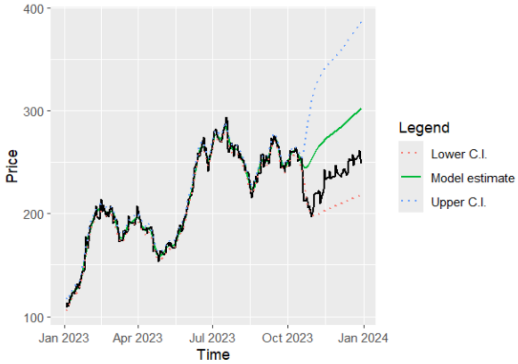

Figure 4. Matérn + Linear kernel model

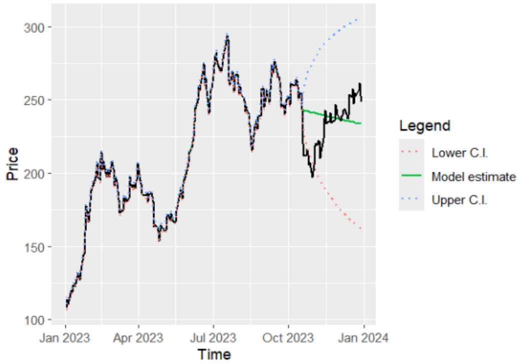

Figure 5. Ornstein-Uhlenbeck kernel model

5.4 MAE results

The results for the MAE from Table 5 are very similar to the results for the MSE with ARIMA(2,1,1) and ARIMA(1,1,2) being the best ARIMA models and the Ornstein-Uhlenbeck model appearing to be the most accurate GPR model. The Ornstein-Uhlenbeck model still follows the ARIMA model closely in accuracy. From theses MAE metrics the relative rankings of accuracy for each window are the same but the scale between the most and least accurate models is smaller. This represents the use of the MAE in determining mean error with no additional penalty for large errors.

5.5 MAPE results

The results for the MAPE from Table 6 are again similar to the MSE and MAE results but metrics for the different models are even closer together. The ARIMA models and Ornstein-Uhlenbeck models still display very similar error measures for all windows. The ranking of accuracy between models is the same with the higher order ARIMA models being the most accurate and the Ornstein-Uhlenbeck and Ornstein-Uhlenbeck X Rational Quadratic being the most accurate GPR models.

Table 5. Table showing rolling MAE results for GPR and ARIMA models for the Tesla share price

	MAE for the following testing days:[4]									
Model	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50
ARIMA(0,1,0)	28.06	9.04	14.75	11.44	7.06	6.28	2.08	4.3	3.4	3.67
ARIMA(1,1,0)	28.44	8.99	14.7	11.44	7.23	6.45	2.02	4.31	3.23	3.69
ARIMA(0,1,1)	28.44	8.99	14.7	11.44	7.21	6.45	2.02	4.31	3.23	3.69
ARIMA(1,1,1)	28.12	9.4	14.68	11.25	9.38	6.45	2.03	4.31	3.23	3.69
ARIMA(2,1,1)	27.68	9.23	14.18	11.5	8.21	6.54	1.6	4.38	3.05	3.58
ARIMA(1,1,2)	27.6	9.17	14.09	11.51	8.2	6.64	1.58	4.4	3.01	3.59
Ma.	21.05	17.33	16.36	14.51	20.64	17.33	3.87	4.08	6.42	15.01
Ma. + Li.	31.15	27.11	7.94	10.12	29.45	6	8.91	7.67	5.35	4.77
OU	27.83	9.2	14.75	11.58	6.21	6.64	1.92	4.07	4.06	3.55
RQ	26.19	14.11	11.01	11.45	10.06	7.88	3.05	4.27	4.51	6.47
OU X RQ	26.15	5.24	18.26	11.91	7.86	4.95	2.54	4.28	3.95	3.43

Table 6. Table showing rolling MAPE results for GPR and ARIMA models for the Tesla share price

	MAPE for the following testing days in %:[5]									
Model	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50
ARIMA(0,1,0)	13.1	4.48	6.69	4.99	3.01	2.58	0.87	1.77	1.34	1.44
ARIMA(1,1,0)	13.28	4.46	6.67	4.99	3.08	2.65	0.85	1.77	1.27	1.45
ARIMA(0,1,1)	13.27	4.46	6.67	4.99	3.07	2.65	0.85	1.77	1.28	1.45
ARIMA(1,1,1)	13.13	4.66	6.66	4.91	3.99	2.65	0.85	1.77	1.27	1.45
ARIMA(2,1,1)	12.92	4.57	6.43	5.01	3.5	2.69	0.67	1.8	1.2	1.4
ARIMA(1,1,2)	12.89	4.55	6.39	5.01	3.49	2.73	0.66	1.81	1.19	1.41
Ma.	9.81	8.58	7.42	6.3	8.75	7.15	1.62	1.66	2.52	5.88
Ma. + Li.	14.54	13.41	3.61	4.5	12.48	2.46	3.73	3.2	2.13	1.85
OU	12.99	4.56	6.69	5.05	2.65	2.73	0.8	1.67	1.6	1.38
RQ	12.22	6.98	5	4.99	4.28	3.24	1.27	1.74	1.77	2.51
OU X RQ	12.2	2.6	8.28	5.18	3.35	2.02	1.07	1.76	1.56	1.34

5.6 Summary metric results

The overall comparative metrics from Table 7 for the whole test period display similar results for the ARIMA models with the most accurate models being ARIMA(1,1,2) followed by ARIMA (2,1,1) and then ARIMA(0,1,0). The ARIMA models are also more accurate over the whole period than all the GPR models including the Ornstein-Uhlenbeck model which performed better than the ARIMA(1,1,1) model in the rolling tests but worse in this 50-day test. The Matérn + Linear, Matérn and Ornstein-Uhlenbeck X Rational Quadratic perform so badly against the ARIMA models that

[4] Where the model is refitted on training and previously tested days

[5] Where the model is refitted on training and previously tested days

Table 7. Table showing comparison statistics for GPR and ARIMA models for the Tesla share price

	Full 50 days			Average rolling measure		
Model	MSE	MAE	MAPE(%)	MSE	MAE	MAPE(%)
ARIMA(0,1,0)	369.78	14.79	6.72	146.72	9.01	4.03
ARIMA(1,1,0)	377.35	14.95	6.79	149.01	9.05	4.19
ARIMA(0,1,1)	377.37	14.95	6.79	148.99	9.05	4.06
ARIMA(1,1,1)	370.91	14.81	6.73	150.79	9.25	4.17
ARIMA(2,1,1)	362.33	14.65	6.65	145.55	8.99	4.37
ARIMA(1,1,2)	360.94	14.62	6.63	144.95	8.98	4.22
Ma.	710.2	24.06	10.11	259.59	13.66	5.97
Ma. + Li.	1871.61	42.68	18.45	339.15	13.85	6.19
OU	390	15.75	7.04	146.40	8.98	4.01
RQ	413.93	17.49	7.66	154.03	9.9	4.4
OU X RQ	628.78	22.89	9.82	144.23	8.86	3.94

using only the 50-day tests they would not be considered as valid models. The average metrics over the rolling windows represent the accuracy of repeated short term estimation without re-specifying or re-optimizing the models. The ARIMA(1,1,2) and ARIMA (2,1,1) perform worse in this metric than the ARIMA(0,1,0) and even ARIMA(1,1,1) this could show that the model is over fitted and possibly only shown as more accurate over the full period due to the over sized effect of the large drop in price at the start of the testing period. The ARIMA(1,1,1) and ARIMA(0,1,0) models are more accurate in a generalisation metric (as the models are applied to different scenarios each window) over other ARIMA models as the other models contain statistically insignificant parameters.

The Ornstein-Uhlenbeck model's performance is closer to the ARIMA models and is equivalent in accuracy to the ARIMA(0,1,0) and in fact is slightly more accurate over the rolling windows. The most accurate model for the average rolling error measures was in fact the Ornstein-Uhlenbeck X Rational Quadratic model beating out all other models in these metrics even while being the 9^{th} most accurate model when using the error metrics for the full test period.

The behaviour of the GPR models' long-term inaccuracy and short-tern accuracy displays the mean-reverting nature of the GPR kernel. Over the long-term this ensures that estimates become the mean of the training data which is likely to be worse than any non mean reverting models unless the data being estimated displays mean reversion.

6. Conclusion and recommendations

Section 6.1 covers key conclusions. Section 6.2 provides some limitations of this paper and GPR. Section 6.3 contains recomendations for possible furture research.

6.1 Key findings

Here are several of the main results:

- Some GPR models display better short-term predictions than the ARIMA models
- All GPR models display worse long-term predictions than the ARIMA models
- The Ornstein-Uhlenbeck X Rational Quadratic is the most accurate short term model
- The ARIMA(1,1,2) is the most accurate long-term model
- Some of the long-term accuracy of the ARIMA models may be due to over-fitting
- The mean reverting property of many GPR kernels makes them inaccurate long-term models.

This paper concludes that GPR could be applicable for short-term forecasting of share prices. The Ornstein-Uhlenbeck model despite being mean reverting displayed short-term accuracy very close to the ARIMA models - including in forecasts where the share price went up. More testing would be required in order to conclude that GPR is a better approach to short-term estimation over ARIMA as the results may be influenced by the particular structure of this data but GPR is worth considering as a possible method of estimation.

6.2 Limitations

The GPR and ARIMA models are constrained by the heavy-tailed nature of the share price changes which makes the confidence intervals inaccurate at the 95% confidence level. The confidence intervals should included values much further away from the estimates than shown in Figures 2, 3, 4, and 5. Another limitation is the mean reverting properties of most GPR kernels making them inaccurate for any non mean reverting data. The *a priori* selection of GPR kernels with no mathematically tractable methods of selection mean that it is difficult to pick the right kernel or combination of kernels to represent the data. This implementation of GPR, while Bayesian in nature, does not represent a truly Bayesian approach as there is no uncertainty estimation for the value of the hyper-parameters.

6.3 Future work

A few possible avenues of future research are:

- The use and study of non mean reverting kernels such as the Linear kernel that might model more complex functions like the Ornstein-Uhlenbeck and Rational Quadratic kernels can for mean reverting data.
- The non-parametric approach used by Ton et al. (2018b) where the kernel function is trained from the data. The same method can be used to represent unspecified kernels that can be derived from copulas which could possibly use recent research into copulas representing share price changes. This method could also be applied to heteroskedastic GPR which is used for volatility modeling to improve the estimation of volatility and hence the estimation of confidence intervals.
- Adopting a truly Bayesian approach could also improve uncertainty estimation and allow for the quantification of uncertainty from the possible miss-specification of the hyper-parameters.
- The multidimensional properties of GPR could be used for better predicting a grouping of shares but estimating each share individually utilising correlation and patterns between shares which ARIMA models can only estimate independently.

References

CHAI, T. AND DRAXLER, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. *Geoscientific Model Development*, **7** (3), 1247–1250.

CHAPADOS, N. AND BENGIO, Y. (2007). Forecasting and trading commodity contract spreads with gaussian processes. *In 13th International Conference on Computing in Economics and Finance*.

DE SOUZA, R. R., TOEBE, M., MELLO, A. C., AND BITTENCOURT, K. C. (2023). Sample size and shapiro-wilk test: An analysis for soybean grain yield. *European Journal of Agronomy*, **142**, 126666.

GONZÁLVEZ, J., LEZMI, E., RONCALLI, T., AND XU, J. (2019). Financial applications of gaussian processes and bayesian optimization. *arXiv preprint arXiv:1903.04841*.

GONZALVEZ, J., LEZMI, E., RONCALLI, T., AND XU, J. (2019). Financial applications of gaussian processes and bayesian optimization. *arXiv preprint arXiv:1903.04841*.

HAN, J., ZHANG, X.-P., AND WANG, F. (2016). Gaussian process regression stochastic volatility model for financial time series. *IEEE Journal of Selected Topics in Signal Processing*, **10** (6), 1015–1028.

JARRETT, J. AND KYPER, E. (2011a). Arima modeling with intervention to forecast and analyze chinese stock prices. *International Journal of Engineering Business Management*, **3**, 17–27.

JARRETT, J. E. AND KYPER, E. (2011b). Arima modeling with intervention to forecast and analyze chinese stock prices. *International Journal of Engineering Business Management*, **3**, 17.

JUAN, N. P., MATUTANO, C., AND VALDECANTOS, V. N. (2023). Uncertainties in the application of artificial neural networks in ocean engineering. *Ocean Engineering*, **284**, 115193.

MA, P. AND KANG, E. L. (2020). A fused gaussian process model for very large spatial data. *Journal of Computational and Graphical Statistics*, **29** (3), 479–489.

PETELIN, D., ŠINDELÁŘ, J., PŘIKRYL, J., AND KOCIJAN, J. (2011a). Financial modeling using gaussian process models. *In 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*, volume 2. IEEE, 672–677.

PETELIN, D., ŠINDELÁŘ, J., PŘIKRYL, J., AND KOCIJAN, J. (2011b). Financial modeling using gaussian process models. *In Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*, volume 2. IEEE, 672–677.

PROIETTI, T. AND LÜTKEPOHL, H. (2013). Does the box–cox transformation help in forecasting macroeconomic time series? *International Journal of Forecasting*, **29** (1), 88–99.

REDDY, C. (2019a). Predicting the stock market index using stochastic time series arima modelling: The sample of bse and nse. *Indian Journal of Finance*, **13** (8), 7–25.

REDDY, C. V. (2019b). Predicting the stock market index using stochastic time series arima modelling: The sample of bse and nse. *Indian Journal of Finance*, **13** (8), 7–25.

SCHULZ, E., SPEEKENBRINK, M., AND KRAUSE, A. (2018a). A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, **85**, 1–16.

SCHULZ, E., SPEEKENBRINK, M., AND KRAUSE, A. (2018b). A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, **85**, 1–16.

TON, J.-F., FLAXMAN, S., SEJDINOVIC, D., AND BHATT, S. (2018a). Spatial mapping with gaussian processes and nonstationary fourier features. *Spatial Statistics*, **28**, 59–78.

TON, J.-F., FLAXMAN, S., SEJDINOVIC, D., AND BHATT, S. (2018b). Spatial mapping with gaussian processes and nonstationary fourier features. *Spatial Statistics*, **28**, 59–78.

WU, Y., HERNÁNDEZ-LOBATO, J. M., AND GHAHRAMANI, Z. (2014). Gaussian process volatility model. *Advances in Neural Information Processing Systems*, **27**.