Journal of Uncertain Systems
Vol. 17, No. 4 (2024) 2450023 (28 pages)

DOI: 10.1142/S1752890924500235

Machine Learning Coffee Price Predictions

Bingzi Jin[*] and Xiaojie Xu[†,‡]

[*]*Advanced Micro Devices (China) Co., Ltd.*
Shanghai, P. R. China

[†]*North Carolina State University*
Raleigh, NC 27695, US
[‡]*xxu6@alumni.ncsu.edu*

Received 7 August 2024
Revised 8 September 2024
Accepted 11 September 2024
Published 11 October 2024

Most market players have found great significance in price projections for basic agricultural commodities for a substantial duration. We look at the daily price of coffee that is released in this research in order to tackle the issue. The analytical sample runs from 2 January 2013 to 10 April 2024, a period of more than 12 years. A significant influence on the business sector comes from the price series under investigation. Specifically, in this particular situation, Gaussian process regression models are developed using Bayesian optimization techniques and cross-validation processes. Thus, this circumstance prompts the development of price forecasting methodologies. Using our empirical forecasting technique, we produce relatively accurate price projections for the out-of-sample assessment period, which runs from 3 January 2022 to 10 April 2024. It was found that price forecasts of coffee had a relative root mean square error of 2.0500%. With the availability of price forecasting models, investors and governments can make educated decisions about the coffee market given that they have access to the required data.

Keywords: Coffee price; time-series prediction; Gaussian process regression; Bayesian optimization; cross validation.

1. Introduction

For the time being, estimates of price time series of major agricultural commodities continue to be a major source of concern for the business and government sectors.[1,2] To aid in the development of well-informed plans and mitigate the current level of uncertainty, estimates of the price time series of certain commodities, like coffee, will need to be obtained.[3] This is because many countries and regions depend significantly on agricultural products to achieve their strategic objectives. The vast majority of businesses involved in the commodities markets must be able to get information about price expectations in order to make decisions that are supported

[‡]Corresponding author.

by reliable information.[4] One crucial component of this is predictive analytics' capacity to provide insightful information about potential future pricing strategies that commodities processors could decide to utilize.[5] Price estimates may be used to provide trade partners with pricing information, enabling them to decide on the conditions of the contract with more knowledge.[6] Participating in the cash and futures markets may be beneficial for those who make use of predictive data, such as traders and investors.[7] It has been shown by the findings of risk analysts that forecast data are an important component of effective risk management programs.[8] Changes in commodity prices that are both big and long-lasting[9,10] have the potential to have a considerable influence on the general well-being and productivity of society[11] when they occur. It would seem that one does not need much motivation in order to appreciate the significance of time series projections for commodity prices, notably those of coffee.[12]

In order to accomplish this goal, a variety of time series strategies for predicting price series of various commodities were investigated and assessed in the papers that were published.[13–35] For the purpose of price series forecasting, previous research made a significant contribution to the creation of vector autoregressive models (VARs), autoregressive integrated moving average models (ARIMAs), and vector error correction models (VECMs)-like approaches. Although there are different procedures that provide more changes and additions,[36–39] the majority of these operations could be much more challenging. Ever since it was originally established, the univariate ARIMA model, which makes use of historical data produced from previously acquired values, has been the driving force behind the effective application of forecasting. This model makes use of past data. The performance of VAR approaches is superior than that of ARIMA models when it comes to the information sets that are used for prediction. This is because VAR techniques take into account the interactions that occur between a number of different integrated economic components. The performance of VECM models is much superior to that of VAR models when it comes to illuminating the long-term correlations that exist between a number of different economic indices. To this goal, the model includes the notion of cointegration directly. Stated differently, a primary driving factor for the construction of the VECM was the inclusion of projections for the far future.

The various technological breakthroughs over the previous many years have enabled us to have access to substantially more powerful processing power than we had just a few years ago.[40–46] In line with this trend, the domains of finance and economics are devoting greater attention to artificial intelligence (AI) and machine learning (ML) technologies.[47] Moreover, recent research has proved the great potential for anticipating patterns in time series of commodity prices by the use of machine learning and artificial intelligence technology.[48] These attempts in price forecasting powered by machine learning have dealt with soybeans,[49] cotton,[50] wheat,[51] soybean oil,[52] coffee,[53–59] corn,[60] sugar,[61] canola,[62] and oats,[63] using the area of the agriculture business as an example. These price prediction studies based upon machine learning have considered electricity,[64] energy securities,[65] heating oil,[66] carbon emission allowances,[67] natural gas,[66] crude oil,[66] and coal[68] in relation to the energy sector. Within the metals sector, platinum,[69] steel,[70] silver,[71] copper,[72] gold,[73] palladium,[74] and lead[75] have been the focuses of several machine learning-based price predicting studies. Multiple machine learning algorithms that are often investigated for commodity price predictions have included K-nearest neighbors, genetic programming, neural networks,[60] extreme learning, support vector regressions, ensembles, deep learning, decision trees, boosting, random forests, and multivariate adaptive regression splines. Previous studies, as reported in the literature, indicate that neural networks are often used by researchers as modeling techniques to solve issues pertaining to commodity price forecasting.[76–78] Despite the limited sample size we looked at, our results generally support previous studies showing the advantages of using AI/ML approaches to financial and economic projections.[79–83] There has not been much in-depth study done on the price predictions made by Gaussian process regressions using time-series data collected for a variety of commodities, such as coffee prices. An attempt is made to address this prediction challenge in this work.

The Gaussian process regressions served as the foundation for the prediction methods that this investigation created. The daily time series of coffee prices is analyzed using the information gathered between 2 January 2013 and 10 April 2024. Model estimation and training are made straightforward by using these methods within the framework of cross-validation and Bayesian optimization methodologies. The models that are being studied show an adequate level of projected accuracy for the out-of-sample time frame from 3 January 2022 to 10 April 2024, with a relative root mean square error of 2.0500%. The empirical data may be used to provide technical predictions for relevant policy research, either alone or in combination with other expected results. Other economic areas with comparable prediction problems might benefit from similar forecasting tools.

The remainder of this work is organized as follows. Section 2 provides a review of the literature. Section 3 describes the data used for analysis. Section 4 discusses the method for constructing forecast models. Section 5 reports the estimated models and forecast results. Section 6 conducts benchmark analysis. Section 7 concludes this study.

2. Literature Review

For a variety of businesses that rely on commodities, economists have had to devote a great deal of time and effort to overcoming the issue of developing accurate and dependable commodity price estimates.[7,84] For instance, despite being extensively used in earlier research, ARIMA models are still in great demand for a variety of prediction tasks, such as for price data obtaining from time series databases. For example, Naveena[85] examined the ARIMA model, generalized auto regressive conditional heteroscedastic model, and artificial neural network model for forecasting different coffee price time series and found that the optimal forecasting model for the price depends on the specific coffee product under consideration. Aduteye *et al.*[86] determined that the ARIMA(0,1,2) model offers optimal forecasts for global coffee prices on a month basis from 1990 to 2023. One econometric tool that is often used for price series forecasts is the VAR technique. It makes use of the connections between the main economic components.[87,88] For instance, Dang *et al.*[89] utilized the VAR model in analyzing the lead-lag relationship between Vietnamese coffee export prices and exchange rates and found that these does not exist long-run relationships between these two economic variables, although there is a short-run lead-lag causal relationship that might help short-run price forecasts. Musumba and Gupta[90] combined the VAR model with the directed acyclic graph for evaluating contemporaneous causal relationships between world coffee prices and prices obtained by Ugandan coffee growers and determined that a positive price shock to London futures prices and indicator prices would have positive and negative influences on prices obtained by Ugandan coffee growers, respectively. Long-term correlations or linkages between many financial elements may be included in VAR-based models, also known as VECMs, due to cointegration.[91–93] This might have a significant impact on long-term estimates for a number of price time series. For example, Milas *et al.*[94] investigated the price forecasting problems for four different coffee types through linear and nonlinear VECMs and found that asymmetric and polynomial VECMs lead to better forecasting accuracy as compared to random walk models. Cariappa and Sinha[95] applied the VECM for investigating price relationships between Arabica and Robusta coffee and found that these two price series are cointegrated with Robusta coffee prices influencing Arabica coffee prices, suggesting predictive information flows.

There appears to have been an upswing in the quantity of studies undertaken in recent years examining the use of machine learning methods on commodity price forecasts.[96–98] For instance, Atsalakis[99] proposed an adaptive neuro fuzzy inference system for forecasting prices of coffee, sugar, cocoa, and wheat, where neural network models and fuzzy logic approaches are combined, and determined that the proposed system leads to better forecast accuracy as compared to a feedforward neural network model and ARIMA models. Mekala *et al.*[4] built a bidirectional long short-term memory convolutional neural network model for forecasting coffee prices and determined that this method outperformed several other machine learning models and traditional time-series models in terms of forecast accuracy. Fofanah[100] explored usefulness of the long short-term memory neural network model, extreme gradient boosting model, and linear regression model for forecasting coffee prices. Efendi and Ardhy[101] found that the backpropagation neural network model could serve as a useful tool for generating forecasts of coffee prices. Herrera-Jaramillo *et al.*[102] proposed the use of the long short-term memory neural network model for the price forecast purpose for Colombian coffee. Le Ngoc *et al.*[103] investigated the potential of several different machine learning models for coffee price forecasts in the Vietnam market, including the long short-term memory neural network model, gated recurrent unit neural network model, random forest model, and support vector regression model, in addition to traditional time-series models, including the ARIMA model and seasonal ARIMA model.

Neal[104] combined the construction of neural networks with Bayesian learning to create a ground-breaking regression technique. The main emphasis of this method is using priors to assess noisy data over Gaussian processes.[105–107] Numerous neural network-based methods for Bayesian regression have shown a tendency to converge towards Gaussian processes inside the infinite network border region.[104,108–110] This modeling technique provides an explanation of how to use Gaussian process regressions in both chaotic[111–113] and noise-free[65,114,115] environments. Brahim-Belhouari and Vesin[116] provide an example of how Bayesian learning may be used in the context of stationary time-series data by using an analysis of the performance of radial basis function neural networks with Gaussian processes. In light of the findings of the study conducted by Bermak and Brahim-Belhouari,[117] a broad range of distinct covariance functions must be used in this investigation. In the end, creating projections from time series that already exhibit nonstationary patterns may greatly benefit from the use of Gaussian process methods.[118–120] Gaussian process regression techniques are likely to provide better results than radial basis function neural networks, according to the research findings of Bermak and Brahim-Belhouari.[117] The Gaussian process formulations also have the benefit of using the precise matrix operations needed to integrate prior and noisy models.[117] We would use model averaging approaches to our instance that are similar to the Gaussian process predictors used by Bermak and Brahim-Belhouari[117] in their multi-model predictions.

The purpose of this research project is to create Gaussian process regression models in order to manage the complicated challenges associated to coffee price forecasts. To authors' knowledge, Gaussian process regression models, as a machine learning technique, have not been sufficiently explored for the purpose of forecasting coffee prices. The price of coffee is an important factor that both policymakers and market participants need to take into account while gathering data about the cost of agricultural commodities. Furthermore, the price of coffee has a significant effect on the economy. For these models to aid in decision making and price trend analysis, the results they provide from their projections must be trustworthy and precise. With the achievement of rather accurate and stable forecast results, our work could contribute to enriching the information set a decision maker could utilize for arriving at prudent choices. This study also investigates the forecasting issue by focusing on price data recorded on a daily basis, which could help decision making in a timely manner.

3. Data

Figure 1 illustrates the daily price of coffee from 2 January 2013 to 10 April 2024, for the sake of this specific inquiry. The data are sourced from Macrotrends. It is also critical to keep in mind that the first price differences are shown on this chart. Figure 1 also includes 50-bin histograms, kernel estimates, and quantile–quantile

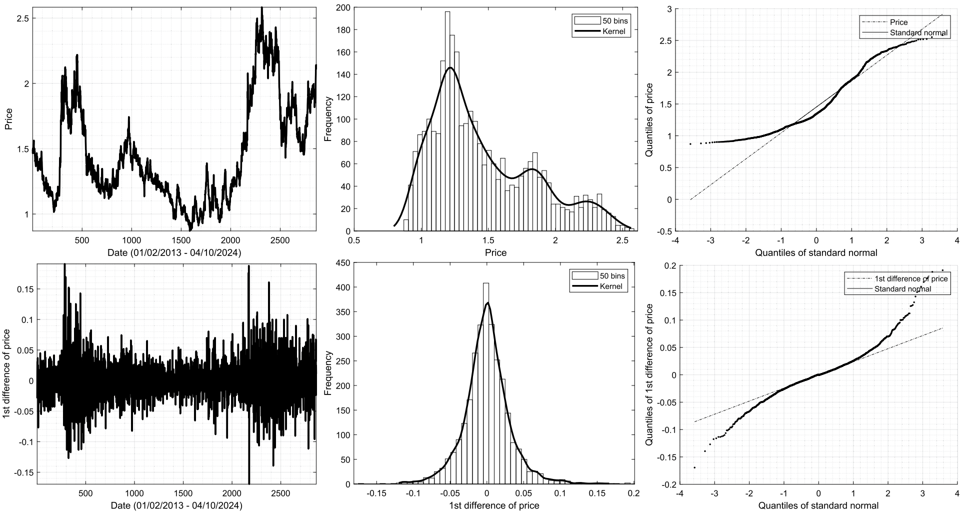


Fig. 1. Daily coffee prices during 2 January 2013–10 April 2024.

plots for the prices as well as their first differences. An overview of the price data's time-series is shown in Table 1. The Jarque–Bera[121,122] and Anderson–Darling[123] test findings unequivocally demonstrate that the prices are not in line with a normal distribution. A finding like this would not be shocking considering the features of numerous financial and economic series.[70,124,125] The platykurtic form and right skewness of the price time series could be determined.

Numerous investigations have been carried out about the appearance of nonlinear features in various time series related to the financial and economic spheres.[126–130] The daily coffee price data are evaluated using the Brock–Dechert–Scheinkman[131] test technique in order to look for any potential nonlinearities. The ϵ distance must be employed in addition to the embedding dimensions, which vary from two to 10, to conclude this inquiry. The ϵ distance, which may be represented by any of the following numbers: 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0, can be multiplied by the price standard deviation to get the distances between the various data points. The p-value was consistently extremely close to zero across all test results. It is important to note that these data support the hypothesis that there are nonlinearities in the price series. In this study, prices were estimated while taking into consideration the nonlinear aspects of the data using Gaussian process regressions.

4. Method

The primary instrument used in the prediction methodology of this work is Gaussian process regressions, or GPRs. The GPR is a sort of probabilistic kernel architecture which is utilized in different scientific domains to anticipate various nonlinear events.[132–139] Equation $\{(x_i, y_i); i = 1, 2, \ldots, T\}$ is used by the model to represent the unknown distribution in relation to the training data. Utilizing the formulae

Table 1. Summary statistics of daily coffee prices during 2 January 2013–10 April 2024.

Series	Minimum	1st percentile	5th percentile	Mean	Median	Standard deviation	95th percentile	99th percentile	Maximum	Skewness	Kurtosis	Jarque–Bera	Anderson–Darling
Price	0.870500	0.921534	0.981500	1.456450	1.345000	0.381639	2.242500	2.409805	2.583500	0.809099	2.761517	< 0.001	< 0.0005
First differences	−0.169500	−0.081760	−0.050000	0.000228	0.000000	0.031992	0.051000	0.094880	0.191000	0.383803	6.548205	< 0.001	< 0.0005

$y_i \in \mathbb{R}$ and $x_i \in \mathbb{R}^d$, respectively, the target variable and the d-dimension predictor are expressed. In this particular scenario, the 10-lag prices will serve as the variables that are used for prediction. To put it another way, the prices found over the course of the previous 10 days would be used in order to make an estimate of the price on the 11th day.

The symbols $y = x^T\beta + \varepsilon$ and $\varepsilon \sim N(0, \sigma^2)$ are the representations of these formulae, which stand for linear regression and error, respectively. The estimates of the target variables for GPRs, on the other hand, are supplied by basis functions and latent variables throughout the process. For the purpose of representing basis functions, the symbol b is used, while the symbol $l(x_i)$ is employed to represent latent variables that satisfy the joint Gaussian requirement. In order to determine whether or not a target variable is smooth, we can find it useful to examine the covariance functions of the latent variables. Through the use of basis functions, the predictors may be mapped to the feature space.

When generating Gaussian processes, it is usual practice to make use of the variance and mean metrics.[140,141] These processes are represented by the notation GPs. In order to display the mean, it is recommended to use the equation $m(x) = E(l(x))$. On the other hand, the covariance may be shown by using the equation $k(x, x') = \mathrm{Cov}[l(x), l(x')]$. The formula $y = b(x)^T\beta + l(x)$ has the potential to be used for the representation of GPRs, provided that $b(x) \in \mathbb{R}^p$ and $l(x) \sim \mathrm{GP}(0, k(x, x'))$. One useful approach to guarantee that the numerical value of θ is created correctly is to apply the $k(x, x'|\boldsymbol{\theta})$ hyper-parameter. It is typical practice in GPR training to employ estimates for β, σ^2, and θ. The model gets the required basis functions (b) and kernels (k) during training.[142,143] The major emphasis of this study is on isotropic and nonisotropic kernels. We analyze five distinct kernel settings for these two sorts of kernels. The explanation of each kernel examined in this study is provided in depth by Eqs. (1) through (10). α represents the scale-mixture parameter, and it is positive. A metric called σ_l is used to represent length scales with isotropic kernel characteristics. The standard deviation of the signal is represented by the metric σ_f, and $r = \sqrt{(x_i - x_j)'(x_i - x_j)}$. The $\theta = (\theta_1, \theta_2) = (\log \sigma_l, \log \sigma_f)$ formula states that σ_f and σ_l are always greater than zero. A nonisotropic kernel's length scale might be represented by the symbol σ_m $(m = 1, 2, \ldots, d)$. Considering this, $\theta = (\theta_1, \theta_2, \ldots, \theta_d, \theta_{d+1}) = (\log \sigma_1, \log \sigma_2, \ldots, \log \sigma_d, \log \sigma_f)$ would represent θ.

$$\text{Isotropic Exponential: } k(x_i, x_j|\theta) = \sigma_f^2 e^{-\frac{r}{\sigma_l}} \tag{1}$$

$$\text{Isotropic Squared Exponential: } k(x_i, x_j|\theta) = \sigma_f^2 e^{-\frac{1}{2}\frac{(x_i - x_j)^T(x_i - x_j)}{\sigma_l^2}} \tag{2}$$

$$\text{Isotropic Matern 5/2: } k(x_i, x_j|\theta) = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right) e^{-\frac{\sqrt{5}r}{\sigma_l}} \tag{3}$$

$$\text{Isotropic Rational Quadratic: } k(x_i, x_j|\theta) = \sigma_f^2 \left(1 + \frac{r^2}{2\alpha\sigma_l^2}\right)^{-\alpha} \tag{4}$$

$$\text{Isotropic Matern 3/2: } k(x_i, x_j|\theta) = \sigma_f^2 \left(1 + \frac{\sqrt{3}r}{\sigma_l}\right) e^{-\frac{\sqrt{3}r}{\sigma_l}} \tag{5}$$

$$\text{Nonisotropic Exponential: } k(x_i, x_j|\theta) = \sigma_f^2 e^{-\sqrt{\sum_{m=1}^{d} \frac{(x_{im}-x_{jm})^2}{\sigma_m^2}}} \tag{6}$$

$$\text{Nonisotropic Squared Exponential: } k(x_i, x_j|\theta) = \sigma_f^2 e^{-\frac{1}{2}\sum_{m=1}^{d} \frac{(x_{im}-x_{jm})^2}{\sigma_m^2}} \tag{7}$$

Nonisotropic Matern 5/2:

$$k(x_i, x_j|\theta) = \sigma_f^2 \left(1 + \sqrt{5\sum_{m=1}^{d} \frac{(x_{im}-x_{jm})^2}{\sigma_m^2}} + \frac{5}{3}\sum_{m=1}^{d} \frac{(x_{im}-x_{jm})^2}{\sigma_m^2}\right) \times e^{-\sqrt{5\sum_{m=1}^{d} \frac{(x_{im}-x_{jm})^2}{\sigma_m^2}}} \tag{8}$$

Nonisotropic Rational Quadratic:

$$k(x_i, x_j|\theta) = \sigma_f^2 \left(1 + \frac{1}{2\alpha}\sum_{m=1}^{d} \frac{(x_{im}-x_{jm})^2}{\sigma_m^2}\right)^{-\alpha} \tag{9}$$

Nonisotropic Matern 3/2:

$$k(x_i, x_j|\theta) = \sigma_f^2 \left(1 + \sqrt{3\sum_{m=1}^{d} \frac{(x_{im}-x_{jm})^2}{\sigma_m^2}}\right) e^{-\sqrt{3\sum_{m=1}^{d} \frac{(x_{im}-x_{jm})^2}{\sigma_m^2}}} \tag{10}$$

Equations (11)–(14) display the four distinct basis functions which are the focus of this investigation. The study of these functions might be comparable to the scenario when several kernels are taken into account. With reference to the four basis functions that are being examined,

$$X = (x_1, x_2, \ldots, x_n)',$$

$$X^2 = \begin{pmatrix} x_{11}^2 & x_{12}^2 & \cdots & x_{1d}^2 \\ x_{21}^2 & x_{22}^2 & \cdots & x_{2d}^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_{T1}^2 & x_{T2}^2 & \cdots & x_{Td}^2 \end{pmatrix}, \quad \text{and}$$

$$B = (b(x_1), b(x_2), \ldots, b(x_n))'.$$

$$\text{Empty: } B = \mathit{Empty\ Matrix} \tag{11}$$

$$\text{Constant: } B = I_{n\times 1} \tag{12}$$

$$\text{Linear: } B = [1, X] \tag{13}$$

$$\text{Pure Quadratic: } B = [1, X, X^2] \tag{14}$$

In the process of developing the GPR, it is possible to take into consideration a wide variety of basis functions and kernels which act as the foundation. It has been discovered that the 10 kernels and four basis functions that are presented here have a high capacity to represent a wide range of complicated data forms. It is common practice to investigate them across a wide range of academic disciplines.[104,144] Because of this, they are taken into account in this work.

Using the expected improvement per second plus ($EIPSP$) method and 10-fold cross-validation procedures, the parameters of the model are determined in Bayesian optimizations. In order to shed some light on the situation, we shall use a GP model with the notation $f(x)$. After the data points that are within the variable's boundaries, or x_is, are randomly selected, Bayesian methods are used to evaluate the associated $y_i = f(x_i)$. In this case, the initial evaluation would utilize the value of $N_s = 4$. The procedure will continue gathering data points in the event that an assessment has faults until N_s examples of successful assessments are found. The next two steps will then be completed in order to put the strategy into practice. One may utilize the updates of the function $f(x)$ to generate a posterior distribution over $Q(f|x_i, y_i$ for $i = 1, \ldots, T)$. Finding a new data point, denoted by x, is the major purpose of the second phase, which tries to minimize the acquisition function, denoted by $a(x)$. To find the magnitude which x gets integrated into Q, $a(x)$ would be used. Rather than concentrating on any value that may raise it, the expected improvement acquisition functions would aim to improve a target function. One may claim that the posterior mean which is produced at x_{best} is the lowest possible value that could have been attained. The lowest value, represented by the symbol $\mu_Q(x_{\text{best}})$, is the one that may be linked to x_{best}. The expected improvement, represented by the symbol EI, would look like this: $\text{EI}(x, Q) = E_Q[\max(0, \mu_Q(x_{\text{best}}) - f(x))]$. Time-weighting techniques may be used by the acquisition functions of Bayesian processes to optimize the benefits for each time unit. This is because, depending on the location, the target assessment time changes. The advancement of optimization methods has made it feasible to use Bayesian models to calculate the time as a function of x for evaluating the goals. The EI per second, or EIPS, of the acquisition functions could be represented as $\text{EIPS}(x) = \frac{\text{EI}_Q(x)}{\mu_S(x)}$. The notation $\mu_S(x)$ represents the posterior mean about the extra-time GP. As a result, these behavioral adjustments stop acquisition functions from enforcing local minimums or functioning on certain target regions. The posterior standard deviation in the case of additive noises would satisfy the $\sigma_Q^2(x) = \sigma_F^2(x) + \sigma_{PN}^2$ requirements. The posterior objective's standard deviation in relation to x is $\sigma_F(x)$. The formula $t_{\sigma_{PN}} > 0$ is used to display the exploration ratio. At the conclusion of each cycle, the acquisition function applies the EIPSP method to confirm that the supplied data point (x) complies with the $\sigma_F(x) < t_{\sigma_{PN}} \sigma_{PN}$ protocol. The kernel function expands by the number of iterations when this condition is met, indicating that x is being overexploited.[145] The value of σ_Q may increase significantly as data points migrate between observations when the EIPSP technique is used. This suggests that one has to use the freshly fitted kernel in order to produce a new data point. If it is shown to be exploitative in the following trial, θ will expand by a factor of 10. This method might be applied up to five times in order to find an x that is not overexploited. Lastly, the revised x — which will be the subsequent exploration ratio — is produced using the EIPSP methodology. To provide a more precise result, the algorithm makes advantage of nearby data points that were later acquired in addition to previously assessed data points. When executing Bayesian optimization, it is essential to assess the condition of basis functions, σ, kernels, and predictor standardization.

The relative root mean square error, or RRMSE for short, is a frequently used statistic for assessing prediction accuracy. This error measure is often used to assess how well various predictions made from various models and goals perform.[146] The RRMSE is shown as follows: $\text{RRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{\text{obs}} - y_i^{\text{for}})^2}}{\frac{1}{n}\sum_{i=1}^{n} y_i^{\text{obs}}}$, where the notations n, y^{obs}, and y^{for} denote the number of observations needed in addition to the target's observed and predicted values to calculate prediction accuracy. Furthermore, the root mean square error (RMSE) and mean absolute error (MAE) are used to evaluate prediction accuracy. The RMSE is stated as follows: $\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{\text{obs}} - y_i^{\text{for}})^2}$, and this is how the MAE is expressed: $\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i^{\text{obs}} - y_i^{\text{for}}|$. Lastly, an additional important indicator is the correlation coefficient (CC), represented as $CC = \frac{\sum_{i=1}^{n}(y_i^{\text{obs}} - \overline{y^{\text{obs}}})(y_i^{\text{for}} - \overline{y^{\text{for}}})}{\sqrt{\sum_{i=1}^{n}(y_i^{\text{obs}} - \overline{y^{\text{obs}}})^2 \sum_{i=1}^{n}(y_i^{\text{for}} - \overline{y^{\text{for}}})^2}}$, where $\overline{y^{\text{obs}}}$ shows the average of y^{obs} and $\overline{y^{\text{for}}}$ expresses the average of y^{for}.

5. Result

The model was trained from 2 January 2013 to 31 December 2021. For the period starting on 3 January 2022 and concluding on 10 April 2024, the predictions are produced one day ahead of schedule. The EIPSP optimizations carried out during the training phase are shown in Fig. 2. Taking into account the optimization outcomes presented in Fig. 2, it was found that the nonisotropic Matern 5/2 kernel expressed in Eq. (8), the empty basis function expressed in Eq. (11), and standardized predictors should be the optimal choice for the coffee price series. 10 different GPR models were created with the use of 10-fold cross-validation. The results of the parameter estimations for the models are shown in Table 2. The parameter estimate results are shown in Table 2 for each column designated as "CV1" through "CV10" (where "CV" stands for "cross validation").

The price prediction technique used 10 GPR models reported in Table 2, labeled "CV1" through "CV10." For the course of the out-of-sample evaluation period, which spans from 3 January 2022 to 10 April 2024, these models supply daily price forecasts. Data collected between 2 January 2013 and 31 December 2021 were used

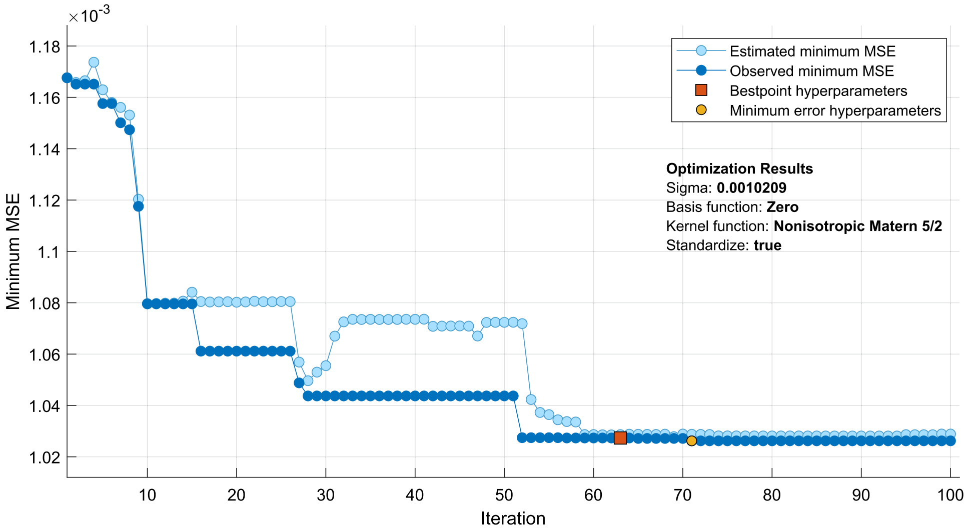


Fig. 2. Bayesian optimization results.

to train these models. Each day's final price projection for the testing period will be based on the average of the 10 forecasts. As a consequence, there is less chance that any of the sub-models would provide unexpected findings, which increases the consistency and reliability of future projections. Some of the features that forecast users search for, as well as the benefits of equal weighting systems, have been studied in the scientific literature.[147] Figure 3 displays the differences in the time-series patterns of the actual and forecasted prices. Figure 4 presents statistics on the percentage of price prediction errors deemed relevant to our inquiry, so supporting all the conclusions that can be made from Fig. 3. Table 3 shows the performance metrics for the predicted results. These indicators are produced by integrating the RRMSE, RMSE, MAE, and CC. More specifically, Table 3 displays performance statistics exactly as they are presented in Figs. 3 and 4. It is evident that the RRMSE of 2.0500% is provided by the GPR models. Numerous criteria have been devised by the research literature to assess the degree of accuracy that different models provide[146]: an excellent level of accuracy achieved by the predictions is defined as a RRMSE of less than 10%; a good level of accuracy achieved by the predictions is defined as a RRMSE spanning from 10% to 20%; a fair level of accuracy achieved by the predictions is defined as a RRMSE spanning from 20% to 30%; and a poor level of accuracy achieved by the predictions is defined as a RRMSE of above 30%. We achieve an excellent level of accuracy for our models by following these grading guidelines.

The results of an error autocorrelation evaluation, which was performed to determine the suitability of the developed models, are shown in Fig. 5. This study, in particular, investigates normalized autocorrelations for delays up to 20. This study confirms the overall appropriateness of the models by eliminating the possibility of

Table 2. Model parameter estimate results.

	CV1	CV2	CV3	CV4	CV5
σ	0.0316024781	0.0321450118	0.0324680023	0.0313472193	0.0330693239
σ_1	21,523.1455417766	141,235.8907672120	98,456.6306385778	151.5809166054	2,004,009.0766677000
σ_2	466.7443903185	512,622.0228384420	43,846.0017693413	229.4578580278	385,798.3346480900
σ_3	404.9469570530	2131.1940293467	2855.4046303594	127.9376647741	90,685.0701361651
σ_4	430,332.9289044760	57,928.4136271582	83,548.1505093005	97,078.9571535160	294,392.0602567220
σ_5	808.8074732810	67,608.4932248984	197,657.5878923900	44,484.4588268253	552,718.5332443360
σ_6	401.3652225063	686,486.7113697110	2,651,933.5188598600	9.3098009220	137,780.5885071630
σ_7	7736.9483902691	12,124.3656909790	410.1910079345	116.6780881167	6,266,298.9647889300
σ_8	2299.7134437913	16,579.8055035651	42,688.3892462008	4799.5646630601	5,844,458.6457370400
σ_9	548.9847363144	937.7346478644	311.4984441179	9.6476386854	40,735.1396079548
σ_{10}	49.0866030342	33.2541187305	39.2880207299	11.7167133959	54.5220211743
σ_f	7.9402054979	5.2567073501	6.2674514094	2.1947489031	8.9401402882
	CV6	CV7	CV8	CV9	CV10
σ	0.0319355492	0.0318325181	0.0315929718	0.0319464588	0.0324463079
σ_1	84,945.7486180479	29,024.2735850609	209,731.4134861370	1,799,896.3230076000	16,611,814.3140678000
σ_2	63,398.2225639125	703.7486368855	108,917.4409989690	297,404.9527725470	21,132.3590023869
σ_3	25,541.2594183395	521.0562785577	1836.7300389640	1386.8756124788	994.9670118928
σ_4	33,675.6784304201	89,958.9535767793	2,674,835.9885092400	721,325.6936906130	41,535.1637090119
σ_5	46,549.0193737037	1904.1122080308	106,519.0507099030	88,643.0355597833	106,091.5920160170
σ_6	30,715.7622163155	36,237.5555412247	4,237,385.5496068600	93,884.2540507468	22,012.3602537443
σ_7	406.2624879432	556.4385397289	2308.4413168355	825.0719810774	476.2957026172
σ_8	209,847.3203260270	71,844.3512669834	45,216.7458859992	21,729.7232810399	36,879.9144546768
σ_9	352.5846018844	618.6670374734	24,053.2603009775	784.7275246961	337.8815328049
σ_{10}	38.1358926762	60.0092112247	75.7120782796	54.2348977144	41.8544887788
σ_f	6.1731458766	9.8224827274	12.3637017981	8.6461782136	6.5537622080

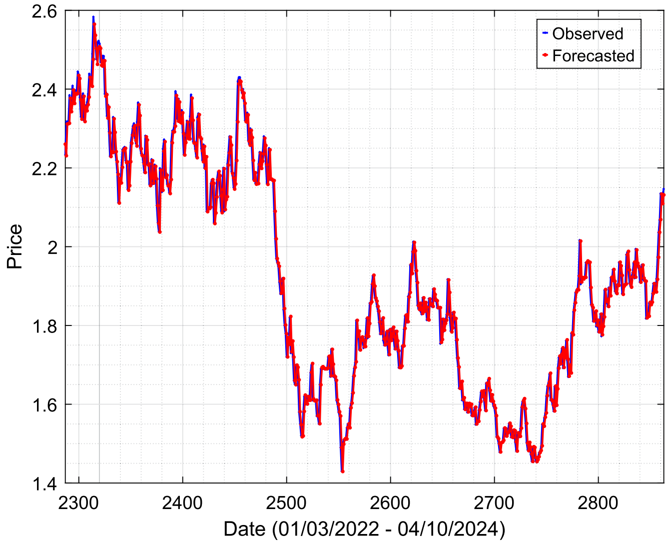


Fig. 3. Out-of-sample forecast results during 3 January 2022–10 April 2024.

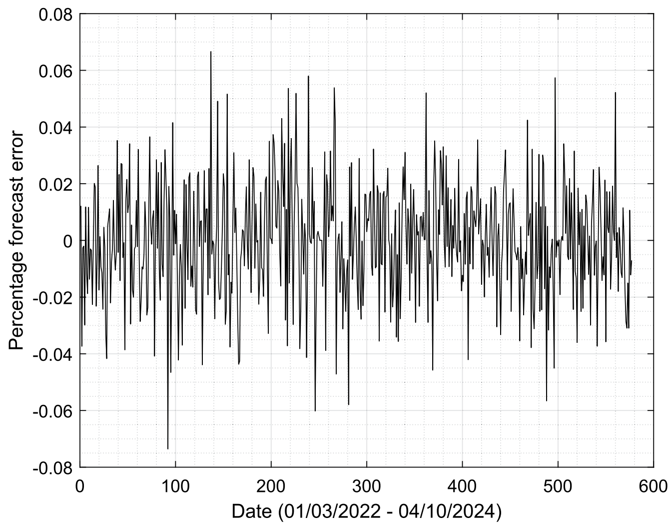


Fig. 4. Out-of-sample percentage forecast errors during 3 January 2022–10 April 2024.

Table 3. Out-of-sample forecast performance during 3 January 2022–10 April 2024.

RRMSE	RMSE	MAE	CC
2.0500%	0.0395	0.0302	99.057%

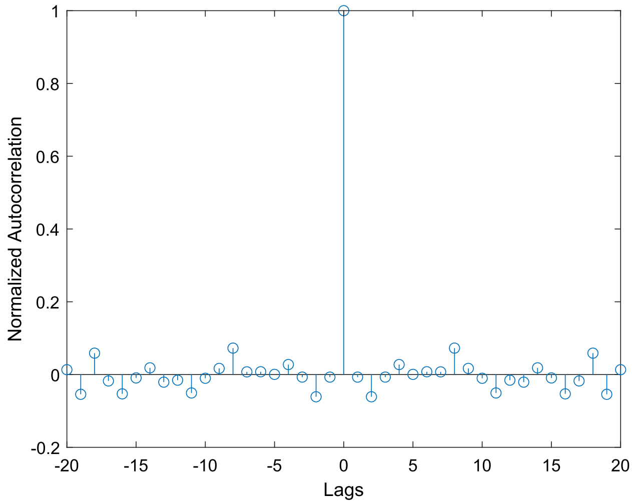


Fig. 5. Error autocorrelations.

major error autocorrelations. It is crucial to remember, however, that real evidence from the literature suggests that, depending on the circumstance, autoregressive conditional heteroskedasticity effects may enhance or degrade the forecasting output of different time-series models e.g., Refs. 148 and 149.

To further test the models' generalization capability, we apply the models reported in Table 2 to a more recent out-of-sample time period of 5/01/2024 – 08/30/2024. The forecast results are shown in Fig. 6, with the RRMSE of 1.1894%, RMSE of 0.0274, MAE of 0.0234, and CC of 98.36%, which are close to those shown in Table 3.

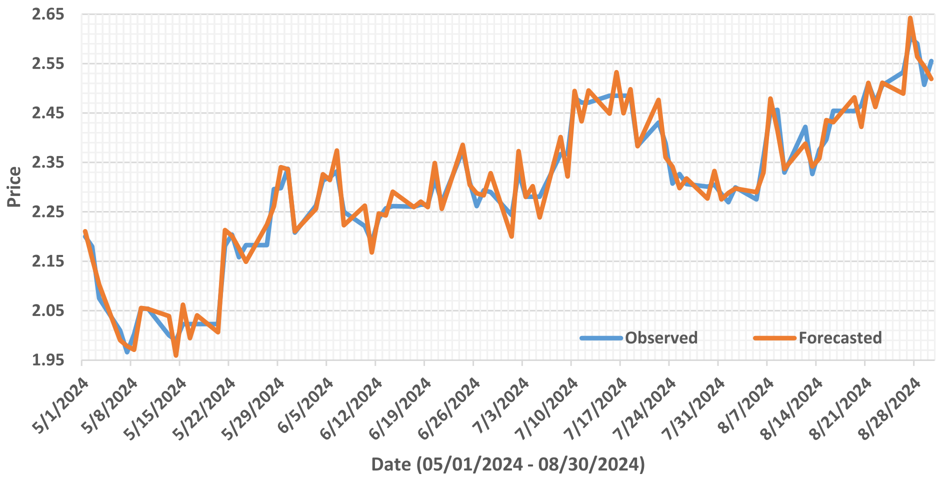


Fig. 6. Out-of-sample forecast results during 1 May 2024–30 August 2024.

6. Benchmark Analysis

The GPR models are compared with several different benchmark models, including the support vector regression model (SVR), regression tree model (RT), autoregressive model (AR), and AR-generalized autoregressive conditional heteroskedasticity model (AR-GARCH). All of these benchmark models use the same predictors as

Table 4. Benchmark analysis: Comparisons of model forecast performance.

Model	RRMSE	RMSE	MAE	CC
GPR	2.0500%	0.0395	0.0302	99.057%
SVR	2.3557%	0.0453	0.0397	98.776%
RT	3.5602%	0.0685	0.0595	97.330%
AR	5.8188%	0.1120	0.0961	93.267%
AR-GARCH	5.4306%	0.1045	0.0893	93.971%

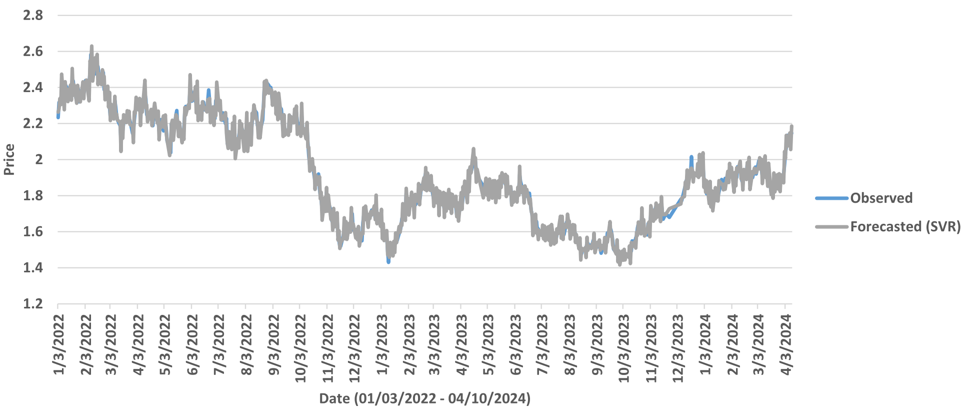


Fig. 7. Benchmark analysis: Plots of forecasted against observed prices based on different models for the out-of-sample period of 3 January 2022–10 April 2024.

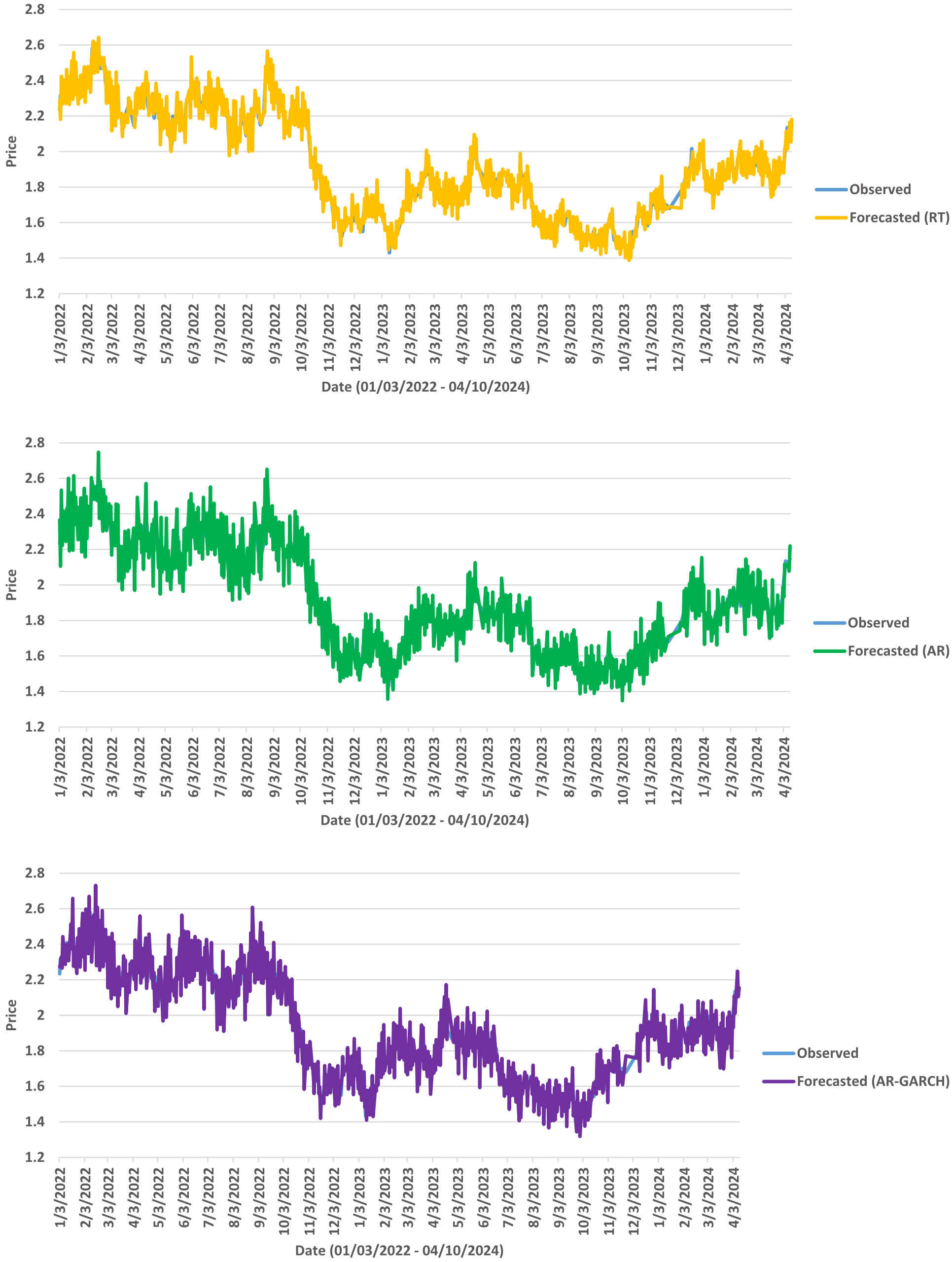


Fig. 7. (*Continued*)

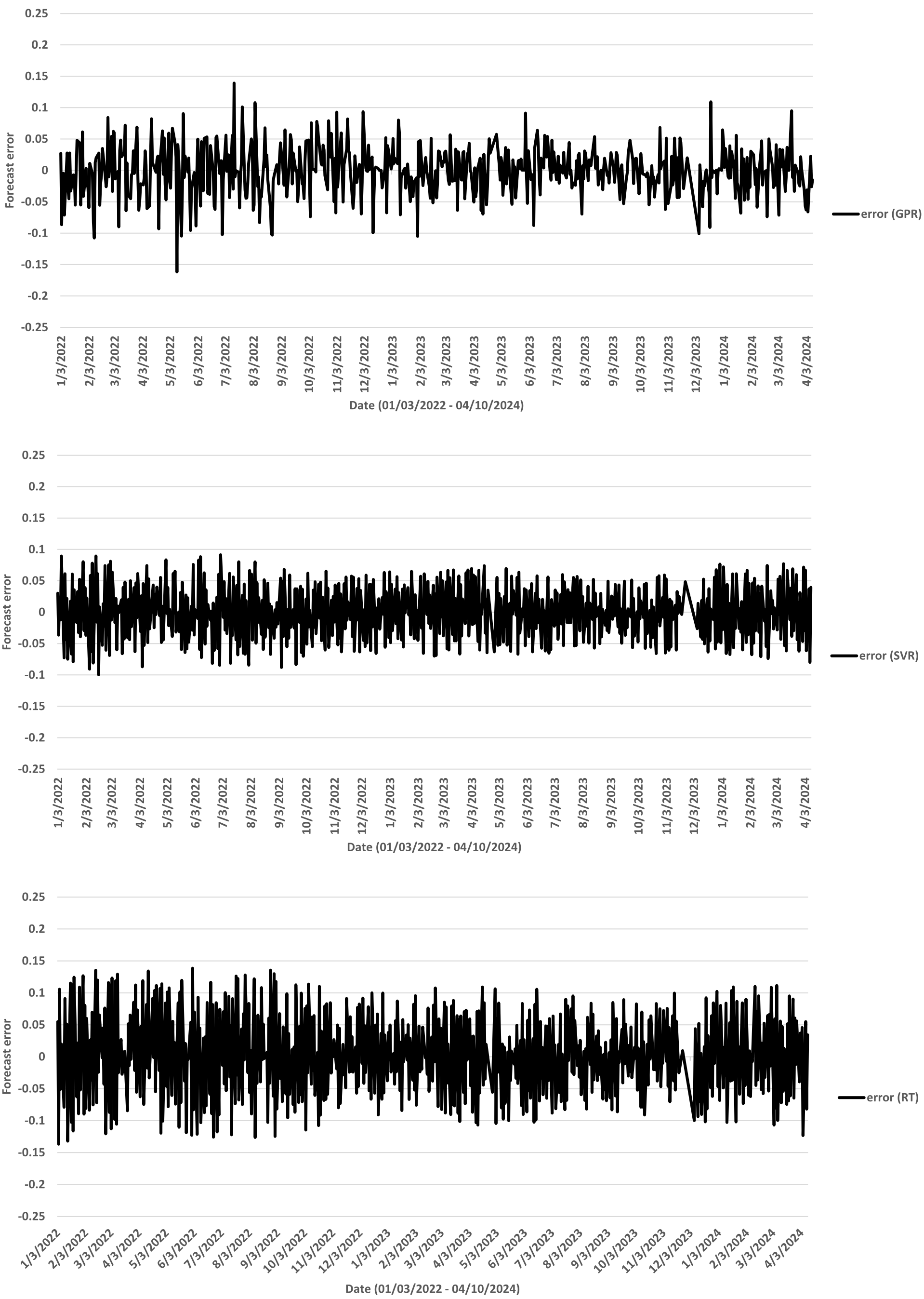
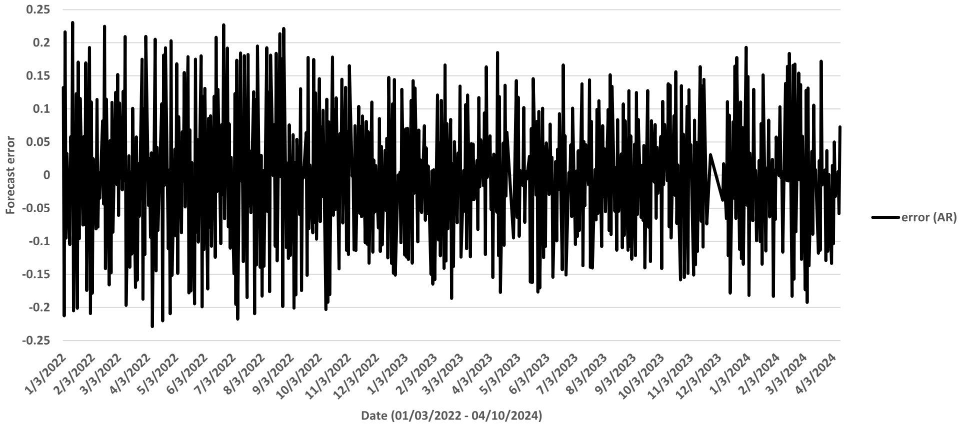


Fig. 8. Benchmark analysis: Plots of price forecast errors based on different models for the out-of-sample period of 3 January 2022–10 April 2024.



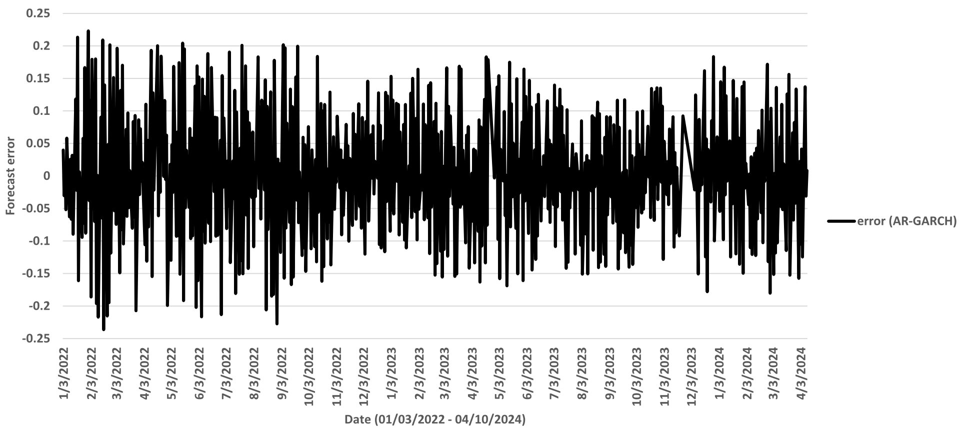


Fig. 8. (*Continued*)

the GPR models. Table 4 shows the summary of model forecast performance comparisons, where it could be seen that the GPR models lead to the lowest RRMSE, RMSE, and MAE, and the highest CC. Detailed visualization of forecasted against observed prices based on different models are shown in Fig. 7. Correspondingly, detailed visualization of forecast errors based upon different models are shown in Fig. 8.

7. Conclusion

Previous research has shown the benefits of decision makers and policymakers having a thorough grasp of commodity price projections. Importance of forecasts of coffee prices is no exception. This study investigates the problem of coffee price projections using daily data from 2 January 2013 to 10 April 2024. In this case, the model parameters are found using the cross-validation approach and Bayesian optimization methods, and the predictions are produced using Gaussian process regressions. The relative root mean square error during the testing period outside of the training sample, which ran from 3 January 2022 to 10 April 2024, was found to be 2.0500%. Technical prediction data may be utilized alone or in combination with other findings for policy research on pricing trends. The methods we used to evaluate these problems may be useful in predicting similar challenges for other economic sectors. This empirical paradigm's versatility and ease of use may be advantageous for a wide range of decision-making procedures. The prices of other commodities which might be important indicators are not taken into consideration in this analysis, nor are macroeconomic movements. Incorporating these extra data into forecasts may help increase accuracy even further. This suggests that the recommended strategy may not work as well in this specific situation. Gaussian process regressions, which may include other exogenous variables to examine these economic factors in more depth, are one method for examining this obstacle. This work concentrates on the expected improvement per second plus algorithm for Bayesian optimizations and future studies could explore other optimization algorithms.

Data Availability Statement

Data are available upon request.

ORCID

Bingzi Jin https://orcid.org/0009-0005-1620-7772
Xiaojie Xu https://orcid.org/0000-0002-4452-1540

References

1. X. Xu and Y. Zhang, Thermal coal price forecasting via the neural network, *Intell. Syst. Appl.* **14** (2022) 200084, 10.1016/j.iswa.2022.200084.
2. X. Xu and Y. Zhang, Coking coal futures price index forecasting with the neural network, *Miner. Econ.* **36** (2023) 349–359, 10.1007/s13563-022-00311-9.
3. Y. A. Herrera-Jaramillo, J. Ortega-Giraldo, A. Acevedo-Amorocho and D. Prada-Marín, Coffee price forecast: A proposal from the tools of logistics 4.0.
4. K. Mekala, V. Laxmi, H. Jagruthi, S. A. Dhondiyal and A. P. Dabral, Coffee price prediction: An application of CNN-BLSTM neural networks, *2023 Int. Conf. Advances in Computing, Communication and Applied Informatics* (*ACCAI*) (IEEE, 2023), pp. 1–7.
5. B. Yashavanth, K. Singh, A. K. Paul and R. K. Paul, Forecasting prices of coffee seeds using vector autoregressive time series model, *Indian J. Agric. Sci.* **87** (2017) 754–758.
6. N. D. Tan, H. C. Yu, L. N. B. Long and S.-S. You, Data analytics and optimised machine learning algorithm to analyse coffee commodity prices, *Int. J. Sustain. Agric. Manag. Inform.* **8**(4) (2022) 345–366, 10.1504/IJSAMI.2022.126799.
7. D. G. Ababu and A. M. Getahun, Time series analysis of price of coffee in case of Mettu town, Ilu ababor zone, oromia regional state, Ethiopia, *Asian J. Dairy Food Res.* **40**(3) (2021) 279–284, 10.18805/ajdfr.DR-204.

8. Y. Kittichotsatsawat, N. Tippayawong and K. Y. Tippayawong, Prediction of arabica coffee production using artificial neural network and multiple linear regression techniques, *Sci. Rep.* **12**(1) (2022) 14488, 10.1038/s41598-022-18635-5.
9. X. Xu and Y. Zhang, Commodity price forecasting via neural networks for coffee, corn, cotton, oats, soybeans, soybean oil, sugar, and wheat, *Intell. Syst. Account., Finance Manag.* **29** (2022) 169–181, 10.1002/isaf.1519.
10. X. Xu and Y. Zhang, Edible oil wholesale price forecasts via the neural network, *Energy Nexus* **12** (2023) 100250, 10.1016/j.nexus.2023.100250.
11. R. Rincón, W. Palacios and H. Paipa, Comparison of statistical forecasting techniques for colombian coffee demand in south korea, *J. Phys.: Conf. Series* **1448**(1) (2020) 012023.
12. C. Fatih, Forecasting models based on fuzzy logic: An application on international coffee prices, *Econometrics* **26**(4) (2022) 1–16, 10.15611/eada.2022.4.01.
13. T. H. Abebe, Modeling time-varying coffee price volatility in ethiopia, *J. Appl. Econ.* **23**(1) (2020) 497–518, 10.1080/15140326.2020.1804304.
14. X. Xu, Cointegration among regional corn cash prices, *Econ. Bull.* **35**(4) (2015) 2581–2594.
15. X. Xu, Short-run price forecast performance of individual and composite models for 496 corn cash markets, *J. Appl. Stat.* **44**(14) (2017) 2593–2620, 10.1080/02664763.2016.1259399.
16. X. Xu, Contemporaneous causal orderings of us corn cash prices through directed acyclic graphs, *Empir. Econ.* **52**(2) (2017) 731–758, 10.1007/s00181-016-1094-4.
17. X. Xu, The rolling causal structure between the chinese stock index and futures, *Financ. Mark. Portf. Manag.* **31**(4) (2017) 491–509, 10.1007/s11408-017-0299-7.
18. X. Xu, Intraday price information flows between the csi300 and futures market: An application of wavelet analysis, *Empir. Econ.* **54**(3) (2018) 1267–1295, 10.1007/s00181-017-1245-2.
19. J. Yang, M. S. Haigh and D. J. Leatham, Agricultural liberalization policy and commodity price volatility: A GARCH application, *Appl. Econ. Lett.* **8**(9) (2001) 593–598, 10.1080/13504850010018734.
20. X. Xu, Using local information to improve short-run corn price forecasts, *J. Agric. Food Indus. Organ.* **16**(1) (2018), 10.1515/jafio-2017-0018.
21. X. Xu, Causal structure among us corn futures and regional cash prices in the time and frequency domain, *J. Appl. Stat.* **45**(13) (2018) 2455–2480, 10.1080/02664763.2017.1423044.
22. X. Xu, Price dynamics in corn cash and futures markets: Cointegration, causality, and forecasting through a rolling window approach, *Financ. Mark. Portf. Manag.* **33**(2) (2019) 155–181, 10.1007/s11408-019-00330-7.
23. X. Xu, Contemporaneous and Granger causality among us corn cash and futures prices, *Eur. Rev. Agric. Econ.* **46**(4) (2019) 663–695, 10.1093/erae/jby036.
24. X. Xu, Contemporaneous causal orderings of csi300 and futures prices through directed acyclic graphs, *Econ. Bull.* **39**(3) (2019) 2052–2077.
25. X. Xu, Corn cash price forecasting, *Am. J. Agric. Econ.* **102**(4) (2020) 1297–1320, 10.1002/ajae.12041.
26. X. Xu and Y. Zhang, Individual time series and composite forecasting of the Chinese stock index, *Mach. Learn. Appl.* **5** (2021) 100035, 10.1016/j.mlwa.2021.100035.
27. X. Xu and Y. Zhang, Contemporaneous causality among one hundred Chinese cities, *Empir. Econ.* **63** (2022) 2315–2329, 10.1007/s00181-021-02190-5.
28. X. Xu and Y. Zhang, Contemporaneous causality among residential housing prices of ten major Chinese cities, *Int. J. Housing Mark. Anal.* **16**(4) (2023) 792–811, 10.1108/IJHMA-03-2022-0039.

29. X. Xu and Y. Zhang, Cointegration between housing prices: Evidence from one hundred chinese cities, *J. Property Res.* **40**(1) (2023) 53–75, 10.1080/09599916.2022.2114926.
30. X. Xu and Y. Zhang, An integrated vector error correction and directed acyclic graph method for investigating contemporaneous causalities, *Decis. Analytics J.* **7** (2023) 100229, 10.1016/j.dajour.2023.100229.
31. J. Yang and T. O. Awokuse, Asset storability and hedging effectiveness in commodity futures markets, *Appl. Econ. Lett.* **10**(8) (2003) 487–491, 10.1080/1350485032000095366.
32. X. Xu and Y. Zhang, Dynamic relationships among composite property prices of major chinese cities: Contemporaneous causality through vector error corrections and directed acyclic graphs, *Int. J. Real Estate Studies* **17** (2023) 148–157, 10.11113/intrest.v17n1.294.
33. X. Xu and Y. Zhang, Contemporaneous causality among regional steel price indices of east, south, north, central south, northeast, southwest, and northwest China, *Miner. Econ.* **37**(1) (2024) 1–14, 10.1007/s13563-023-00380-4.
34. X. Xu and Y. Zhang, Spatial-temporal analysis of residential housing, office property, and retail property price index correlations: Evidence from ten chinese cities, *Int. J. Real Estate Studies* **17** (2023) 1–13, 10.11113/intrest.v17n2.274.
35. X. Xu and Y. Zhang, Contemporaneous causality among office property prices of major Chinese cities with vector error correction modeling and directed acyclic graphs, *J. Model. Manag.* **19**(4) (2024) 1079–1093, 10.1108/JM2-08-2023-0171.
36. X. Xu and Y. Zhang, Network analysis of corn cash price comovements, *Mach. Learn. Appl.* **6** (2021) 100140, 10.1016/j.mlwa.2021.100140.
37. X. Xu and Y. Zhang, Network analysis of comovements among newly-built residential house price indices of seventy Chinese cities, *Int. J. Housing Mark. Anal.* **17**(3) (2024) 726–749, 10.1108/IJHMA-09-2022-0134.
38. X. Xu and Y. Zhang, Network analysis of price comovements among corn futures and cash prices, *J. Agric. Food Ind. Organ.* **22**(1) (2024) 53–81, 10.1515/jafio-2022-0009.
39. X. Xu and Y. Zhang, Network analysis of housing price comovements of a hundred chinese cities, *Natl. Inst. Econ. Rev.* **264** (2023) 110–128, 10.1017/nie.2021.34.
40. I. O. Alade, Y. Zhang and X. Xu, Modeling and prediction of lattice parameters of binary spinel compounds (am_2x_4) using support vector regression with bayesian optimization, *New J. Chem.* **45**(34) (2021) 15255–15266, 10.1039/d1nj01523k.
41. B. Jin and X. Xu, Machine learning-based scrap steel price forecasting for the northeast chinese market, *Int. J. Empir. Econ.* (2024), 10.1142/S2810943024500112.
42. I. O. Alade, M. O. Oyedeji, M. A. A. Rahman and T. A. Saleh, Prediction of the lattice constants of pyrochlore compounds using machine learning, *Soft Comput.* **26**(17) (2022) 8307–8315, 10.1007/s00500-022-07218-1.
43. X. Xu and Y. Zhang, Predicting the open interest of the thermal coal futures using a neural network, *Financ. Mark. Portf. Manag.* (2024), 10.1007/s11408-024-00458-1.
44. I. O. Alade, M. A. A. Rahman, A. Hassan and T. A. Saleh, Modeling the viscosity of nanofluids using artificial neural network and Bayesian support vector regression, *J. Appl. Phys.* **128**(8) (2020), 10.1063/5.0008977.
45. B. Jin and X. Xu, Office real estate price index forecasts through Gaussian process regressions for ten major Chinese cities, *Adv. Comput. Intell.* **4** (2024) 8, 10.1007/s43674-024-00075-5.
46. A. A. Adewumi, T. O. Owolabi, I. O. Alade and S. O. Olatunji, Estimation of physical, mechanical and hydrological properties of permeable concrete using computational intelligence approach, *Appl. Soft Comput.* **42** (2016) 342–350, 10.1016/j.asoc.2016.02.009.

47. B. Jin and X. Xu, Carbon emission allowance price forecasting for China Guangdong carbon emission exchange via the neural network, *Global Finance Rev.* **6**(1) (2024) 3491, 10.18282/gfr.v6i1.3491.
48. Y. Luo, Soybean futures price prediction based on CNN-LSTM model of Bayesian optimization algorithm, *Highlights Bus.*, *Econ. Manag.* **16** (2023) 6–17.
49. A. K. Mahto, M. A. Alam, R. Biswas, J. Ahmad and S. I. Alam, Short-term forecasting of agriculture commodities in context of indian market for sustainable agriculture by using the artificial neural network, *J. Food Qual.* **2021** (2021) 1–13, 10.1155/2021/9939906.
50. M. Ali, R. C. Deo, N. J. Downs and T. Maraseni, Cotton yield prediction with Markov chain monte carlo-based simulation model integrated with genetic programing algorithm: A new hybrid copula-driven approach, *Agric. Forest Meteorol.* **263** (2018) 428–448, 10.1016/j.agrformet.2018.09.002.
51. Y. Fang, B. Guan, S. Wu and S. Heravi, Optimal forecast combination based on ensemble empirical mode decomposition for agricultural commodity futures prices, *J. Forecast.* **39**(6) (2020) 877–886, 10.1002/for.2665.
52. G. Li, W. Chen, D. Li, D. Wang and S. Xu, Comparative study of short-term forecasting methods for soybean oil futures based on LSTM, SVR, ES and wavelet transformation, *J. Phys.*: *Conf. Ser.* **1682**(1) (2020) 012007.
53. L. Kouadio, R. C. Deo, V. Byrareddy, J. F. Adamowski, S. Mushtaq *et al.*, Artificial intelligence approach for the prediction of robusta coffee yield using soil fertility properties, *Comput. Electron. Agric.* **155** (2018) 324–338, 10.1016/j.compag.2018.10.014.
54. Y. Abreham, Coffee price prediction using machine-learning techniques, Ph.D. Thesis, ASTU (2019).
55. H. T. Huy, H. N. Thac, H. N. T. Thu, A. N. Nhat and V. H. Ngoc, Econometric combined with neural network for coffee price forecasting, *J. Appl. Econ. Sci.* **14**(2) (2019) 378.
56. W. A. Degife and A. Sinamo, Efficient predictive model for determining critical factors affecting commodity price: The case of coffee in ethiopian commodity exchange (ECX), *Int. J. Inf. Eng. Electron. Bus.* **11**(6) (2019) 32–36, 10.5815/ijieeb.2019.06.05.
57. K. Naveena, S. Subedar *et al.*, Hybrid time series modelling for forecasting the price of washed coffee (arabica plantation coffee) in india, *Int. J. Agric. Sci.* **9** (2017) 4004–4007.
58. L. P. Lopes, Prediction of the brazilian natural coffee price through statistical machine learning models, *SIGMAE* **7**(1) (2018) 1–16.
59. C. Deina, M. H. do Amaral Prates, C. H. R. Alves, M. S. R. Martins, F. Trojan, S. L. Stevan Jr. and H. V. Siqueira, A methodology for coffee price forecasting based on extreme learning machines, *Inf. Process. Agric.* **9** (2021) 556–565, 10.1016/j.inpa.2021.07.003.
60. X. Xu and Y. Zhang, Corn cash price forecasting with neural networks, *Comput. Electron. Agric.* **184** (2021) 106120, 10.1016/j.compag.2021.106120.
61. J. Zhang, Y. Meng, J. Wei, J. Chen and J. Qin, A novel hybrid deep learning model for sugar price forecasting based on time series decomposition, *Math. Probl. Eng.* **2021** (2021) 1–9, 10.1155/2021/6507688.
62. G. Wen, B.-L. Ma, A. Vanasse, C. D. Caldwell, H. J. Earl and D. L. Smith, Machine learning-based canola yield prediction for site-specific nitrogen recommendations, *Nutr. Cycling Agroecosyst.* **121**(2) (2021) 241–256, 10.1007/s10705-021-10170-5.
63. J. J. Harris, A machine learning approach to forecasting consumer food prices, Faculty of Graduate Studies Online Theses (2017).

64. Y. Zou, M. Tu, X. Teng, R. Cao and W. Xie, Electricity price forecast based on stacked autoencoder in spot market environment, *2019 9th Int. Conf. Power and Energy Systems* (*ICPES*) (IEEE, 2019), pp. 1–6.
65. X. Xu and Y. Zhang, China mainland new energy index price forecasting with the neural network, *Energy Nexus* **10** (2023) 100210, 10.1016/j.nexus.2023.100210.
66. B. Jin and X. Xu, Price forecasting through neural networks for crude oil, heating oil, and natural gas, *Meas.*: *Energy* **1**(1) (2024) 100001, 10.1016/j.meaene.2024.100001.
67. Y. Huang, X. Dai, Q. Wang and D. Zhou, A hybrid model for carbon price forecasting using GARCH and long short-term memory network, *Appl. Energy* **285** (2021) 116485, 10.1016/j.apenergy.2021.116485.
68. K. Zhang, H. Cao, J. Thé and H. Yu, A hybrid model for multi-step coal price forecasting using decomposition technique and deep learning algorithms, *Appl. Energy* **306** (2022) 118011, 10.1016/j.apenergy.2021.118011.
69. J. Zhou and Z. Xu, A novel three-stage hybrid learning paradigm based on a multi-decomposition strategy, optimized relevance vector machine, and error correction for multi-step forecasting of precious metal prices, *Resour. Policy* **80** (2023) 103148, 10.1016/j.resourpol.2022.103148.
70. X. Xu and Y. Zhang, Price forecasts of ten steel products using gaussian process regressions, *Eng. Appl. Artif. Intell.* **126** (2023) 106870, 10.1016/j.engappai.2023.106870.
71. P. Sadorsky, Predicting gold and silver price direction using tree-based classifiers, *J. Risk Financ. Manag.* **14**(5) (2021) 198, 10.3390/jrfm14050198.
72. Z. Li, Y. Yang, Y. Chen and J. Huang, A novel non-ferrous metals price forecast model based on LSTM and multivariate mode decomposition, *Axioms* **12**(7) (2023) 670, 10.3390/axioms12070670.
73. S. Aruna, P. Umamaheswari, J. Sujipriya *et al.*, Prediction of potential gold prices using machine learning approach, *Ann. Rom. Soc. Cell Biol.* **25**(4) (2021) 1385–1396.
74. T. L. D. Huynh, The effect of uncertainty on the precious metals market: New insights from transfer entropy and neural network var, *Resour. Policy* **66** (2020) 101623, 10.1016/j.resourpol.2020.101623.
75. Z. Gligorić, M. Gligorić, D. Halilović, Č. Beljić and K. Urošević, Hybrid stochastic-grey model to forecast the behavior of metal price in the mining industry, *Sustainability* **12**(16) (2020) 6533, 10.3390/su12166533.
76. X. Xu and Y. Zhang, Canola and soybean oil price forecasts via neural networks, *Adv. Comput. Intell.* **2** (2022) 32, 10.1007/s43674-022-00045-9.
77. X. Xu and Y. Zhang, Steel price index forecasting through neural networks: The composite index, long products, flat products, and rolled products, *Miner. Econ.* **36** (2023) 563–582, 10.1007/s13563-022-00357-9.
78. X. Xu and Y. Zhang, Platinum and palladium price forecasting through neural networks, *Commun. Stat.-Simul. Comput.* (2024), 10.1080/03610918.2024.2330700.
79. X. Xu and Y. Zhang, Yellow corn wholesale price forecasts via the neural network, *EconomiA* **24** (2023) 44–67, 10.1108/ECON-05-2022-0026.
80. X. Xu and Y. Zhang, House price forecasting with neural networks, *Intell. Syst. Appl.* **12** (2021) 200052, 10.1016/j.iswa.2021.200052.
81. B. Liu, I. Kiskin and S. Roberts, An overview of gaussian process regression for volatility forecasting, *2020 Int. Conf. Artificial Intelligence in Information and Communication* (*ICAIIC*) (IEEE, 2020), pp. 681–686.
82. X. Xu and Y. Zhang, Residential housing price index forecasting via neural networks, *Neural Comput. Appl.* **34** (2022) 14763–14776, 10.1007/s00521-022-07309-y.

83. X. Xu and Y. Zhang, Scrap steel price forecasting with neural networks for east, north, south, central, northeast, and southwest China and at the national level, *Ironmaking Steelmaking* **50**(11) (2023) 1683–1697, 10.1080/03019233.2023.2218243.
84. S. K. Hundie and B. Biratu, Response of ethiopian coffee price to the world coffee price: Evidence from dynamic ARDL simulations and nonlinear ARDL cointegration, *Cogent Econ. Finance* **10**(1) (2022) 2114168.
85. K. Naveena, Statistical modelling for forecasting of price and export of Indian coffee, Ph.D. Thesis, Institute of Agricultural Sciences, Banaras Hindu University (2016).
86. E. K. Aduteye, T. T. Sete and Y. N. Chi, Time series analysis of global prices of coffee: Insights into a complex market, *Int. J. Bus. Econ.* **8**(2) (2023) 138–151.
87. L. Hamzah, S. Nabilah, E. Russel, M. Usman, E. Virginia *et al.*, Dynamic modelling and forecasting of data export of agricultural commodity by vector autoregressive model, *J. Southwest Jiaotong Univ.* **55**(3) (2020), 10.35741/issn.0258-2724.55.3.41.
88. R. K. Paul and K. Sinha, Spatial market integration among major coffee markets in india, *J. Indian Soc. Agric. Stat.* **69** (2015) 281–287.
89. T. T. Dang, C. Zhang, T. H. Nguyen and N. T. Nguyen, Assessing the influence of exchange rate on agricultural commodity export price: Evidence from Vietnamese coffee, *J. Econ. Dev.* **22**(2) (2020) 297–309, 10.1108/JED-02-2020-0014.
90. M. Musumba and R. S. Gupta, Transmission of world prices to ugandan coffee growers in a liberalised economy, *Dev. Policy Rev.* **31**(2) (2013) 219–234, 10.1111/dpr.12004.
91. W. S. Nugroho, A. B. Astuti and S. Astutik, Vector error correction model to forecasting spot prices for coffee commodities during COVID-19 pandemic, *J. Phys.: Conf. Ser.* **1811**(1) (2021) 012076.
92. B. Jin and X. Xu, Contemporaneous causality among price indices of ten major steel products, *Ironmaking & Steelmaking* **51**(6) (2024) 515–526, 10.1177/03019233241249361.
93. C. Milas and J. Otero, Smooth transition vector error correction models for the spot prices of coffee, *Appl. Econ. Lett.* **9**(14) (2002) 925–928, 10.1080/13504850210138513.
94. C. Milas, J. Otero and T. Panagiotidis, Forecasting the spot prices of various coffee types using linear and non-linear error correction models, *Int. J. Finance Econ.* **9**(3) (2004) 277–288, 10.1002/ijfe.245.
95. A. Cariappa and M. Sinha, Are the Indian arabica and robusta coffee prices cointegrated? A vector error correction approach, *Indian J. Agric. Market.* **35**(3) (2021) 44–59.
96. P. L. Brignoli, A. Varacca, C. Gardebroek and P. Sckokai, Machine learning to predict grains futures prices, *Agric. Econ.* **55** (2024) 479–497, 10.1111/agec.12828.
97. A. L. Xavier, B. J. Fernandes and J. F. De Oliveira, A hybrid swarm-based system for commodity price forecasting during the COVID-19 pandemic, *IEEE Access* **11** (2023) 74379–74387, 10.1109/ACCESS.2023.3293738.
98. R. Murugesan, E. Mishra and A. H. Krishnan, Forecasting agricultural commodities prices using deep learning-based models: Basic LSTM, bi-LSTM, stacked LSTM, CNN LSTM, and convolutional LSTM, *Int. J. Sustain. Agric. Manag. Inform.* **8**(3) (2022) 242–277, 10.1504/IJSAMI.2022.125757.
99. G. S. Atsalakis, Agriculture commodity prices forecasting using a fuzzy inference system, in *Agricultural Cooperative Management and Policy: New Robust, Reliable and Coherent Modelling Tools* (Springer, 2014), pp. 353–368.
100. A. J. Fofanah, Machine learning model approaches for price prediction in coffee market using linear regression, XGB, and LSTM techniques, *Int. J. Sci. Res. Sci. Technol.* (6) (2021) 10–48.

101. D. M. Efendi and F. Ardhy, Prediction of coffee prices with backpropagation neural networks, in *Proc. Int. Conf. Information Technology and Business* (2019), pp. 153–163.
102. Y. A. Herrera-Jaramillo, J. C. Ortega-Giraldo, A. Acevedo-Amorocho and D. Prada-Marin, Colombian coffee price forecast via LSTM neural networks, in *Technological and Industrial Applications Associated with Intelligent Logistics* (Springer, Cham, 2021), pp. 501–517, 10.1007/978-3-030-68655-0_25.
103. T. N. Le Ngoc, D. T. Lam, T. N. H. Minh, T. C. Doan, N. P. Nguyen, H. M. Nguyen, T. N. Nguyen, L. D. Tran and N.-Q. Tran, Machine learning for agricultural price prediction: A case of coffee commodity in vietnam market, *2023 IEEE/ACIS 8th Int. Conf. Big Data, Cloud Computing, and Data Science (BCD)* (IEEE, 2023), pp. 38–41.
104. R. M. Neal, *Bayesian Learning for Neural Networks* (Springer Science & Business Media, 2012).
105. X. Xu and Y. Zhang, Rent index forecasting through neural networks, *J. Econ. Studies* **49** (2022) 1321–1339, 10.1108/JES-06-2021-0316.
106. X. Xu and Y. Zhang, Second-hand house price index forecasting with neural networks, *J. Prop. Res.* **39**(3) (2022) 215–236, 10.1080/09599916.2021.1996446.
107. X. Xu and Y. Zhang, High-frequency csi300 futures trading volume predicting through the neural network, *Asian J. Econ. Banking* **8**(1) (2024) 26–53, 10.1108/AJEB-05-2022-0051.
108. X. Xu and Y. Zhang, Forecasting the total market value of a shares traded in the Shenzhen stock exchange via the neural network, *Econ. Bull.* **42**(3) (2022) 1266–1279.
109. G.-X. Xu, B.-C. Shia, Y.-B. Hsu, P.-C. Shen and K.-H. Chu, To integrate text mining and artificial neural network to forecast gold futures price, *2009 Int. Conf. New Trends in Information and Service Science* (IEEE, 2009), pp. 1014–1020.
110. X. Xu and Y. Zhang, Neural network predictions of the high-frequency csi300 first distant futures trading volume, *Financ. Mark. Portf. Manag.* **37** (2023) 191–207, 10.1007/s11408-022-00421-y.
111. X. Xu and Y. Zhang, Corn cash-futures basis forecasting via neural networks, *Adv. Comput. Intell.* **3** (2023) 8, 10.1007/s43674-023-00054-2.
112. X. Xu and Y. Zhang, Retail property price index forecasting through neural networks, *J. Real Estate Portf. Manag.* **29** (2023) 1–28, 10.1080/10835547.2022.2110668.
113. X. Xu and Y. Zhang, Office property price index forecasting using neural networks, *J. Financ. Manag. Prop. Construct.* **29**(1) (2024) 52–82, 10.1108/JFMPC-08-2022-0041.
114. X. Xu and Y. Zhang, A high-frequency trading volume prediction model using neural networks, *Decis. Anal. J.* **7** (2023) 100235, 10.1016/j.dajour.2023.100235.
115. R. M. Neal, Monte carlo implementation of Gaussian process models for Bayesian regression and classification (1997), arXiv:physics/9701026.
116. S. Brahim-Belhouari and J.-M. Vesin, Bayesian learning using Gaussian process for time series prediction, in *Proc. 11th IEEE Signal Processing Workshop on Statistical Signal Processing (Cat. No. 01TH8563)* (IEEE, 2001), pp. 433–436.
117. S. Brahim-Belhouari and A. Bermak, Gaussian process for nonstationary time series prediction, *Comput. Stat. Data Anal.* **47**(4) (2004) 705–712, 10.1016/j.csda.2004.02.006.
118. X. Xu and Y. Zhang, Regional steel price index forecasts with neural networks: Evidence from east, south, north, central south, northeast, southwest, and northwest China, *J. Supercomput.* **79** (2023) 13601–13619, 10.1007/s11227-023-05207-1.

119. X. Xu and Y. Zhang, Wholesale food price index forecasts with the neural network, *Int. J. Comput. Intell. Appl.* **22** (2023) 2350024, 10.1142/S1469026823500244.
120. X. Xu and Y. Zhang, Composite property price index forecasting with neural networks, *Prop. Manag.* **42**(3) (2024) 388–411, 10.1108/PM-11-2022-0086.
121. C. M. Jarque and A. K. Bera, Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Econ. Lett.* **6**(3) (1980) 255–259, 10.1016/0165-1765(80)90024-5.
122. C. M. Jarque and A. K. Bera, A test for normality of observations and regression residuals, *Int. Stat. Rev./Rev. Int. Stat.* **55** (1987) 163–172, 10.2307/1403192.
123. T. W. Anderson and D. A. Darling, A test of goodness of fit, *J. Am. Stat. Assoc.* **49**(268) (1954) 765–769, 10.2307/2281537.
124. B. Jin and X. Xu, Forecasting wholesale prices of yellow corn through the gaussian process regression, *Neural Comput. Appl.* **36**(15) (2024) 8693–8710, 10.1007/s00521-024-09531-2.
125. X. Xu and Y. Zhang, A Gaussian process regression machine learning model for forecasting retail property prices with Bayesian optimizations and cross-validation, *Decis. Anal. J.* **8** (2023) 100267, 10.1016/j.dajour.2023.100267.
126. X. Xu, Linear and nonlinear causality between corn cash and futures prices, *J. Agric. Food Ind. Organ.* **16**(2) (2018) 20160006, 10.1515/jafio-2016-0006.
127. J. Yang, J. Cabrera and T. Wang, Nonlinearity, data-snooping, and stock index ETF return predictability, *Eur. J. Oper. Res.* **200**(2) (2010) 498–507, 10.1016/j.ejor.2009.01.009.
128. X. Xu, Cointegration and price discovery in us corn cash and futures markets, *Empir. Econ.* **55**(4) (2018) 1889–1923, 10.1007/s00181-017-1322-6.
129. X. Xu and Y. Zhang, House price information flows among some major chinese cities: Linear and nonlinear causality in time and frequency domains, *Int. J. Hous. Mark. Anal.* **16**(6) (2023) 1168–1192, 10.1108/IJHMA-07-2022-0098.
130. X. Xu and Y. Zhang, Soybean and soybean oil price forecasting through the nonlinear autoregressive neural network (NARNN) and NARNN with exogenous inputs (NARNN–x), *Intell. Syst. Appl.* **13** (2022) 200061, 10.1016/j.iswa.2022.200061.
131. W. A. Brock, J. A. Scheinkman, W. D. Dechert and B. LeBaron, A test for independence based on the correlation dimension, *Econom. Rev.* **15**(3) (1996) 197–235, 10.1080/07474939608800353.
132. B. Jin and X. Xu, Machine learning price index forecasts of flat steel products, *Miner. Econ.* (2024), 10.1007/s13563-024-00457-8.
133. B. Jin and X. Xu, Forecasts of coking coal futures price indices through gaussian process regressions, *Miner. Econ.* (2024).
134. B. Jin and X. Xu, Wholesale price forecasts of green grams using the neural network, *Asian J. Econ. Bank.* (2024), 10.1108/AJEB-01-2024-0007.
135. B. Jin and X. Xu, Pre-owned housing price index forecasts using Gaussian process regressions, *J. Model. Manag.* (2024), 10.1108/JM2-12-2023-0315.
136. B. Jin and X. Xu, Machine learning predictions of regional steel price indices for east China, *Ironmaking Steelmaking* (2024), 10.1177/03019233241254891.
137. B. Jin and X. Xu, Predictions of steel price indices through machine learning for the regional northeast Chinese market, *Neural Comput. Appl.* **36** (2024) 20863–20882, 10.1007/s00521-024-10270-7.
138. B. Jin and X. Xu, Forecasts of thermal coal prices through Gaussian process regressions, *Ironmaking & Steelmaking* (2024), 10.1177/03019233241265194.
139. B. Jin and X. Xu, Regional steel price index predictions for north China through machine learning, *Int. J. Min. Miner. Eng.* **15**(3) (2024) 314–350, 10.1504/IJMME.2024.10065342.

140. B. Jin, X. Xu and Y. Zhang, Thermal coal futures trading volume predictions through the neural network, *J. Model. Manag.* (2024), 10.1108/JM2-09-2023-0207.
141. B. Jin and X. Xu, Gaussian process regression based silver price forecasts, *J. Uncertain Syst.* **17**(3) (2024) 2450013, 10.1142/S1752890924500132.
142. B. Jin and X. Xu, Forecasts of China mainland new energy index prices through Gaussian process regressions, *J. Clean Energy Energy Storage* **1** (2024) 2450006, 10.1142/S2811034X24500060.
143. B. Jin and X. Xu, Palladium price predictions via machine learning, *Mater. Circ. Econ.* **6** (2024) 32, 10.1007/s42824-024-00123-y.
144. C. Williams and C. Rasmussen, Gaussian processes for regression, in *Advances in Neural Information Processing Systems*, Vol. 8 (MIT Press, 1995), pp. 514–520.
145. A. D. Bull, Convergence rates of efficient global optimization algorithms, *J. Mach. Learn. Res.* **12**(10) (2011) 2879–2904.
146. M. Despotovic, V. Nedic, D. Despotovic and S. Cvetanovic, Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation, *Renew. Sustain. Energy Rev.* **56** (2016) 246–260, 10.1016/j.rser.2015.11.058.
147. A. Timmermann, Forecast combinations, in *Handbook of Economic Forecasting*, Vol. 1 (Elsevier, 2006), pp. 135–196, 10.1016/S1574-0706(05)01004-9.
148. P. Ou and H. Wang, Volatility prediction by treed Gaussian process with limiting linear model, *Int. J. Model. Simul.* **31**(2) (2011) 166–174, 10.2316/Journal.205.2011.2.205-5498.
149. J. Han and X.-P. Zhang, Financial time series volatility analysis using Gaussian process state-space models, *2015 IEEE Global Conf. Signal and Information Processing* (*GlobalSIP*) (IEEE, 2015), pp. 358–362.

About the Authors

Bingzi Jin obtained her Bachelor of Science in Engineering in June 2011 from Zhejiang University of Technology, Bachelor of Computing in July 2011 from University of Tasmania, and Master of Science in Engineering in June 2013 from Zhejiang University. Bingzi Jin is currently working for Advanced Micro Devices (China) Co., Ltd., Shanghai, China.

Xiaojie Xu obtained his B.S. in Management (minor in Finance) in June 2011 from Zhejiang University, and Master of Financial Mathematics in December 2011, Master of Economics in May 2012, and Ph.D. in Economics in July 2015 from North Carolina State University, whose research areas include applied econometrics and statistics, economic, econometric and statistical forecasting and predictions, financial, commodity and real estate markets, and agricultural and resource economics.