

Jin, Bingzi; Xu, Xiaojie

Article

Predicting wholesale edible oil prices through Gaussian process regressions tuned with Bayesian optimization and cross-validation

Asian Journal of Economics and Banking (AJEB)

Provided in Cooperation with:

Ho Chi Minh University of Banking (HUB), Ho Chi Minh City

Suggested Citation: Jin, Bingzi; Xu, Xiaojie (2025) : Predicting wholesale edible oil prices through Gaussian process regressions tuned with Bayesian optimization and cross-validation, Asian Journal of Economics and Banking (AJEB), ISSN 2633-7991, Emerald, Leeds, Vol. 9, Iss. 1, pp. 64-82, <https://doi.org/10.1108/AJEB-06-2024-0070>

This Version is available at:

<https://hdl.handle.net/10419/334137>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Predicting wholesale edible oil prices through Gaussian process regressions tuned with Bayesian optimization and cross-validation

Bingzi Jin

Advanced Micro Devices (China) Co., Ltd, Shanghai, China, and

Xiaojie Xu

North Carolina State University at Raleigh, Raleigh, North Carolina, USA

Abstract

Purpose – Developing price forecasts for various agricultural commodities has long been a significant undertaking for a variety of agricultural market players. The weekly wholesale price of edible oil in the Chinese market over a ten-year period, from January 1, 2010 to January 3, 2020, is the forecasting issue we explore.

Design/methodology/approach – Using Bayesian optimisations and cross-validation, we study Gaussian process (GP) regressions for our forecasting needs.

Findings – The produced models delivered precise price predictions for the one-year period between January 4, 2019 and January 3, 2020, with an out-of-sample relative root mean square error of 5.0812%, a root mean square error (RMSEA) of 4.7324 and a mean absolute error (MAE) of 2.9382.

Originality/value – The projection's output may be utilised as stand-alone technical predictions or in combination with other projections for policy research that involves making assessment.

Keywords Wholesale edible oil, Price forecasting, Gaussian process regression, Bayesian optimization, Cross-validation, Chinese market

Paper type Research paper

1. Introduction

Food price projections from the agriculture sector are crucial for a range of market participants, such as processors, speculators, hedgers and policymakers (Raihan *et al.*, 2023). Producers, for example, often need price forecast data to set sales prices before production begins, exporters and processors to fulfil their contractual duties, speculators to profit, hedgers to control risks, and policymakers to develop, track and evaluate strategic plans and policies (Dacha *et al.*, 2021). China's significant agricultural market share (Ji *et al.*, 2022), close linkages to the energy sector (Ranguwal *et al.*, 2023), and the influence of financial markets and macroeconomic factors (Raihan, 2023) make edible oil price forecasting there no exception. Because of the physical limitations on the food supply—such as land, agricultural technology, environmental sustainability and climate change—policymakers see food prices as strategic issues. This is especially true, given China's massive population and expanding economy. A number of macroeconomic and financial factors, such as interest



JEL Classification — C22, C53, C63, Q11, Q13

© Bingzi Jin and Xiaojie Xu. Published in *Asian Journal of Economics and Banking*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

Conflict of interest: There is no conflict of interest.

Data availability statement: Data are available upon request.

rates, stock prices, exchange rates and financialisation levels, as well as changes in the energy markets, such as the price of oil, ethanol and the demand for biofuels, could put food price stability at risk. Price forecasting may not need to be motivated a lot since agricultural commodity prices often show erratic volatility patterns (Yeasin *et al.*, 2020), have a significant impact on market participants' decisions (Rahoveanu *et al.*, 2018), and ultimately affect resource allocations and overall economic wellbeing (Wulandari *et al.*, 2021).

Many research studies have been conducted in the literature on a variety of time series approaches for price prediction (Jin and Xu, 2024a). In these early studies described below, models like vector autoregressive (VAR) models, autoregressive integrated moving average models, vector error correction models and many variants of these models are often mentioned. The autoregressive integrated moving average (ARIMA), for example, has been shown in previous studies to be a highly preferred choice for a variety of time series forecasting applications. It was shown that ARIMA performs noticeably better than expert views and structural model-based forecasts for the US hog and cattle markets. The accuracy of hog price projections will only be slightly improved by moving from the ARIMA to models that include more data from the sow's farrowing price, according to another research. A number of distinctions exist between this empirical data and the wheat price data, which demonstrated that the ARIMA model's forecast accuracy may be enhanced with adding exchange rate series data. Prior studies have shown that combining the ARIMA with various model types may increase prediction accuracy more than depending just on one data source. One well-liked econometric method for price series forecasts is the VAR approach, which emphasises the connections between various economic factors. The comparison's conclusions demonstrate that the VAR predicts US cotton prices better than structural models at times of normal price volatility. In distinguishing the predictive content of a set of wheat futures prices from various countries and in differentiating between the prices of soybeans and soy in various US regions, the VAR was shown to be useful. Long-term relationships between economic variables are also taken into consideration by the vector error correction model (VECM), which is closely related to the VAR, via cointegration. It may be particularly helpful for long-term price forecasts. According to studies, for example, the VECM often performs better than the VAR in predicting global wheat prices.

In price predicting studies, the previously outlined econometric models have proven useful, especially in edible oil studies. Karia *et al.* (2016), for example, used the auto-regressive fractionally integrated moving average (ARFIMA) and ARIMA for palm, rapeseed, soybean, linseed and sunflower oil to study price forecasts. Resolving the over-differencing issue had little influence on the prediction accuracy of any model, and they found contradictory results about the efficacy of several models. According to Priyanga *et al.* (2019), the ARIMA might be useful for forecasting the price of coconut oil in Kerala. The ARIMA model was used by Darekar and Reddy (2017) to forecast the pricing of an oil seed, namely groundnuts, in India. They found that farmers, legislators and marketers may benefit from the projections. There are minor differences between the model's projected and actual values, according to Meena *et al.* (2014)'s analysis of oil and mustard seed prices in India using the ARIMA model. A multivariate ARIMA, which combines the ARIMA with an econometric equation predetermined for Malaysian palm oil price estimates, was proposed by Shamsudin and Arshad (2000) to improve forecasts based on the ARIMA. Khin *et al.* (2011) found similar empirical evidence to back their Malaysian palm oil price predictions. The ARIMA, exponential generalised auto-regressive conditional heteroscedastic (GARCH) model (EGARCH) and GARCH model were studied by Lama *et al.* (2015) in order to forecast edible oil prices both domestically and internationally. Because it better captures the volatility pattern, they found that the EGARCH model outperforms the other two. In order to anticipate the prices of soybean and rapeseed oil, Wang *et al.* (2013) demonstrated the effectiveness of the seasonal VECM in China. In order to demonstrate the availability of possible predictive information from crude oil prices to those of edible oil, Hasanov *et al.* (2016) used the GARCH-in-mean model and volatility impulse response function analysis. Their findings

showed that crude oil prices might aid in forecasting edible oil prices. This empirical evidence may vary depending on the historical periods in question. Using the VECM and directed acyclic graph approach, [Yu et al. \(2006\)](#) examined the price correlations between crude oil and various food oils and concluded that there was no discernible impact of crude oil prices on edible oil prices.

Researchers have lately shown a great lot of interest in examining the uses of machine learning algorithms for agricultural commodity price forecasts because of the ease with which computer resources and technology are now accessible ([Alade et al., 2021](#); [Jin and Xu, 2024b](#)). Therefore, a variety of commodities—such as soybeans, sugar, corn, wheat, soybean oil, coffee, cotton, green beans, canola, edible oil and peanut oil—have been the subject of research using neural networks, genetic programming, deep learning, support vector regressions, random forests, K-nearest neighbours, multivariate adaptive regression splines, decision trees, ensembles, and boosting. Neural networks may be the most popular machine learning model for predicting the price of agricultural commodities, according to these and other findings from earlier studies, however this is by no means an exhaustive analysis. Furthermore, previous empirical studies demonstrating the effectiveness of machine learning methods for financial and economic forecasting are often in agreement with these evaluations.

The study demonstrates that machine learning techniques are being used more and more to forecast edible oil prices. In their investigation of price projections for Malaysian palm, soybean, coconut, olive, rapeseed and sunflower oils, for example, [Kanchymalay et al. \(2017\)](#) found that the sequential minimum approach to support vector regression (SVR) improves forecast accuracy. The random forest may be useful in forecasting Myanmar's edible oil price, claim [Myat and Tun \(2019\)](#). In key Indian markets, [Singh \(2021\)](#) and [Jha and Sinha \(2014\)](#) examined neural networks with ARIMA for price forecasts of mustard, groundnut, rapeseed, and soybean oil. They found that, on average, neural networks are more accurate than ARIMA. [Mishra and Singh \(2013\)](#) focused on predicting groundnut oil prices in Delhi using neural networks and ARIMA; they found mixed findings on the two models' efficacy. [Lama et al. \(2016\)](#) suggest that combining the neural network with GARCH might improve the forecasts of the individual model for edible oil prices in both domestic and international economies. The price forecasting challenges for maize and palm oil were examined by [Jaiswal et al. \(2022\)](#) using the ARIMA, deep long short-term memory neural network and time-delay neural network. They discovered that the most accurate predictions were made by the deep long short-term memory neural network. According to [Jaiswal et al. \(2023\)](#), a nonlinear autoregressive neural network containing exogenous variables could be a helpful tool for soybean oil price forecasting. [Silalahi \(2013\)](#) discovered that the neural network model that was optimised by the genetic algorithm could provide a reasonable level of accuracy in price forecasting for soybean and palm oil. [Salman et al. \(2018\)](#) proposed tuning the backpropagation neural network for palm oil price forecasting using particle swarm optimisation, which increases prediction accuracy compared to the conventional backpropagation neural network. [Amal \(2021\)](#) discovered that the long short-term memory neural network model could be tuned using adaptive moment estimate optimisation to provide a palm oil price prediction with a high degree of accuracy.

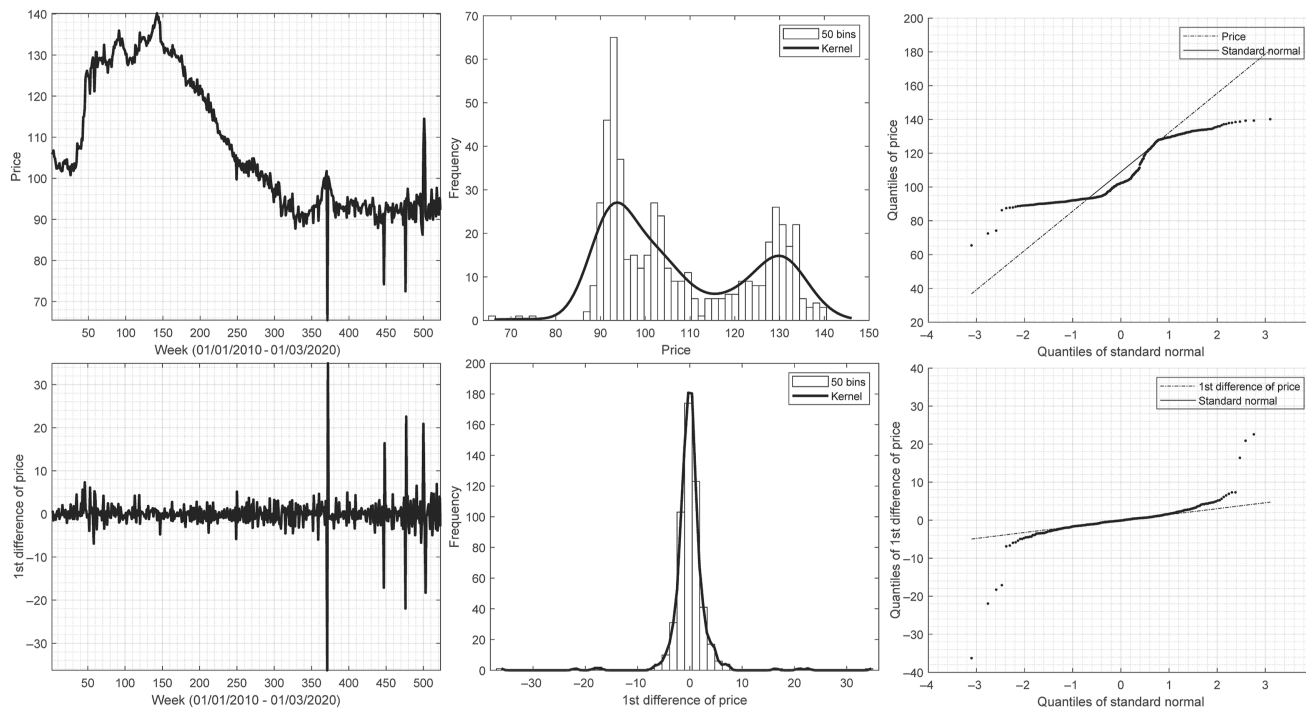
For time-series data gathered on edible oil prices, however, the predictions generated by the Gaussian process (GP) regression have not gotten much attention. A novel regression technique is put forward, drawing on Neal's work on Bayesian learning for neural networks ([Neal, 2012](#)). Since the technique relies on priors over functions of Gaussian processes, simulating noisy data makes sense ([Jin and Xu, 2024c](#)). The convergence of several neural network-based Bayesian regression models to Gaussian processes close to the edge of an infinite network was demonstrated ([Neal, 2012](#)). Research has shown the effectiveness of GP regressions in replicating data that is either noisy ([Williams and Rasmussen, 1995](#)) or noise-free ([Neal, 1997](#)). In their study of Gaussian processes using radial basis function neural networks for forecasting problems involving stationary time-series data, Brahim-Belhouari and Vesin discovered that Bayesian learning yields superior prediction outcomes

(Brahim-Belhouari and Vesin, 2001). According to the results of Brahim-Belhouari and Bermak (2004), this study looked at a wide range of covariance functions. GP prediction techniques are also helpful for forecasting problems with non-stationary time-series data. Based on the work of Brahim-Belhouari and Bermak, GP regressions are more effective than radial basis function neural networks overall (Brahim-Belhouari and Bermak, 2004). The exact matrix operations required to integrate prior and noisy models are also the reason for the GP formulation's effectiveness and success (Brahim-Belhouari and Bermak, 2004). Additionally, Brahim-Belhouari and Bermak proposed the use of GP predictors in the multi-model forecasting technique (Brahim-Belhouari and Bermak, 2004), which function similarly to how we would use model averaging.

Over a ten-year period, from January 1, 2010 to January 3, 2020, we use the weekly wholesale edible oil price to conduct our inquiry. The forecast model we use is GP regression. The edible oil price index for the Chinese market has a significant underlying economic relevance as it is meant to reflect the wholesale market trend for all types of edible oil throughout the country. This price index may provide useful projections for policymakers and other market participants. To the best of our knowledge, none of the few earlier studies—including the ones mentioned above—have examined the pertinent prediction issue for this price index. Our prediction exercise is based on this price index, and we follow the literature on commodity price forecasts in order to fill this research gap. Therefore, our results serve to provide useful forecast information of the important price index to different forecast consumers. We examine how well models trained using cross-validation and Bayesian optimisations provide forecasts. As it turns out, our models are rather straightforward and provide accurate and reliable forecasts. As far as we are aware, and considering the aforementioned studies, this is the first study to predict China's wholesale edible oil pricing using the machine learning technique of GP regression. Model prediction performance may be improved by using Bayesian optimisation to provide GP regressions more flexibility, especially when dealing with underlying data that exhibits nonlinear properties. On the one hand, since it doesn't take a lot of computing time to implement, this prediction framework is rather efficient. However, this approach produces forecasts that are quite accurate. There is no doubting the importance of accurate and timely agricultural commodity price projections for both policymakers and market players. Risk management, market evaluations and portfolio adjustments may benefit from such forecasts. This research assists decision-makers in making timely judgements by looking at difficulties related to projecting agricultural commodity prices and utilising weekly data, which are rather high-frequency for the wholesale market. Our results may be utilised as independent technical price projections, on the one hand. Nonetheless, they might be used in combination with other (basic) prediction findings for policy research and the creation of hypotheses about pricing trends. This method may be helpful for both market participants and policymakers as it facilitates the extension of the framework to potential commodity price projections in other business sectors.

2. Data

Over the course of ten years, from January 1, 2010, to January 3, 2020, we examine weekly wholesale edible oil costs for the Chinese market. These prices are taken from China's National Wholesale Price Information System. Figure 1's top panel displays the price series plot on the left, the quantile-quantile plot against the standard normal distribution on the right, and the 50-bin histogram with kernel estimates in the centre. For the price series' differences, the relevant data visualisation is shown in the bottom panel of Figure 1. For the pricing data, Table 1 offers crucial summary information. The price series definitely deviates from normal distributions, as shown by the results of the Jarque-Bera and Anderson–Darling tests. This may not come as a surprise considering the nature of economic data (Jin and Xu, 2024d). In particular, the price series exhibits platykurtic behaviour and right skew. As can be seen



Source(s): Figure by authors

Figure 1. Visualization of weekly wholesale edible oil prices and their first differences during the period of January 1, 2010–January 3, 2020, together with associated 50-bin histograms with kernel estimates and quantile-quantile plots against the standard normal distribution

Table 1. Data summary of weekly wholesale edible oil prices during the period of January 1, 2010–January 3, 2020

Series	Minimum	1st percentile	5th percentile	Mean	Median	Standard deviation	95th percentile	99th percentile	Maximum	Skewness	Kurtosis	Jarque-Bera	Anderson-Darling
Price	65.4700	87.6898	90.1130	107.4965	102.3900	16.1876	133.6985	138.1588	140.1600	0.4619	1.7983	<0.001	<0.0005
First difference	−36.2600	−6.7388	−3.4080	−0.02600	−0.1000	3.6374	3.8600	7.3028	34.9400	−0.0096	46.2291	<0.001	<0.0005

Source(s): Table by authors

from Figure 1, the price series has four notable surges. Keep in mind that the base period price, which is 100, is derived from the average weekly price for June 1994.

Previous studies on the emergence of nonlinear features at higher moments over a broad range of time-series data have been published in significant numbers by the financial and economic domains (Yang *et al.*, 2008). To find any possible nonlinear trends, the price series is put through the Brock–Dechert–Scheinkman (BDS) test (Brock *et al.*, 1996). As can be seen, the test yielded almost zero p -values. These results imply the presence of nonlinearities in the price series. The purpose of this study is to predict the price series using GP regressions while taking these particulars into consideration.

3. Method

The primary focus of this work is on the forecasting approach of GP regressions, a kind of probabilistic kernel model that has shown predictive ability in predicting a range of nonlinear patterns across several scientific fields (Jin and Xu, 2024e). The training data used to estimate model parameters with an uncertain distribution is represented by the model as follows: $\{(x_i, y_i); i = 1, 2, \dots, T\}$. These are the expressions for the d – dimension predictors: $x_i \in \mathbb{R}^d$, and the reflection of the target occurs via $y_i \in \mathbb{R}$. By use of twelve-lag prices as predictors, the price estimate is produced. Prices from the previous twelve weeks, for instance, will be used as predictors to determine the price for the thirteenth week.

The approach below may be used to express a linear regression: $y = x^T \beta + \varepsilon$, with $\varepsilon \sim N(0, \sigma^2)$ reporting the error item. However, in GP regressions, an explicit basis and latent variables are used to define the target variable. $l(x_i)$ represents the latent variables from Gaussian processes that satisfy the requirements for a joint Gaussian distribution, whereas b represents the basis function. The basis function's purpose is to project different predictors onto the feature space; the objective's smoothness will be shown by the latent variables' covariance function (Zhang and Xu, 2020; LI *et al.*, 2015).

Usually, a Gaussian process (GP) is described by two metrics: mean and covariance. We would adopt $k(x, x') = \text{Cov}[l(x), l(x')]$ to report the covariance and $m(x) = E(l(x))$ to report the mean. It would be reported that $y = b(x)^T \beta + l(x)$ express the GP regression, where $l(x) \sim GP(0, k(x, x'))$ and $b(x) \in \mathbb{R}^p$. Via a hyper-parameter named θ , $k(x, x')|\theta$ would be able to be parameterized. A GP regression is often trained using the following variables, which are often calculated using a particular approach: σ^2 , θ , and β . Furthermore, we would provide the basis functions (called b 's) and kernels (called k 's) that would be used throughout the model's training processes. This paper would examine two categories of kernels: one is the nonisotropic kernel (also known as automatic relevance determination kernel) and the other is the isotropic kernel. To explore both isotropic kernels and nonisotropic kernels, five distinct kernels of each category would be applied. Equations (1) through (10) provide the details for each kernel under discussion. To indicate the scale-mixture parameter, we would use the notation $\alpha > 0$. We would utilize σ_l for showing the characteristic length scale associated with isotropic kernels. We would use σ_f for showing the standard deviation associated with the signal. $r = \sqrt{(x_i - x_j)'(x_i - x_j)}$. Applying $\theta = (\theta_1, \theta_2) = (\log \sigma_b, \log \sigma_f)$ suggests an approach for the purpose of guaranteeing that σ_l and σ_f are above zero. The length scale associated with each predictor that corresponds to a nonisotropic kernel would be unique and would be reflected thorough the use of σ_m ($m = 1, 2, \dots, d$). Consequently, θ would be reflected via the use of $\theta = (\theta_1, \theta_2, \dots, \theta_d, \theta_{d+1}) = (\log \sigma_1, \log \sigma_2, \dots, \log \sigma_d, \log \sigma_f)$.

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

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

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

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

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

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

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

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

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

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

Analyses that are close to those of various kinds of kernels of four different possible basis functions (indicted in [Equations \(11\)–\(14\)](#)) would be carried out in our work as well. From

$$\text{Equations (11)–(14), } X = (x_1, x_2, \dots, x_n)', \quad X^2 = \begin{pmatrix} x_{11}^2 & x_{12}^2 & \cdots & x_{1d}^2 \\ x_{21}^2 & x_{22}^2 & \cdots & x_{2d}^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_{T1}^2 & x_{T2}^2 & \cdots & x_{Td}^2 \end{pmatrix}, \quad B = (b(x_1),$$

$b(x_2), \dots, b(x_n))'$, and an empty matrix is referring to a matrix who has at least one of its dimensions that is zero.

$$\text{Empty: } B = \emptyset, \text{ which is Empty Matrix} \quad (11)$$

$$\text{Constant: } B = I_{n \times 1} \quad (12)$$

$$\text{Linear: } B = [1, X] \quad (13)$$

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

Numerous basis functions and a great number of kernels might be taken into consideration. These four basis functions and ten kernels are typical examples of several types of functions that are often examined in a wide range of topics. Their potential for modelling a wide variety of complex underlying data types has been shown. Consequently, this study takes them into account.

In order to estimate the model parameters, Bayesian optimisation (Shahriari *et al.*, 2015) based on the expected improvement per second plus (EIPSP) approach is paired with ten-fold cross validation. Let us examine a GP model, whose representation w $f(x)$. The associated $y_i = f(x_i)$ would be assessed through using N_s randomly chosen data points of x_i 's within the variable boundaries via applying the Bayesian approach. In this case, $N_s = 4$ data points would be adopted for the purpose of the preliminary analysis. If the evaluation involves occurrences of mistakes, the algorithm would continue to gather more data until it reaches N_s successful evaluation occurrences. As can be seen from below, the procedure would then be repeated in the first phase and second phase. After $f(x)$ is updated, the posterior distribution over $Q(f|x_i, y_i \text{ for } i = 1, \dots, T)$ would be produced at first. To made determination of the reduction target associated with the acquisition function (denoted as $a(x)$), the next step would be choosing a new data point (x). One hundred is the most iterations that would be applied. With respect to Q , we would utilise $a(x)$ for the purpose of assessing the goodness of fit of x . Expected improvement (EI) acquisition functions would make comparisons of the expected degree of progress to values that would possibly increase the target function. The data point x_{best} would be indicating that the lowest posterior mean that would have been attained, and the lowest mean's associated numerical value would be reported as $\mu_Q(x_{best})$. This formula that might be applied to denote the expected improvement (EI) would be reported as $EI(x, Q) = E_Q[\max(0, \mu_Q(x_{best}) - f(x))]$. Since location would alter targets' assessment time, the Bayesian strategy would provide more benefits per unit of time through including a time-weighting technique into the acquisition function. During the optimisation procedures, another Bayesian model corresponding to the time that is needed for assessing the goal as a function of x would need to be tracked. Having this knowledge in mind, the acquisition function's EI per second (denoted as EIPS) would be reported as follows: $EIPS(x) = \frac{EI_Q(x)}{\mu_S(x)}$, where $\mu_S(x)$ would be used to report the posterior mean that is associated with the additional GP model that corresponds to time. For avoiding the case that the acquisition function is utilising a specific area and for preventing a local minimum of the target, the behaviour of the function would need to be changed as follows. We would now report the posterior standard deviation that is associated with the additive noises satisfying $\sigma_Q^2(x) = \sigma_F^2(x) + \sigma_{PN}^2$ as well as the posterior goal $\sigma_F(x)$ that is corresponding to x . For the purpose of accounting for the exploration ratio, $t_{\sigma_{PN}} > 0$ would need to be applied. The acquisition function, which is applying the EIPSP algorithm, would make determination after each iteration on whether the particular data point, namely x , would meet the requirements indicated as $\sigma_F(x) < t_{\sigma_{PN}} \sigma_{PN}$. The kernel function would accordingly be changed, x would be deemed as being overexploited, and θ would be increased by the number of repetitions provided that this condition would be met (Bull, 2011). The EIPSP approach would make adjustments for successfully raising σ_Q for all data points between the observations. Next, a new data point would then be created via the use of the kernel that is newly fitted. In the subsequent trials, θ would then be boosted ten times

provided that it would discover that the new data point would be exploitative. This process would then be repeated with a maximum of five times for the purpose of obtaining one data point, x , that would not be reported as overexploited. The modified x would then be recognized as the subsequent exploration ratio based upon the use of the *EIPSP* algorithm. The algorithm would aim at producing results that are in general more accurate through arriving at a balance between examining data points that are recently acquired and surrounding data points that are previously researched. By determining the model parameters, which in our case are shown in Table 2, one may utilise historical prices to get ten estimates for a certain week within the forecast duration. These ten projections are then averaged to generate the final price projection for that particular week.

The Bayesian optimization techniques would be applied across kernels, basis functions, how predictors are standardized, and σ . Forecast performance would be reported with the use of the relative root mean square error (*RRMSE*), which would be allowing comparisons of different kinds of anticipated outcomes over a variety of models or objectives. Here is one

example of how the *RRMSE* would be reported: $RRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^{obs} - y_i^{for})^2}}{\frac{1}{n} \sum_{i=1}^n y_i^{obs}}$, where the notation y^{for} would report the target's numerical value that is predicted; y^{obs} would represent the target variable's numerical value that is observed; and n would denote the total number of observations that would be utilized for the purpose of performance assessments. The prediction accuracy would then be examined via the use of two other performance measures. They are the mean absolute error (*MAE*) and the root mean square error (*RMSE*), and their magnitude and units would be corresponding to the target variable under consideration. The *RMSE* would be reported as: $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^{obs} - y_i^{for})^2}$. The *MAE* would be reported as: $MAE = \frac{1}{n} \sum_{i=1}^n |y_i^{obs} - y_i^{for}|$.

4. Result

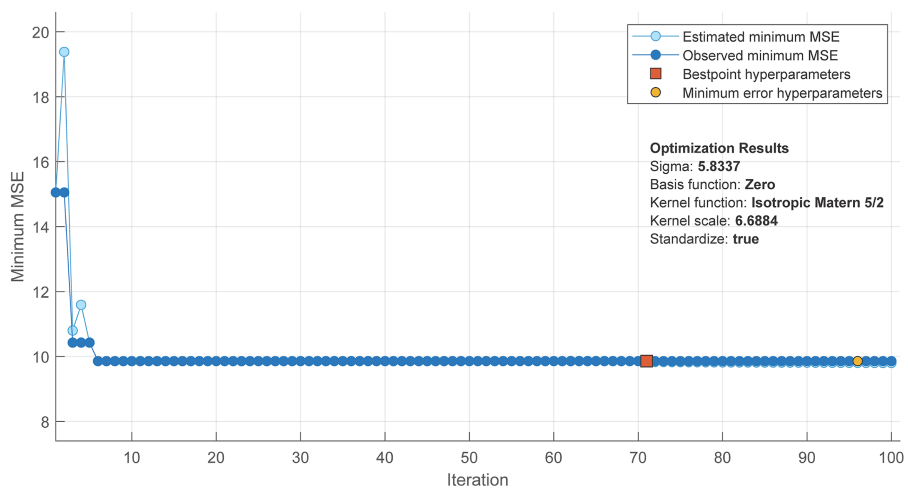
Price data that are associated with the out-of-sample testing phase, which include making predictions one week in advance, would span from January 4, 2019 to January 3, 2020, while those that are associated with the training phase would span from January 1, 2010 to December 28, 2018. Figure 2 would report the output that is associated with *EIPSP* optimizations which are based upon the training set. According to these findings, the isotropic Matern 5/2 kernel (reported in Equation (3)), empty basis function (reported in Equation (11)) and standardized predictors would be applied for estimating the price series. Table 2 would report the findings that correspond to the parameter estimates based upon the 10 GPR models that are created through the application of the ten-fold cross validation framework. The initials, reported as “CV1,” “CV2,” . . . , and “CV10,” which are standing for cross validation (“CV”), would be reflecting the outcomes of these parameter estimations.

The ten GPR models (in other words, Models “CV1,” “CV2,” . . . and “CV10”) that are reported in Table 2 would be in charge for the purpose of producing price projections. They were constructed via the use of the training data that are spanning from January 1, 2010 to

Table 2. Estimated parameters associated with ten GPR models for weekly edible oil prices

	CV1	CV2	CV3	CV4	CV5	CV6	CV7	CV8	CV9	CV10
σ	2.9184	2.5293	2.7982	2.9563	2.9816	2.9292	2.9336	2.8983	2.3807	2.9410
σ_l	10.8819	10.8951	10.0539	13.9821	15.2768	11.0737	12.3127	11.0935	9.6654	11.5593
σ_f	55.9051	53.7552	55.5117	60.2831	56.8791	56.9511	55.4756	56.1214	57.5565	58.9208

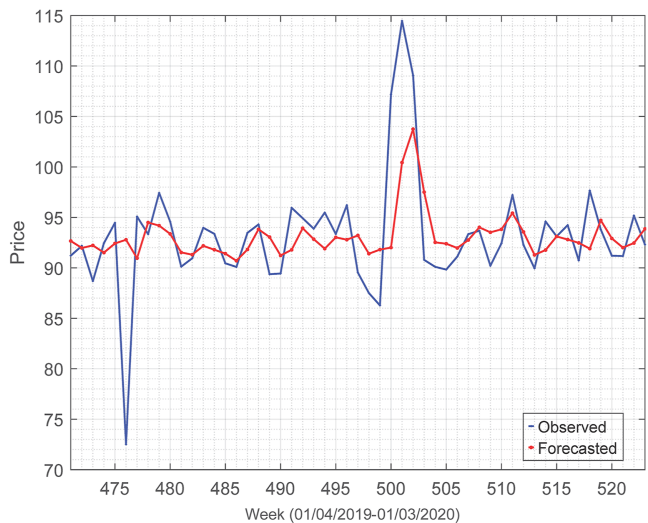
Source(s): Table by authors



Source(s): Figure by authors

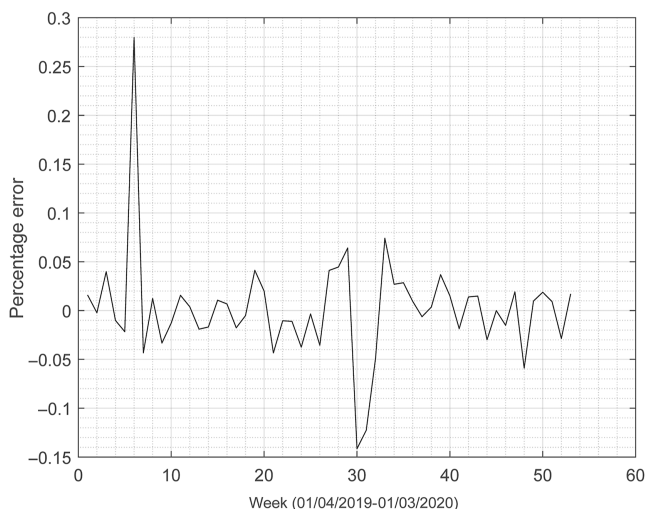
Figure 2. EIPSP optimization processes for weekly wholesale edible oil prices

December 28, 2018. This would mean that for every week that is in the testing period, ten price values that are predicted would be available. The final price estimate that is associated with that particular week would be determined via the averaging process of the 10 different estimations. This essentially would lead to the removal of any possible idiosyncratic estimates that are stemming from a given sub-model, which would be making way for more reliable and consistent predicted results for the future. The literature has already mentioned the benefits together with ideal qualities of adopting equal weighting techniques. The observed prices and the forecasted prices would be contrasted in Figure 3. Figure 4 would display the percentage



Source(s): Figure by authors

Figure 3. Forecast results and actual observations of weekly wholesale edible oil prices during the out-of-sample testing period of 01/04/2019–01/03/2020



Source(s): Figure by authors

Figure 4. Percentage forecast errors for weekly wholesale edible oil prices during the out-of-sample testing period of 01/04/2019–01/03/2020

forecast errors that are in accordance with what is reported through Figure 3. There is unquestionably a strong correlation between the expected and actual pricing. For the forecast findings that are reported in Figures 3 and 4, Table 3 would then provides an overview of the results that are associated with *RMSE*, *RRMSE* and *MAE*. In particular, 5.0812% would be the generated result for the *RRMSE*. Previous research (Jamieson *et al.*, 1991) has already made the recommendations of applying the following techniques for the purpose of grading model prediction accuracy levels: excellent provided that $RRMSE < 10\%$, good provided that $10\% < RRMSE < 20\%$, fair provided that $20\% < RRMSE < 30\%$ and poor provided that $RRMSE \geq 30\%$. These results would demonstrate that excellent prediction accuracy through the GPR models that are produced here is reached. Three significant forecast mistakes occur during the out-of-sample testing phase. The first one occurs on February 8, 2019 on which the observed price drops to 72.51 from 94.46 on February 1, 2019, representing a price reduction of 23.24%. The second and third ones occur on July 26, 2019 and August 2, 2019 on which the observed prices elevate to 107.17 and 114.47, respectively, from 86.28 on July 19, 2019, representing a price increase of 24.21 and 32.67%. These significant price increases that have been noticed should be linked to the government's use of policy shocks that are intended to modify prices for brief periods of time and often have no lasting effects. The model cannot instantly account for such quick and significant one-time policy changes since it relies on historical prices as predictors. However, the long-term performance of the model will not be impacted by these one-time shocks.

Table 3. Forecast performance during the out-of-sample testing period of 01/04/2019–01/03/2020 for weekly wholesale edible oil prices

<i>RRMSE</i>	<i>RMSE</i>	<i>MAE</i>
5.0812%	4.7324	2.9382

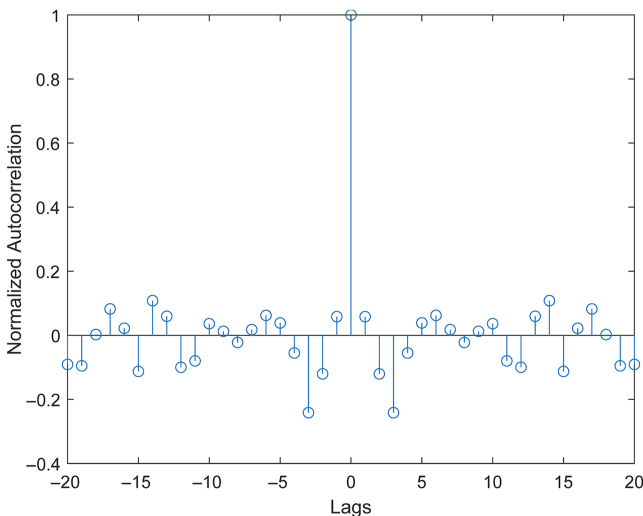
Source(s): Table by authors

The findings that are associated with error autocorrelation analysis, which was carried out for the purpose of evaluating the suitability of the built models, would be reported in [Figure 5](#). The study would be focusing on normalized autocorrelation results and would be carried out for a maximum of 20 delays. These results would then lead to the exclusion of any discernible autocorrelations and the validation of the built models' general validity. Despite potential conflicting empirical data reported in the literature, it is important to remember that the autoregressive conditional heteroskedasticity effect would possibly help improve a prediction model's performance.

5. Benchmark analysis

Up until this point, the majority of our analysis has been focusing on the building of GP regressions. In this section, we would then be examining four models that would be serving as benchmarks. These models would be the autoregressive model (AR), the AR-generalized autoregressive conditional heteroskedasticity model (AR-GARCH), the SVR model and the regression tree (RT) model (RT). Two metrics would be applied in the process of making evaluations of the performance of these various different models from the perspective of forecasting. These two different metrics would be the RMSE and the modified Diebold-Mariano (MDM) test ([Diebold and Mariano, 2002](#); [Harvey et al., 1997](#)). The MDM test would be applied for determining the significance of the difference that is in the mean squared errors (MSEs) of two different models' individual forecast results. In the MDM test, the null hypothesis states that the MSEs generated by two distinct models are identical. The MDM test would be exhibiting a *t*-distribution with *T* – 1 degrees of freedom under this null hypothesis.

The information regarding the four benchmark models would be given below. Both the AR and the AR-GARCH would make use of the same amount of lags for being applied as their predictors as the GPR model does. In the AR-GARCH model, the GARCH component would take the form GARCH(1,1). For the purpose of this present research, the linear ϵ -insensitive SVR would be the primary emphasis. The classification analysis and regression tree (CART) technique ([Breiman, 2017](#)) would be serving as the foundation for the RT. Same as the GPR model does, both the SVR and RT would be using lags of one to twelve price series for



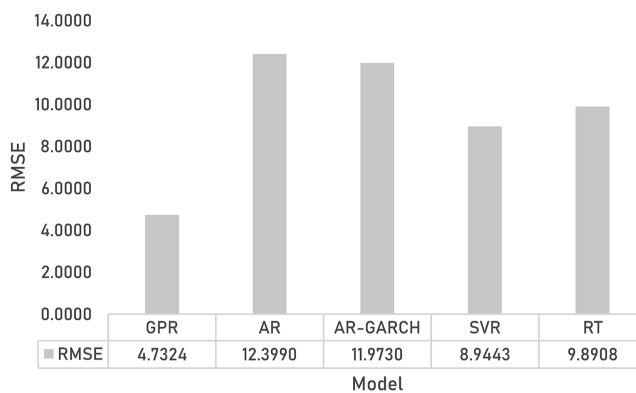
Source(s): Figure by authors

Figure 5. Normalized error autocorrelations based upon GPR models for weekly wholesale edible oil prices

predictors. Figure 6 would be displaying the findings of the benchmark research that is associated with the RMSE conducted during the testing phase. When being compared to the four different benchmark models under consideration, the chosen GPR would be having the best accuracy as well as the lowest RMSE. When making comparisons of the chosen GPR model to each of the four different benchmark models, the corresponding p -value generated from the MDM test would be less than 0.001, reporting that the chosen GPR model performs statistically significantly better than the four different benchmark models. Detailed p -values of MDM tests for comparisons of the GPR model with the AR, AR-GARCH, SVR and RT models would be presented in Figure 7.

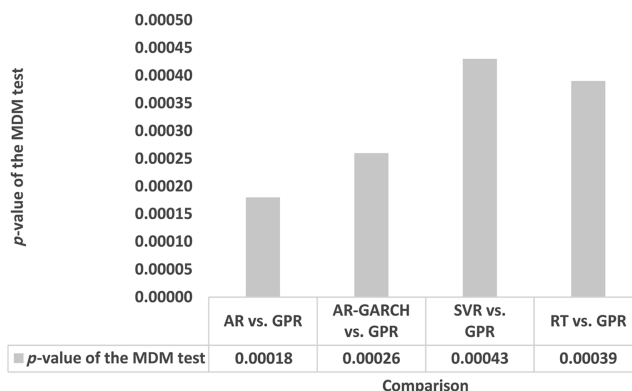
6. Policy implication

Estimating commodity prices may be very difficult for governments and investors, particularly when it comes to important agricultural goods. For investors to effectively manage risks, make strategic plans, and allocate and modify their portfolios in a suitable way, they must have



Source(s): Figure by authors

Figure 6. Benchmark analysis: comparisons of the GPR model with the AR, AR-GARCH, SVR and RT models in terms of the RMSE



Source(s): Figure by authors

Figure 7. Benchmark analysis: p – values of MDM tests for comparisons of the GPR model with the AR, AR-GARCH, SVR and RT models

precise estimates of commodity prices. Price projections are required for policymakers to know in order to conduct out market evaluations, create policies, put such policies into operation and make continual changes. Proper forecasting is crucial for a number of reasons, including maintaining a positive corporate environment and averting market collapse. To the best of their knowledge, econometric models—in particular, time-series econometric models—are the basis of the forecasting approach that the government and many investors use when a significant proportion of commodity prices are at risk. In the meanwhile, price estimates are nevertheless often based on expert opinions. This is supported by the possibility that developing, implementing, and maintaining econometric models and expert judgements will be comparatively simple. And because many of these models have been extensively used for decades by a broad range of forecast clients, it is likely that many of them have a respectable prediction accuracy. Given the growing affordability of computer resources and the solid foundation for expected nonlinear properties in price series of many commodities, machine learning models are widely recognised to have promise and worth more investigation. It may be difficult for some investors and policymakers to properly examine these models, however, since some decision-makers may still see them as too complicated forecasting tools. Actually, advanced investors and some governments have been increasingly interested in machine learning technologies in recent years. The study conducted here is a component of a broader approach that explores the possible uses of GP regression as a machine learning technique to solve the edible oil forecasting issue. The findings presented suggest that machine learning models may be worth looking into, maybe for a wider range of commodities, given the suggested method of building such a model and the shown strong forecast accuracy and stabilities.

7. Conclusion

Commodity price forecasts are important to a number of agricultural industry stakeholders. In this research, we use weekly wholesale edible oil prices over a ten-year period, from January 1, 2010, to January 3, 2020, to predict the Chinese market. Not enough attention has been paid to the projections for this significant price series in the literature. As a forecasting tool, the GP regression is investigated using cross-validation and Bayesian optimisations, yielding accurate and dependable findings. More precisely, for the period spanning from January 4, 2019 to January 3, 2020, the generated models were capable of generating predicting results of the prices with an out-of-sample RRMSE of 5.0812%, RMSE of 4.7324 and MAE of 2.9382. The benefit of the generated models is shown by our benchmark study, which compares the GP regression models with a number of other time-series and machine learning models. This data may be used for independent technical predictions or combined with other estimations when doing policy research that calls for price trend views. The modelling methodology that underlies these predictions may also be used to forecast problems of a similar kind in other economic areas. The simplicity and convenience of use of this framework may be essential for a number of decision-making processes. The technique presented here may be limited by the absence of potentially useful predictive data from other economic aspects, which might improve prediction performance. Examining regression models using Gaussian processes and exogenous inputs might help resolve this potential limitation. This is an important area for further study if data on other economic factors could be gathered. Future research on additional Bayesian optimisation procedures may be worthwhile, since the current study focusses on the EI per second plus method for Bayesian optimisations. Given that the time period under examination concludes in January 2020, it may also be a valuable direction for future research that considers more current time periods for analysis. We use the Bayesian optimisation technique to guide our model development procedure. The method determines the best GPR model based on the training sample by experimenting with different kernels, basis functions, and matching parameters. Following the construction of the optimum model, out-of-sample forecasting is performed. A promising direction for future research is the notion of building

several models and selecting significant ones based on the model confidence set. The price projection problem for China's wholesale edible oil market has been investigated in this paper. Investigating the possibilities of this empirical forecasting framework based on other commodity prices would be interesting.

Abbreviation

ARIMA	–	Autoregressive Integrated Moving Average
VAR	–	Vector Autoregressive
VECM	–	Vector Error Correction Model
BDS	–	Brock–Dechert–Scheinkman
GP	–	Gaussian Process
EIPSP	–	Expected Improvement Per Second Plus
EI	–	Expected Improvement
EIPS	–	Expected Improvement Per Second
RRMSE	–	Relative Root Mean Square Error
RMSE	–	Root Mean Square Error
MAE	–	Mean Absolute Error
CV	–	Cross Validation
AR	–	Autoregressive
GARCH	–	Generalized Autoregressive Conditional Heteroskedasticity
SVR	–	Support Vector Regression
RT	–	Regression Tree
MSE	–	Mean Square Error
MDM	–	Modified Diebold-Mariano
GPR	–	Gaussian Process Regression
CART	–	Classification Analysis and Regression Tree

References

- Alade, I.O., Zhang, Y. and Xu, X. (2021), “Modeling and prediction of lattice parameters of binary spinel compounds (am_2x_4) using support vector regression with Bayesian optimization”, *New Journal of Chemistry*, Vol. 45 No. 34, pp. 15255-15266, doi: [10.1039/d1nj01523k](https://doi.org/10.1039/d1nj01523k).
- Amal, I. (2021), “Crude palm oil price prediction using multilayer perceptron and long short-term memory”, *Journal of Mathematical and Computational Science*, Vol. 11, pp. 8034-8045, doi: [10.28919/jmcs/6680](https://doi.org/10.28919/jmcs/6680).
- Brahim-Belhouari, S. and Bermak, A. (2004), “Gaussian process for nonstationary time series prediction”, *Computational Statistics and Data Analysis*, Vol. 47 No. 4, pp. 705-712, doi: [10.1016/j.csda.2004.02.006](https://doi.org/10.1016/j.csda.2004.02.006).
- Brahim-Belhouari, S. and Vesin, J.-M. (2001), “Bayesian learning using Gaussian process for time series prediction”, *Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing (Cat. No. 01TH8563)*, IEEE, pp. 433-436, doi: [10.1109/SSP.2001.955315](https://doi.org/10.1109/SSP.2001.955315).
- Breiman, L. (2017), *Classification and Regression Trees*, Routledge.
- Brock, W.A., Scheinkman, J.A., Dechert, W.D. and LeBaron, B. (1996), “A test for independence based on the correlation dimension”, *Econometric Reviews*, Vol. 15 No. 3, pp. 197-235, doi: [10.1080/07474939608800353](https://doi.org/10.1080/07474939608800353).
- Bull, A.D. (2011), “Convergence rates of efficient global optimization algorithms”, *Journal of Machine Learning Research*, Vol. 12.
- Dacha, K., Cherukupalli, R. and Sinha, A. (2021), “Food index forecasting”, in *Applied Advanced Analytics*, Springer, pp. 125-134, doi: [10.1007/978-981-33-6656-5_11](https://doi.org/10.1007/978-981-33-6656-5_11).

- Darekar, A. and Reddy, A. (2017), "Forecasting oilseeds prices in India: case of groundnut, forecasting oilseeds prices in India: case of groundnut (december 14, 2017)", *Journal of Oilseeds Research*, Vol. 34, pp. 235-240, doi: [10.2139/ssrn.3237483](https://doi.org/10.2139/ssrn.3237483).
- Diebold, F.X. and Mariano, R.S. (2002), "Comparing predictive accuracy", *Journal of Business and Economic Statistics*, Vol. 20 No. 3, pp. 134-144, doi: [10.2307/1392185](https://doi.org/10.2307/1392185).
- Harvey, D., Leybourne, S. and Newbold, P. (1997), "Testing the equality of prediction mean squared errors", *International Journal of Forecasting*, Vol. 13 No. 2, pp. 281-291, doi: [10.1016/S0169-2070\(96\)00719-4](https://doi.org/10.1016/S0169-2070(96)00719-4).
- Hasanov, A.S., Do, H.X. and Shaiban, M.S. (2016), "Fossil fuel price uncertainty and feedstock edible oil prices: evidence from mgarch-m and virf analysis", *Energy Economics*, Vol. 57, pp. 16-27, doi: [10.1016/j.eneco.2016.04.015](https://doi.org/10.1016/j.eneco.2016.04.015).
- Jaiswal, R., Jha, G.K., Kumar, R.R. and Choudhary, K. (2022), "Deep long short-term memory based model for agricultural price forecasting", *Neural Computing and Applications*, Vol. 34 No. 6, pp. 4661-4676, doi: [10.1007/s00521-021-06621-3](https://doi.org/10.1007/s00521-021-06621-3).
- Jaiswal, R., Jha, G.K., Kumar, R.R. and Lama, A. (2023), "Agricultural price forecasting using narx model for soybean oil", *Current Science*, pp. 79-84.
- Jamieson, P., Porter, J. and Wilson, D. (1991), "A test of the computer simulation model arcwheat1 on wheat crops grown in New Zealand", *Field Crops Research*, Vol. 27 No. 4, pp. 337-350, doi: [10.1016/0378-4290\(91\)90040-3](https://doi.org/10.1016/0378-4290(91)90040-3).
- Jha, G.K. and Sinha, K. (2014), "Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India", *Neural Computing and Applications*, Vol. 24 Nos 3-4, pp. 563-571, doi: [10.1007/s00521-012-1264-z](https://doi.org/10.1007/s00521-012-1264-z).
- Ji, M., Liu, P., Deng, Z. and Wu, Q. (2022), "Prediction of national agricultural products wholesale price index in China using deep learning", *Progress in Artificial Intelligence*, Vol. 11 No. 1, pp. 121-129, doi: [10.1007/s13748-021-00264-0](https://doi.org/10.1007/s13748-021-00264-0).
- Jin, B. and Xu, X. (2024a), "Contemporaneous causality among price indices of ten major steel products", *Ironmaking and Steelmaking*, Vol. 51 No. 6, pp. 515-526, doi: [10.1177/03019233241249361](https://doi.org/10.1177/03019233241249361).
- Jin, B. and Xu, X. (2024b), "Forecasts of coking coal futures price indices through Gaussian process regressions", *Mineral Economics*, doi: [10.1007/s13563-024-00472-9](https://doi.org/10.1007/s13563-024-00472-9).
- Jin, B. and Xu, X. (2024c), "Price forecasting through neural networks for crude oil, heating oil, and natural gas", *Measurement: Energy*, Vol. 1, 100001, doi: [10.1016/j.meane.2024.100001](https://doi.org/10.1016/j.meane.2024.100001).
- Jin, B. and Xu, X. (2024d), "Forecasting wholesale prices of yellow corn through the Gaussian process regression", *Neural Computing and Applications*, Vol. 36 No. 15, pp. 8693-8710, doi: [10.1007/s00521-024-09531-2](https://doi.org/10.1007/s00521-024-09531-2).
- Jin, B. and Xu, X. (2024e), "Wholesale price forecasts of green grams using the neural network", *Asian Journal of Economics and Banking*, doi: [10.1108/AJEB-01-2024-0007](https://doi.org/10.1108/AJEB-01-2024-0007).
- Kanchymalay, K., Salim, N., Sukprasert, A., Krishnan, R. and Hashim, U.R. (2017), "Multivariate time series forecasting of crude palm oil price using machine learning techniques", *IOP Conference Series: Materials Science and Engineering*, Vol. 226, IOP Publishing, doi: [10.1088/1757-899X/226/1/012117](https://doi.org/10.1088/1757-899X/226/1/012117).
- Karia, A.A., Abd Hakim, T. and Bujang, I. (2016), "World edible oil prices prediction: evidence from mix effect of ever difference on box-jerkins approach", *Journal of Business and Retail Management Research*, Vol. 10.
- Khin, A.A., Mohamed, Z. and Malarvizhi, C. (2011), *Forecasting Methods of Spot Palm Oil Prices: Comparative Techniques*.
- Lama, A., Jha, G.K., Paul, R.K. and Gurung, B. (2015), "Modelling and forecasting of price volatility: an application of garch and egarch models", *Agricultural Economics Research Review*, Vol. 28 No. 1, pp. 73-82, doi: [10.5958/0974-0279.2015.00005.1](https://doi.org/10.5958/0974-0279.2015.00005.1).
- Lama, A., Jha, G.K., Gurung, B., Paul, R.K., Bharadwaj, A. and Parsad, R. (2016), "A comparative study on time-delay neural network and garch models for forecasting agricultural commodity price volatility", *Journal of the Indian Society of Agricultural Statistics*, Vol. 70, pp. 7-18.

- Li, F., Gao, F. and Kou, P. (2015), "Integrating piecewise linear representation and Gaussian process classification for stock turning points prediction", *Journal of Computer Applications*, Vol. 35, p. 2397, doi: [10.11772/j.issn.1001-9081.2015.08.2397](https://doi.org/10.11772/j.issn.1001-9081.2015.08.2397).
- Meena, D.C., Singh, O. and Singh, R. (2014), "Forecasting mustard seed and oil prices in India using arima model", *Annals of Agri Bio Research*, Vol. 19, pp. 183-189.
- Mishra, G. and Singh, A. (2013), "A study on forecasting prices of groundnut oil in Delhi by arima methodology and artificial neural networks", *Agris on-line Papers in Economics and Informatics*, Vol. 5, pp. 25-34, doi: [10.22004/ag.econ.157527](https://doi.org/10.22004/ag.econ.157527).
- Myat, A.K. and Tun, M.T.Z. (2019), "Predicting palm oil price direction using random forest", *2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE)*, IEEE, pp. 1-6, doi: [10.1109/ICTKE47035.2019.8966799](https://doi.org/10.1109/ICTKE47035.2019.8966799).
- Neal, R.M. (1997), "Monte Carlo implementation of Gaussian process models for Bayesian regression and classification", arXiv preprint physics/9701026.
- Neal, R.M. (2012), "Bayesian learning for neural networks", *Springer Science and Business Media*, Vol. 118.
- Priyanga, V., Lazarus, T.P., Mathew, S. and Joseph, B. (2019), "Forecasting coconut oil price using autoregressive integrated moving average (arima) model", *Journal of Pharmacognosy and Phytochemistry*, Vol. 8, pp. 2164-2169.
- Rahoveanu, A.T., Rahoveanu, M.M.T. and Ion, R.A. (2018), "Energy crops, the edible oil processing industry and land use paradigms in Romania—an economic analysis", *Land Use Policy*, Vol. 71, pp. 261-270, doi: [10.1016/j.landusepol.2017.12.004](https://doi.org/10.1016/j.landusepol.2017.12.004).
- Raihan, A. (2023), "The dynamic nexus between economic growth, renewable energy use, urbanization, industrialization, tourism, agricultural productivity, forest area, and carbon dioxide emissions in the Philippines", *Energy Nexus*, Vol. 9, 100180, doi: [10.1016/j.nexus.2023.100180](https://doi.org/10.1016/j.nexus.2023.100180).
- Raihan, A., Muhtasim, D.A., Farhana, S., Hasan, M.A.U., Pavel, M.I., Faruk, O., Rahman, M. and Mahmood, A. (2023), "An econometric analysis of greenhouse gas emissions from different agricultural factors in Bangladesh", *Energy Nexus*, Vol. 9, 100179, doi: [10.1016/j.nexus.2023.100179](https://doi.org/10.1016/j.nexus.2023.100179).
- Ranguwal, S., Sidana, B.K., Singh, J., Sachdeva, J., Kumar, S., Sharma, R.K. and Dhillon, J. (2023), "Quantifying the energy use efficiency and greenhouse gas emissions in Punjab (India) agriculture", *Energy Nexus*, Vol. 11, 100238, doi: [10.1016/j.nexus.2023.100238](https://doi.org/10.1016/j.nexus.2023.100238).
- Salman, N., Lawi, A. and Syarif, S. (2018), "Artificial neural network backpropagation with particle swarm optimization for crude palm oil price prediction", *Journal of Physics: Conference Series*, Vol. 1114, 012088, doi: [10.1088/1742-6596/1114/1/012088](https://doi.org/10.1088/1742-6596/1114/1/012088).
- Shahriari, B., Swersky, K., Wang, Z., Adams, R.P. and De Freitas, N. (2015), "Taking the human out of the loop: a review of Bayesian optimization", *Proceedings of the IEEE*, Vol. 104, pp. 148-175, doi: [10.1109/jproc.2015.2494218](https://doi.org/10.1109/jproc.2015.2494218).
- Shamsudin, M.N. and Arshad, F.M. (2000), "Short term forecasting of Malaysian crude palm oil prices", URL: econ1, available at: upm.edu.my/fatimah/pipoc.html
- Silalahi, D.D. (2013), "Application of neural network model with genetic algorithm to predict the international price of crude palm oil (cpo) and soybean oil (sbo)", in *12th National Convention on Statistics (NCS)*, pp. 1-2, Mandaluyong City, Philippine, October.
- Singh, A. (2021), "Comparison of artificial neural networks and statistical methods for forecasting prices of different edible oils in Indian markets", *International Research Journal of Modernization in Engineering Technology and Science*, Vol. 3, pp. 1044-1050.
- Wang, J., Dharmasena, S. and Bessler, D.A. (2013), "Price dynamics and forecasts of world and China vegetable oil markets". doi: [10.22004/ag.econ.151150](https://doi.org/10.22004/ag.econ.151150).
- Williams, C. and Rasmussen, C. (1995), "Gaussian processes for regression", *Advances in Neural Information Processing Systems*, Vol. 8.
- Wulandari, R., Surarso, B., Irawanto, B. and Farikhin, F. (2021), "The forecasting of palm oil based on fuzzy time series-two factor", *Journal of Soft Computing Exploration*, Vol. 2, pp. 11-16.

- Yang, J., Su, X. and Kolari, J.W. (2008), "Do euro exchange rates follow a martingale? Some out-of-sample evidence", *Journal of Banking and Finance*, Vol. 32 No. 5, pp. 729-740, doi: [10.1016/j.jbankfin.2007.05.009](https://doi.org/10.1016/j.jbankfin.2007.05.009).
- Yeeasin, M., Singh, K., Lama, A. and Paul, R.K. (2020), "Modelling volatility influenced by exogenous factors using an improved garch-x model", *Journal of the Indian Society of Agricultural Statistics*, Vol. 74, pp. 209-216.
- Yu, T.-H.E., Bessler, D.A. and Fuller, S.W. (2006), "Cointegration and causality analysis of world vegetable oil and crude oil prices". doi: [10.22004/ag.econ.21439](https://doi.org/10.22004/ag.econ.21439).
- Zhang, Y. and Xu, X. (2020), "Curie temperature modeling of magnetocaloric lanthanum manganites using Gaussian process regression", *Journal of Magnetism and Magnetic Materials*, Vol. 512, 166998, doi: [10.1016/j.jmmm.2020.166998](https://doi.org/10.1016/j.jmmm.2020.166998).

Corresponding author

Xiaojie Xu can be contacted at: xxu6@alumni.ncsu.edu